Probabilistic Methods (PM - 330725)

TOPIC 1: Foundations of Probability

Lecture 1

April 2025

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Politècnica Superior d'Enginyeria de Manresa

Roadmap

Learning Outcomes

By the end of this topic, you will be able to:

- Grasp foundational concepts like probability spaces, random variables, and distributions.
- Compute and interpret expectation and variance for various probability distributions.
- Use discrete and continuous distributions to model and analyze data.

Course Setup

Welcome to the course on Probabilistic Method, with a focus on its ties to Statistics and Machine Learning!

1. Instructor: Cristian Rodriguez Rivero
2. Syllabus of the Course: website
3. We will use ATENEA for announcements and discussions of this module

Prerequisites

Welcome to the course on Probabilistic Methods, with a focus on its ties to Statistics and Inference!

- no formal ones, but assume previous ML class, and comfort with
  - Ability to perform integration and differentiation (including multivariable calculus for advanced topics).
  - familiarity with sets, subsets, functions, mappings, and notations like unions, intersections, and complements.
  - descriptive statistics (mean, median, variance) and the concept of randomness.
  - exposure to basic probability ideas (e.g., discrete vs. continuous variables, probability rules).
  - have some context where probability is applied—such as machine learning, data science, engineering, or economics—can deepen understanding and relevance.

Evaluation

1. Homeworks (not graded but useful)
2. 1 Labs – 20% (LAB)
3. 2 Quizzes - 10% (QUIZ)
4. Final Written Exam - 40% (FIN)
5. Project: something useful/interesting with programming probabilistic methods. Groups of 2 and/or 3 – 30% (PRJ)

The final grade is calculated with the following weights:

$$\text{Overall grade} = 0.1 * \text{QUIZ} + 0.4 * \text{FIN} + 0.2 * \text{LAB} + 0.3 * \text{PRJ}$$

Readings

- required readings most weeks
  - Research articles, chapters, ranging from applied to theoretical
  - Required response to readings (short questions; fast) that you submit for your QUIZZES.

- Background videos (optional)
  - probability spaces, events, and axioms of probability
  - random variables (discrete and continuous) and their probability distributions.
  - expectation, variance, and other statistical measures.
  - Assume that you have watched these before the relevant lecture.

Projects

- This will be the most interesting part of class, and where you will learn the most
- Teams of 2-3 students
- real-world data!
- Two types of projects:
  - 8 projects proposed, some from different disciplines.
  - Your own design, using publicly available data the relevant lecture.

What is expected in this first topic

- Building a deep understanding of probability spaces, events, and axioms of probability.
- Introducing random variables (discrete and continuous) and their probability distributions.
- Demonstrating expectation, variance, and other statistical measures.
- Highlighting practical uses in data analysis and decision-making contexts.

Historical Context and Motivation

Probability theory has its origins in the study of games of chance, dating back to the 17th century. Mathematicians such as Blaise Pascal and Pierre de Fermat explored the mathematics of gambling problems, laying the groundwork for what would become a robust field underpinning much of modern science, finance, engineering, and beyond. In subsequent centuries, luminaries such as Jacob Bernoulli, Pierre-Simon Laplace, and Andrey Kolmogorov refined probability into a rigorous mathematical discipline.

Statistics, on the other hand, arose from the need to interpret data about populations—first for governance and economics, and eventually expanding into areas such as experimental design, psychometrics, biostatistics, and machine learning. The synergy between probability and statistics allows for predictive modeling and inference, turning raw data into actionable knowledge.

Why Study Probability and Statistics?

Probability and statistics provide the mathematical language and tools for dealing with uncertainty in diverse fields—from physics and engineering to finance, data science, biology, and social sciences.

- **Uncertainty Modeling:** many real-world processes have inherent randomness. Probability allows us to model, quantify, and predict outcomes.
- **Data Analysis:** statistics leverages probability to interpret data, build models, and make inferences about populations or processes.
- **Decision-Making:** by understanding risks, expectations, and distributions of possible outcomes, decision-makers can optimize strategies in uncertain conditions.

