

Sprint 2:

Probability & Discrete Probability Distributions

Experiment, Event, Sample space, Probability, Counting rules, Conditional probability, Bayes's rule, random variables, moments generator function



Agenda

Probability: different definitions

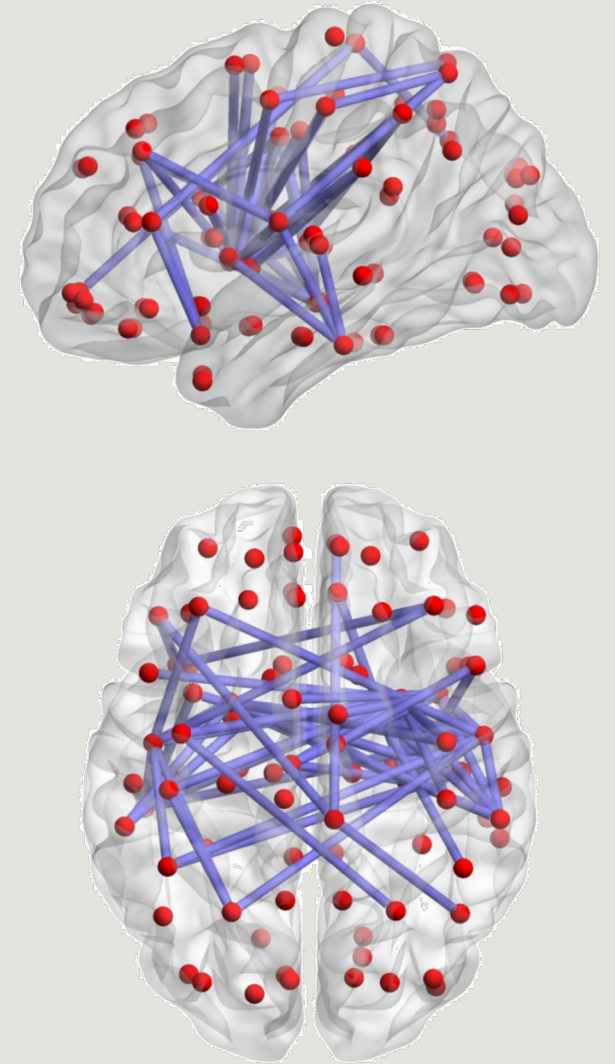
Experiments & Events

Conditional Probabilities

Bayes' Theorem

Random Variables

Probability Distributions



This Photo by Unknown Author is licensed under CC BY-SA

Why Learn Probability?

- **Probability** is a way to describe **how likely** something is to happen.
It's a measure of uncertainty — a number between **0** and **1**, where:
 - **0** means the event **cannot happen** (impossible),
 - **1** means the event **will definitely happen** (certain),
 - and numbers in between (like 0.3 or 0.75) show *degrees of likelihood*.
- For example:
 - The probability of flipping heads on a fair coin is **0.5** (50%).
 - The probability of rolling a 6 on a fair die is **$1/6 \approx 0.167$** (about 17%).

Probability: The Classical Definition (When All Outcomes Are Equally Likely)

- If an experiment has a fixed number of equally likely outcomes, then:

$$P(\text{event}) = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

- **Example:**

- A Fair die has 6 outcomes.
- The event "rolling an even number" has 3 favorable outcomes (2, 4, 6).

$$P(\text{even}) = \frac{3}{6}$$

Probability: The Frequentist View (Based on Observations)

- If you repeat an experiment many times, the **relative frequency** of an event A

$$\text{Relative Frequency} = \frac{\text{Number of times event occurs}}{\text{Total number of trials}}$$

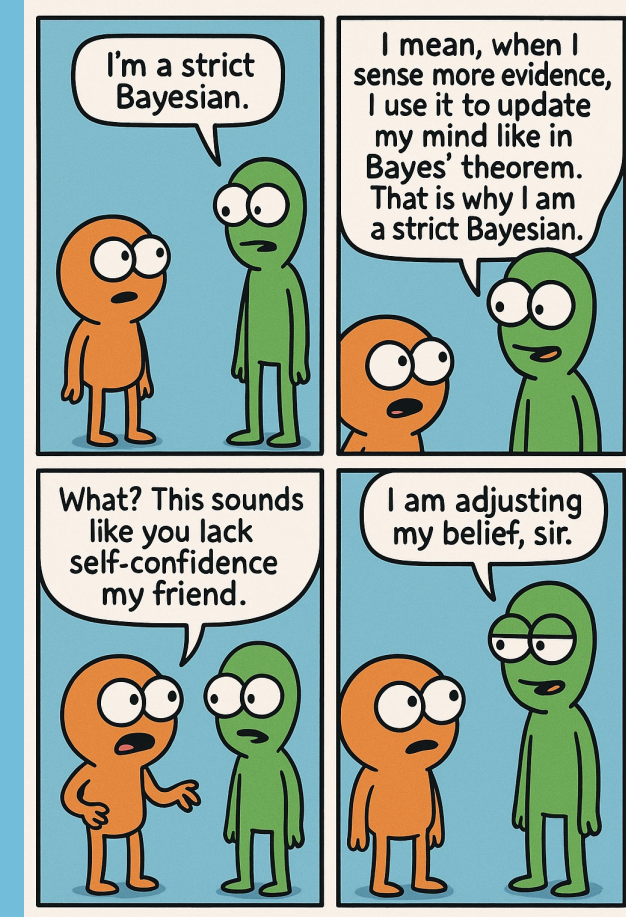
- As the number of trials n grows large ($n \rightarrow \infty$), this relative frequency *converges* to the real probability — that's the **Law of Large Numbers**.
- Therefore, we assume that relative frequency approximates the true probability

$$p(A) = \lim_{n \rightarrow \infty} \frac{f}{n}$$

Probability: The Bayesian View (Probability as Belief)

Sometimes we talk about probability when we don't repeat experiments, like:

- “What’s the probability that this satellite component will fail in orbit?”
- Here, probability expresses **degree of belief** — how confident we are in a statement, given what we know which could be just a rough estimate.
- This is the **Bayesian interpretation** — it uses probability to quantify uncertainty in our knowledge.



AI generated image

Probability goes from theory → data

You know something about the population, so you predict what data should look like.

