Lecture 5 - Probability

Theodore Landsman

Georgetown University

August 19th - 2022

Morning challenge!

1. Calculate the area under the curve:
   1.1 $\int_2^2 (ln(x^5 + \sqrt[3]{x^5}) + \frac{4x^3}{x-5})dx$
2. Calculate the gradient and Hessian of the following functions:
   2.1 $f(x,y) = 3x^4y^5 - \frac{4}{x^2y}$
   2.2 $f(x,y,z) = 2x + 4xy + 5y^3z^2 + x^3y^45^5$
3. For $x = (1,2,3,4,5)$ and $y = (1.5,4,4,9,14)$
   3.1 Calculate β and α and for the OLS regression line using the formulas:
   $$\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}))}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$
   $$\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$$
   3.2 What is the Residual Sum of Squares (RSS) for this model?

Agenda

Agenda

Probability

Distributions

Frequentist vs. Bayesian

There are two schools of thought in statistics: frequentists and bayesian.

Frequentist: The parameters of interest is constant but unknown.

Bayesian: The parameters of interest is a random variable.

In general, PhD programs have more courses related with the frequentist school. Bayesian statistics often comes up in terms of advanced models and forecasting. In broad strokes we might say that academia is more focused on causal identification (was x an important factor in producing y), where Bayesian methods are not always necessary, while industry tends to be more focused on prediction (can we use x to predict the behavior of y), where Bayesianism is more relevant.

Definitions:

- **Outcomes** are anything that might happen in the world.
- **Events** are composed of one or more outcomes.

Example: When one rolls a die there are six possible outcomes. So each of these outcomes can be defined as an event. But also, you can have an event with more outcomes (the event that one rolls an even number).

In political science we are interested in **random events** or, in other words, in events that are probabilistic (as opposed to deterministic events).

Specifically, for a random event one might be able to specify causal processes that alter the chance that it occurs, but one cannot specify causal processes that guarantee the event will occur.

Sample space: is the set of all possible outcomes: it is a list of each event we might observe. For example: Coin flip: head or tail.

Calculating Probability:

$$Pr(e) = \frac{\text{No. of outcomes in event e}}{\text{No. of outcomes in sample space}}$$

e.g. What is the probability of getting a tail when you flip a coin?

Relationships between events

Independence: The probability of one event does not change the probability of another event.

Example: Attended Math Camp today and vote choice in the next presidential election.

Mutual exclusivity: If one events occurs, the other event cannot occur.

Example: You were accepted in two PhD programs. If you choose one, you cannot choose the other at the same time.

Collective Exhaustivity: Every possible event fits into one of the categories.

Example: You are a human being or you are not.

Conditional events:: The probability of one event occurring is affected by whether another event occurs.

Example: The probability to be successful in the PhD in government is affected by having attended math camp.

Notation!

- $Pr(A)$ is the probability that an event A occurs. It is an unconditional probability: it assumes that the probability that A will occur is completely independent of everything else in the sample space.
- All probabilities lie between zero and one, so $Pr(A) \in [0;1]$.
- $Pr(A|B)$ is the conditional probability of A on B. In other words, it is the probability that A occurs given that B has already occurred.
- For independent events, $Pr(A|B) = Pr(A)$
- $A \cup B$ is the compound event where either A or B happens, or both. Thus, we read $A \cup B$ as "A or B".
- $A \cap B$ is the compound event where both A and B happen. Thus, we read $A \cap B$ as "A and B".

More on notation and definitions

- $Pr(A \cap B) = Pr(B|A)Pr(A) = Pr(A|B)Pr(B)$
- When A and B are independent: $Pr(A \cap B) = Pr(A)Pr(B)$
- When A and B are mutually exclusive: $Pr(A \cap B) = 0$ because $Pr(A|B) = Pr(B|A) = 0$.
- $A \cup B = Pr(A) + Pr(B) - Pr(A \cap B)$
- We will define $\sim A$ as "not A".

Bayes theorem:

The *economist* (2004) offers the following illustration of Bayes' rules:

The canonical example is to imagine that a precarious newborn observes his first sunrise, and wonders whether the sun will rise again or not. He assigns equal prior probability to both possible outcomes, and represents this by placing one white and one black marble into a bag. The following day, when the sun rises, the child places another white marble in the bag. The probability that a marble plucked randomly from the bag will be white (i.e., the child's degree of belief in future sunrises) has thus gone from a half to two-thirds. After sunrise the next day, the child adds another white marble, and the probability (and thus the degree of belief) goes from two-thirds to three-quarters. And so on. Gradually, the initial belief that the sun is just as likely to rise as not to rise each morning is modified to become a near-certainty that sun will always rise.