Relevance in Modern Data Analysis

For today's data-centric world, whether you're working in academic research, data science, or industry R&D, these foundational topics have ubiquitous relevance. You'll see them in everything from reliability engineering (e.g., exponential distribution for failure times) to A/B testing in tech companies (e.g., binomial models for success rates).

Today, probability is omnipresent in:

- Quantum mechanics (modeling inherently probabilistic events at the atomic scale).
- Financial models (stock price fluctuations, risk assessment).
- Machine learning (Bayesian inference, probabilistic graphical models).
- Public health and epidemiology (spread of diseases, success rates of treatments).

The Concept of Uncertainty

At its core, probability attempts to mathematically describe uncertainty. Events in the real world—like whether it will rain tomorrow, or the outcome of a random experiment—cannot always be predicted with certainty. Probability theory provides a framework to assign numerical values between 0 and 1 to the likelihood of these events, enabling rational decision-making even when outcomes are not guaranteed.

Real-World Examples of Uncertainty

- Weather Forecasting: A prediction stating there is a 70% chance of rain tomorrow.
- Financial Markets: The probability that a stock's price will exceed a certain threshold by next month.
- Manufacturing: The probability that a part produced on an assembly line is defective.

Probability Spaces, Events, and Axioms

A probability space is a triplet $(\Omega, \mathcal{F}, P)$:

- $(\Omega$ (the sample space) is the set of all possible outcomes of an experiment.
- $\mathcal{F}$ (the sigma-algebra of events) is a collection of subsets of Ω that includes the empty set $\emptyset$ and is closed under complement and countable union.
- P (the probability measure) is a function that assigns each event $A \in \mathcal{F}$ a real number $P(A)$ satisfying the Kolmogorov Axioms:

  1. Non-negativity: $P(A) \geq 0$ for all $A \in \mathcal{F}$
  2. Normalization: $P(\Omega) = 1$
  3. Countable Additivity: For any countable collection of pairwise disjoint events $A_1, A_,2, \ldots$
     $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

Example: A Simple Coin Toss

- Sample Space: $\Omega = H, T$
- Events: $\mathscr{F}$ might include $\emptyset$, H, T and Ω
- Probability Measure: If the coin is fair, $P(\{H\}) = P(\{T\}) = 0.5$

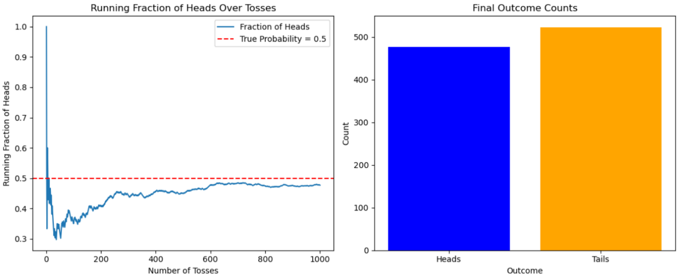
Execute the Python script in Examples_Lecture1 and observe the outputs for different values of $n(e.g., 100, 1000, 10, 000)$.
See how the empirical probabilities get closer to 0.5 for each outcome as n increases.

Questions:

What happens to the proportion of Heads as n becomes very large? Explain why the law of large numbers suggests that $\frac{Heads\_count}{n} \longrightarrow 0.5$. If the coin were biased (say, $P(H) = 0.3$, how would you modify the code?

Hint: Consider storing each outcome in a list to later compute statistics such as streaks or runs of heads/tails. Expand to multiple coins tossed at once and compare their distributions.

Example: A Simple Coin Toss



Homework

Experiment with different values of tosses (e.g., 100, 10000, 100000, 1000000) to see how the running fraction converges. Change the bias of the coin:

This exercise demonstrates both the simulation of a fair coin toss and how to visualize probability empirically in Python.