- Suppose I know exactly the failure rates of different components in a high-integrity avionics system during a mission is 0.001 (that's 1 in 1,000).
- Then you can use **probability** to predict or simulate outcomes.
 - “What’s the probability that in a batch of 1000 flights, at least one failure occurs?”
 - “What’s the expected number of failures over 10,000 flights?”
- This is forward reasoning — from the known model to possible data.
- This is probabilistic reasoning as I know the population (failure rates) and predict the performance of a specific system

Example: Statistical Reasoning — Inferring from Observed Data

- Suppose you *don't know* the component's true failure probability.
- You can monitor the performance of a random sample of avionics systems over time and analyze the data to estimate the failure rates of the components.

Example:

- You collect data from 10,000 flights and see that 8 failures occurred. From this **sample**, you can estimate the underlying (population) probability:

$$\hat{p} = \frac{8}{10,000} = 0.0008$$

This is statistical reasoning as I infer the population characteristics (failure rates) from sample observations. **You can also** quantify uncertainty (for example, build a confidence interval). That's **statistical inference** — going from data back to the model.

Of course, the larger your sample is the more certain you can be about your estimate.

Why Learn Probability?

Probability underpins:

- **Statistics:** Inferring population characteristics from samples.
- **Risk Analysis:** Estimating the likelihood and cost of system failures or accidents.
- **Machine Learning:** Modeling uncertainty and predicting outcomes.
- **High-Integrity Systems:** Measuring reliability and safety margins.

Basic terminology: Experiment and Event

- An **Experiment** is the process through which observations (or measurements) are obtained.
- An **event** is an outcome of an experiment, typically denoted by a capital letter.
- **Probability** is applied to events, assessing the likelihood of a specific event occurring.
- **Experiment: System Integrity Check in a Secure Database**
 - Event A: The system passes all security compliance tests.
 - Event B: A data breach is detected during the check.
- **Experiment: Monitoring Server Uptime**
 - Event A: The server remains operational without downtime for 24 hours.
 - Event B: The server experiences several unexpected downtime within the 24-hour period

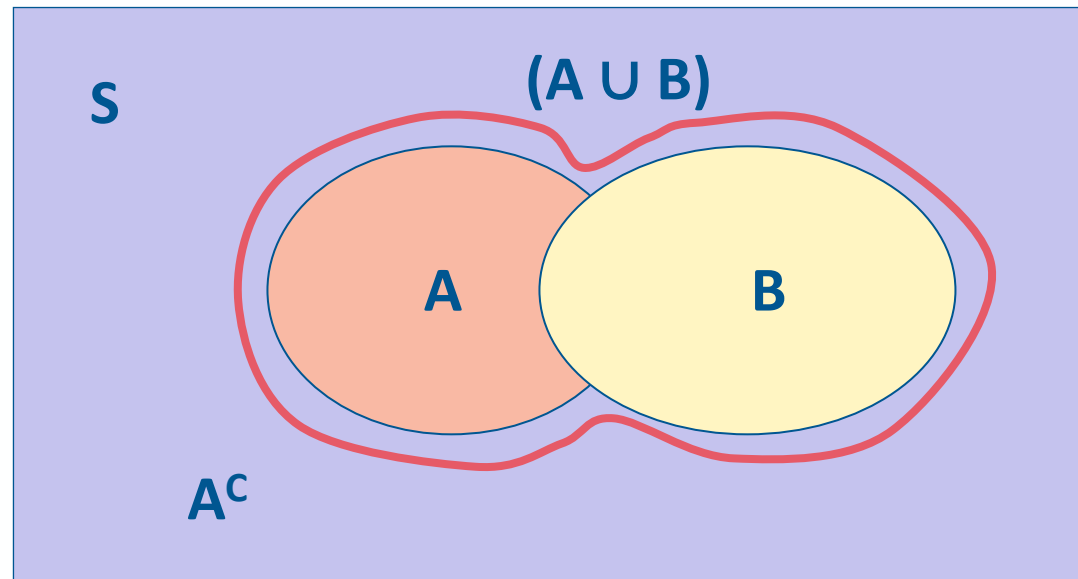
Basic terminology: Sample Space and Simple Events

- A **simple event** is a single, specific outcome from a random experiment. It's the most basic type of event because it cannot be broken down into a combination of other events.
- For example, rolling a "3" on a standard six-sided die is a simple event because it is only one possible outcome out of the six possibilities.
- An **event** can be a collection of one or more simple events.
 - A: an even number
 - B: a number < 5
- All the possible outcomes of an experiment are called **sample space**, S .

Event Relations: Union of events

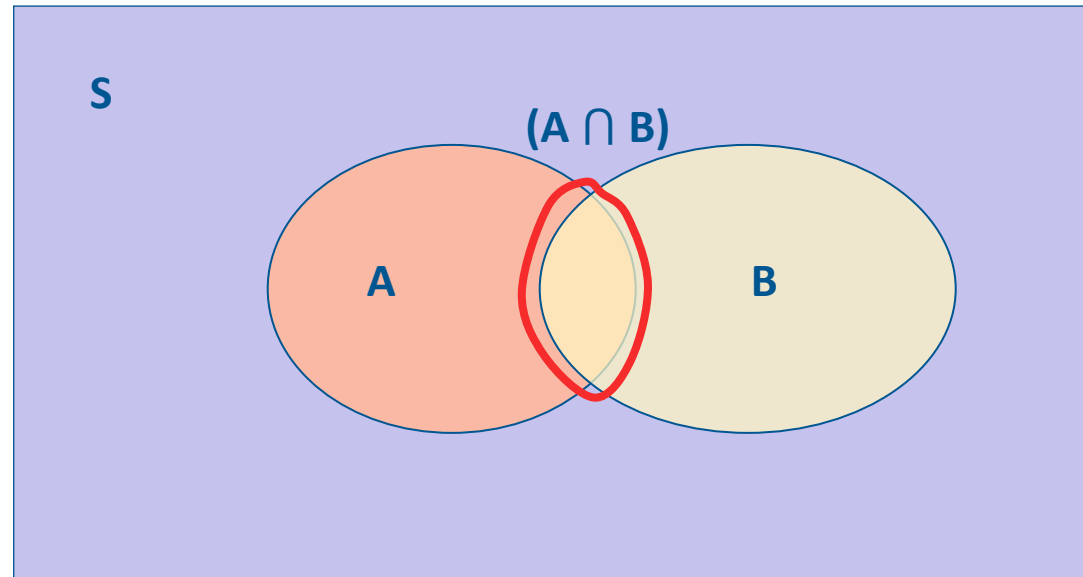
- The **union** of two events, **A** and **B**, is the event that either **A or B or both** occur when the experiment is performed.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Event Relations: complementary event

