Lecture 5 — Probability

Henry Watson

Georgetown University

8/18/23

Morning challenge!

1. Calculate the area under the curve:
   1. $\int_1^5 (8x+3)(4x^2+3x+1)^2$
2. Calculate the gradient and Hessian of the following function:
   1. $f(x,y,z) = 2x + 4xy + 5y^3z^2 + 6x^3y^4$
3. For $x = (1,2,3,4,5)$ and $y = (1.5,4,4,9,14)$
   1. Calculate β and α and for the OLS regression line using the formulas: $\hat{\beta} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}))}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$
      $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$

Agenda

Frequentist vs. Bayesian

- There are two schools of thought in statistics: Frequentist and Bayesian
- **Frequentist**: The parameters of interest is constant but unknown
  - There is a truth, a right answer, and I can be ***confident*** to some degree than I have found the truth
- **Bayesian**: The parameters of interest is a random variable.
  - I have an opinion, which I will update based on data to form a new opinion
- Frequentists care about truth, whereas Bayesians care about prediction and perspective
- In academia, we tend to focus more on Frequentist statistics as it is more relevant to causal identification

Frequentist vs. Bayesian

- If I am going to flip a fair coin, what is the probability that it will land heads-side up?
  - Both a frequentist and a bayesian will tell you 50%
- If I have already flipped a coin, but not looked at it yet, what is the probability that that specific coin is heads-side up?
  - The bayesian will still say 50%, but the frequentist will say it's no longer a matter of probability

Definitions:

- **Outcomes**: anything that might happen in the world
- **Events**: composed of one or more outcomes
- **Sample Space**: the set of all possible outcomes
- **Random Events**: events that are probabilistic (as opposed to deterministic
    - When we say "random" here, we mean probabilistic
    - Can identify causal processes that alter the probability, but not causal processes that guarantee the event will occur

Calculating Probability:

$$Pr(e) = \frac{\text{No. of outcomes in event e}}{\text{No. of outcomes in sample space}} \tag{1}$$

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
- **Mutual exclusivity:**If one events occurs, the other event cannot occur
- **Collective Exhaustivity:** Every possible event fits into one of the categories
- **Conditional events:**: The probability of one event occurring is affected by whether another event occurs

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.

Notation!

- $Pr(A)$: the unconditional probability that an event A occurs
  - All probabilities lie between zero and one, so $Pr(A) \in [0;1]$
- $Pr(A|B)$: the conditional probability of A given B
  - If A is independent of B, $Pr(A|B) = Pr(A)$
- $A \cup B$: the compound event where either A or B happens, or both
- $A \cap B$: the compound event where both A and B happen

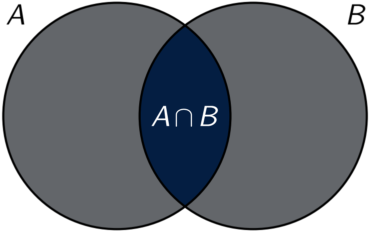
More on notation and definitions

- $Pr(A \cap B) = Pr(B|A)Pr(A) = Pr(A|B)Pr(B)$
  - The probability of A and B happening is equal to the conditional probability of B given A times the unconditional probability of A (or vice versa)
  - The probability of someone voting and donating to a candidate is equal to the probability of a donor voting times the probability of donating
- When A and B are independent: $Pr(A \cap B) = Pr(A)Pr(B)$
- Example:
  - $\Pr(\text{voting}) = 0.5$
  - $\Pr(\text{donating}) = 0.2$
  - If voting is independent of donating, $\Pr(\text{voting} \cap \text{donating}) = 0.5 \times 0.2 = 0.10$
  - If $\Pr(\text{voting}|\text{donating}) = 0.9$, $\Pr(\text{voting} \cap \text{donating}) = 0.9 \times 0.2 = 0.18$

More on notation and definition

- When A and B are mutually exclusive: $Pr(A \cap B) = 0$ because $Pr(A|B) = Pr(B|A) = 0$.
- $A \cup B = Pr(A) + Pr(B) - Pr(A \cap B)$
  - The probability of A or B happening is equal to the sum of their unconditional probabilities minus the probability of both happening
  - Using our example from before, Pr(voting $\cup$ donating) $=$ $0.5 + 0.2 - 0.18 = 0.52$
- We will define $\sim A$ as "not A".