- If you want a biased coin with $P(H) = p \neq 0.5$, just replace if random.random() < 0.5: with if $random.random() < p$:.
- Observe how the final proportion converges to p rather than 0.5.

Class Example

- Let's say there are 100 students in the class
- Let's say 10 of them are industrious, 90 are Not industrious
- Probability of a randomly picked student being industrious
  - $P(industrious) = 0.1$
- We know that 70% of the industrious students got an A.
  - $P(a|industrious) = 0.7$
  - 7 industrious students got an A; 3 did not get an A.
- What is $P(industrious|a) = ?$
  - Depends on $P(a)$

We'll dive into this later!!!

Class Example

- Total students = 100
- Industrious students = 10 $\longrightarrow P(industrious) = 0.1.$
- Not industrious = 90 $\longrightarrow P(Notindustrious) = 0.9.$
- $P(a|industrious) = 0.7$
- We now assume that $P(a|notindustrious) = 0.3$

Apply the Rule of Total Probability:

$$P(a) = P(a|industrious) * P(industrious) + P(a|notindustrious) * P(Notindustrious)$$

$$P(a) = 0.7 * 0.1 + 0.3 * 0.9 = 0.34$$

Then

$$P(industrious|a) = \frac{P(a|industrious) * P(industrious)}{P(a)}$$

$$P(industrious|a) = \frac{0.7 * 0.1}{0.34} = 0.20$$

Introducing Conditional Probability

- A conditional probability is the chance that one thing happens, given that some other thing has already happened.
- A great example is a weather forecast: if you look outside this morning and see gathering clouds, you might assume that rain is likely and carry an umbrella.
- We express this judgment as a conditional probability: e.g. “the conditional probability of rain this afternoon, given clouds this morning, is 60%.”

Introducing Conditional Probability

In statistics we write this a bit more compactly:

- P(rain this afternoon | clouds this morning) = 0.6
- That vertical bar means "given"or çonditional upon."
- The thing on the left of the bar is the event we're interested in.
- The thing on the right of the bar is our knowledge, also called the "conditioning event" or "conditioning variable": what we believe or assume to be true.

$P(A|B)$: "the probability of A, given that B occurs."

A really important fact is that conditional probabilities are not symmetric:

$$P(A|B) \neq P(B|A)$$

Some technical points

1. the "non-baby"version of Kolmogorov's third axiom is actually something called countable additivity (versus "finite additivity".)
   - Consider a countable sequence of events $A_1, A_2, \ldots$, each $A_i \subset \Omega$
   - Suppose the events are all disjoint: $A_i \bigcap A_j = \emptyset$ for $i \neq j$.
   - we saw tht Countable additivity says that

   $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

   where

   $$\bigcup_{i=1}^{\infty} A_i = A_1 \bigcup A_2 \bigcup A_3 \bigcup \ldots$$

2. our definition of probability assumes that A is well defined for any $A_i \subset \Omega$. If is a finite or countable set (e.g. the integers), then is always well defined.

Some technical points

But if is uncountably infinite (e.g. the real numbers), then $P(A)$ is not necessary well defined for all possible subsets $A_i \subset \Omega$.

1. Wild but true! It is possible to define bizarre sets for which there is no meaningful notion of that set's size. (If you care: Google "nonmeasurable set"and "Banach-Tarski paradox").

2. But it's hard to construct such crazy sets. Every "normal" set you might care about (intervals, unions of intervals) has a well-defined size. Technically speaking, these are called the Borel sets.

Quick summary so far...

1. **Uncertain outcome/"random process":** we know the possibilities ahead of time, just not the specific one that occurs.
2. **Sample space:** the set of possible outcomes.
3. **Event** a subset of the sample space.
4. **Probability:** a function that maps events to real numbers and that obeys Kolmogorov's axioms.