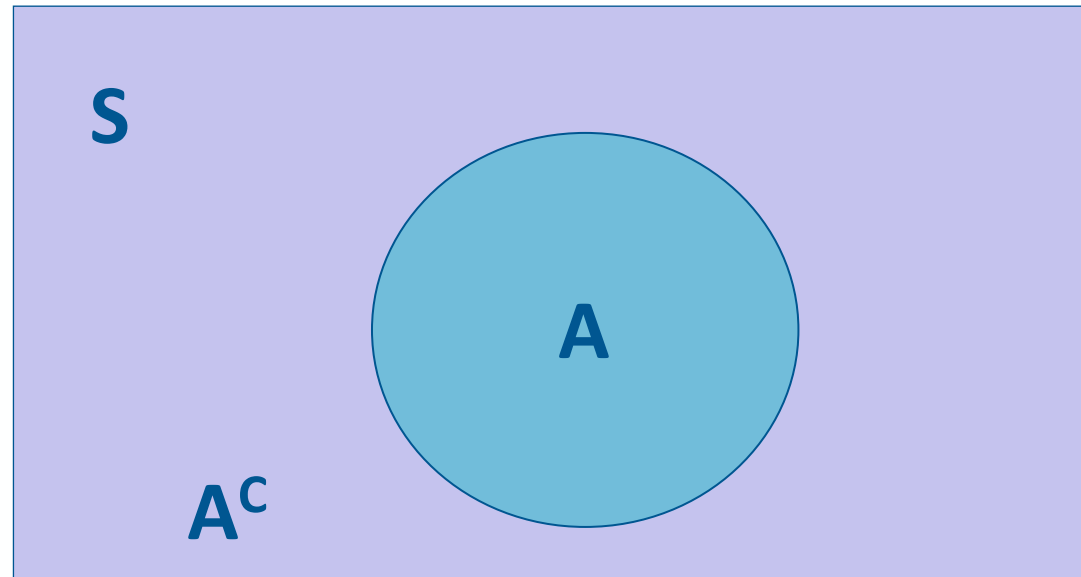
- The **intersection** of two events, **A** and **B**, is the event that both A **and** B occur when the experiment is performed. We write $A \cap B$.



Event Relations: complementary event

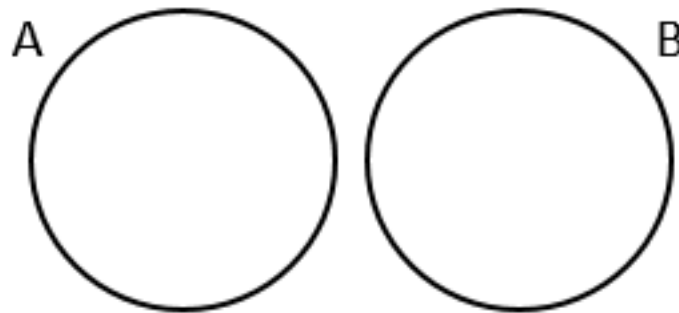
- The **complement** of an event **A** consists of all outcomes of the experiment that do not result in event **A**.
- We write **A^c** .

$$P(A^c) = 1 - P(A)$$



Event Relations: Mutually exclusive events

- Two events are mutually exclusive if, when one event occurs, the other cannot, and vice versa.
- **Occurrence of one event will result in the non-occurrence of the other.**
- Let A, B be mutually exclusive events: $P(A \cap B) = 0$ & $P(A \cup B) = P(A) + P(B)$



$$P(A \text{ or } B) = P(A) + P(B)$$

Event Relations: Mutually exclusive events

- Two events are mutually exclusive if, when one event occurs, the other cannot, and vice versa.
- **Occurrence of one event will result in the non-occurrence of the other.**
- Let A, B be mutually exclusive events: $P(A \cap B) = 0$ & $P(A \cup B) = P(A) + P(B)$

Experiment: stop a student at FraUAS

–A: observe an HIS student

–B: observe an Inclusive Design student

The probability of A and B is the 0 (it is impossible for a student to enroll in both programs):

$$P(A \text{ and } B) = P(\text{HIS} \cap \text{ID}) = 0$$

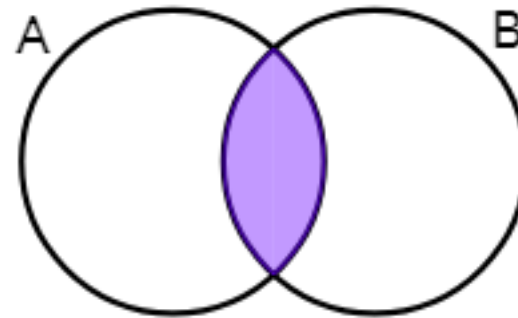
The probability of A or B is the sum of the individual probabilities:

$$P(A \text{ or } B) = P(\text{HIS} \cup \text{ID}) = P(\text{HIS}) + P(\text{ID})$$

Event Relations: Not Mutually exclusive events

- Not mutually exclusive if, when one event occurs, the other can still occur.

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$$



$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

Not Mutually exclusive events

- Not mutually exclusive if, when one event occurs, the other can still occur.

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \text{ and } B)$$

- Altogether there are **180 HIS students**.
- Prof. Marouf supervises **17 people on “Project HIS”**
- He teaches **170 students “IDA”**.
- How many students study **both IDA and Project HIS** under the supervision of Prof. Marouf?

Quiz

How many students study **both IDA and Project HIS** under the supervision of Prof. Marouf?

- Altogether there are **180 HIS students**.
- Prof. Marouf supervises **17 people on “Project HIS”**
- He teaches **170 students “IDA”**.

- a. **7**
- b. **10**
- c. **18**

Quiz

- How many students study **both IDA and Project HIS** under the supervision of Prof. Marouf?
 - Altogether there are **180 HIS students**.
 - Prof. Marouf supervises **17 people on “Project HIS”**
 - He teaches **170 students “IDA”**.

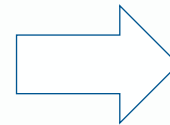
This is a case of **not** Mutually Exclusive (a student can study IDA **and** project).

Let's say **b** is how many students both subjects:

- people studying IDA Only must be $170-b$
- people studying project Only must be $17-b$

And we know there are **180** people, so:

$$(170-b) + b + (17-b) = 180 \rightarrow b = 7$$



So, we know all this now:

$$P(\text{IDA}) = 170/180$$

$$P(\text{project}) = 17/180$$

$$P(\text{IDA Only}) = 163/180$$

$$P(\text{project Only}) = 10/180$$

$$P(\text{IDA or project}) = 180/180 = 1$$

$$P(\text{IDA and project}) = 7/180$$

Lastly, let's check with our formula:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

$$180/180 = 170/180 + 17/180 - 7/180$$

Scenario: Aircraft Flight Control System Failure

System Failure (Event A): Occurs if at least one of the following software modules fails. These modules depend on each other and failure in one module increase the chance of failure in other modules.

