Introduction to Probability: Patterns and Chance

- Probability is the logic of uncertainty — used daily for weather, health, and business decisions.
- Course path: uncertainty → sample spaces → events → long-run frequency → basic rules.
- Deterministic thinking assumes cause → effect; probabilistic thinking reasons with likelihood and evidence.
- No advanced math required; we build on everyday reasoning and real examples.
- This training sharpens risk communication and data-informed decision-making under uncertainty.

What Does 'Uncertainty' Really Mean?

- Uncertainty as a state of knowledge, not just physical randomness
- Aleatoric: intrinsic randomness; cannot be reduced (e.g., coin toss)
- Epistemic: lack of knowledge; reducible with more data or learning
- Short-term unpredictability often hides stable long-run patterns
- Knowing which type you face shapes how you decide and act

Sample Spaces: Mapping What Can Happen

- Sample space: the set of all possible outcomes of an experiment or observation
- List what can happen before calculating any probabilities
- Examples: coin toss $\{H, T\}$, die roll $\{1, 2, 3, 4, 5, 6\}$, stock $\{\text{up, flat, down}\}$
- Finite vs. countably infinite vs. continuous sample spaces
- Common errors: incomplete spaces and overlapping outcomes

Events: Grouping Outcomes That Matter

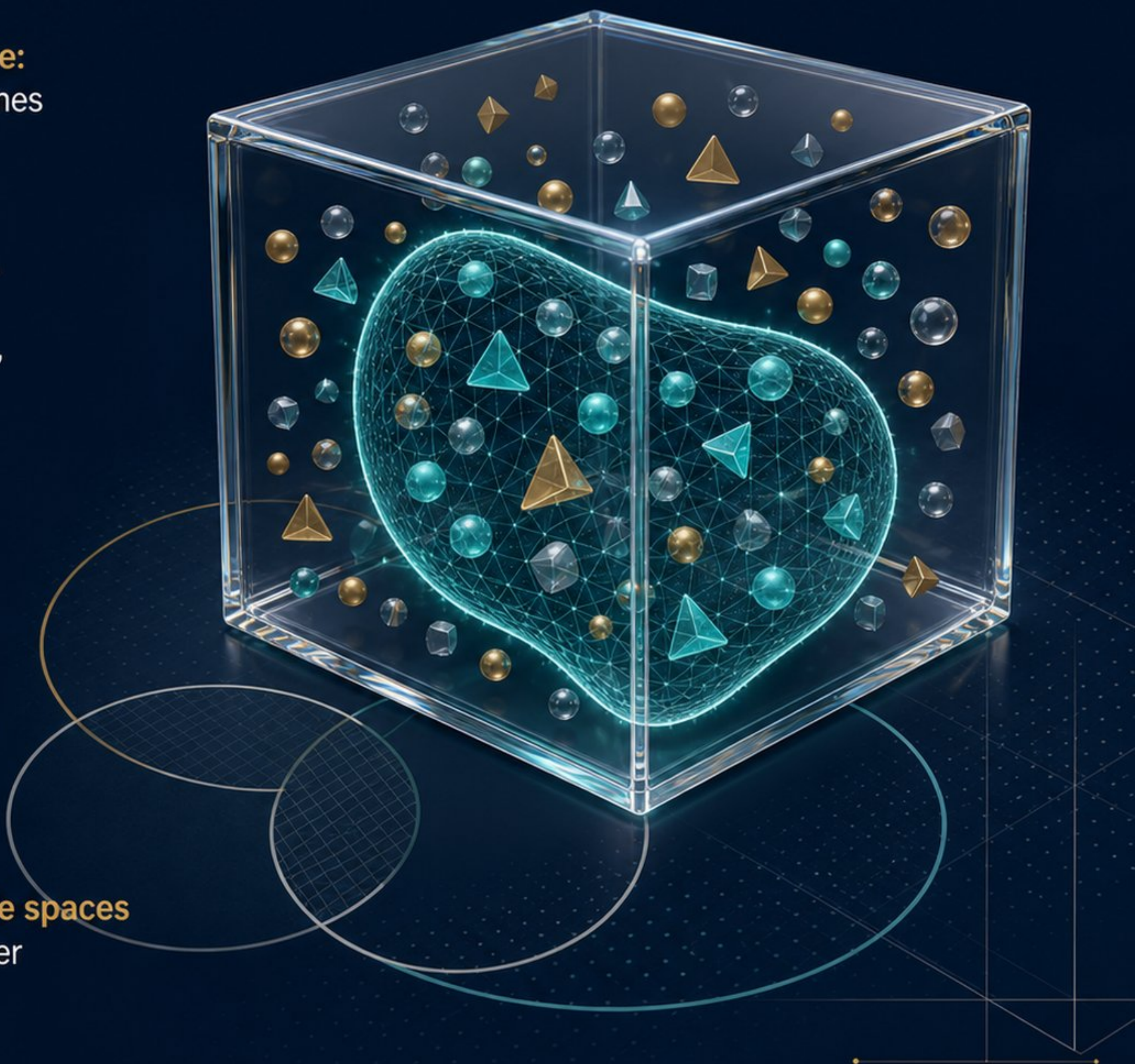
Event ⊆ sample space:
any collection of outcomes you care about

