Stat C131A: Statistical Methods for Data Science

Fall 2025

Lecture 4: Probability Distributions

Sep 9 2025

Good news: two

- **Course Website:** https://stat131a.berkeley.edu/fall-2025
  - Update issues resolved !
- Switched to **Keynote** !

Jacob **Bernoulli** *1655-1705*

what is the **probability** that a person might **live to a thousand years?**


JACOBI BERNOULLI,
Profeſſ. Baſil. & utriuſque Societ. Reg. Scientiar.
Gall. & Pruſſ. Sodal.
MATHEMATICI CELEBERRIMI,
ARS CONJECTANDI,
OPUS POSTHUMUM.
Accedit
TRACTATUS
DE SERIEBUS INFINITIS,
Et EPISTOLA Gallicè ſcripta
DE LUDO PILÆ
RETICULARIS.
FESTINA LENTE
BASILEÆ,
Impenſis THURNISIORUM, Fratrum.
cIↄ Iↄcc xiii.

- Interested in the question not literally (knew that the chances are practically none), but from the perspective of **quantifying the virtually improbable**
- *Ars Conjectandi* (The Art of Conjecturing)
  - Law of large numbers
  - First systematic treatise on probability

$$p\left(live\ a\ 1000\ years\right) \sim 10^{-10^{10}}$$

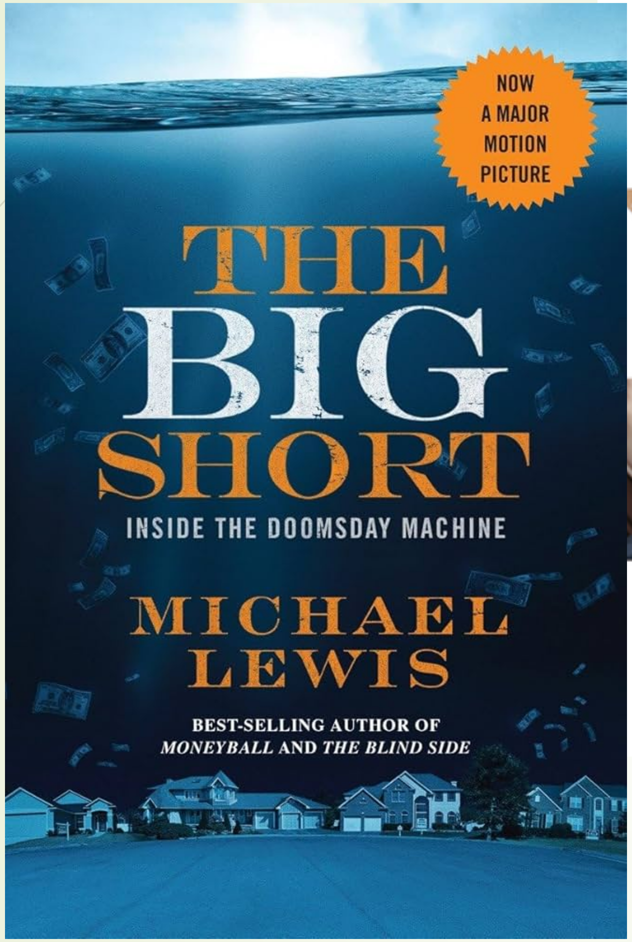
The improbable

- We don’t (mostly) care about **practically** improbable events
- How about **seemingly** improbable ones ?

2008 Financial crisis: subprime mortage

- Home foreclosures on a massive scale
- Household savings and retirement accounts wiped out
- Millions of jobs lost and unemployment surged
- Many families left with negative equity in their homes
- Small businesses closed as credit dried up
- Major banks and financial institutions collapsed or needed bailouts
- Global recession triggered, spreading beyond the US
- Public trust in financial and regulatory systems severely damaged

Over 99% of professionals, including Wall Street bankers, investment analysts, credit rating agencies, regulators, and most academic economists, **failed to forecast** the 2008 mortgage crisis



Over 99% of professionals, including Wall Street bankers, investment analysts, credit rating agencies, regulators, and most academic economists, **failed to forecast** the 2008 mortgage crisis

But … some did, by **NOT** doing the **flawed statistical inference that the majority did**

