Chapter 5

DISCRETE PROBABILITY

5.1 An Introduction to Discrete Probability
5.2 Probability Theory
5.3 Expected Value and Variance
5.1  INTRO TO DISCRETE PROBABILITY

The study of probability uses special jargon.

**DEF:** A *sample space* is a nonempty set.

**DEF:** An *experiment* is a process that produces a point in a sample space.

**DEF:** An *event* is a subset of a sample space.

**DEF:** The *event space* is the power set of the sample space.

**Experiment 5.1.1:** Toss a coin.
Sample space: $U = \{H, T\}$
Event space: $P(U) = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$.

**Remark:** Often, some of the more interesting events have special names.

**Experiment 5.1.2:** Toss two coins.
Sample space: $U = \{HH, HT, TH, TT\}$
Event space: subsets of U
Named events: *match* = \{HH, TT\}
*at least one head* = \{HT, TH, HH\}
UNIFORM PROBABILITY MEASURE

DEF: The uniform probability measure on a finite sample space $S$ assigns to each event $E$ the probability

$$p(E) = \frac{|E|}{|S|}$$

Experiment 5.1.1, continued: coin toss.
Sample space: $U = \{H, T\}$

<table>
<thead>
<tr>
<th>Event</th>
<th>$\emptyset$</th>
<th>${H}$</th>
<th>${T}$</th>
<th>${H, T}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2}$</td>
<td>1</td>
</tr>
</tbody>
</table>

This example is readily generalized.

Experiment 5.1.2, continued: two coins
Sample space: $U = \{HH, HT, TH, TT\}$

<table>
<thead>
<tr>
<th>Event Name</th>
<th>at least one head</th>
<th>one of each</th>
</tr>
</thead>
<tbody>
<tr>
<td>Event</td>
<td>${HT, TH, HH}$</td>
<td>${HT, TH}$</td>
</tr>
<tr>
<td>Probability</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{2}{4}$</td>
</tr>
</tbody>
</table>
Experiment 5.1.3: roll a die  
Sample space: \( U = \{1, 2, 3, 4, 5, 6\} \)  
Event Name \quad odd \quad even \quad 2 \mod 3  
Event \quad \{1, 3, 5\} \quad \{2, 4, 6\} \quad \{2, 5\}  
Probability \quad \frac{3}{6} = \frac{1}{2} \quad \frac{3}{6} = \frac{1}{2} \quad \frac{2}{6} = \frac{1}{3}  

Experiment 5.1.4: roll two dice  
Sample space: \( U = \{11, \ldots, 16, \ldots, 61, \ldots, 66\} \)  
Event Name \quad doubles \quad sum is nine  
Event \quad \{11, 22, \ldots, 66\} \quad \{36, 45, 54, 63\}  
Probability \quad \frac{6}{36} = \frac{1}{6} \quad \frac{4}{36} = \frac{1}{9}  

Experiment 5.1.5: roll three dice  
Sample space: \( U = \{111, \ldots, 116, \ldots, 661, \ldots, 666\} \)  
Event Name \quad triples \quad doubles \quad singles  
Event \quad \{111, \text{ etc.}\} \quad \{112, \text{ etc.}\} \quad \{123, \text{ etc.}\}  
Probability \quad \frac{6}{216} = \frac{1}{36} \quad \frac{90}{216} = \frac{5}{12} \quad \frac{120}{216} = \frac{5}{16}
Experiment 5.1.6: A base-10 numeral is randomly chosen from the range 000...999.

Q1: What is the probability that the numeral contains no 3’s or 5’s?

Ans. There are $8^3$ base-10 numerals containing no 3’s or 5’s and $10^3$ three-digit numerals altogether. Thus, the probability is

$$\frac{8^3}{10^3} = \frac{512}{1000} = 0.512$$

Q2: What is the probability that the numeral contains one 3 and no 5’s?