Real-world:
“stock rises tomorrow,” rain on the weekend,” “at least one customer”

Venn operations:
complement (not), intersection (AND), union (OR)

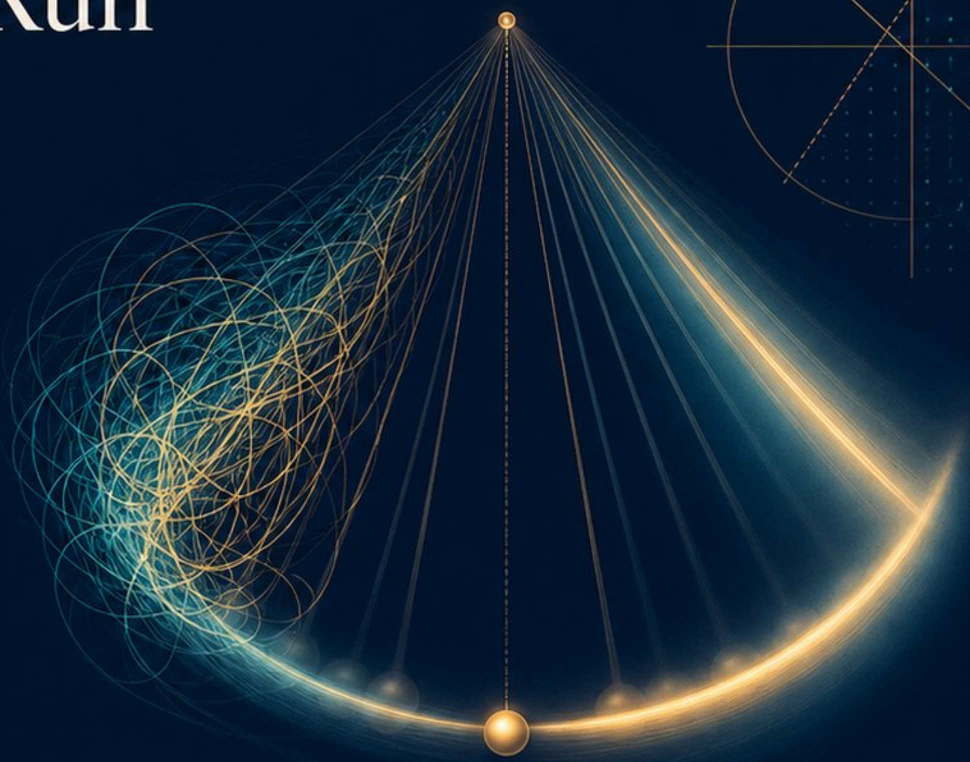
Quick examples:
rolling an even number, drawing a face card, positive test result