Subtract the intersection so you don't double-count it



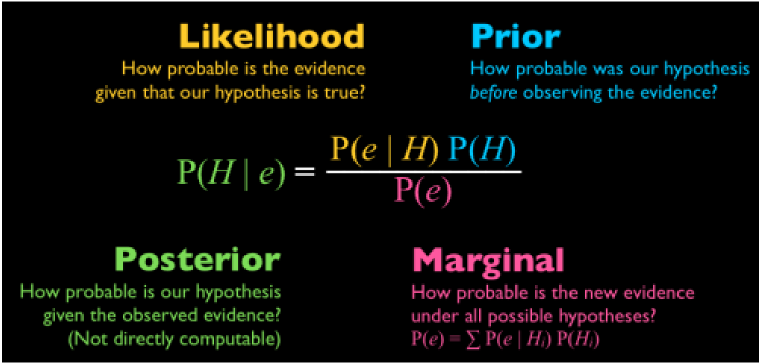
Bayes theorem:

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{P(A)} \tag{2}$$

What is equivalent to:

$$Pr(B|A) = \frac{Pr(A|B)Pr(B)}{P(A|B)Pr(B) + Pr(A| \sim B)Pr(\sim B)} \tag{3}$$

Bayes theorem



Bayes theorem example

- Covid tests can have false positive (say you're sick when you aren't) and false negative (say you're not sick when you are) results
- 1,000 people have symptoms, and take a test. 50 of them are actually sick.
- 45 people test positive, but 5 of those are false positives
- 955 people test negative, but 10 of those are false negatives

Bayes theorem example

- We can use Bayes Theorem to determine the probability of testing positive if you are sick
- $Pr(Positive|Sick) = \frac{Pr(Sick|Positive)Pr(Positive)}{P(Sick)}$
- $Pr(Sick|Positive) = 40/45 = 0.89$
- $Pr(Positive) = 45/1000 = 0.045$
- $Pr(Sick) = 50/1000 = 0.050$
- $Pr(Positive|Sick) = \frac{(0.89)(0.045)}{0.050} = 0.80$
- **Main lesson:** $Pr(Positive|Sick) \neq Pr(Sick|Positive)$

The Monty Hall Problem

- You're on a game show, hosted by Monty Hall. There are 3 doors, one of which has a prize behind it. You state to Monty which door you want to open.
- Before revealing what is behind your door, Monty opens one of the other two doors, revealing that there is no prize behind it.
- Monty then gives you the option to switch your original choice.
- Switching doors will, statistically, increase your chances of winning. What?!

The Monty Hall Problem

- The key to this problem is that Monty knows where the prize is. He will not open the door with the prize, and he will not open your door.
- If you chose the right door initially, you will always lose if you switch
- If you chose the wrong door initially, Monty will open the only other door without a prize, meaning that switching your answer will give you the prize.
    - If you picked incorrectly the first time, you will always win if you switch
- The start of the game is random: you had a $\frac{1}{3}$ chance of guessing right, and a $\frac{2}{3}$ chance of guessing wrong.
    - Chances of winning if you switch: $(1.00 \times 0.66) + (0.00 \times 0.33)$

Odds

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 $1:5$.

$$Odds = \frac{Pr(A)}{1 - Pr(A)} \tag{4}$$

Odds Ratios

- Odds ratios: the ratio of the odds of an event occurring under the condition A and the odds of that event occurring under the condition B

$$\text{Odds ratio} = \frac{\#Events(A)/\#NonEvents(A)}{\#Events(B)/\#NonEvents(B)} \tag{5}$$

