Understanding and Quantifying Uncertainty.
How Likely is That? Exploring Probability Through Stories.

Prepared By: Siman Giri, Instructor: Ronit and Shiv for Herald Center for AI.

Summer, 2025

1 Learning Objectives.

- Focus: Probability Foundations, Conditional Probability & Bayes Rule
- Skills Targeted: Justify Business Decisions using probabilistic reasoning.
- By the end of this worksheet, you will be able to:
  - Define and apply the axioms of probability to real - world datasets.
  - Compare marginal, joint and conditional probabilities from data.
  - Explain the difference between independent and dependent events using the dataset.
  - Derive Baye's Rule from the definition of Conditional probability.
  - Calculate the components of Bayes Rule from real - world data.
  - Interpret how prior probabilities affect posterior conclusions.

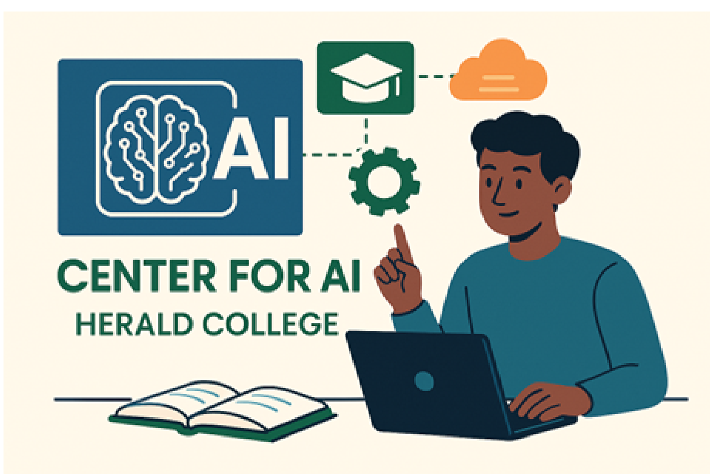

image generated via copilot.

2 Interpreting Probability.

2.1 Distinguish between classical, empirical, and subjective probability through real-world examples:

Case 1: Predictive Maintenance.

- **Scenario:** A factory uses sensor data to predict machine failures:
  - **Model A:** Claims **"85% accuracy"** based on past data.
  - **Model B:** A technicial says **"I'm 70% sure it will fail by Friday"** based on noise/vibration.
- **Tasks:**
  1. Label the probability type for each model.
  2. Which would you trust more? Justify using data quality.
  3. How would you combine both approaches into a hybrid model?

Sample Solution:

1. Label the Probability Type for each Model:
   - **Model A - 85% accuracy:** This represents an empirical probability because it is based on past data and outcomes from historical sensor readings. Accuracy is calculated from actual observed frequencies of correct predictions.
   - **Model b - 70% certainty:** This is subjective probability because it reflects the technician's personal judgment or belief, informed by experience, intuition, and interpretation of machine behavior like noise or vibration.
2. Which Model Would you Trust More? Why?
   Trust depends on context and use case, but in general:
   - **Model A** is more objectively reliable as it is trained and validated on a large dataset, offering reproducibility and a known performance metric (**85% accuracy**. It benefits from scale, statistical rigor, and non biased input.
   - **Model B** leverages expert intuition, which may catch early warning signs that are not yet detectable by the sensor model. Human expertise can be more sensitive to context-specific signals or unusual conditions not present in training data.
   - **Balanced View** If the sensor data is recent, high-quality, and the model has been well-validated across different failure types, Model A should generally be trusted more.
     However, if the technician has years of experience with that specific machine or has previously caught issues the model missed, Model B might be valuable too.
3. How would You combine Both Approaches into a Hybrid Model?
   To build a hybrid predictive maintenance model, you can integrate both empirical and subjective sources:
   - **Hybridization Strategy:**

(a) **Bayesian Updating:** Use the technician's subjective probability (prior belief) as a prior, and update it with the model's empirical prediction (likelihood) using Bayes' Theorem. This mathematically fuses expert belief with data evidence.