OK, so how do we actually calculate probabilities?

The discrete uniform distribution and the counting rule

Suppose our sample space is a finite set consisting of N elements $\omega_1, \ldots, \omega_n$

Suppose further that $P(\omega_i) = \frac{1}{N}$ each outcome is equally likely, i.e. we have a discrete uniform distribution over possible outcomes.

Then for each set $A_i \subset \Omega$,

$$A = \frac{|A|}{N} = \frac{numofelementsinA}{numofelementsin\Omega}$$

That is, to compute $P(A)$, we just need to count how many elements are in A.

The counting rule: example 1

Let's use the counting rule.

- Our sample space has $N = \binom{52}{5} = 2,598,960$ possible poker hands, each one equally likely.

$$\binom{52}{5} = \frac{52!}{5!(52-5)!} = \frac{52 \times 51 \times 50 \times 49 \times 48}{5 \times 4 \times 3 \times 2 \times 1} = 2{,}598{,}960$$

- We only care about which cards you get, not the order. number of distinct 5-card hands is the number of combinations of 5 cards from 52.

- How many possible flushes are there? Let's start with hearts:
  - There are 13 hearts.
  - To make a flush with hearts, you need any 5 of these 13 cards.
  - Thus there are $\binom{13}{5} = 1287$ possible flushes with hearts.

- The same argument works for all four suits, so there are $4 \times 1287 = 5,148$ flushes. Thus

$$P(flush) = \frac{|A|}{\Omega} = \frac{5148}{2598960} = 0.00198079$$

Probability calculus

The "probability calculus"provides a set of rules for calculating probabilities. These aren't axioms: they can be derived from Kolmogorov's axioms.

1. $P(A^C) = 1 - P(A)$
   (Why? Because $A \bigcup A^C = \Omega$ and $P(\Omega) = 1$
2. if $A \subset B$, then $P(A) \leq P(b)$.
   (Why? Write B as $B = A \bigcup (B|A)$ and use finite additivity.)
3. $P(B|A) = P(B) - P(A \bigcap B)$
   (Why? try yourself as homework!)
4. Addition rule: $P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$
   (Why? try yourself as homework!)

Class Example 2

- Let's say there are 100 students in the class
- Let's say 10 of them are industrious, 90 are industrious
- Probability of a randomly picked student being industrious
  - $P(industrious) = 0.1$
- We know that 70% of the industrious students got an A.
  - $P(a|industrious) = 0.7$
  - 7 industrious students got an A; 3 did not get an A.
- What is $P(industrious|a) = ?$
  - Depends on $P(a)$?

Class Example 2

- Total students = 100
- Industrious students = 10 $\longrightarrow P(industrious) = 0.1$
- Not industrious = 90 $\longrightarrow P(Notindustrious) = 0.9$
- $P(a|industrious) = 0.7$
- We'll assume $P(a|Notindustrious) = 0.3$

Apply the Rule of Total Probability:

$$P(a) = P(a|industrious) \cdot P(industrious) + P(a|Notindustrious) \cdot P(Notindustrious)$$

$$P(a) = 0.7 \cdot 0.7 + 0.3 \cdot 0.9$$

$$P(a) = 0.34$$

Then we apply

$$P(a|industrious) = \frac{P(a|industrious) \cdot P(industrious)}{P(a)}$$

$$P(a|industrious) = \frac{0.7 * 0.1}{0.34}$$

$$P(a|industrious) = 0.2059$$

Exercise 1

In a province of Spain, 60% of the hospitalized patients are vaccinated.

- $P(a|h) = 0.6$

What does this number tell you about the effectiveness of the vaccines in preventing hospitalizations?

Demontrate using your intuition and knowledge!

TIPS: there might be a wrong conclusion some may draw:
Ïf most hospitalized people are vaccinated, the vaccine must not work."
This is incorrect because it ignores base rates - i.e., how many people in the population are vaccinated in the first place!!!