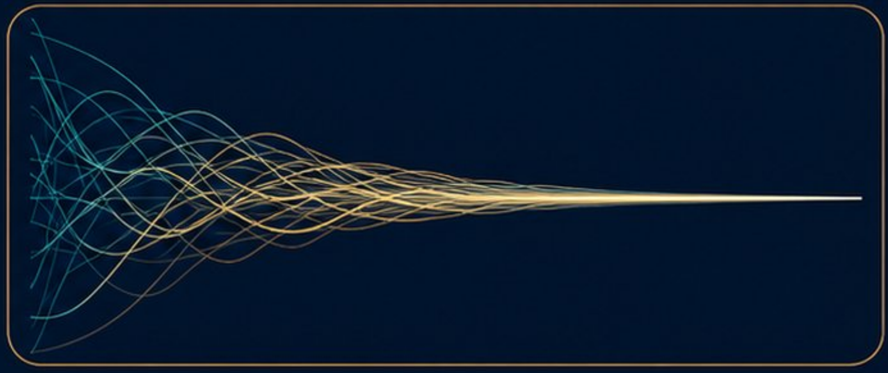
Events link raw sample spaces
to actual decisions under uncertainty



Probability as Long-Run Relative Frequency

- Probability is the proportion of times an event occurs over many independent trials
- Short-term results are noisy; long-run patterns become stable and predictable
- Law of Large Numbers: observed frequency converges to true probability as trials increase
- Simulations show convergence in real time—try 10 vs. 10,000 coin flips
- Limitation: one-off events (elections, startup success) have no repeatable long run

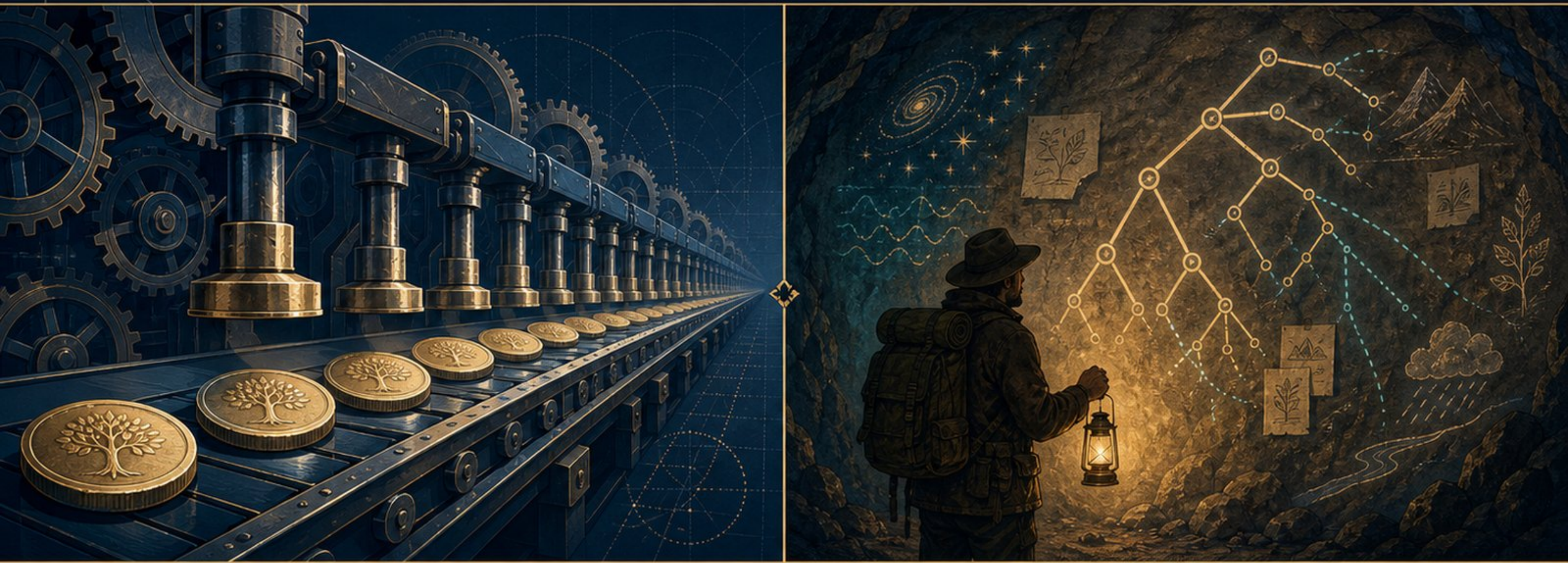




Axioms and Basic Rules of Probability

- Non-negativity: probabilities are never less than zero.
- Unit measure: the probability that something happens is 1 (or 100%).
- Additivity for disjoint events: P(A or B) = P(A) + P(B) when they can't both occur.
- Complement rule: P(not A) = 1 − P(A) — a shortcut for 'at least one' problems.
- General addition rule: P(A or B) = P(A) + P(B) − P(A and B) to avoid double-counting overlaps.