	Small Landlord	Large Landlord
Received Eviction Filing	4,890	13,040
Did not Receive Eviction Filing	86,390	58,860

- Do larger landlords evict at a higher rate?
- $\frac{13040/58860}{4890/86390} = 3.91$

Frequentist Methods

- Frequentist statistics relies on parameters of interest being an unknown fixed quantity
- Our methods are built around estimating that truth, and expressing uncertainty
- Examples:
  - A true share of people support Donald Trump for president, and we can estimate that share with a poll
  - That poll will have a "margin of error"
  - There is a true relationship between Democracy and Conflict, and we can estimate that relationship with tools such as linear regression
  - Our estimates will have a "confidence interval"
  - The difference in vote choice between registered Democrats and registered Republicans is *statistically significant*, in that it was unlikely to have occurred by chance

Frequentist Uncertainty

- What is a confidence interval?
- In straightforward terms, a margin of error around an estimate of a parameter which expresses uncertainty
- "Were this procedure to be repeated on numerous samples, the proportion of calculated 95% confidence intervals that encompassed the true value of the population parameter would tend toward 95%."
- We are aiming at an unknown, but constant, parameter. We'll rarely hit it straight on, but if we try many times, our 95% confidence interval will contain the truth 95% of the time. 5% of the time, we'll miss completely, and our confidence interval won't contain the truth.
  - Confidence intervals do not have to be set at 95%, but that is conventional
- A confidence interval should ***NOT*** be interpreted as "there is a 95% chance that this particular interval contains the true parameter"
  - We live in a frequentist world with constant but unknown parameters; a specific confidence interval either contains the truth or it doesn't

Confidence Interval Demonstration in R

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.

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

Draws with Replacement: Choose an object k times from a set of n objects when the order **does** matter and you **replace** the object each time

$$\binom{n}{k} = n^k$$

Review of Factorials

- $n! = n \times (n-1) \times \cdots \times 1$
- Example: $3! = 3 \times 2 \times 1 = 6$
- Meaning: possible combinations of how a dataset can be arranged
- 1,2,3 ; 1,3,2 ; 2,1,3 ; 2,3,1 ; 3,1,2 ; 3,2,1
    - First number in the combination has three possibilities
    - Second has two possibilities (after the first has been chosen)
    - Third has one possibility (after the first two have been chosen)
- $0! = 1$ (One possible way to arrange a dataset of zero numbers)
- $\frac{n!}{(n-k)!} = n \times (n-1) \cdots \times (n-k+1)$
    - $\frac{5!}{(5-3)!} = 5 \times 4 \times 3$
    - $\frac{5 \times 4 \times 3 \cancel{\times 2 \times 1}}{\cancel{2 \times 1}}$

Why do these formulas work?

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

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

- $n!$ gives you the total number of ways you can arrange all of your objects
- We use factorials because each time you select an object you have one fewer to choose from. If $n = 5$, the first object you draw has 5 possibilites, the second 4, and so on
- Dividing by $(n-k)!$ reduces the number of permutations based on how many objects you're choosing.
- We only need to consider the first k elements of the factorial in the numerator, because you are only choosing k objects from the set n

Why do these formulas work?

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

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

- The additional $k!$ in the denominator eliminates permutations that are simply re-orderings of other permutations
- Logic of this is that you can arrange a set of k numbers in $k!$ factorial ways
- So if we divide by $k!$, we're left with only one permutation of each original combination

Why do these formulas work?

Draws with Replacement: Choose an object k times from a set of n objects when the order **does** matter and you **replace** the object each time

$$\binom{n}{k} = n^k$$

- There are n possible options every time we draw
- We don't need factorials because replacement means our number of options isn't reduced
- If I have 5 objects, there are 5 possibilities for my first draw. **If I replace that object**, there are still 5 possibilities for my second draw.