Exercise 2

Let's play!

Someone deals you a five-card poker hand. What is the probability of either a straight (five cards in a row, e.g. 3-4-5-6-7) or a flush (all cards the same suit)?

Note: these aren't mutually exclusive, since you might draw a hand that is both a straight AND a flush (e.g. 5-6-7-8-9 of clubs).

Exercise 2

If all 2,598,960 poker hands are equally likely, then using the counting rule:

- $P(\mathit{flush}) = 0.00198079$ (5,148 possible flushes)
- $P(\mathit{straight}) = 0.00392465$ (10,200 possible straights.)
- $P(\mathit{straight} \bigcap \mathit{flush}) = 0.0000153908$ (40 possible straight flushes)

So by the addition rule:

Note: these aren't mutually exclusive, since you might draw a hand that is both a straight AND a flush (e.g. 5-6-7-8-9 of clubs).

$$P(\mathit{straight} \bigcup \mathit{flush}) = P(\mathit{straight}) + P(\mathit{flush}) - P(\mathit{straight} \bigcap \mathit{flush}) = 0.00392465 + 0.00198079 - 0.0000153908 = 0.005890049$$

The multiplication rule

We've met Kolmogorov's three axioms, together with several rules we can derive from these axioms.

There's one final axiom for conditional probability, often called the multiplication rule. Let $P(A, B) = P(A \bigcap B)$ be the joint probability that both A and B happen. Then:

$$P(A|B) = \frac{P(A,B)}{P(B)}, P(B) > 0$$

Or equivalently:

$$P(A, B) = P(A|B) \cdot P(B)$$

This is an axiom: it cannot be proven from Kolmogorov's rules.

Consider the story of Abraham Wald and the "missing"WWII bombers.

Introducing Conditional probabilities from data

- B-17 bombers returning from their missions in WWII often had damage: on the fuselage, across the wings, on the engine block, and sometimes even near the cockpit.
- At some point, a clever data-minded person had the idea of analyzing the distribution of these hits over the surface of the returning planes.
- The thinking was that, if you could find patterns in where the B-17s were taking enemy fire, you could figure out where to reinforce them with extra armor, to improve survivability.
- You couldn't reinforce them everywhere, or they would be too heavy to fly.

Introducing Conditional probabilities from data

Suppose we saw data on returning bombers that looked like this:

Location	Number of planes
Engine	53
Cockpit area	65
Fuel system	96
Wings, fuselage, etc.	434

Naive answer: of 648 returning planes, 434 (68%) were hit on the fuselage.

$$P(\text{hit on wings or fuselage} \mid \text{returns home}) = 0.68$$

So bulk up the fuselage!

Introducing Conditional probabilities from data

But that's the right answer to the wrong question! We need the inverse probability

$$P(\text{ returns home} \mid \text{hit on wings or fuselage}) = 0.68$$

Wald used some fancy math to reconstruct the joint frequency distribution over damage type and mission result:

Location	Number of planes
Engine	53
Cockpit area	65
Fuel system	96
Wings, fuselage, etc.	434

Introducing Conditional probabilities from data

This gives us:

$$P(\text{ returns safely } | \text{ hit on wings or fuselage}) = \frac{434}{434+33} \approx 0.93$$

We would argue Pretty high!

On the other hand, of the 110 planes that had taken damage to the engine, only 53 only returned safely. Therefore

$$P(\text{ returns safely } | \text{ hit on engine}) = \frac{53}{53+57} \approx 0.48$$

Similarly,

$$P(\text{ returns safely } | \text{ hit on cockpit area}) = \frac{65}{65+46} \approx 0.59$$

Moral of the story: make sure you're focusing on the right conditional probability. And remember that $P(A|B) \neq P(B|A)$

Introducing Conditional probabilities from data

The same math that Abraham Wald used to analyze bullet holes on B-17s also underpins the modern digital economy of films, television, music, and social media.