Ans. The 3 could occur as each of the three digits. There would be $8^2$ possibilities for the other two digits. Thus, the probability is

$$\frac{3 \cdot 8^2}{10^3} = \frac{192}{1000} = 0.192$$
DEF: *(standard) deck of cards*

\[
\begin{align*}
2\spadesuit & \quad 3\spadesuit \quad \ldots \quad 10\spadesuit & \quad J\spadesuit & \quad Q\spadesuit & \quad K\spadesuit & \quad A\spadesuit \\
2\diamondsuit & \quad 3\diamondsuit \quad \ldots \quad 10\diamondsuit & \quad J\diamondsuit & \quad Q\diamondsuit & \quad K\diamondsuit & \quad A\diamondsuit \\
2\heartsuit & \quad 3\heartsuit \quad \ldots \quad 10\heartsuit & \quad J\heartsuit & \quad Q\heartsuit & \quad K\heartsuit & \quad A\heartsuit \\
2\clubsuit & \quad 3\clubsuit \quad \ldots \quad 10\clubsuit & \quad J\clubsuit & \quad Q\clubsuit & \quad K\clubsuit & \quad A\clubsuit
\end{align*}
\]

**Experiment 5.1.7:** deal one card from deck

Sample space: standard 52-card deck

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Event</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>heart seven</td>
<td>{2\heartsuit, \ldots, A\heartsuit}</td>
<td>{7\spadesuit, 7\diamondsuit, 7\heartsuit, 7\clubsuit}</td>
</tr>
<tr>
<td>Probability</td>
<td>(\frac{13}{52} = \frac{1}{4})</td>
<td>(\frac{4}{52} = \frac{1}{13})</td>
</tr>
</tbody>
</table>

**Experiment 5.1.8:** deal five cards from deck

Sample space: all possible 5-card hands.

<table>
<thead>
<tr>
<th>Event Name</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 of a kind</td>
<td>(\frac{\binom{13}{1} \cdot \binom{48}{4}}{\binom{52}{5}})</td>
</tr>
<tr>
<td>full house</td>
<td>(\frac{\binom{13}{1} \cdot \binom{4}{3} \cdot \binom{12}{1} \cdot \binom{4}{2}}{\binom{52}{5}})</td>
</tr>
<tr>
<td>3 of a kind</td>
<td>(\frac{\binom{13}{1} \cdot \binom{4}{3} \cdot \binom{48}{4} \cdot \binom{44}{1}}{\binom{52}{5}})</td>
</tr>
</tbody>
</table>
INFINITE SAMPLE SPACES (optional)

If a sample space is finite, it is sufficient to specify the probabilities of singletons, and to infer the probabilities of compound events by addition. However, this does not work for infinite sample spaces.

**Experiment 5.1.9:** pick a natural number

Probabilities of singletons

\[ 0 = p(0) = p(1) = p(2) = \cdots \]

do not yield probabilities of infinite subsets:

\[ p(\text{even}), p(\text{power of two}) \]

Uniform probability for an arbitrary set \( E \) of numbers is defined by the rule

\[ p(E) = \lim_{n \to \infty} \frac{|E \cap \{0, 1, \ldots, n-1\}|}{n} \]

This illustrates why the most general kind of probability measures (discussed elsewhere, but not here) are assigned to the event space, rather than to the sample space.
PROBABILITY of the COMPLEMENTARY EVENT

Proposition 5.1.1. Let $E$ be an event in a finite sample space $S$, under the uniform probability distribution. Then

$$p(\overline{E}) = 1 - p(E)$$

Proof:

1. $|S| = |E| + |\overline{E}|$ by Rule of Sum
2. $\frac{|S|}{|S|} = \frac{|E|}{|S|} + \frac{|\overline{E}|}{|S|}$
3. $1 = p(|E|) + p(\overline{E})$
4. $p(\overline{E}) = 1 - p(E)$ \(\diamondsuit\)

Experiment 5.1.10: A (putatively fair) coin is tossed 10 times. Find the probability of at least one tail.