Probability: in data science

- Models uncertainty in real-world data
- Provides foundation for statistical inference
- Enables prediction with confidence (e.g., confidence intervals, p-values)
- Powers machine learning models (e.g., Naive Bayes, Hidden Markov Models)
- Guides decision-making under risk and variability
- Connects raw data to patterns through distributions and likelihoods

Today

- 2.2 Probability Distributions
  - 2.2.1 Definition of a Probability Distribution
  - 2.2.2 More Examples of Probability Distributions
  - 2.2.3 Conditional Probability and Independence

Random Variables

- A **random variable** assigns a number to an experiment outcome.
  - Example: dice roll (X=1,…,6).
  - Example: coin toss (X=0 for tails, 1 for heads).
- Can represent real data like salaries or waiting times.
- Foundation of probability modeling.
- Two main types: **discrete** and **continuous**

Sample Space (Ω)

- **Sample space** Ω is the set of **all possible outcomes**.
  - Dice: Ω = {1,2,3,4,5,6}.
  - Coin toss: Ω = {H,T}.
- Salary example: Ω = all possible employee salaries.
- **Events are subsets** of Ω !
- Important: probabilities are defined relative to Ω

Events

- An **event** is a **set** of outcomes from Ω
  - Example: Odd roll = {1,3,5}
  - Example: Salary < 72K
- Subsets of Ω
- Events can overlap or be disjoint.
- Events allow us to ask specific probability questions.
- Events: building blocks of probability.

Types of Events

- **Mutually exclusive:** cannot happen together.
  - Example: rolling a 2 and rolling a 5.
- **Complement:** A^c = all outcomes not in A.
- Operations
  - **Union:** $A \cup B$ = outcomes in A or B or both.
  - **Intersection:** $A \cap B$ = outcomes in both A and B.
- These operations follow set rules and identities.

Probability Distribution

- A **probability distribution assigns probabilities to events**
- Must satisfy rules:
  - P **gives a value for all subsets of** Ω
    - This ensures that all possible events have a probability (the probability could be zero!)
  - P **gives values only in** [0,1] This ensures we don't have negative probabilities or probabilities greater than 1. This is pretty intuitive to our idea of probability.
  - $P(X \in \Omega)=1$ aka "nothing is missing"
- For discrete X, probability distribution is defined via "**PMF**"
  - Example: fair die → each outcome has probability 1/6
- Provides **link** between **sample space** and **event likelihood**

Axioms of Probability

- Three
  - Non-negativity: $P(A) \geq 0$ for all events A.
  - Normalization: $P(\Omega) = 1$
  - Additivity: if $A \cap B = \varnothing$, then $P(A \cup B) = P(A)+P(B)$.
- These three axioms define probability rigorously
- All probability laws can be derived from these axioms.
- Foundation for valid models of randomness.

Union Rule

- Formula: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
  - Adjusts for double-counting overlap
- If A and B are disjoint, then $P(A \cup B) = P(A)+P(B)$.
  - Example: P(odd or >4) in dice.
- Union is one of the most common operations in practice !
  - Generalizes to more than 2 events

Intersection Rule

- Formula: $P(A \cap B)$
  - Represents probability that **both events occur**
- If **independent**: $P(A \cap B) = P(A)P(B)$
- Intersection highlights dependencies in data
  - Used extensively in joint probability modeling.