- Netflix, Hulu, and other video-streaming services all use this same math to examine what shows their users are watching, and apply the results of their number-crunching to recommend new shows.
- Many companies do the same:
  - ** Amazon for products
  - ** Google for web pages
  - ** etc., etc.,..

- **Conditional Probability:** For events A and B with $P(B) \geq 0$

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

- **Independence:** Events A and B are said to be independent if

$$P(A \bigcap B) = P(A)P(B)$$

Intuitively, knowing that B occurred provides no information about whether A occurred.

The rule of total probability

Consider the following data on complication rates at a maternity hospital in Cambridge, England:

	Easier deliveries	Harder deliveries	Overall
Senior doctors	0.052	0.127	0.076
Junior doctors	0.067	0.155	0.072

Would you rather have a junior or senior doctor? Simpson's paradox. Senior doctors have:

1. lower complication rates for easy cases.
2. lower complication rates for hard cases.
3. higher complication rates overall! (7.6% versus 7.2%.) **Why?**

The rule of total probability

Let's see the table with number of deliveries performed (in parentheses):

	Easier deliveries	**Harder deliveries**	**Overall**
Senior doctors	0.052	0.127	0.076
Junior doctors	0.067	0.155	0.072

Would you rather have a junior or senior doctor? Now we see what's going on:

1. Most of the deliveries performed by junior doctors are easier cases, where complication rates are lower overall.
2. The senior doctors, meanwhile, work a much higher fraction of the harder cases.

The rule of total probability

1. It turns out the math of Simpson's paradox can be understood a lot more deeply in terms of something called the rule of total probability, or the mixture rule.

2. This rule sounds fancy, but is actually quite simple. It says to divide

   and conquer: **the probability of any event is the sum of the probabilities for all the different ways in which the event can happen**.

3. Really just Kolmogorov's third rule in disguise!

The rule of total probability: example 3

Let's see this rule in action for the **hospital data**.

There are two types of deliveries: easy and hard. So:

$$P(complication) = P(easy \& complication) + P(hard \& complication)$$

Now use the rule for conditional probabilities to each joint probability on the right-hand side:

$$P(complication) = P(easy) \cdot P(complication|easy) + P(hard) \cdot P(complication|hard)$$

The rule of total probability says that overall probability is a weighted

average -a mixture - of the two conditional probabilities

The rule of total probability

For senior doctors we get

$$P(complication) = \frac{213}{315} \cdot 0.052 + \frac{102}{315} \cdot 0.027 = 0.076$$

And for junior doctors, we get

$$P(complication) = \frac{3169}{3375} \cdot 0.067 + \frac{206}{3375} \cdot 0.155 = 0.072$$

This is a lower **marginal** or **overall** probability of a complication, even though junior doctors have higher conditional probabilities of a complication in all scenarios.

Synonyms: **overall probability = total probability = marginal probability**

The rule of total probability

You can see why these are called marginal probabilities if you go back to the Abraham Wald example:

	Returned	Shot down
Engine	53	57
Cockpit area	65	46
Fuel system	96	16
Wings, fuselage, etc.	434	33

Original table

	Returned	Shot down
Engine	0.066	0.071
Cockpit area	0.081	0.058
Fuel system	0.120	0.020
Wings, fuselage, etc.	0.542	0.042

Divide by the total number (800) to turn these into joint probabilities...

	Returned	Shot down	Marginal
Engine	0.066	0.071	0.137
Cockpit area	0.081	0.058	0.139
Fuel system	0.120	0.020	0.140
Wings, fuselage, etc.	0.542	0.042	0.584
Marginal	0.809	0.191	1

Now calculate the overall probabilities for each individual type of event, and put those in the margins of the table.