- **Event S1:** Failure of the autopilot software. $P(S1)=0.005$
- **Event S2:** Failure of the navigation software. $P(S2)=0.007$
- **Event S3:** Failure of the sensor data processing software. $P(S3)=0.004$
- **Calculate the the probability of system failure:**

$$P(A)=\dots\dots\dots$$

Scenario: Aircraft Flight Control System Failure

System Failure (Event A): Occurs if at least one of the following software modules fails. These modules depend on each other and failure in one module increase the chance of failure in other modules.

- **Event S1:** Failure of the autopilot software. $P(S1)=0.005$
- **Event S2:** Failure of the navigation software. $P(S2)=0.007$
- **Event S3:** Failure of the sensor data processing software. $P(S3)=0.004$

Actually, we need more information to be able to calculate the probability. Assume:

- $P(S1 \text{ and } S2)=0.002$
- $P(S1 \text{ and } S3)=0.001$
- $P(S2 \text{ and } S3)=0.001$
- $P(S1 \text{ and } S2 \text{ and } S3)=0.0001$

So, probability of system failure:

$$P(A)=P(S1)+P(S2)+P(S3)-P(S1 \cap S2)-P(S1 \cap S3)-P(S2 \cap S3)+P(S1 \cap S2 \cap S3)$$

$$P(A)=(0.005+0.007+0.004)-(0.002+0.001+0.001)+(0.0001) = 0.0121$$

independent events

- Two events, **A** and **B**, are said to be **independent** if the occurrence or nonoccurrence of one of the events does not change the probability of the occurrence of another event.
- Occurrence of one event will have no influence on the occurrence of the other.
- The probability that either event **A** or **B** occurs is given by:

$$P(\text{A and B}) = P(A \cap B) = P(A) P(B)$$

The probability of A or B is :

$$P(\text{A or B}) = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A) P(B)$$

Quiz: which of the following statements is true?

1. Mutually exclusive events **cannot be independent** unless one of them has probability 0
2. Mutually exclusive events **cannot be independent**
3. Mutually exclusive events **are by default independent**

Conceptually, mutually exclusive events are dependent. If one event occurs, the other cannot.

The probability of both events occurring together is zero ($P(A \cap B) = 0$). However, if both events have a non-zero probability individually, then $P(A) * P(B)$ would be greater than zero, contradicting the condition for independence, where $P(A \cap B)$ should equal $P(A) * P(B)$.

The only scenario where mutually exclusive events can be considered independent is if at least one of the events has a probability of zero. In such cases, $P(A) * P(B) = 0$, which satisfies the condition for independence. Therefore, mutually exclusive events cannot be independent unless one of them has a probability of zero.

Quiz

- In a certain factory, 10% of the products are known to have some minor production defects and needs manual improvements. Three products were randomly selected from on Monday. What is the probability that exactly one of the three has a defect?
- a. 0.1%
 - b. 0.234
 - c. 0.333

Define D: Defect N: No defects

It is either the first one or second one or third one that has defects

$$\begin{aligned} P(\text{exactly one has defect}) &= P(DNN) + P(NDN) + P(NND) \\ &= P(D)P(N)P(N) + P(N)P(D)P(N) + P(N)P(N)P(D) \\ &= (.1)(.9)(.9) + (.9)(.1)(.9) + (.9)(.9)(.1) = 3(.1)(.9)^2 = .243 \end{aligned}$$

Example Scenario: Aircraft Flight Control System Failure

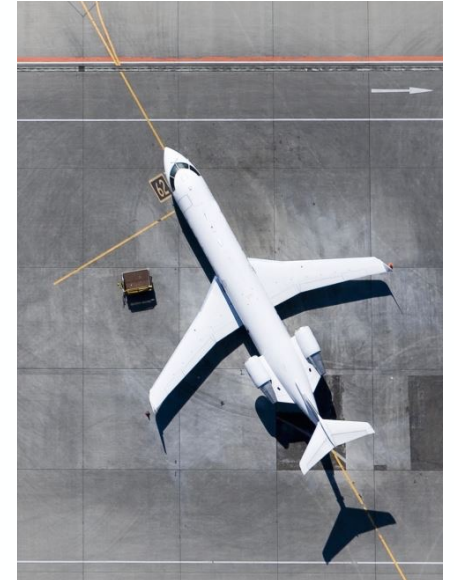
- Consider a high-integrity system such as an aircraft's flight control system. Suppose we are interested in the probability of a system failure event (Event A) due to software faults.

The Failure scenarios are independent from each and they can be:

- Failure of the autopilot software (Event S1)
 - Failure of the navigation software (Event S2)
 - Failure of the sensor data processing software (Event S3)
- The probability of system failure ($P(A)$) is the sum of the probabilities of these simple events:

$$P(A) = P(S1 \cup S2 \cup S3) = P(S1) + P(S2) + P(S3) - P(S1 \cap S2) - P(S1 \cap S3) - P(S2 \cap S3) + P(S1 \cap S2 \cap S3)$$

$$P(A) = P(S1) + P(S2) + P(S3) - P(S1) \times P(S2) - P(S1) \times P(S3) - P(S2) \times P(S3) + P(S1) \times P(S2) \times P(S3)$$



Conditional Probabilities

- Conditional probability measures the likelihood of event A occurring given that event B has occurred.

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) \neq 0$$

“given”



AI generated comic

Example: Conditional Probabilities

- Consider a flight control system with two components:
 - Event A: Failure of the autopilot system
 - Event B: degradation of the sensor system
- Suppose:
 - $P(A \cap B) = 0.0001$ (degradation and failure together)
 - $P(B) = 0.002$ (sensor degradation)
- Then:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{0.0001}{0.002} = 0.05$$

Interpretation: given that the sensor has failed, there is a 5% chance that the autopilot will also fail or the probability of sensor failure increases from 0.2% to 5%.

What about independent events

We can redefine independence in terms of conditional probabilities:

Two events A and B are **independent** if and only if

$$P(A|B) = P(A) \quad \text{or} \quad P(B|A) = P(B)$$

$$P(A \cap B) = P(A) P(B \text{ given that } A \text{ occurred}) = P(A)P(B|A) = P(A) P(B)$$

Otherwise, they are **dependent**.