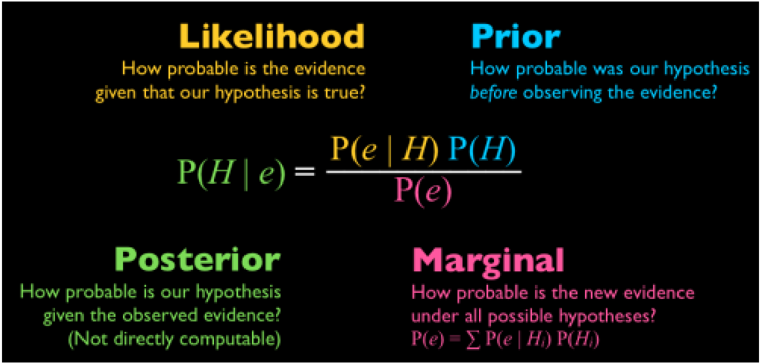
Bayes theorem:

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

What is equivalent to:

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{P(A|B)Pr(B)+Pr(A|\sim B)Pr(\sim B)}$$

Bayes theorem



Bayes theorem example

Suppose you are going to scan your suitcase at an airport, but before doing so an alarm has been triggered because the previous suitcase appears to have a bomb in it.

Some important facts to consider:

- These x-ray tests are not perfect, so when there IS a bomb, it gets detected 98% of the time.
- When there IS NOT a bomb, the staff will think they see one about 1% of the time.
- Only 1 person in a million actually DOES have a bomb in their bag.

Assuming that 100,000,000 people use that airport we have:

	Bomb	No Bomb	
Positive	98	999,999	
Negative	2	98,999,901	
TOT PEOPLE:	100	99,999,900	100,000,000

Bayes theorem example

What is the probability that the previous bag ACTUALLY contains a bomb?

We can use Bayes Theorem!

- Using the theorem: $Pr(Bomb|Alarm) = \frac{Pr(Alarm|Bomb)Pr(Bomb)}{P(Alarm)}$
- $Pr(Alarm/Bomb) = 0.98$
- $Pr(Bomb) = 0.000001)$
- $Pr(Alarm) = \frac{\text{Alarm activated}}{\text{Tot People}} = \frac{99+999999}{100000000} = 0.01$

Replacing in our initial Equation:

$$Pr(Bomb|Alarm) = \frac{Pr(Alarm|Bomb)Pr(Bomb)}{P(Alarm)} = \frac{0.98 \cdot 0.000001}{0.01} = 0.00009 = 0.009\%$$

Main lesson: $Pr(Alarm|Bomb) \neq Pr(Bomb|Alarm)$

Odds ratios

The odds of an event happening are defined as the ratio of the probability of the event occurring divided by the probability of the event not occurring. Example: The odds of rolling a four on a dice are $1/5 = 20\%$.

$$Odds = \frac{Pr(x)}{Pr(\sim x)}$$

Odds ratios: they refer to the ratio of the odds of an event occurring in the exposed group versus the unexposed group.

$$\text{Odds ratio} = \frac{Pr(x_1)/Pr(\sim x_1)}{Pr(x_2)/Pr(\sim x_2)}$$

Combinations and permutations

Combinations: Choose k objects from a set of n objects when the order **does not** matter.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Permutations: Choose k objects from a set of n objects when the order **does** matter.

$$\binom{n}{k} = \frac{n!}{(n-k)!}$$

By definition, there will be more permutations that combinations.

Combinations and permutations: Example:

12 PhD students decide to host a thanksgiving potluck:

- 3 students will be asked to bring appetizers
- 5 students will be asked to bring main courses
- 3 students will be asked to bring desserts
- The remaining student will be asked to bring drinks

How many different ways can the students be divided into these groups?

$$\binom{12}{3} \cdot \binom{9}{5} \cdot \binom{4}{3} \cdot \binom{1}{1} = \frac{12!}{3!9!} \cdot \frac{9!}{5!4!} \cdot \frac{4!}{3!1!} \cdot \frac{1!}{1!0!} = 220 \cdot 125 \cdot 4 \cdot 1 = 110,880$$

Agenda

Random variables

Random variables can take on an array of values, with the probability that it takes on any specific value determined by the distribution.

A random variable is **realized** when it takes on any specific value from the set of possible values.

We say that random variables are **stochastic**: Their probability distribution can be analyzed statistically, but cannot be precisely predicted.

Discrete versus continuous

A random variable X is **discrete** if it takes finite or countably infinite number of values.