(b) **Ensemble Model:** Treat the technician's prediction as a separate feature or input in a machine learning ensemble (e.g., Random Forest, Gradient Boosting), combining it with sensor-based features to improve performance.

Case 2: Election Forecasting.

- **Scenario:** A political analyst says: “Our model gives a **68% chance** that Candidate A will win the upcoming election.”
- **Tasks:**
  1. What type of probability is this?
  2. What data and assumptions drive the model?
  3. How could different models yield different probabilities?

Case 3: AI System Failure.

- **Scenario:** An AI safety researcher claims: “There is a 10% chance that a powerful AI system will behave unsafely by 2030.”
- **Tasks:**
  1. Is this a classical, empirical, or subjective probability?
  2. What evidence or reasoning might support such a claim?
  3. What makes this probability controversial or uncertain?

Case 4: Insurance Pricing.

- **Scenario:** An actuary sets car insurance rates based on:
  - **Initial Policy Based on Historical Crash Data -** E.g. 5% of 20 year old drivers file claims.
  - **New policy -** Drivers who pass a safe-driving course get a 10% discount.
- **Tasks:**
  1. Are the initial rates classical, empirical, or subjective?
  2. Is the discount’s probability adjustment empirical or subjective? Why?

Case 5: Dynamic Pricing.

- **Scenario:** A ride-sharing app charges ”2x surge pricing” during rain. The algorithm claims:
  - ”There’s a 60% probability of rain at 5 PM,” using weather APIs and historical demand.
- **Tasks:**
  1. Deconstruct the 60%: Is it empirical (weather data), subjective (demand heuristics), or both?
  2. How might riders perceive this probability differently than the company?
  3. How would you test if the surge price is truly demand-driven vs. exploitative?

3 Q.E.D: Quantitative Exercises in Probability.

Q.E.D = Proof Completed.

1. Proof Question.

1. Using Kolmogorov's axioms, prove that:

$$\mathbf{P(A^c) = 1 - P(A).}$$

   `Hint: Use` $\mathbf{A \cup A^c = \Omega}$ `and Additivity Axiom.`

2. Derive the multiplication rule:

$$\mathbf{P(A \cap B) = P(A|B) \cdot P(B).}$$

   `Hint: Start from the definition of` $\mathbf{P(A|B)}$.

3. If A and B are independent, show that:

$$\mathbf{P(A|B) = P(A).}$$

   `Hint: Use the definition of independence` $\mathbf{P(A \cap B) = P(A) \cdot P(B)}$.

4. Derive **Bayes Rule** from the definition of conditional probability.

5. Derive the **law of total probability:**

$$\mathbf{P(A) = P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)}$$

   `Hint: Decompose` $\mathbf{A = (A \cap B) \cup (A \cap B^c)}$.

4 Probability Puzzlers: Computational Casebook.

Objective:

Solve real-world scenarios by applying probability axioms and computations

Case 1: E - Commerce Purchase Behavior.

- **Scenario:** A website records the following user behavior:
  - 60% of visitors view a product.
  - 25% add it to their cart.
  - 15% make a purchase.
  - 10% of visitors both add to cart and make a purchase.
- **Tasks:**
  1. What is the probability that a visitor adds to cart but does not purchase?
  2. Are the events "adding to cart" and "making a purchase" independent?
  3. Use Axiom 3 to verify the relationship between the probabilities.