Quiz

- In a factory, 49% of the products come from a new, fully automated production line, while the remaining 51% come from the old production line. Products from the new line have a defect rate of 8%, and those from the old line have a defect rate of 10%.
- If one product is randomly selected by the quality team, what is the probability that it **is defective and comes from the new production line?**

Define D: Defect NL: New production line

From the example, $P(NL) = .49$ and $P(D | NL) = .08$.

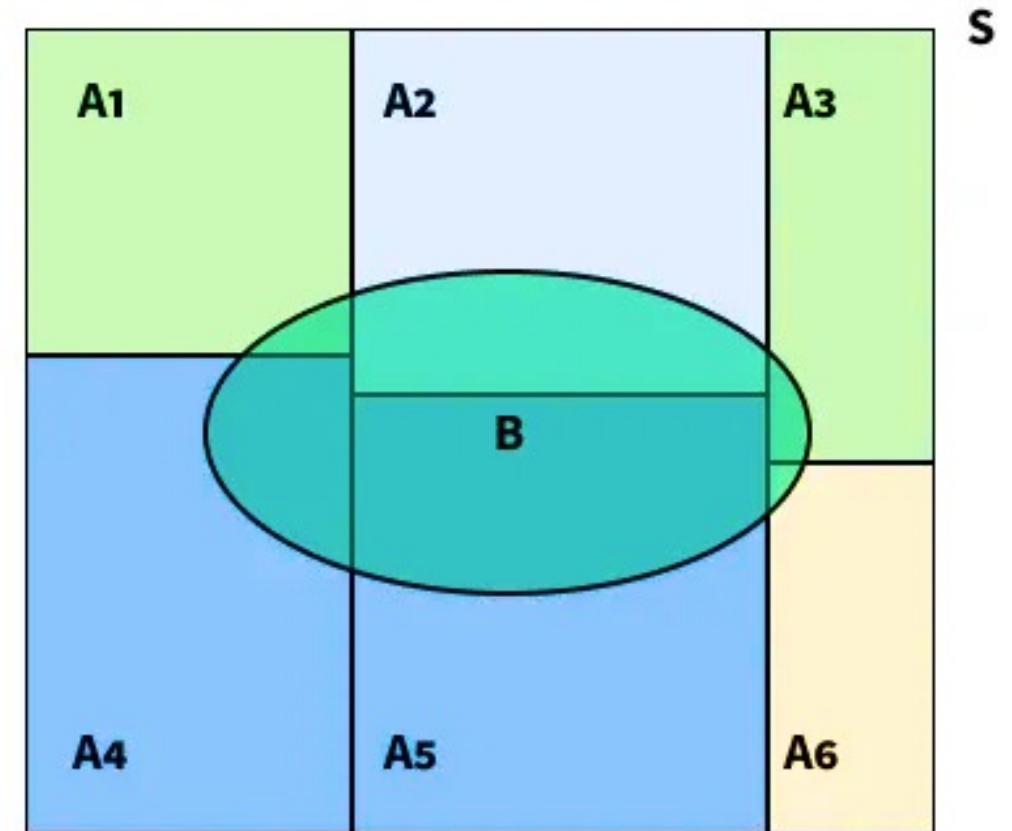
Use the Multiplicative Rule to calculate the joint probability:

$$P(D \text{ and new line}) = P(D \text{ and NL}) = P(NL)P(D | NL) = .49(.08) = .0392$$

The Law of Total Probability

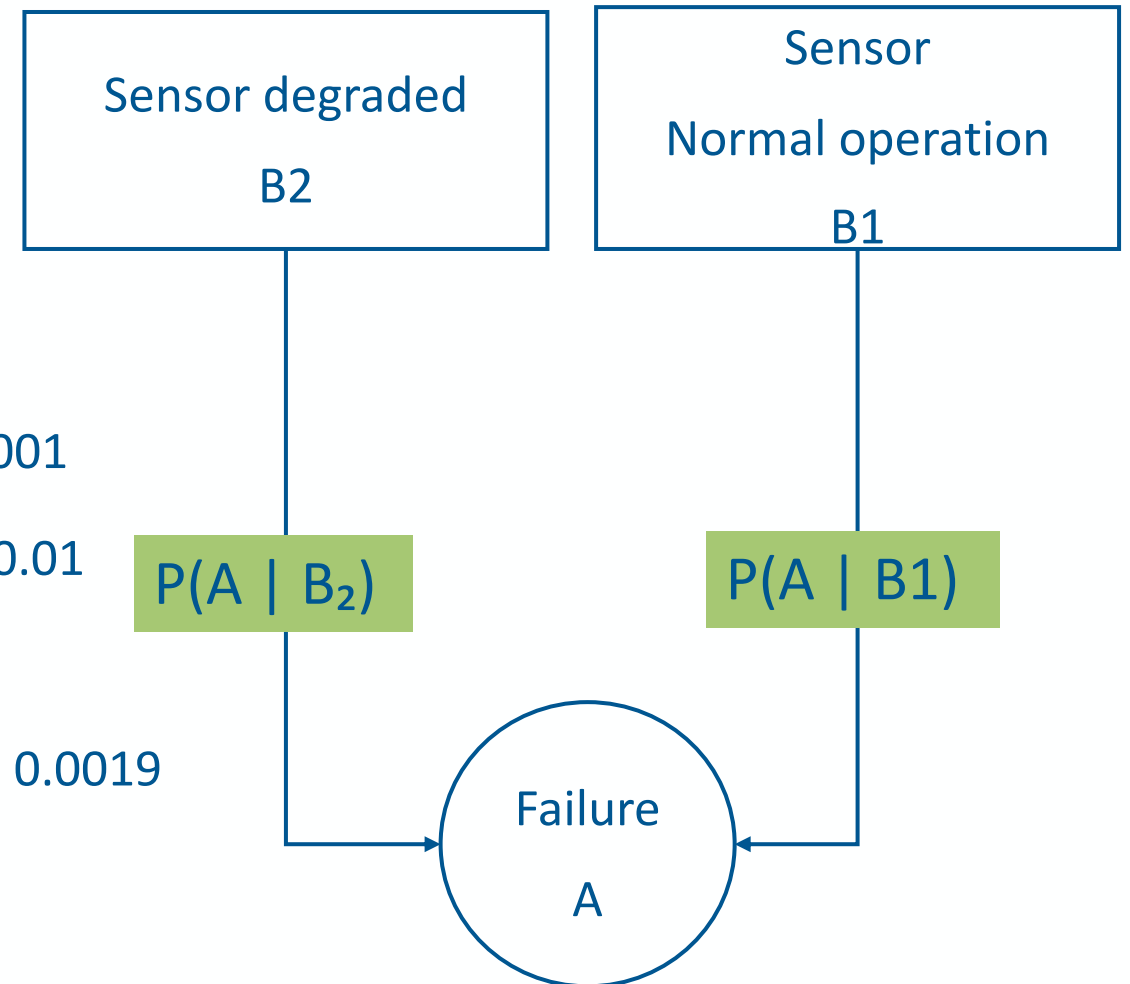
- The Law of Total Probability is used to find the probability of an event **B** by considering all possible ways **B** can happen through a partition of the sample space.
- Let $A_1, A_2, A_3, \dots, A_k$ be **mutually exclusive and exhaustive** events (that is, one and only one must happen). Then the probability of any event **B** can be written as:

$$\begin{aligned} P(B) &= P(B \cap A_1) + P(B \cap A_2) + \dots + P(B \cap A_k) \\ &= P(A_1)P(B | A_1) + P(A_2)P(B | A_2) + \dots + P(A_k)P(B | A_k) \end{aligned}$$



Example: The Law of Total Probability

- Let A = 'System Failure'
- Partition the system into:
 - B_1 : Sensor normal Operation ($P(B_1) = 0.9$)
 - $B_2 = B_1'$: Sensor degraded ($P(B_2) = 0.1$)
- Given:
 - System failure upon normal operation $P(A | B_1) = 0.001$
 - System failure upon sensor degradation $P(A | B_2) = 0.01$
- Then:
 - $P(A) = (0.001 \times 0.9) + (0.01 \times 0.1) = 0.0009 + 0.001 = 0.0019$
 - → Total system failure probability is 0.0019



General Form of Bayes' Theorem

- Bayes' Rule:

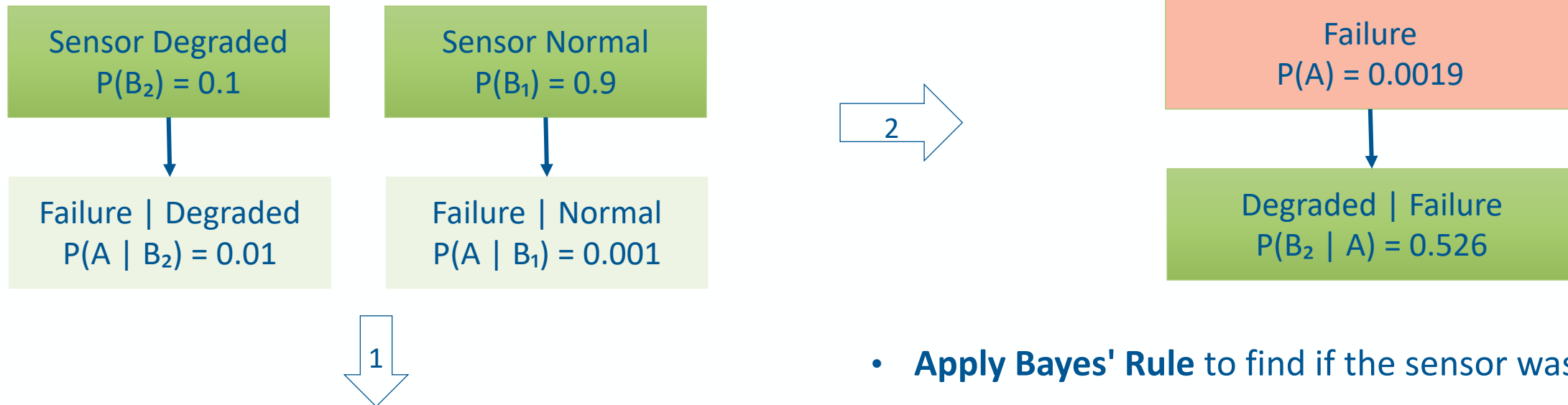
$$P(B|A) = \frac{P(A | B) \times P(B)}{P(A)}$$

- Let $B_1, B_2, B_3, \dots, B_k$ be mutually exclusive and exhaustive events with prior probabilities $P(B_1), P(B_2), \dots, P(B_k)$. If an event A occurs, the posterior probability of B_i given that A occurred is:

$$P(B_i|A) = \frac{P(A | B_i) \times P(B_i)}{\sum_i P(A | B_j) \times P(B_j)}$$

Example: Bayes' Theorem

Bayes' Rule helps us estimate: $P(\text{Sensor Degraded} \mid \text{System Failure})$ in the previous example. In other words, the system has failed, is it because of the sensor?



Probability of failure **using the law of total probability**

$$P(A) = 0.01 \times 0.1 + 0.001 \times 0.9 = 0.0019$$

- **Apply Bayes' Rule** to find if the sensor was the culprit

$$P(B_2 \mid A) = (0.01 \times 0.1) / 0.0019 \approx 0.526$$

- Interpretation: If a failure occurs, there's a 52.6% chance the sensor was degraded.

General Form of Bayes' Theorem

Posterior Probability \longrightarrow $P(H|E) = \frac{\overset{\text{Likelihood}}{P(E | H)} \times \overset{\text{Prior Probability}}{P(H)}}{\underset{\text{Marginal Probability}}{P(E)}}$

Marginal probability refers to the probability of a single event occurring, independent of other events, and is derived from a joint probability distribution.

General Form of Bayes' Theorem

$$P(H|E) = \frac{P(E | H) \times P(H)}{P(E)}$$

Prior Probability (hypothesis) → P(H)

- Your belief or estimate of the probability of a hypothesis before seeing the evidence.
- “What’s the chance the sensor is degraded before we observe a failure?” or P(Sensor Degraded)

Likelihood → P(E | H)

- Probability of the evidence assuming the hypothesis is true.
- “If the sensor is degraded, how likely is a system failure?” or P(System Failure | Sensor Degraded)

Marginal Probability → P(E)

- The total probability of the evidence (regardless of which hypothesis is true).
- This is computed as a weighted sum over all hypotheses:

$$P(E) = \sum_i P(E | H_i) \times P(H_i)$$

Posterior Probability → P(H | E)

- The updated belief in the hypothesis after observing the evidence.
- “Given that the system failed, what’s the chance the sensor was degraded?” P(System Failure | Sensor Degraded)

Quiz

- Suppose a rare disease infects **one** out of every 1000 people in a population. Some company published a test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces some false positives: 2% of uninfected people(test positive although healthy).
- If someone just tested positive, what are his chances of having this disease?