The rule of total probability

Here's the formal statement of the rule. Let be any sample space, and let be a partition of —that is, a set of events such that:

$$P(B_i, B_j) = 0 \text{ for any } i \neq j \text{, and } \sum_N^{i=1} P(B_i) = 1$$

Now consider any event A. Then

$$P(A) = \sum_N^{i=1} P(A, B_i) = \sum_N^{i=1} P(B_i) \cdot P(A|B_i)$$

Independence

Two events A and B are independent if

$$P(A|B) = P(A|notB) = P(A)$$

In words: A and B convey no information about each other:

1. P(flip heads second time | flip heads first time) = P(flip heads second time)
2. P(stock market up | bird poops on your car) = P(stock market up)
3. P(God exists | Longhorns win title) = P(God exists)

So if A and B are independent, then $P(A, B) = P(A) \cdot P(B)$

Independence

Two events A and B are conditionally independent, given C, if

$$P(A, B|C) = P(A|C) = P(A|C) \cdot P(B|C)$$

In words: A and B convey no information about each other once we know $C : P(A, B|C) = P(A|C)$

Neither independence nor conditional independence implies the other.

1. It is possible for two outcomes to be dependent and yet conditionally independent.

2. Less intuitively, it is possible for two outcomes to be independent and yet conditionally dependent.

Independence

Let's see an example. Alice and Brianna live next door to each other and both commute to work on the same metro line.

- A = Alice is late for work.
- B = Brianna is late for work.

A and B are are dependent: if Brianna is late for work, we might infer that the metro line was delayed or that their neighborhood had bad weather. This means Alice is more likely to be late for work:

$$P(A|B) \geq P(A)$$

Independence

Now let's add some additional information:

- A = Alice is late for work.
- B = Brianna is late for work.
- C = The metro is running on time and the weather is clear.

A and B are conditionally dependent given C. If Brianna is late for work but we know that the metro is running on time and the weather is clear, then we don't really learn anything about Alice's commute:

$$P(A|B, C) \geq P(A|C)$$

Independence

Again, let's add some additional information.

- A = Alice has blue eyes.
- B = Brianna has blue eyes.

A and B are dependent given C. Alice's eye color can't give us information about Brianna's.

Independence

Again, let's add some additional information.

- A = Alice has blue eyes.
- B = Brianna has blue eyes.
- C = Alice and Brianna are sisters.

A and B are conditionally dependent given C. if Alice blue eyes, and we know that Brianna is her sister, then we know something about Brianna's genes. It is now more likely that Brianna has blue eyes.

Checking independence from data

Suppose we have two random outcomes A and B and we want to know if they're independent or not.

Solution:

- Check whether B happening seems to change the probability of A happening.
- That is, verify using data whether $P(A|B) = P(A|notB) = P(A)$.
- These probabilities won't be exactly alike because of statistical fluctuations, especially with small samples.
- But with enough data they should be pretty close if A and B are independent.

When independence goes wrong!!!

It can sound a bit silly, right?

But you'd be surprised at how often people make this mistake! We might call this the "fallacy of mistaken compounding": assuming events are independent and naively multiplying their probabilities.

Out of class, I'm asking you to read two short pieces that illustrate this unfortunate reality:

- How likely is it that birth control could let you down? from the New York Times newspaper.

Wrapping up

Today we discussed:

- Probability and statistics provide the mathematical language and tools for dealing with uncertainty in diverse fields—from physics and engineering to finance, data science, biology, and social sciences.
- It involves:
  1. Uncertainty Modeling
  2. Many real-world processes have inherent randomness.
  3. Probability allows us to model, quantify, and predict outcomes.

References

Online Tutorials, Courses, and other books

- Chapte 1 - Ross, S. Introduction to probability models. 13th edition. Amsterdam: Academic Press, 2023. ISBN 9780443187612.
- MIT OpenCourseWare:
  Probabilistic Systems Analysis and Applied Probability
- Other books:
  All of Statistics: A Concise Course in Statistical Inference

Thank you very much!

ANY QUESTIONS OR COMMENTS?