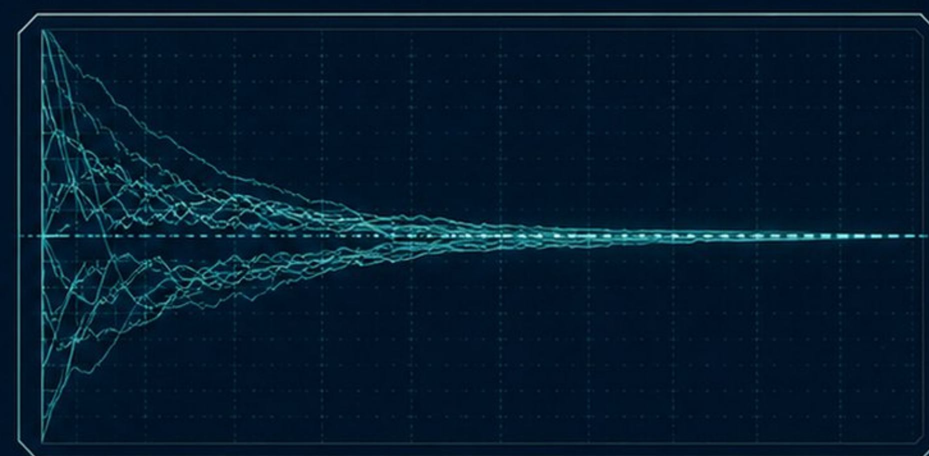
From Frequency to Subjective Probability



- - Frequentist: probability = long-run frequency in repeatable experiments
- - Bayesian: probability = degree of belief, updated with evidence
- - Use frequentist for repeatable events; Bayesian for unique predictions
- - Examples: elections, startup success, project completion dates
- - Preview: subjective priors combined with new data via Bayes' rule

Patterns in Chance: The Law of Large Numbers

- As sample size grows, observed frequency converges toward true probability
- LLN promises long-run stability, not short-run correction or balance
- Gambler's fallacy: believing past independent outcomes predict future ones
- Monte Carlo 1913: 26 consecutive blacks; gamblers lost millions betting on red
- Randomness produces streaks—they do not imply momentum or compensation