Sample Solution:

- **Given Probabilities:**
  - **V = views product: $\mathbf{P(V) = 0.60}$**
  - **A = adds to cart: $\mathbf{P(A) = 0.25}$**
  - **P = makes purchase: $\mathbf{P(P) = 0.15}$**
  - **Probability of add cart and make purchase: $\mathbf{P(A \cap P) = 0.10}$**
  - **Tasks:**
    1. **Probability of adding to cart but not purchasing:**

       $$\mathbf{P(A \text{ and not } P) = P(A) - P(A \cap P) = 0.25 - 0.10 = 0.15}$$

    2. **Check independence of A and P:**
       Two events are independent if: $\mathbf{P(A \cap P) = P(A) \times P(P)}$
       * Calculate: $\mathbf{P(A) \times P(P) = 0.25 \times 0.15 = 0.0375}$
       * Compare to given: $\mathbf{P(A \cap P) = 0.10}$
       * Since $\mathbf{0.10 \neq 0.0375}$, the events are NOT independent.
    3. **Verify using Axiom 3 i.e. Addition Rules $\mathbf{P(A \cup P) = P(A) + P(P) - P(A \cap P)}$**

Case 2: College Course Enrollment.

- **Scenario:** Out of 1,000 students:
  - 600 enroll in Data Science (DS)
  - 450 enroll in AI
  - 250 enroll in both courses
- **Tasks:**
  1. What is the probability that a student is enrolled in at least one course?
  2. What is the probability that a student is enrolled in neither course?
  3. Use the inclusion-exclusion principle to verify your answer.

Case 3: University Library Usage.

- **Scenario:** Survey of 1,200 students:
  - 700 use digital resources
  - 500 use physical books
  - 300 use both
- **Tasks:**
  1. Find the probability a student uses neither
  2. Is the event "uses digital resources" disjoint from "use physical books"? Why or Why not?
  3. Use axioms to explain the logical structure.

Case 4: Online Streaming Habits.

- **Scenario:** A sample of users from a streaming platform showed:
  - 60% watch movies
  - 40% watch series
  - 20% watch both
- **Tasks:**
  1. Calculate:
     - P(only movies)
     - P(only series)
     - P(neither)
  2. Are movie - watching and series - watching mutually exclusive?

Case 5: Smartphone Failure Report.

- **Scenario:** A company finds the following over one year:
  - 5% of phones have battery issues
  - 3% have screen issues
  - 1% have both
- **Tasks:**
  1. What's the probability a phone has either issue?
  2. Explain which axioms of probability this relates to.

5 Data Dilemmas: Solving Through Bayes.

When intuition fails, let Probability Guide your Decisions.

5.0.1 Exercise 1 - Conceptual Understanding - Bayes Rule:

- **Match each description with Terms.**

S.No	Description	Your Answer
1	How common is the condition in the population?	____________
2	If the condition is true, how likely is this test result?	____________
3	If we observed this result, how likely is the condition?	____________
4	What's the overall chance of seeing this result?	____________

Options:

- **Posterior, Prior, Evidence, Likelihood.**

5.1 Exercise 2 - Case Studies:

Case 1: Fraud or False Alarm.

- **Context:** You are a data analyst at NeoBank, a fast-growing digital bank. The bank uses a machine learning–based alert system to flag potentially fraudulent credit card transactions.
  Your team needs to assess the reliability of this system and guide the fraud investigation team on how many alerts actually correspond to real fraud.

Table 1: Data from Past 100,000 Transactions

Metric	**Value**
Transactions flagged as suspicious	2,000
Confirmed fraudulent transactions	500
Of those 500 fraud cases, flagged	490 (i.e., sensitivity = 98%)
Of the 99,500 non-fraud cases, flagged	1,510 (i.e., false positive rate ≈ 1.5%)

- **Tasks:**

  1. Estimate the probability that a **flagged transaction** is actually fraudulent i.e. compute:

$$\mathbf{P}(\mathbf{Fraud}|\mathbf{Alert}) = ?$$

Hint:

(a) **Identify the Components:**
  - **Prior:** $\mathbf{P}(\mathbf{Fraud})$
  - **Likelihood:** $\mathbf{P}(\mathbf{Alert}|\mathbf{Fraud})$
  - **False Positive Rate:** $\mathbf{P}(\mathbf{Alert}|\mathbf{NoFraud})$
  - **Evidence:** $\mathbf{P}(\mathbf{Alert})$