$$p(\#\text{tails} \geq 1) = 1 - p(10 \text{ heads})$$

Answer:

$$= 1 - \frac{1}{1024}$$

$$= \frac{1023}{1024}$$
PROBABILITY of a UNION of EVENTS

**Proposition 5.1.2.** Let $E_1$ and $E_2$ be events in a finite sample space $S$, under the uniform probability distribution. Then

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

**Proof:**

$$|E_1 \cup E_2| = |E_1| + |E_2| - |E_1 \cap E_2| \quad \text{by Incl-Excl}$$

$$\frac{|E_1 \cup E_2|}{|S|} = \frac{|E_1|}{|S|} + \frac{|E_2|}{|S|} - \frac{|E_1 \cap E_2|}{|S|}$$

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2) \quad \diamondsuit$$

**Experiment 5.1.11:** An integer is chosen from the interval $[1, \ldots, 100]$. Find the probability that it is divisible either by 6 or by 15.

**Solution:**

$$p(6 \mid n \vee 15 \mid n) = p(6 \mid n) + p(15 \mid n) - p(30 \mid n)$$

$$= \frac{16}{100} + \frac{6}{100} - \frac{3}{100}$$

$$= \frac{19}{100}$$
5.2 PROBABILITY THEORY

§5.1 defines the uniform probability measure for an event E in a finite sample space S:

\[ pr(E) = \frac{|E|}{|S|} \]

§5.2 provides a less restrictive form of probability measure for a finite sample space S. For each possible outcome \( s \in S \), there is a probability \( p(s) \), such that

\[ 0 \leq p(s) \leq 1 \quad \text{and} \quad \sum_{s \in S} p(s) = 1 \]

The most general form of probability measure for infinite sample spaces assigns probabilities to all events, rather than to outcomes. In the present context, we define the probability of an event E:

\[ p(E) = \sum_{s \in E} p(s) \]
Example 5.2.1: Uniform prob measure (§5.1) is a probability measure. If the space $S$ has $n$ possible outcomes, then assign each outcome the probability $p(s) = \frac{1}{n}$.

Prop 5.2.1. For any finite sample space $S$, assign singleton probabilities that add to one. Then define the probability of any event to be the sum of the probabilities of the singletons in the event. The result is a probability measure.

Proof: Immediate from the definition.  

Here are two non-uniform probability measures for some familiar sample spaces.

TWO STANDARD EXAMPLES

DEF: The standard loaded coin has probability $p(H) = 0.8$ and $p(T) = 0.2$.

DEF: The standard loaded die has sample space $U = \{1, 2, 3, 4, 5, 6\}$. The singleton $\{j\}$ has probability $\frac{j}{21}$.
PROPERTIES of PROBABILITY MEASURES

Prop 5.2.2. Let \( p \) be a probability measure on a sample space \( S \). Then \( p(\emptyset) = 0 \).

Proof: \( p(S) = p(S \cup \emptyset) = p(S) + p(\emptyset) \). \( \diamond \)

Prop 5.2.3. Let \( p \) be a probability measure on a sample space \( S \), and let \( E \) be an event. Then \( p(\overline{E}) = 1 - p(E) \).

Proof: \( 1 = p(S) = p(E \cup \overline{E}) = p(E) + p(\overline{E}) \). \( \diamond \)

BINOMIAL DISTRIBUTIONS

DEF: A Bernoulli distribution is a probability measure on a sample space with exactly two points.

Example 5.2.2: Flip standard loaded coin.
Sample space = \( \{H, T\} \)
Event space = \( \{\emptyset, \{H\}, \{T\}, \{H, T\}\} \)
\( p(\{H\}) = \frac{4}{5}, p(\{T\}) = \frac{1}{5} \)
Example 5.2.3:  Roll a fair die.
Sample space = $U = \{1, 2, 3, 4, 5, 6\}$
$p(3) = \frac{1}{6}; \quad p(\neg 3) = p(\{1, 2, 4, 5, 6\}) = \frac{5}{6}$

Example 5.2.4:  The standard loaded die may be considered a Bernoulli distribution with
$p(\text{even}) = \frac{4}{7}$ and $p(\text{odd}) = \frac{3}{7}$.