Define P: positive test D: having disease H or D': Healthy

- a. 4.72%
- b. 2%
- c. 99%

Quiz

- Suppose a rare disease infects **one** out of every 1000 people in a population. Some company published a test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces some false positives: 2% of uninfected people(test positive although healthy).
- And someone just tested positive. What are his chances of having this disease?

Define P: positive test D: having disease H or D': Healthy

$$P(D) = .001$$

$$P(H) = 0.999$$

$$P(P | D) = .99$$

$$P(P | H) = 0.02$$

$$\begin{aligned} P(D|P) &= \frac{P(P | D) \times P(D)}{P(P)} \\ &= \frac{P(P | D) \times P(D)}{P(P | D) \times P(D) + P(P | H) \times P(H)} \\ &= \frac{.99 \times .001}{.99 \times .001 + .002 \times .999} = .0472 \end{aligned}$$

Quiz

- Suppose a rare disease infects **one** out of every 1000 people in a population. Some company published a test for this disease: if a person has the disease, the test comes back positive 99% of the time. On the other hand, the test also produces some false positives: 2% of uninfected people(test positive although healthy).
- 100.000 passengers travel though an airport, how many positive cases can we expect?

Define P: positive test D: having disease H or D': Healthy

- a. 4720
- b. 2097
- c. 1500

Solution

$$P(D)=1/1000=0.001$$

$$P(D') = 1-0.001 = 0.999$$

$$P(T+|D)=0.99$$

$$P(T+|D')=0.02$$

$$P(T+)=P(T+|D) \cdot P(D) + P(T+|D') \cdot P(D') = 0.02097$$

This means that approximately 2097 people would test positive among the 100.000 passengers.

Example

- A rare disease affects approximately 1 out of every 1,000 people in a population. A medical company has developed a diagnostic test for this disease with the following properties:
 - If a person (regardless of gender) is infected, the test correctly returns a positive result 99% of the time.
 - If a person (regardless of gender) is not infected, the test still returns a positive result 2% of the time.
- Demographic background:
 - 49% of the population is female, and 51% is male.
- Suppose a sample of 100 people is randomly selected from the population, and all 100 test positive for the disease.

Question:

Based on this information, what is the likely percentage of males in this group of 100 individuals who tested positive?

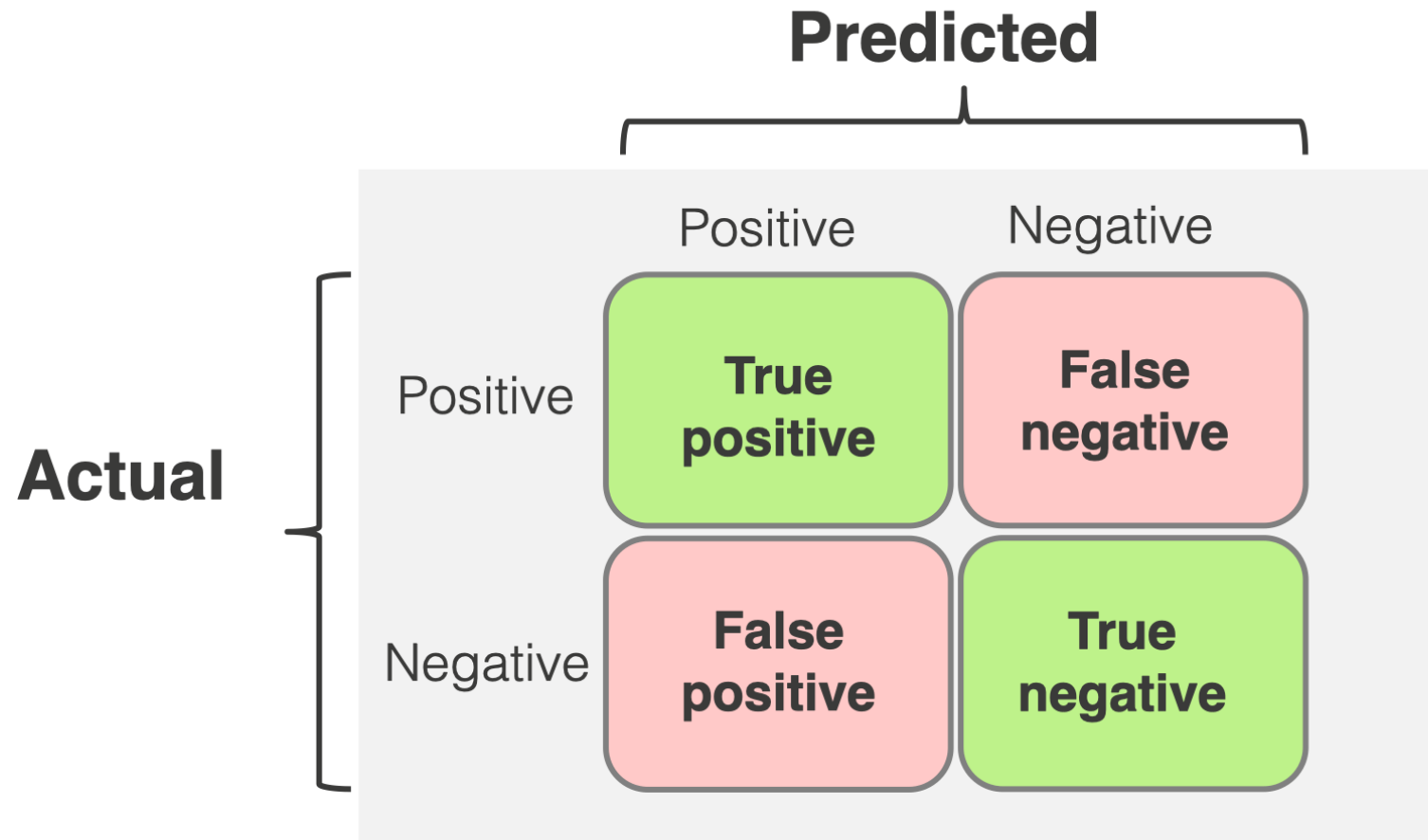
Solution

Since gender and test results are independent, we leverage the definition of independence:

$$P(M \mid T+) = P(M) = 0.51$$

This means that approximately 51% of those who test positive are male.