(b) **Compute:**

i. $\mathbf{P(Alert) = P(Alert|Fraud) \cdot P(Fraud) + P(Alert|NoFraud) \cdot P(NoFraud)}$

ii.
$$\mathbf{P(Fraud|Alert)} = \frac{\mathbf{P(Alert|Fraud) \cdot P(Fraud)}}{\mathbf{P(Alert)}}$$

2. **Discussion Questions:**
   (a) Even though the system detects 98% of fraud cases, why is the final probability of fraud after an alert not close to 98%?
   (b) What's the probability a transaction is not fraud given it was flagged?
   (c) What does this tell you about using alerts to automate fraud investigation decisions?
   (d) How would your recommendation to the fraud team change if the prior fraud rate increased?
   (e) How can the system be improved — through better modeling, threshold tuning, or adding new signals?
   (f) Bayesian updating if multiple evidence sources (e.g., location mismatch, IP risk) are available.

Case 2: Is this a Problem Product?

- **Scenario:** You work as a data analyst at ShopSmart, a popular e-commerce platform. The company is concerned about products with high return rates and suspiciously positive reviews. Your job is to help determine:
  - **What is the probability that a product is defective, given that it has a large number of returns and unusually positive reviews?**

  This information will guide the product quality assurance team and help filter out misleading listings.

- **Background Data:**

Table 2: Background Data

Observation	**Value**
Percentage of products that are defective	4% (i.e., prior = 0.04)
Probability of high return rate given defective	90% (i.e., likelihood = 0.90)
Probability of high return rate given non - defective	5%
Probability a Product has a high return rate overall	?

- **Your Task:** Compute the probability that a product is **defective given it has high return rate:**

$$\mathbf{P(Defective|HighReturn)} = \frac{\mathbf{P(HighReturn|Defective) \cdot P(Defective)}}{\mathbf{P(HighReturn)}}$$

- **Hint:**

1. **Identify the Bayes Rule Components.**

2. **Compute the Evidence.** Using Total Probability i.e.

$$\begin{aligned}\mathbf{P}(\mathbf{HighReturn}) = \mathbf{P}(\mathbf{HighReturn}|\mathbf{Defective}) \cdot \mathbf{P}(\mathbf{Defective})\\ +\mathbf{P}(\mathbf{HighReturn}|\mathbf{NotDefective}) \cdot \mathbf{P}(\mathbf{NotDefective})\end{aligned}$$

3. **Compute the Posterior.**

- **Discussion Questions:**

1. Why is the posterior probability not close to 90%, even though most defective products have high return rates?
2. How does the rarity of defective products affect your interpretation?
3. What actions should the company take if this probability is above 50%?
4. How would additional evidence (e.g., suspicious review patterns) improve this inference?
5. How can you use Bayes' Rule to create a product risk score across multiple signals?

Case 3: Is Clicking the Email a Good Predictor of Conversion?

A marketing team recently initiated an email-based campaign and wants to assess its effectiveness. As a data analyst, you are provided with information on 25 customers, including whether they clicked the email and whether they ultimately converted (i.e., responded positively to the campaign).

The central question is: *What is the probability that a customer converted, given that they clicked on the marketing email?*

Bayesian Setup:

- **Hypothesis (H):** The customer converted.**Success** = 1
- **Evidence (E):** The customer clicked the marketing email. **ClickedEmail** = 1
- **Goal:** Estimate $\mathbf{P}(\mathbf{Success} \mid \mathbf{ClickedEmail})$

Customer Data:

CustomerID	Success (1 = Yes)	ClickedEmail (1 = Yes)
C01	0	0
C02	1	1
C03	1	1
C04	0	1
C05	0	0
C06	0	1
C07	0	0
C08	0	0
C09	0	0
C10	1	1
C11	1	1
C12	0	0
C13	0	0

C14	1	0
C15	1	1
C16	0	0
C17	1	1
C18	1	1
C19	0	0
C20	0	0
C21	0	1
C22	0	0
C23	0	0
C24	0	0
C25	0	0

Using this dataset, apply **Bayes' Theorem** to compute:

$$P(\text{Success} \mid \text{ClickedEmail}) = \frac{P(\text{ClickedEmail} \mid \text{Success}) \cdot P(\text{Success})}{P(\text{ClickedEmail})}$$

You are encouraged to compute the necessary probabilities either manually (by counting frequencies) or programmatically using Python or Excel.