Combinations Example:

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

- How many ways can I make a team of 3 out of 5 people?
- $\frac{5!}{3!(5-3)!}$
- $\frac{120}{6(2)}$
- 10

Combinations Example:

- $\frac{5!}{3!(5-3)!}$
- 5! is the number of ways we can order 5 people (120)
- Dividing by $(5-3)!$ reduces this numerator to $5 \times 4 \times 3$, because we only want combinations of three people: 60 permutations
- Additionally dividing by 3! eliminates duplicated (reordered) combinations
    - Example: I can arrange the numbers 3,4,5 in 3! (6) different ways: 3,4,5 ; 3,5,4 ; 4,3,5 ; 4,5,3 ; 5,3,4 ; 5,4,3
    - But these are not original combinations; we have six times as many permutations as unique combinations
    - Dividing by 3! (6) solves this: 10 unique combinations

Combinations Example:

A, B, C, D, E

1. A, B, C
2. A, B, D
3. A, B, E
4. A, C, D
5. A, C, E
6. A, D, E
7. B, C, D
8. B, C, E
9. B, D, E
10. C, D, E

Combinations 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?

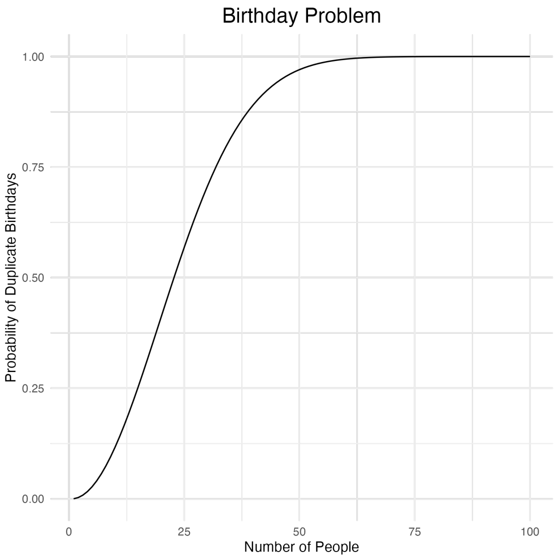
$$\binom{12}{3}\cdot\binom{9}{5}\cdot\binom{4}{3}\cdot\binom{1}{1} = \frac{12!}{3!9!}\cdot\frac{9!}{5!4!}\cdot\frac{4!}{3!1!}\cdot\frac{1!}{1!0!} = 220\cdot 125\cdot 4\cdot 1 = 110,880$$

Calculating probabilities

- Common problem: Given a known $Pr(A)$, what is the $Pr(A)$ over many attempts/trials/rolls/etc. if each attempt is independent?
- For example, I play a game of chance with a 10% chance of success. If I play 5 times, what is the probability that I win at least once?
- The probability of winning 5 times in a row is intuitive: $0.1 \times 0.1 \times 0.1 \times 0.1 \times 0.1$ or $(0.1)^5 = 0.00001$
- For the probability of winning *at least once*, we consider the probability of **losing** 5 times in a row, and take the inverse
    - $1-(1-0.1)^5 \approx 0.41$
    - After 10 tries: $1-(1-0.1)^10 \approx 0.65$

The Birthday Problem

- What is the probability that, in a set of n randomly chosen people, at least two will share a birthday?
- How many people are needed for that probability to exceed 50%?
- Think in terms of permutations: how many different combinations of 365 possible birthdays could k people have *without* any overlaps? Since we want unique birthdays, we don't consider "replacement".
  - $\frac{365!}{(365-k)!}$
- That gives us the number of permutations of *unique* birthdays (draws without replacement). How many combinations of birthdays are there in total, including ones with duplicates (draws with replacement)?
  - 365^k
- $\frac{365!}{(365-k)!} \div 365^k$ gives us the share of possible birthday combinations where all are unique
- $1 - (\frac{365!}{(365-k)!} \div 365^k)$ gives us the share of possible birthday combinations with at least one duplicate


Birthday Problem
Probability of Duplicate Birthdays
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Number of People

