3. Discrete Probability

Discrete probabilities is probably one of the most useful topics of MACM 201. You are very likely to have to deal with probabilities in further classes (e.g. probabilistic algorithms, or probabilistic methods in graph theory), but also in everyday life.
The main idea to remember is that *computing a probability reduces to counting the size of sets of possible events*.
Topics we will cover are:

- Basics of probabilities: ratios of the sizes of sets.
- The axioms of probability.
- Conditional probabiliy.
- Bayes Theorem.
- Independent events.
- Random variables.
- Expected value.
- The geometric distribution.
- The binomial distribution.
- Balls in bins.
- The coupon collector.

3.1 Basics of Discrete Probability

3.1.1 Lecture 7 (Grimaldi 3.4, 3.5)

Example 3.1 Suppose we pick a binary string x of length 6 at **random**.
What is the **probability** that x has two 1's in it?

Definition 3.1.1 — Probability of an Event.
Hypothesis:
S is a **set** of possible outcomes called the **sample space**, all having equal likelihood.
Each subset $A \subseteq S$ is called an **event**, it is a set of *considered outcomes*.
Each element of S determines an **outcome**.

Experiment:
We generate an experiment by "drawing" at random an outcome x from S.
Note: Other words used for "drawing" are "choosing", "selecting" and "picking".

Event: Let $Pr(A)$ denote the probability that $x \in A$.

Question: What is $Pr(A)$?
Answer: If each outcome is equally likely and $|S|$ is finite

$$Pr(A) = \frac{|A|}{|S|}$$

Fundamental Principle. Calculating $Pr(A)$ requires defining two sets:
- the sample space S, and
- the considered outcomes A.

If all outcomes are equally likely, we just need to calculate $|S|$ and $|A|$.

Example 3.2 What is the probability that a random binary string of size $n \geq 2$ starts with 11 ?

Example 3.3 What is the probability that the sum of two rolls of a dice is 7 ?

Definition 3.1.2 — Axioms of Probability. Let S be a sample space and let A and B be subsets of S.

1. $0 \leq Pr(A) \leq 1$
2. $Pr(S) = 1$
3. If $A \cap B = \phi$ then $Pr(A \cup B) = Pr(A) + Pr(B)$.

Note: These axioms hold whether the outcomes of S have equal likelihood or not.

Theorem 3.1.1 — The rule of complement. Let S be a sample space, $A \subseteq S$, and $\overline{A} = S - A$ the **complement** of A. Then

$$Pr(\overline{A}) = 1 - Pr(A).$$

Proof:

Example 3.4 Illustration of the third axiom:
What is the probability that a random binary string of size $n \geq 3$ contains either exactly two 1's or exactly three 1's?

Example 3.5 Let $S = \{1,2,3,\ldots,12\}$. If x is chosen from S at random what is the probability that x is divisible by 2 OR 3 ?

Theorem 3.1.2 — The additive rule. Let S be a sample space, $A, B \subseteq S$ be two events from S. Then

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B).$$

Proof:

Exercise 3.1 Let $S = \{1, 2, 3, \ldots, 60\}$. If x is chosen from S at random what is the probability that x is divisible by 2 **or** divisible by 3 **or** divisible by 5 ?
Generalize the additive rule to three subsets A, B, C.

3.2 Conditional Probability and Independence

3.2.1 Lecture 8 (Grimaldi 3.6)

Example 3.6 Let S be the set of binary sequences of length 8.
Let $A \subset S$ be the sequences starting with 111.
Let B be the sequences in S with five 1's.
Suppose we pick x from A at random. What is $Pr(B)$?

Definition 3.2.1 — Conditional Probability. Let S be a sample space and A and B two subsets of S. The **conditional probability** of B given/knowing A, denoted by

$$Pr(B|A)$$

is the probability that a random outcome from A also belongs to B. It can be obtained by the formula

$$Pr(B|A) = \frac{Pr(B \cap A)}{Pr(A)}.$$

Example 3.7 Assume two dice are rolled. What is the probability that, if they sum up to at least 9 (event A) that both dice have the same value (event B).

Four consequences of $Pr(B|A) = \dfrac{Pr(B \cap A)}{Pr(A)}$**:**

1. Switching A and B:

2. Multiplicative rule:

3. Law of total probability:

4. Bayes' Theorem:

Example 3.8 Suppose 10% of olympic cyclists use steroid Z and the IOC develops a test for Z with the following properties.

1. If a cyclist is taking Z the probability they test positive is 0.99.
2. If they are not taking Z the probability they test positive is 0.05.

Question: If a randomly chosen cyclist tests positive for Z, what is the probability they are taking steroid Z.

Independence

Definition 3.2.2 — independent events. Two events A and B are **independent** if either one of them has probability 0 or both have positive probability and

$$Pr(B|A) = Pr(B) \text{ and } Pr(A|B) = Pr(A).$$

For example, if we toss a coin twice, the first toss is independent of the second.