Complement Rule

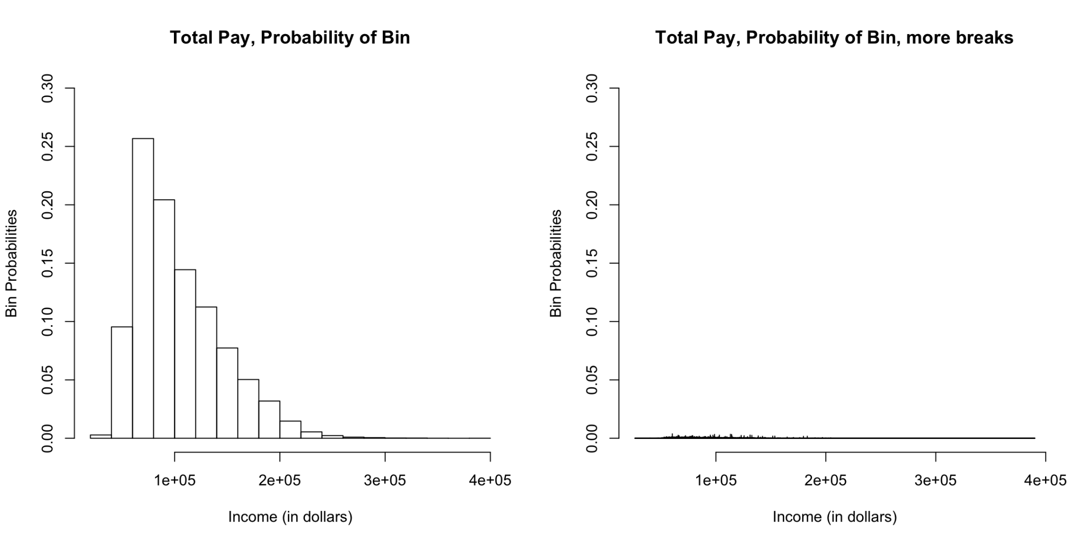
- Formula: $P(A^c) = 1 – P(A)$
  - Every event has a complement.
- Useful for computing probabilities indirectly.
  - Example: P(not even) = 1 – P(even)
  - Important when direct computation of P(A) is difficult.
- Complement is widely used in problem-solving

Example: Dice Distribution

- **Uniform distribution** across 6 faces
  - $P(X=k) = 1/6$ for $k=1,\ldots,6$
- Represents a fair and unbiased die.
- Illustrates the idea of discrete probability.
- Forms the simplest probability model

Histograms as Approximations

- Histograms approximate probability distributions.
- Each bar corresponds to a bin of values.
- Frequency histogram: raw counts.
- Density histogram: relative frequency, sums to 1
- Large sample → histogram approaches true distribution.
- Bridge between data and probability.



Frequency vs Density

- Frequency histogram: counts of outcomes.
- Density histogram: normalized to integrate to 1.
- Density allows comparison across samples.
- Probability = area under density histogram.
- Key concept for connecting data to theory.
- Used in both discrete and continuous settings.

Probability Mass Function (PMF)

$$\Omega \xrightarrow{P} [0,1]$$

- Definition: $p(k) = P(X=k)$
- Maps discrete outcomes to probabilities
  - Domain: possible values of X.
  - Range: values between 0 and 1.
  - $\Sigma\ p(k)$ over domain = 1
- Foundation of discrete probability

PMF Properties

- **Properties**
  - **Non-negativity:** $p(k) \geq 0$.
  - **Normalization:** $\Sigma\, p(k) = 1$.
- Each $p(k)$ corresponds to **one outcome**
- Distribution **fully described** by its PMF
- Valid PMF must satisfy both conditions
- Invalid PMFs can be detected easily.

Valid PMF Example

- Conference championship
  - **p(49ers)=0.4, p(kc chiefs)=0.15, p(ravens)=0.25, p(eagles)=0.2**
- Is this a valid PMF ?

PMF Example

- **p(0)=0.55, p(1)=0.35, p(2)=0.20**
- How about this ?