6 Coding Exercise:

Use Python Programming to answer the following Questions.

Identifying Defective Products in an E-Commerce Platform.

Should we Recall This Product?

Background:

You are a data analyst at Quick-shop, an e-commerce company. The customer support team has flagged "Product X" for unusually high return rates. Before taking costly actions (e.g., recalls, supplier penalties), you need to determine whether the returns are due to defects (quality issues) or other factors (e.g., customer preferences).
Your team has provided a dataset with 10,000 products, including:
`Dataset: "product\_review\_dataset.csv"`

- `Defective` (Binary: 1 = defective, 0 = not defective)
- `HighReturn` (Binary: 1 = high return rate, 0 = low return rate)
- `ReviewRating` (Numeric: 1–5 star rating)
- `HasComplaint` (Boolean: True/False)
- `VerifiedPurchase` (Boolean: True/False)

Sample row:
`Defective=0, HighReturn=1, ReviewRating=3.8, HasComplaint=False, VerifiedPurchase=True`

Tasks

1. Exploratory Analysis

1. Compute the prior probability that a product is defective ($P(\text{Defective})$)
2. Compare the average review rating for defective vs. non-defective products
3. Calculate the return rate separately for:
   - Defective products
   - Non-defective products

2. Bayesian Inference

1. Calculate:
   - $P(\text{HighReturn} \mid \text{Defective})$
   - $P(\text{HighReturn} \mid \text{Not Defective})$
   - $P(\text{Defective} \mid \text{HighReturn})$ (Posterior probability)

2. Interpretation:
   - If $P(\text{Defective} \mid \text{HighReturn}) = 27\%$, what does this imply?
   - Should QuickShop prioritize recalls based solely on this probability? Justify your answer.

3. Multi-Feature Risk Scoring

1. Create a risk score combining:
   - High returns (`HighReturn`=1)
   - Low ratings (`ReviewRating` ≤ 2)
   - Complaints (`HasComplaint`=True)
2. Write Python code to identify the top 10 highest-risk products

4. Decision-Making (10 minutes)

1. For a product with:
   - `HighReturn`=1
   - `ReviewRating`=1.5
   - `HasComplaint`=True
   - `VerifiedPurchase`=False

   Would you recommend a recall? Justify using your calculations.
2. Propose what additional data would improve this analysis

5. Deliverables

Submit a 1-page report containing:

- Key probabilities and their interpretations
- A ranked list of high-risk products (from Task 3)
- Your decision and justification for the scenario in Task 4

7 Appendix.

Computing Risks:

Key Concepts:

1. **Prior Probability $\mathbf{P}(\textbf{Defective})$:**
   - The base rate of defective products in your dataset.
   - Hint for Calculation:

$$\mathbf{P}(\textbf{Defective}) = \frac{\texttt{Number of Defective Products}}{\texttt{Total Products}}$$

2. Create a Feature - **Low Rating**: To flag products with poor reviews, define a binary feature **Low Rating** where:

$$\texttt{Low Rating} = \begin{cases} 1 & \texttt{if Review Rating} \leq 2, \\ 0 & \texttt{otherwise.} \end{cases}$$

3. **Conditional Probabilities:** likelihood of observing a risk factor i.e. (High Return, Low Rating, and High Complaint) given a product is defective. Compute:

$$P(\text{High Return} \mid \text{Defective}) = \frac{\text{Defective products with HighReturn=1}}{\text{Total defective products}}$$

$$P(\text{Low Rating} \mid \text{Defective}) = \frac{\text{Defective products with LowRating=1}}{\text{Total defective products}}$$

$$P(\text{High Complaint} \mid \text{Defective}) = \frac{\text{Defective products with Complaint=1}}{\text{Total defective products}}$$