Interpreting Test Results: Sensitivity, False Positives, and More



Random Variables

- A **random variable** is a mathematical concept representing a quantity that can have different outcomes from a random event, like the outcome of a dice roll
- Random variables can be:
 - **discrete** (countable and finite)
 - **continuous** with values that varies along intervals, like real numbers
- **Examples:**
 - ✓ **Discrete variable**
 - ✓ x = number on the upper face of a randomly tossed die
 - ✓ x = The number of undetected faults in a safety-critical software module
 - ✓ **Continuous variable**
 - ✓ x = The time until the next failure of a crucial sensor in an aircraft's flight control system.
 - ✓ x = temperature of CPU at any given time

Probability Distributions: An Overview

- A **probability distribution** shows how the probabilities are distributed over the possible values of a random variable.
- **Probability distribution** fundamentally describes the **population**— that is, the *theoretical behavior* of a random variable across **all possible outcomes**.
- The **sample distribution** (e.g., the histogram of your sample) *approximates* the true population distribution.

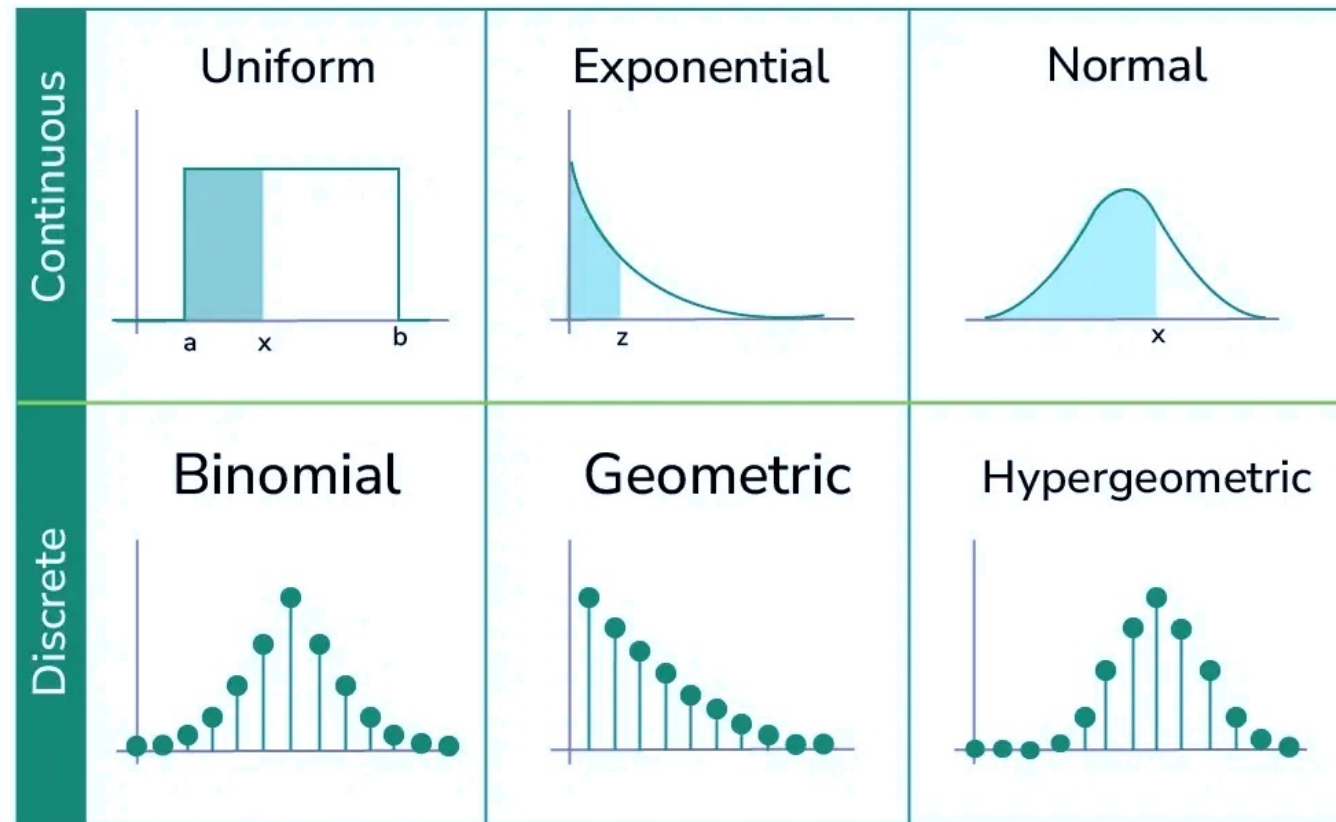
Sample Distribution vs. Population Distribution

- Let x be a discrete random variable with probability distribution $p(x)$. Then the mean, variance and standard deviation of x (*of a population*) are denoted as μ and σ .
- However, we usually work with samples so we can use the **sample distribution** (e.g., the histogram of your sample) to *approximate and make assumptions about* the true population distribution.
- Sample's mean and standard deviation (denoted as \bar{x} , s) estimate the population's μ and σ .
- As sample size increases \uparrow , the sample distribution better represents the population.

Probability Distributions: An Overview

- We fit or model data with probability distributions to **compress reality into something we can reason about**.
- Raw data points are messy; a distribution gives us a *simplified, structured description* of how the data behaves.
- Once we have that structure, we can:
 - **Estimate** probabilities of events
 - **Generalize** beyond the sample
 - **Compare** groups or processes using a common language
 - **Simulate** systems (or scenarios)
 - **Model uncertainty** in real-world events

Intuition: A distribution is a “shape” that summarizes your data. If your data roughly forms a bell shape, a normal distribution tells you where most of the mass is and how spread out it is, without storing every data point.



Examples on some commonly used distributions

Example: Distributions

- Suppose you measure the daily demand for a product for 60 days. The numbers bounce around, but they cluster near 120 units with some variability. If you **model demand** as **Normal(120, 15²)**, you can now answer questions like:
 - “What’s the chance demand exceeds 150 units?”
 - “How much inventory should we keep to satisfy 95% of days?”
 - “If we change pricing and the mean shifts, what happens to stockouts?”
- The distribution gives you the *generalizable, predictive picture*.

Probability Distributions: An Overview

- **Key Features:**
 - **Shape:** Can be symmetric, skewed, bi-modal, mound-shaped...etc
 - **Outliers:** Represent unusual or unlikely measurements.
 - **Center & Spread:** Defined by the mean (μ) and the standard deviation (σ).
- **We can use these features to tell something about data:**
 - Raw data → hard to compare.
Distributions → standardized explainable summaries that make differences clear.



Questions



Answers