- e.g. Coin flip: Heads or tails

A random variable X is **continuous** if it can take any real numbers in the domain.

- e.g. height of a randomly selected person

Sample distribution

Sample distributions are empirical constructs. They are representations of the number of cases that take each of the values in a sample space for a given portion of the population of cases.

Frequency Distribution: A count of the number of cases with any value, often used to rank order.

Relative Frequency Distribution:A count of the number of cases with any value, divided by the total number of cases.

Table 10.3: Militarized Interstate Dispute Initiators, 1816–2002

MID Initiator	No. of Countries	% of Countries
No	67	31
Yes	147	69

Contingency tables

A matrix that shows the joint frequency distribution for two variables.

Also known as the infamous "2 x 2" tables:

	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

Probability distribution

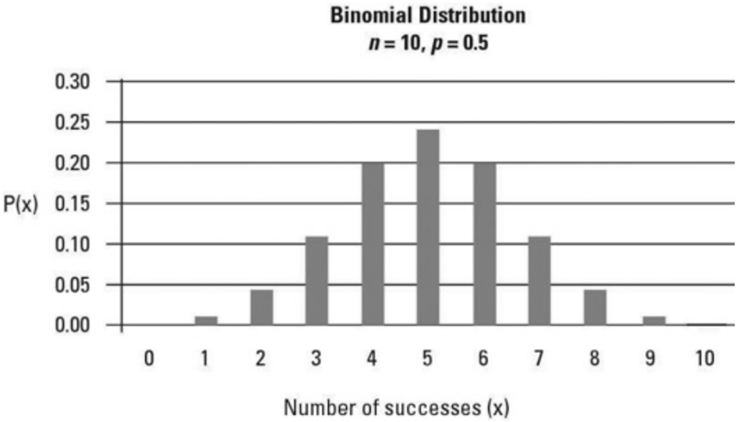
The function that describes the likelihood of getting any specific value of a random variable.

The sum of all probabilities must equal 1. The probability of any one value or any range of values must be between 0 and 1.

These distributions will be different if the variable is discrete or continuous.

Probability Mass Function

For discrete variables the probability distribution will be called the Probability Mass Function.



Binomial distribution: ten trials with $p = 0.5$.

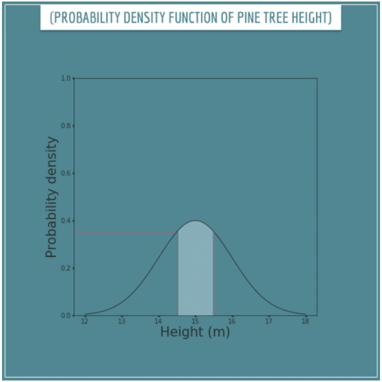
Probability Density Function

For continuous variables the probability distribution will be called the Probability Density Function.

What is the likelihood that a random pine tree is any specific height?

It is not the value of the function, the probability is the area under the curve!

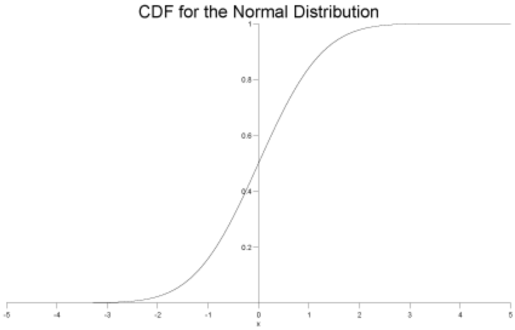
This is why we need integration.



Probability Density Function

The integral, or antiderivative, of the Probability Density Function.

It tell us the area under the curve until any specific value.

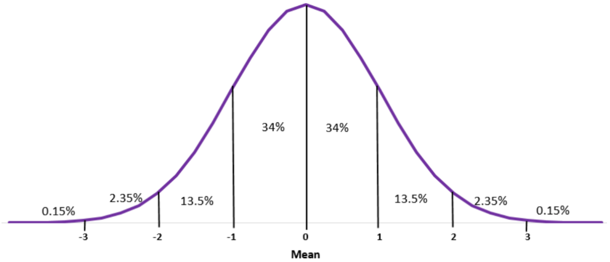


Normal/Gaussian Distribution (the bell curve)

Important features:

- This distribution is symmetric around the center.
- Standard deviations are measures of how spread out the observations are: 68% of values are within one SD, 95% are within two, 99.7% are within three.

Very typical in political science!



Other types of distributions:

- Poisson
- Binomial
- Negative binomial
- t distribution
- F distribution
- Exponential distribution
- Gamma distribution
- And many more...

Ovid ♡

THE END!

THANK YOU!