Theorem 3.2.1 Two events A and B are independent if and only if

$$Pr(A \cap B) = Pr(A)Pr(B).$$

Proof:

Example 3.9 Suppose Alex tosses a fair coin 3 times. Here the sample space

$$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Consider the events

A: The first toss is a H: $A = \{HHH, HHT, HTH, HTT\}$.
B: The second toss is a H: $B = \{HHH, HHT, THH, THT\}$.
C: There are 2 or 3 heads: $C = \{HHH, HHT, HTH, THH\}$.

Are A and B independent?

Are A and C independent?

Are A and $\bar{B}$ independent?

3.3 Discrete Random Variables

3.3.1 Lecture 9 (Grimaldi 3.7)

Definition 3.3.1 — Random Variable. Let S be a sample space. A **random variable** X on S is a function $X : S \to \mathbb{R}$ that associates a numerical value to each possible outcome.

Example 3.10 If S is the set of all binary sequences of size 4. The function that counts the number of 1's in a random variable.

Example 3.11 If S is the set of all rolls of two dice. The function that adds the values of the dice is a random variable.

Example 3.12 Suppose we throw m balls into n bins randomly. Let X be the number of empty bins.

The **range** $r(X)$ of X is the set of all values it can take.

Definition 3.3.2 Let S be a sample space and X a random variable on S. Let x be a value from the range of X. The **probability of** x, denoted by

$$Pr(X = x)$$

is the sum of the probabilities of all outcomes s of S such that $X(s) = x$.

Example 3.13 (example 3.10 cont.) Let $X(s)$ be the number of 1 bits in a binary string s with 4 bits. Here $r(X) = \{0, 1, 2, 3, 4\}$.

$Pr(X = 0) =$
$Pr(X = 1) =$
$Pr(X = 2) =$
$Pr(X = 3) =$
$Pr(X = 4) =$

Definition 3.3.3 — Expected Value. The **expected value** of a random variable X on a sample space S is defined by

$$E(X) = \sum_{x \in r(X)} xPr(X = x) = \sum_{s \in S} X(s)Pr(s).$$

Example 3.14 (example 3.10 cont.) Let $X(s)$ be the number of 1 bits in a binary string s with 4 bits. Here $r(X) = \{0,1,2,3,4\}$ and

x	0	1	2	3	4
$Pr(X = x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

$$E(X) = \sum_{x \in r(X)} xPr(X = x) =$$

The Geometric Distribution

(reference Example 9.18 on page 428 Grimaldi)

Example 3.15 On average, how many times must we roll a fair die before we get a 6?

Example 3.15 continued.

Summary: We say T is geometrically distributed with parameter p and $\Pr(T = k) = p(1-p)^{k-1}$ for $k \geq 1$ and $E(T) = 1/p$.

The Binomial Distribution

Example 3.16 Suppose we toss a biased coin n times. Let the probability of getting heads be $p = 0.7$ and tails be $q = 0.3$. Let H be the number of heads. What is $\Pr(H = k)$ and $E(H)$?

Summary: We say X is binomially distributed with parameters p and n when $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ for $0 \leq k \leq n$ and $E(X) = np$.

3.4 Applications of Discrete Random Variables

3.4.1 Lecture 10

The bins and balls problem

Suppose we throw m balls into n bins randomly. On average, how many bins will be empty? On average, how many bins will have one ball in them?

The coupon collectors problem

Suppose we have a bin containing n types of coupons and we draw coupons one at a time from the bin at random. Assume the probability of drawing each type of coupon is $1/n$ and the bin has a very large number of coupons. On average, how many draws do we need to make until we get all n coupons?

We first review the *expected value* of a random variable.

Definition 3.4.1 Let S be a sample space and X a random variable on S. Let x be a value from the range of X. The probability of x, denoted by $Pr(X = x)$ is the sum of the probabilities of all outcomes s of S such that $X(s) = x$.

Example 3.17 Let S be the set of all binary sequences of size $n = 3$ bits.
Let $X(s)$ be the number of 1 bits in a binary string $s \in S$.
Here the range of X denoted $r(X)$ is $\{0, 1, 2, 3\}$.

$\Pr(X = 0) =$

$\Pr(X = 1) =$

$\Pr(X = 2) =$

$\Pr(X = 3) =$

$\Pr(X = k) =$

Definition 3.4.2 The **expected value** of a random variable X on a sample space S is defined by

$$E(X) = \sum_{x \in r(X)} xPr(X = x) = \sum_{s \in S} X(s)Pr(s).$$

Example 3.18 — (3.17 continued).

x	0	1	2	3
$Pr(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$E[X] = \sum_{x \in r(X)} xPr(X = x) =$$

Theorem 3.4.1 — Linearity of Expection. Let X and Y be two random variables on the same sample space S and $a \in \mathbb{R}$. Then

(1) $E(aX) = aE(X)$ and

(2) $E(X+Y) = E(X) + E(Y)$.