Coin Toss Distribution

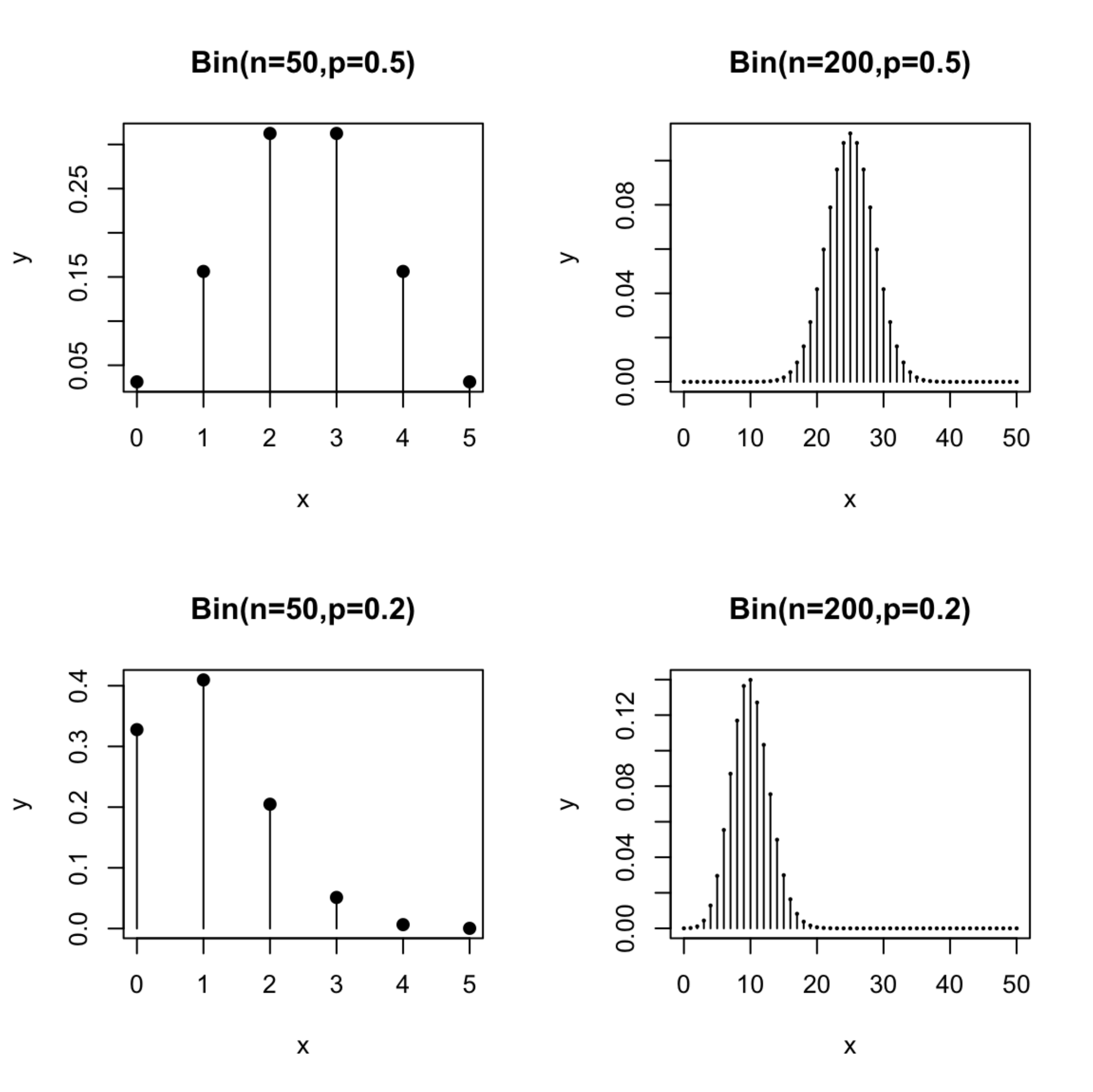
- Flip coin 10 times
- X = number of heads
- Follows a **Binomial distribution**
- Model for repeated independent trials

Binomial Distribution

$$p(k) = P(X = k) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$

- Random variable = number of successes in n trials
- Parameters: n (trials), p (success probability).
- **Support:** the set of values for which probability of success is > 0
- Models fixed trial experiments.
- Key discrete distribution.

