Probability Models

Johannes Wissel

School of Operations Research & Information Engineering
Cornell University

Spring 2024

Reading: Devore 2.1–2.3

Probability Models

A **probability model** describes an experiment or phenomenon where the outcome isn't known in advance, or can't be predicted with certainty, e.g.,

- flipping a coin;
- who will be the next president;
- whether two students in 2700 have the same birthday.

Probability models are defined by *three elements*:

1. sample space $\mathcal{S}$: contains all the possible outcomes;
2. set of events $\mathcal{E}$: subsets of $\mathcal{S}$;
3. probability function P: assigns probabilities (i.e., numbers between 0 and 1) to events.

Probability Models: Example

Probability model for flipping a fair coin:

1. sample space $\mathcal{S} = \{\text{heads}, \text{tails}\}$
2. set of events $\mathcal{E} = \{\emptyset, \{\text{heads}\}, \{\text{tails}\}, \mathcal{S}\}$
3. probability function P:
   - $P(\emptyset) = 0$
   - $P(\{\text{heads}\}) = 0.5$
   - $P(\{\text{tails}\}) = 0.5$
   - $P(\mathcal{S}) = 1$

Events

An **event** is a subset of possible outcomes that has a probability assigned to it.

Examples of Events:

- (throwing a die) $\{1, 3, 5\}$ = set of all odd outcomes
- (value of the Dow Jones) $\{x \in \mathbb{R} : x > 15,000\}$ = set of all outcomes exceeding 15,000

Set manipulation can make computing probabilities easier.

Set Manipulation

Let A and B be events (which are sets).

- A and B occur: $A \cap B =$ set of all outcomes that are in both A and B
- A or B occurs: $A \cup B =$ set of all outcomes that are in A or B
- A doesn't occur: $A^c =$ set of all outcomes that are not in A

For a sequence $A_1, A_2, \ldots$ of events,

- $\bigcap_{i=1}^{\infty} A_i =$ set of all outcomes in A_1 and A_2 and $\ldots$
- $\bigcup_{i=1}^{\infty} A_i =$ set of all outcomes in A_1 or A_2 or $\ldots$

Set Manipulation

De Morgan's Laws: For any sequence $A_1, A_2, \ldots$ of events,

$$\left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c$$

$$\left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c$$

Probability Functions

A **probability function** is a function $P(\cdot)$, which assigns a number between 0 and 1 to every event, that satisfies the *axioms of probability*:

1. $P(A) \geq 0$ for every event A.
2. If the events $A_1, A_2, \ldots$ are *mutually exclusive* (i.e., $A_i \cap A_j = \emptyset$ for all i and j), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1) + P(A_2) + \cdots .$$

3. $P(\mathcal{S}) = 1$.

Properties of Probability Functions

$$P(\emptyset) = 0$$

For *finitely* many mutually exclusive events $A_1, \ldots, A_n$,

$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$$

For any event A,

$$P(A) \leq 1 \qquad P(A^c) = 1 - P(A)$$

Properties of Probability Functions

For any two events A and B,

$$P(A) \leq P(B) \qquad \text{if } A \text{ is a subset of } B$$

Inclusion-Exclusion Formula:

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

Defining Probability Functions: Discrete $\mathcal{S}$

When the sample space $\mathcal{S}$ is discrete (i.e., finite or countably infinite), any assignment of numbers p_s for each outcome $s \in \mathcal{S}$ where

$$p_s \geq 0 \quad \text{for all } s, \qquad \text{and} \qquad \sum_{s \in \mathcal{S}} p_s = 1$$

defines a probability function P as follows:

$$P(A) = \sum_{s \in A} p_s \qquad \text{for each event } A.$$

Example: Throwing a Fair Die

Sample Space: $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$

Set of Events: $\mathcal{E} =$ collection of all subsets of $\mathcal{S}$

Probability Function: Let $p_s = \frac{1}{6}$ for $s = 1, 2, 3, 4, 5, 6$.

