Probability (1)

STA 211: The Mathematics of Regression

Yue Jiang

February 7, 2023

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.

An important disclaimer

This is not a probability class. There are semester-long (and multiple semester-long) courses on probability, and so what we cover in just two lectures scarcely touches on even the basics.

However, familiarity with some concepts from probability, such as probability distributions and certain aspects of them (e.g., expectation, variance, etc.) are needed to more fully grasp linear models. As such, we will be presenting a very abridged treatment of some of the fundamentals needed to proceed.

Probabilities come up all the time

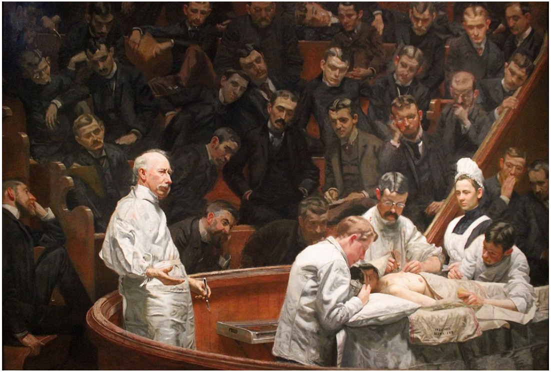
How do we interpret the following statements?

- 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

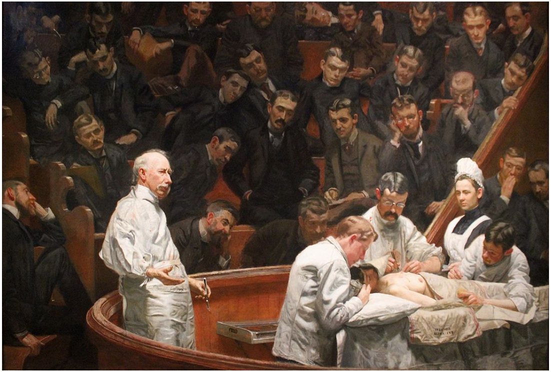
"There is a 1 in 3 chance of selecting a white ball"

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.

A probability space consists of three components:

1. A sample space, the set of all possible outcomes, denoted Ω
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

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!*

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

Events as (sub)sets

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 A occurring or B occurring, but are there "other" events we can define?

Set operations

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

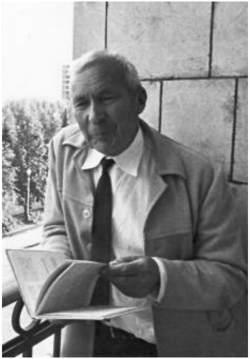
Two sets A and B are said to be disjoint if $A \cap B = \emptyset$

How do probabilities "work"?

Kolmogorov axioms

1. The probability of any event in the sample space is a non-negative real number: $P(A) \geq 0$ for every event A.
2. The probability of the entire sample space is 1: $P(\Omega) = 1$.
3. If A_1 and A_2 are disjoint events (mutually exclusive), then the probability of A_1 or A_2 occurring is the sum of the individual probabilities that they occur. Generally:

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$$



Finite or countable sample spaces

For sample spaces Ω that have either a finite or countably infinite number of elements, then the probability function can be completely specified by the probabilities of individual outcomes in this space.

Example, countably infinite sample space: a fair coin is tossed until the first heads; the random experiment in question is the number of tosses. Then the sample space is $\Omega = \mathbb{Z}_+$.

- Show that $P(\Omega) = 1$, as in the second Kolmogorov axiom
- What is the probability that the number of tosses it takes is a multiple of 3?

Mass function

For finite or countable sample spaces (this is to say, for **discrete random variables**), the **probability mass function** provides the probability that a discrete random variable equals some value:

$$p_X(x) = P(X = x)$$

Distribution functions

Remember that random variables can be thought of as functions that "tell us about" outcomes of random experiments. The **distribution function** of a random variable X is defined as the following probability:

$$F_X(x) = P(X \leq x)$$

we're skipping some nuance, but this is an acceptable definition for STA 211 purposes.

Distribution functions are non-decreasing with $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

Probability mass functions and distribution functions

(See board)

Uncountably infinite sample spaces

But how might we enumerate events in an *uncountable* sample space? (unfortunately, we can't, or at least not in a way that "makes sense")

(See board - notion of length on number line)

- How might we find the probability of a finite or countable union of disjoint open intervals?

For these types of spaces, we can't assign probabilities to "every subset of Ω" in a meaningful way. Instead, we work with a class of subsets that we "know what to do with" - I'll leave it at that.

Distribution functions

However, we can still use the notion of a distribution function to define **continuous random variables** (ones with uncountably infinite sample spaces) as well.

$$F_X(x) = P(X \leq x)$$

we're skipping some nuance, but this is an acceptable definition for STA 211 purposes.

Distribution functions are non-decreasing with $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.

Density functions

For continuous, differentiable distribution functions, the derivative of the distribution function is called the **density function**, denoted $f_X(x)$. We can write the distribution function in terms of the density function as follows:

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz$$

$$f_X(x) = \frac{d}{dx}F_X(x)$$

We can define probabilities as follows:

$$P(a < X \leq b) = \int_{a}^{b} f_X(z)dz,$$

noting that $P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f_X(z)dz = 1$.

we're again skipping some nuance once more, but this is again an acceptable definition for STA 211 purposes.

Distribution functions and density functions

(See board)

Expectations of (functions of) random variables

For a discrete random variable:

$$E(X) = \sum_{all\ \ x} xP(X = x),$$

$$E(g(X)) = \sum_{all\ \ x} g(x)P(X = x)$$

"Similarly," for a continuous random variable:

$$E(X) = \int_{-\infty}^{\infty} xf_X(x)dx,$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$$

Variance (and covariance)

The **variance** of a function is defined as

$$Var(X) = E\left((X - E(X))^2\right)$$

The square root of the variance is known as the **standard deviation**

The **covariance** of two random variables X and Y is defined as

$$Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right)$$

- What do the variance and covariance mean in words?
- What is $Cov(X, X)$?

The normal (Gaussian) distribution

For $X \sim N(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

(See board re: density and distribution function)

$E(X) = \mu$ and $Var(X) = \sigma^2$ (can you show this?).

Homework 4: Due Valentine's Day! (sorry)

1. For any random variable X and constants a and b, show that $E(aX + b) = aE(X) + b$ and $Var(aX + b) = a^2 Var(X)$
2. For any random variable X, show that $Var(X) = E(X^2) - (E(X))^2$
3. Compute the following integral using basic algebra and any fact stated in the slides:

$$\int_{-\infty}^{\infty} e^{-x^2} dx$$

In particular, this means **without** using power series expansions or any other asymptotic approaches, change of variables, integration in polar coordinates and/or by parts, the Leibniz integral rule, or the residue theorem. Even if you successfully obtain the answer using these approaches, you'll get zero credit!