Coin Flips: Binomial distribution



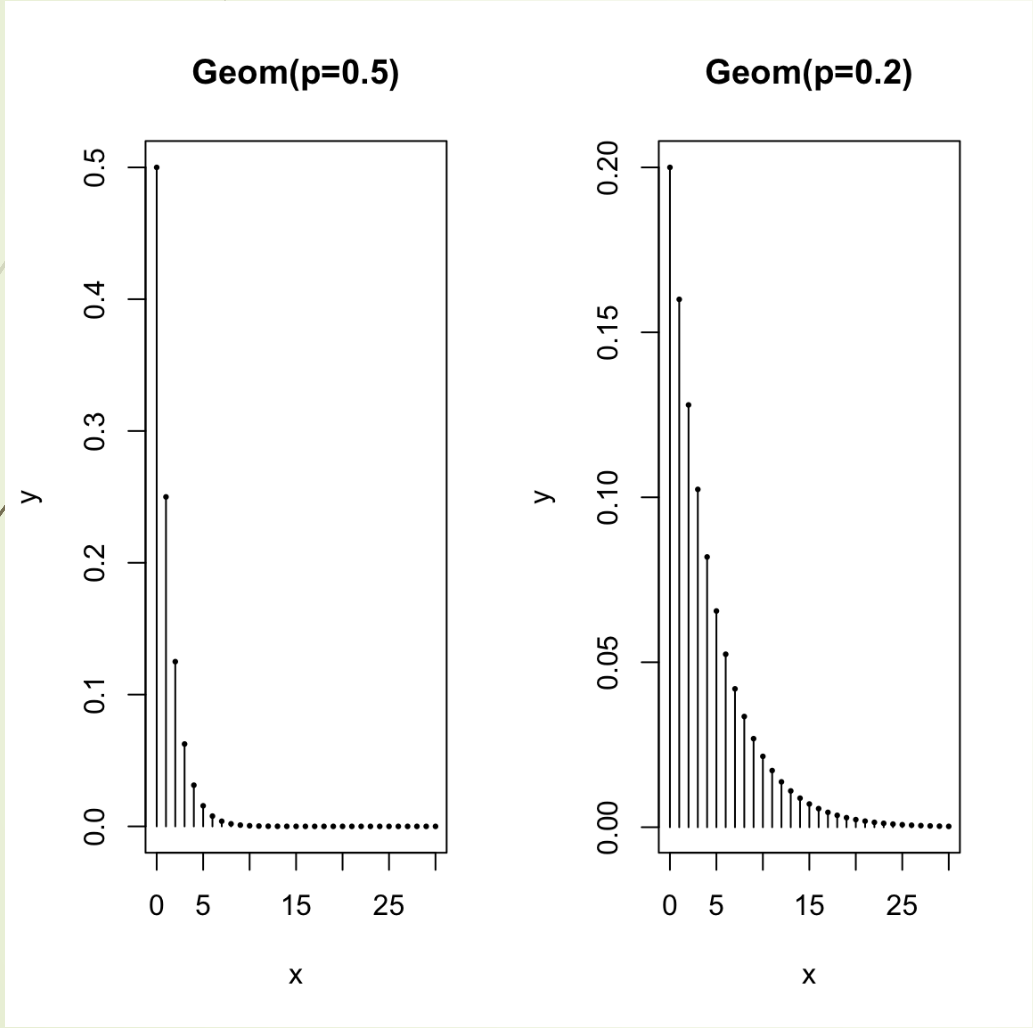
- Experiment: n independent trials; each trial is success/failure with probability p of success.
- Parameter n: number of trials (not histogram bins).
- Parameter p: success probability per trial.
- x-axis: k = 0,1,2,...,n (every integer count is a point).
- y-axis: P(X = k)
- Note: Seeing **only a few** x-tick labels (such as 0-5 in top-left) is an axis choice; otherwise the pmf has n+1 discrete points.

Geometric Distribution

$$p(k) = P(Y = k) = (1 - p)^{k-1} p$$

- Random variable = trials until first success.
- **Support**
  - the probability that we get the first Head on the kth flip
- Models waiting time.
- Decreasing shape.
- Key for sequential experiments.

Coin Flips: Geometric distribution



- Experiment: repeat independent trials (success probability p) until the first success occurs.
- Parameter p: success probability per trial.
- x-axis: k = 1,2,3,... (first success at k th trial)
- y-axis: P(Y = k)

Comparing Binomial vs Geometric

- Binomial: fixed n, count successes.
- Geometric: variable n, wait for first success.
- Both based on independent trials.
- Different support and interpretation.
- Both widely used in probability modeling.
- Illustrate two perspectives of randomness.

Conditional Probability

- P(49ers win superbowl)
- P(49ers win Superbowl | against Baltimore Ravens)
- **Conditional** distribution
- Note one but **two** random events
- Independent events
  - $P(A \cap B) = P(A)P(B)$
- $P(A \cap B) = P(B)\, P(A \mid B)$
- $P(A \mid B) = P(A \cap B) / P(B)$