Unit 1

Basic Probability and simulation

Ethan Levien

Math 50

September 2, 2025

Outline

1. Introduction

2. Random variables and probability distributions

3. Independence, conditioning and marginalization

What is a Model?

- Models are simplified representations of the world.
- Examples of models include
  - Astrology. this is a model of human behavior
  - Newton's laws
- We focus on **statistical models** (also known as **probability models**): These are models where the variables can not be predicted exactly.
- **Goal of unit 1**: develop the language and notation for describing randomness mathematically.

Example: The Bernoulli random variable

How do we model a binary event such as: The flip of a coin, whether it rains etc.
We introduce a random variable called a Bernoulli random variable

$$X \sim \text{Bernoulli}(q)$$

This reads as “X is a Bernoulli random variable”, which means

1) **Sample space:** X is either one or zero
2) **Probability distribution** $X = 1$ is q, which we write as

$$P(X = 1) = q \text{ or } P_X(1) = q$$

Note that $P_X(1) + P_X(0) = 1$, hence $P_X(0) = 1 - q$. We could also write

$$P(X = x) = \begin{cases} q & x = 1 \\ 1 - q & x = 0 \end{cases}$$

Generating random variables with Python

```python
import numpy as np

# Bernoulli with probability p = 0.5
p = 0.5
samples = np.random.binomial(1, p, 100)
print(samples)
```

Tips for using ChatGPT to help with coding

- Include “In Python using numpy” in the prompt
- Ask it to break down the syntax if you don’t understand what it returns (for simple things)
- If you know another language, try writing the code in that language and ask it to convert it to Python. This is a code way to see how syntax you know maps to python.

General definitions

- The **Sample space** S is the set of all outcomes of an experiment, **events** are subsets of the sample space.
- We write the probabilities as

$$P(X = x) = P_X(x), \quad x \in S \tag{1}$$

  Sometimes we use $\mathbb{P}$ instead of P.
- If the distribution has a name, then we write

$$X \sim \mathrm{NameOfDistribution}(\text{parameters}) \tag{2}$$

  where **parameters** are numbers that effect the probabilities. E.g. q for the Bernoulli distribution.
- If we have multiple random variables, $X, Y, Z, \ldots$ we write

$$P(X = x, Y = y) = P_{X,Y}(x, y) \tag{3}$$

  to represent the probability that $X = x$ and $Y = y$ at the same time.

Examples

- Let X_1, X_2 bet the outcome of two independent coin flips. What is the sample space?
- A coin is flipped and then a dice is rolled. What is the sample space? If they are fair, what is the probability distribution?
- A survey is given to 3 students with two yes know questions. What is the sample space?

Properties of probability distributions

1) Probabilities of disjoint events add together. For example, the probability of rolling 1 of 2 is $2/6$.
2) Probability
3) Probabilities multiple.

How likely are there aliens?

Probability distributions from data

Conditional probability

Conditional probabilities are probabilities we obtain when we restrict ourselves to certain events in the sample space.

Conditional probability examples

Marginal probabilities