Example: The probability that the outcome is divisible by 3 is

$$P(\{3, 6\}) = p_3 + p_6 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Defining Probability Functions: Continuous $\mathcal{S}$

When $\mathcal{S}$ is $\mathbb{R}$ (the set of all real numbers), we consider a *density functions* $f(x)$ where

$$f(x) \geq 0 \quad \text{for all } x \in \mathbb{R}, \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)dx = 1.$$

To define a probability function P, for

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

let

$$P([a, b]) = \int_a^b f(x)dx.$$

Example: Service Times

In *queueing theory*, the following probability model is often used for the time that it takes a server (e.g., a bank teller) to serve a customer. Take some $\lambda > 0$ and consider the density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{when } x \geq 0, \\ 0 & \text{when } x < 0. \end{cases}$$

Question: What's the probability that the service time is at least 3 time units?

A. $1 - \lambda e^{-3\lambda}$

B. $e^{-3\lambda}$

C. $\lambda e^{-3\lambda}$

D. $1 - e^{-3\lambda}$

Picking a Probability Function

Three common ways to decide what probability function to use (i.e. what p_s's or what $f(x)$ to use):

1. Symmetry
2. Estimate the probabilities from the data
3. Theory

Symmetry

When we want all outcomes to be equally likely.

Examples:

- *Discrete $\mathcal{S}$:* Toss 2 fair dice.

$$\mathcal{S} = \{(i,j) : i,j \in \{1,2,3,4,5,6\}\}$$

$$p_{(i,j)} = 1/36 \qquad \text{for all } \ (i,j) \in \mathcal{S}$$

- *Continuous $\mathcal{S}$:* Pick a number at random from $[0,1]$.

$$\mathcal{S} = [0,1] = \{x \in \mathbb{R} : 0 \leq x \leq 1\}.$$

$$f(x) = \begin{cases} 1 & \text{if } \ 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Estimate Probabilities from Data

Example: To estimate the probability that a Cornell student was born in New York state vs. other places, we take a poll!
Caution: The folks in this room are a biased sample!

A. Born in New York state.

B. Born elsewhere in the U.S.

C. Born outside of the U.S.

Theory

Example: The *Poisson distribution* is appropriate for modeling the number of calls that a call center receives per hour when receiving one call doesn't affect the time that another call's received.

In this case, $\mathcal{S} = \{0, 1, 2, \dots\}$ and the probability that s calls are received during a given hour is

$$p_s = \frac{e^{-\lambda}\lambda^s}{s!},$$

where the parameter λ needs to be estimated.

Some other random phenomena:

- Number of patients arriving in an emergency room between midnight and 1am.
- Number of decay events per second from a radioactive source.
- Number of search queries Google receives per second.

Birthday Problem

Question: What's the probability that at least 2 students in this class have the same birthday?

Assumptions:

1. No leap year (i.e., every year has 365 days).
2. Births occur on every day with equal likelihood.

Sample Space: Let N be the number of students in this class, and

$$\mathcal{S} = \{(b_1, \ldots, b_N) : b_n \in \{1, \ldots, 365\},\ n = 1, \ldots N\}$$

Birthday Problem

Let A be the event that at least 2 students have the same birthday.

It's easier to calculate the probability of A^c, the event that everyone has a different birthday, and use $P(A) = 1 - P(A^c)$:

Birthday Problem

Note that $P(A^c)$ decreases (so $P(A)$ increases) as the number of students N increases.

Question: Based on your intuition, how many students do you think are needed to make $P(A) \approx 0.5$?

A. 2
B. 25
C. 250
D. 2,500

Summary

- A **probability model** consists of:
  1. sample space $\mathcal{S}$
  2. set of events $\mathcal{E}$
  3. probability function P satisfying the *axioms of probability*
- Many useful properties of probability functions can be derived from the axioms of probability.
- Three common ways to pick a probability function to use are:
  1. Symmetry
  2. Estimate the probabilities from the data
  3. Theory