Avoiding the Gambler's Fallacy and the 'Hot Hand' Trap

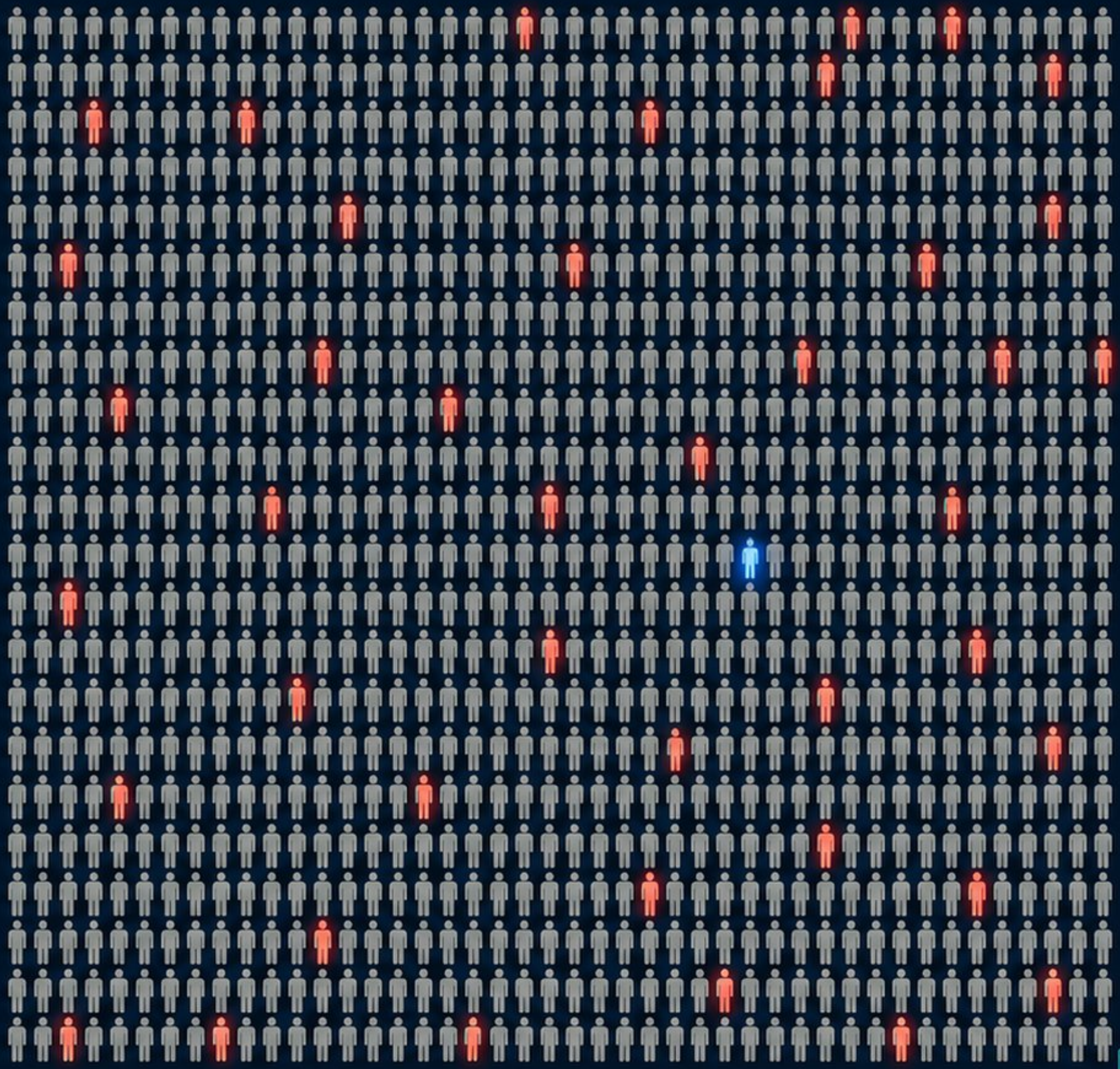
- **Gambler's fallacy:** mistaking past outcomes as influencing future independent events
- The **'hot hand' belief:** recent success predicts continued success in independent events
- Fallacy persists even when people know the correct probability
- Focus on **independence**, use data instead of intuition, frame with **natural frequencies**
- Randomness produces streaks; streaks do not imply momentum or compensation

The Base-Rate Paradox: Why Positive Tests Can Mislead

- The false positive paradox: rare conditions + highly accurate tests = mostly false alarms
- Disease 1/10,000, test 99% accurate → a positive result is <1% likely to be real
- Base-rate neglect: we ignore prevalence when interpreting test evidence
- Consequences: overdiagnosis, false accusations, excessive security alarms
- Fix: always ask "What is the base rate?" and use natural frequencies