Random variables

- Distributions determine the probability that a **random variable** takes on any specific value
- 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

- Remember from the first day:
- Discrete: finite, countable
- Continuous: infinite, uncountable

Sample distribution

- **Sample distribution**: representation of the number of cases that take each value in a sample space
- **Frequency Distribution**: The number of observations/cases for each potential value of a variable
- **Relative Frequency Distribution**: a frequency distribution represented as the share of observations/cases

Contingency tables

- A matrix that shows the joint frequency distribution for two variables
- Also known as "cross-tabulations" or "cross-tabs"

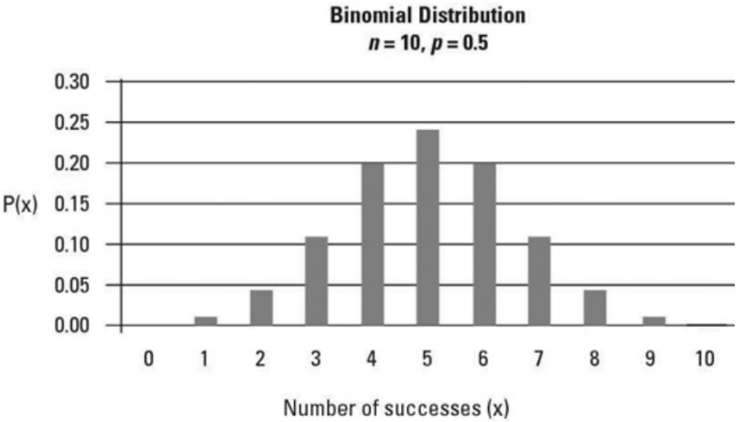
	Dog	Cat	Total
Male	42	10	52
Female	9	39	48
Total	51	49	100

Probability distribution

- The function that describes the likelihood of getting any specific value of a random variable
- Probability of any one value (or range of values) is between 0 and 1
- Sum of probabilities for the full range of values equals 1

Probability Mass Function

For discrete values, we use the Probability Mass Function (effectively a bar

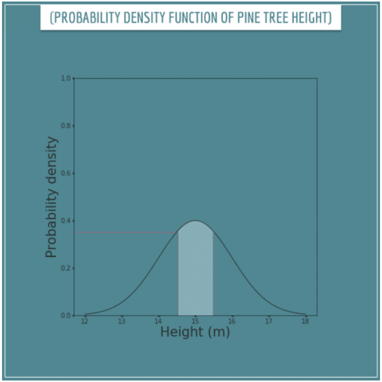


chart

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

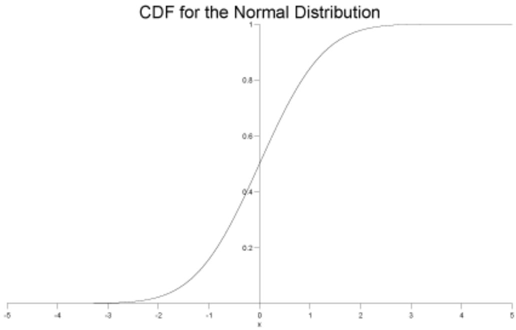
Probability Density Function

- For continuous variables, we use the Probability Density Function
- To find the likelihood that an observation will be between two given values, we look at the area under the curve (integration)



Cumulative Density Function

The integral, or antiderivative, of the Probability Density Function

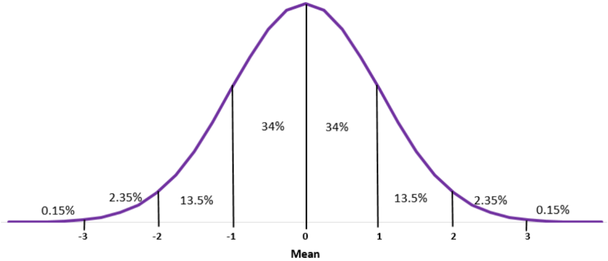


Normal/Gaussian Distribution (the bell curve)

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...