Basics of Probability

STA 198: Introduction to Health Data Science

Yue Jiang

May 22, 2020

The following material was used by Yue Jiang during a live lecture.

Without the accompanying oral comments, the text is incomplete as a record of the presentation.

Announcements

- HW/Lab 01 due next Tuesday, May 26 on Gradescope.
- No lecture on Monday, May 25, due to Memorial Day Holiday (I'll still be available if you have questions, though there will NOT be official office hours on Monday).

What's the use of probability?

- Last time: how descriptive statistics are used to *describe* data
- Goal: Make *inferences* about a population based on a sample

To do this, we need a solid foundation of probability theory.

Probabilities come up all the time

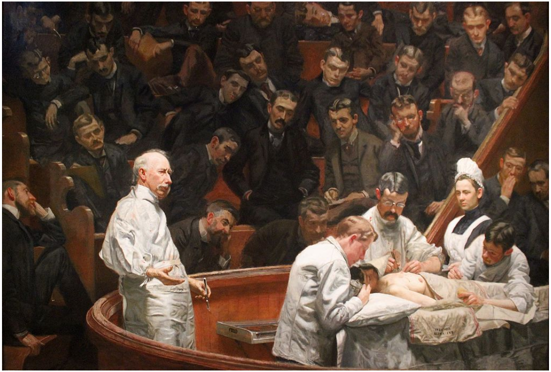
How do we interpret the following statements?

- There is a moderate chance of drought in North Carolina during the next year
- The surgery has a 50-50 probability of success
- The ten-year survival probability of invasive breast cancer among U.S women is 83%

Interpretations of probability

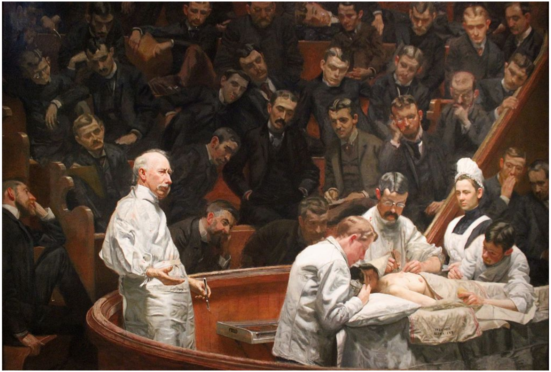
"There is a 1 in 3 chance of selecting a white ball"

Interpretations of probability



"The surgery has a 50% probability of success"

Interpretations of probability



Long-run frequencies vs. degree of belief

Probability spaces

Mathematical objects that model random experiments, real-world processes involving states that occur randomly

A probability space consists of three components:

1. A sample space, the set of all possible outcomes
2. Subsets of the sample space, called events, which comprise any number of possible outcomes (including none of them!)
3. A function that assigns probabilities to events

An event occurs if the outcome of the random experiment is contained in that event

Sample spaces

Sample spaces depend on the random experiment in question

- Tossing a single fair coin
- Tossing two fair coins
- Sum of rolling two fair six-sided dice
- Survival (years) after cancer diagnosis

Sample spaces

Sample spaces depend on the random experiment in question

- Tossing a single fair coin $\{H, T\}$
- Tossing two fair coins $\{HH, HT, TH, TT\}$
- Sum of rolling two fair six-sided dice $\{2, 3, 4, \cdots, 12\}$
- Survival (years) after cancer diagnosis $[0, \infty)$

Events

Subsets of the sample space that comprise possible outcomes. Essentially, these are all the ‘plausibly reasonable’ events we’re interested in calculating probabilities for*:

- Tossing a single fair coin
- Tossing two fair coins
- Sum of rolling two fair six-sided dice
- Survival (years) after cancer diagnosis*

**there are some nasty mathematical details behind this seemingly simple task. Don’t worry about them!*

Events

Subsets of the sample space that comprise possible outcomes. Essentially, these are all the ‘plausibly reasonable’ events we’re interested in calculating probabilities for*:

- Tossing a single fair coin A head
- Tossing two fair coins At least one head
- Sum of rolling two fair six-sided dice An odd number
- Survival (years) after cancer diagnosis* $>$one year

**there are some nasty mathematical details behind this seemingly simple task. Don’t worry about them!*

Probabilities

A number describing the likelihood of each event's occurrence. This maps events to a number between 0 and 1, inclusive:

- Tossing a single fair coin A head
- Tossing two fair coins At least one head
- Sum of rolling two fair six-sided dice An odd number
- Survival (years) after cancer diagnosis >one year

Probabilities

A number describing the likelihood of each event's occurrence.
This maps events to a number between 0 and 1, inclusive:

- Tossing a single fair coin A head 0.5
- Tossing two fair coins At least one head 0.75
- Sum of rolling two fair six-sided dice An odd number 0.5
- Survival (years) after cancer diagnosis >one year ...harder

How did we come up with those answers?

- For the first three, you probably would intuit your way through using a discrete probability distribution (more on this in just a few lectures)
- For the last one, by making some assumptions on the survival process, we can use similar probability tools to arrive at an answer

Events as (sub)sets

Remember, events are subsets of the entire sample space. Let's take for now the example of tossing a single fair coin and recording the outcome.

There are only two elements in the outcome space:

- A: getting a head
- B: getting a tail

We can define the simple events of just A or B occurring, but are there "other" events we can define?

Set operations

Sets can be related to each other in different ways. For two sets (or events) A and B, the most common relationships are:

- Intersection ($A \cap B$): A and B both occur
- Union ($A \cup B$): A or B occur (including when both occur)
- Complement (A^c): A does not occur
- Difference ($A \setminus B$): A occurs, but B does not occur; equivalent to ($A \cap B^c$) (why?)

Two sets A and B are said to be disjoint if $A \cap B = \emptyset$

Those "other" events

What are the intersection, union, complement, and difference of events A (getting a head) and B (a tail) as applied to our coin-toss example? Are the two events A and B disjoint?

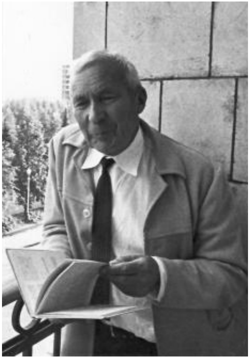
What are the probabilities that assigned to those events:

- $P(A \cap B) = ?$
- $P(A \cup B) = ?$
- ...etc.

How do probabilities "work"?

Kolmogorov axioms

1. The probability of any event in the sample space is a non-negative real number (could be zero!)
2. The probability of the entire sample space is 1
3. If A and B are disjoint events (mutually exclusive), then the probability of A or B occurring is the sum of the individual probabilities that they occur



How do probabilities "work"?

For two events A and B with probabilities $P(A)$ and $P(B)$ of occurring, the Kolmogorov axioms give us two important rules:

- Complement Rule: $P(A^c) = 1 - P(A)$
- Inclusion-Exclusion: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

How do we extend inclusion-exclusion to more than two events?

DeMorgan's laws

- Complement of union:
  $(A \cup B)^c = A^c \cap B^c$
- Complement of intersection:
  $(A \cap B)^c = A^c \cup B^c$

How do we interpret these in plain English?

How do we extend DeMorgan's laws to more than two events?

Gunter et al. study (2017)

Original Research

Annals of Internal Medicine

Coffee Drinking and Mortality in 10 European Countries

A Multinational Cohort Study

What was the probability a randomly selected person in the trial...

	Died		
Coffee drinking	Yes	No	Total
None	1039	5438	6477
Med-Low	4440	29712	29809
High	3601	24934	28535
Total	9080	60084	64821

- ...did not drink coffee?
- ...died during the study or did not drink coffee?
- ...did not die during the study and was a high coffee drinker?