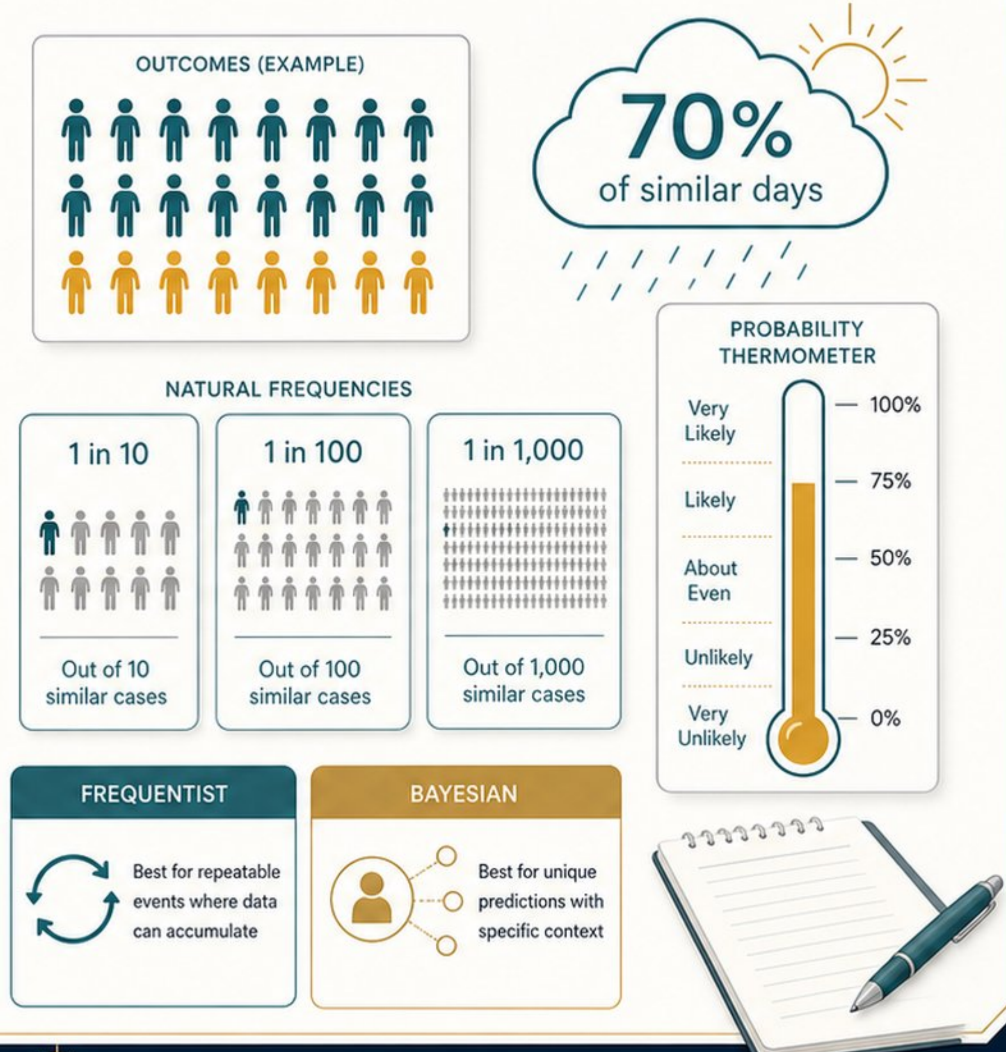
Natural Frequencies: Making Probability Intuitive

- People reason better with counts (e.g., "1 in 100") than percentages ("1%").
- Ex: Out of 10,000 people, 1 has disease. About 100 test positive, only 1 real.
- Makes base rates visible; avoids confusing P(A|B) with P(B|A).
- Use for medical results, risk communication, and teaching probability.
- Practice: reframe classic puzzles in natural frequency format.



Communicating Uncertainty Clearly

- Use natural frequencies: "1 in 100" instead of "1% chance"
- Always state the reference class and baseline risk
- Distinguish absolute risk (2% to 4%) from relative risk (100% increase)
- Avoid vague words like "likely" or "rare"—they mean different things to everyone
- Choose frequentist framing for repeatable events, Bayesian for unique predictions



Applying Probability to Everyday Decisions

- Screening tests: most positives are false alarms for rare diseases—always ask for the base rate
- Business: forecast demand and project timelines using base rates of similar projects (planning fallacy)
- Personal: weigh extended warranty cost against breakdown probability; ignore vivid but low-risk events
- Framework: define the sample space -> identify the event -> estimate base rate -> update with evidence
- Bias shield: use natural frequencies ('30 out of 100') instead of percentages to make risks intuitive

Common Biases in Probabilistic Thinking

- **Overconfidence:** we are more certain than justified by the evidence.
- **Conjunction fallacy:** judging a specific combination as more likely than a single event.
- **Availability heuristic:** overestimating vivid, recent, or emotional risks.
- **Neglect of sample size:** trusting patterns from small samples.
- **Debiasing:** use base rates, calibration training, and the outside view.

Summary and Next Steps

- Core concepts: sample spaces, events, long-run frequency, subjective probability
- Aleatoric uncertainty: irreducible randomness (accept); epistemic: lack of knowledge (reduce)
- Mental checklist: sample space → event → base rate → update with evidence
- Use natural frequencies: '1 in 10,000' beats '0.01% chance'
- Avoid gambler's fallacy and base-rate neglect in decisions