Risks Score Formula:

1. Naive Bayes Approach:

Assuming independence between features, the posterior probability of a product being defective given the observed features is:

$$\texttt{Risk Score} = \texttt{P(Defective)} \times \prod_{\mathbf{i}} \mathbf{P}(\mathbf{Feature_i} | \mathbf{Defective})$$

Here: $\prod_{\mathbf{i}} \mathbf{P}(\mathbf{Feature_i} | \mathbf{Defective})$: Likelihood of observing the feature given the product is defective. If features are Binary, this can be written as:

$$\textbf{RiskScore} = \mathbf{P}(\textbf{Defective}) \times \prod_{\mathbf{i}} \begin{cases} P(\textbf{Feature}_i \mid \textbf{Defective}) & \text{if Feature}_i = 1 \\ 1 - P(\textbf{Feature}_i \mid \textbf{Defective}) & \text{if Feature}_i = 0 \end{cases}$$

For our case:

$$\begin{aligned}\text{RiskScore} = P(\text{Defective}) \times & \\ &\begin{cases} P(\text{HighReturn} \mid \text{Defective}) & \text{if HighReturn} = 1 \\ 1 - P(\text{HighReturn} \mid \text{Defective}) & \text{if HighReturn} = 0 \end{cases} \\ &\times \\ &\begin{cases} P(\text{LowRating} \mid \text{Defective}) & \text{if LowRating} = 1 \\ 1 - P(\text{LowRating} \mid \text{Defective}) & \text{if LowRating} = 0 \end{cases} \\ &\times \\ &\begin{cases} P(\text{HasComplaint} \mid \text{Defective}) & \text{if HasComplaint} = 1 \\ 1 - P(\text{HasComplaint} \mid \text{Defective}) & \text{if HasComplaint} = 0 \end{cases}\end{aligned}$$

2. Log - Probability version:

For numerical stability and to avoid tiny numbers we use log probabilities:

$$\texttt{Risk Score} = \texttt{log P(Defective)} + \sum_{\mathbf{i}} \mathbf{logP(Feature_i|Defective)}$$

3. Normalized Risk Score (0 to 1):

To make scores interpretable as probabilities:

$$\mathbf{NormalizedRiskScore} = \frac{\mathbf{RiskScore}}{\mathbf{Riskscore + P(NotDefective) \times \prod_i P(Feature_i|NotDefective)}}$$

Example Calculation:

Given:

$$\begin{aligned} P(Defective) &= 0.05 \\ P(HighReturn \mid Defective) &= 0.6 \\ P(LowRating \mid Defective) &= 0.4 \\ P(Complaint \mid Defective) &= 0.3 \end{aligned}$$

Product A: HighReturn=1, LowRating=0, Complaint=1

$$\begin{aligned} \text{RiskScore} &= 0.05 \times P(\text{HighReturn} \mid \text{Defective}) \\ &\quad \times (1 - P(\text{LowRating} \mid \text{Defective})) \\ &\quad \times P(\text{Complaint} \mid \text{Defective}) \\ &= 0.05 \times 0.6 \times (1 - 0.4) \times 0.3 \\ &= 0.0054 \end{aligned}$$

Product B: HighReturn=0, LowRating=1, Complaint=0

$$\begin{aligned}
\text{RiskScore} &= 0.05 \times (1 - P(\text{HighReturn} \mid \text{Defective})) \\
&\quad \times P(\text{LowRating} \mid \text{Defective}) \\
&\quad \times (1 - P(\text{Complaint} \mid \text{Defective})) \\
&= 0.05 \times 0.4 \times 0.4 \times 0.7 \\
&= 0.0056
\end{aligned}$$

Interpretation:

- Product A has a risk score of 0.0054
- Product B has a risk score of 0.0056
- In this case, Product B has slightly higher risk than Product A
- These scores can be normalized to probabilities (see Normalized Risk Score)

The - End