Proof:

Example 3.19 The bins and balls problem.
Suppose we throw m balls into n bins randomly.
Question 1: What is the probability that bin i has k balls?

Question 2: On average, how many bins are empty?

Exercise 3.2 On average, how many bins have one ball? ■

Example 3.20 The coupon collectors problem
Suppose a large bin contains many copies of $n = 10$ coupons. Assuming there are an equal number of each coupon, if we draw coupons at random from the bin, on average, how many draws will it take to get all n coupons?

Exercise 3.3 On average, how many times must you toss a fair coin before you get a head and a tail? ■

3.5 Summary

This chapter was dedicated to discrete probabilities. A fundamental step in answering a probability question is to properly set up the problem in terms of the *sample space* and the *events*, which are subsets of the sample space.

- If outcomes are equally likely, the ratio of the event size over the sample space size gives the probability. Events and sample space sizes are obtained using the counting techniques seen in Section 3.1.
- Otherwise the axioms of probability are used to compute probabilities.

A conditional probability $Pr(B|A)$ is the probability of an event B happening knowing (conditional to) the event A having happened. Computing a conditional probability $Pr(B|A)$ can be done by redefining the sample space as being A and then computing the product $Pr(A \cap B)\,\mathrm{Pr}(A)$. An important part of conditional probability theory is Bayes' Theorem.

Finally, a random variable is a function from a sample space. It associates a number (in general an integer but not always) to an outcome.

- the probability of a random value X taking value x is the sum of the probabilities of the outcomes s such that $X(s) = x$.
- The expected value of a random variable is the average of all possible values, each value being weighted by its probability.
- Expected values behave linearly: the sum of a set of expected values is the expected value of their sum.

Throughout this chapter we saw many examples based on binary sequences, dice, urns, and you should take to time to make sure you understand them very well.

We also did see an example of testing for a disease in which the surprising outcome illustrates the need for being well educated in probability theory.

Vocabulary

- sample space, outcome, event, equal likelihood, probability, complement
- conditional probability, Bayes' theorem, independent events
- random variable, expected value

Skills to acquire

- For a given probability question, identify clearly its set-up (sample space, events, probabilities of outcomes) and use the appropriate formula to compute the required probability.
- Know and know how to use the axioms of probability.
- Computing a conditional probability.
- Manipulating conditional probability formulas, including Bayes' theorem.
- Setting up proper random variables when one is asked to compute some expected property of a set of outcomes.
- Computing probability and expected value for a given random variable.

Basic Facts

- An *outcome* s is a possibility.
- A *sample space* S is a set of outcomes.
- An *event* A is a subset of the sample space.
- The *complement* $\overline{A}$ of an event A is the set of all outcomes not in A.
- A *probability* is a number between 0 and 1: $0 \Rightarrow$ impossible, $1 \Rightarrow$ certainty.
- The probability that an event B has occurred given that event A has occurred is called the *conditional probability* of B given A.
- Events A and B are said to be *independent events* if $Pr(A|B) = Pr(A)$.

Basic Formulas for Computing Probabilities

- The probability of an event A, denoted $Pr(A)$ is: $Pr(A) := \sum_{x\in A} Pr(s)$.
- If each outcome in a probability model is equally likely then $Pr(a) = \frac{|A|}{|S|}$.
- The axioms of probability:
- 1. $\forall A \subseteq S,\ 0 \leq Pr(A) \leq 1$
  2. $Pr(S) = 1$
  3. If A and B are *disjoint* subsets of S then $Pr(A \cup B) = Pr(A) \cup Pr(B)$.
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- $Pr(A \cap B) = 0$ if A and B are disjoint
- $Pr(A \cap B) = Pr(A)Pr(B)$ if A and B are independent
- $Pr(A \cap B) = Pr(A)Pr(B|A) = Pr(B)Pr(A|B)$
- $Pr(\overline{A}) = 1 - Pr(A)$
- $Pr(B|A) = \dfrac{Pr(B \cap A)}{Pr(A)}$
- $Pr(B) = Pr(B \cap A) + Pr(B \cap \overline{A})$
- Total Probability: $Pr(B) = Pr(A)Pr(B|A) + Pr(\overline{A})Pr(B|\overline{A})$
- Bayes' theorem: $Pr(B|A) = \dfrac{Pr(A|B)Pr(B)}{Pr(A)}$

Basic Facts about Random Variables

- A random variable is a function $X : S \to \mathbb{R}$
- $Pr(X = x) = \sum_{s\in S, X(s)=x} Pr(s)$
- The expected value $E(X) = \sum_{x\in \mathrm{range}(X)} xPr(X = x) = \sum_{s\in S} X(s)Pr(s)$
- (linearity of expectation) For X, Y two random variables, and a, c two numbers:

$$E(X + c) = E(X) + c, \qquad E(X + Y) = E(X) + E(Y), \qquad E(aX) = aE(X)$$