NOTATION: Let $U = \{x, y\}$ be a binary sample space. Then the set
$\{x\ldots xxx, x\ldots xxy, \ldots, y\ldots yyy\}$
of length-$n$ sequences in $U$ is denoted $U^n$.

Prop 5.2.4. A Bernoulli distribution
$p(x) = p, \quad p(y) = 1 - p$ on $U = \{x, y\}$
induces a probability measure on the sample space $U^n$ in which an $n$-string with $k$ occurrences of $x$ and $n - k$ of $y$ has probability $p^k(1 - p)^{n-k}$.

REVIEW:  The number of $n$-strings in $\{x, y\}$ with $k$ occurrences of $x$ and $n - k$ occurrences of $y$ is
\[
\binom{n}{k}
\]
Prop 5.2.5. A Bernoulli distribution
\[ p(x) = p, \quad p(y) = 1 - p \text{ on } U = \{x, y\} \]
induces a probability measure on the sample space \( \{0, 1, \ldots, n\} \) in which
\[ p(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n \]
is the probability of obtaining \( k \) occurrences of \( x \) and \( n - k \) occurrences of \( y \).

Def: A binomial probability distribution on the sample space \( \{0, 1, \ldots, n\} \) is a probability induced by a Bernoulli distribution as in the Prop 5.2.5.

Example 5.2.5: Flip standard loaded coin 10 times. Then
\[ p(k \text{ heads}) = \binom{10}{k} 0.8^k 0.2^{10-k}, \quad \text{for } k = 0, \ldots, 10 \]

CLASSROOM EXERCISE
Seven standard loaded dice are rolled. What is the probability of three 5’s ??
Example 5.2.6: A fair dime and a fair nickel are flipped. An oracle tells you that at least one of them is heads. What is the probability that the other also is heads?

Analysis: If the oracle says the dime is heads, then there are two equally likely cases for the nickel. Similarly, if the oracle says that the nickel is heads, then there are two equally likely cases for the dime.

Question: How many cases are there if the oracle doesn’t say which coin?

Suggestion: Program a computer to initialize the variables $n$ and $h2$ at 0. Run 10,000 trials. On each trial, generate two random real numbers between 0 and 1. If at least one is less than 0.5, then increment the variable $n$ by 1. If both are less than 0.5, then also increment $h2$ by 1. Calculate the fraction $\frac{h2}{n}$. 
DEF: Let $p$ be a probability distribution on a sample space $U$, and let $Y$ be an event. Then the conditional probability of event $E$ given that event $Y$ has occurred is

$$p(E \mid Y) = \frac{p(E \cap Y)}{p(Y)}$$

**Example 5.2.6:** continued.

$$p(HH \mid \neg TT) = \frac{p(HH \land \neg TT)}{p(\neg TT)} = \frac{p(HH)}{p(\neg TT)} = \frac{1/4}{3/4} = \frac{1}{3}$$

**Example 5.2.7:** Two fair dice are rolled. At least one is a four. What is the probability that both are fours?

$$p(44 \mid 1\text{-or-2} \text{ 4's}) = \frac{p(44 \land 1\text{-or-2} \text{ 4's})}{p(1\text{-or-2} \text{ 4's})} = \frac{1/36}{11/36} = \frac{1}{11}$$
DEF: Let \( p \) be a probability distribution on a sample space \( U \). Event \( E \) is \textbf{probabilistically independent} of event \( Y \) if \( p(E \mid Y) = p(E) \).

**Prop 5.2.6.** Let \( p \) be a probability distribution on a sample space \( U \), and let event \( E \) be probabilistically independent of event \( Y \). Then event \( Y \) is probabilistically independent of event \( E \).

**Proof:** \( p(E) = p(E \mid Y) = p(E \cap Y)/p(Y) \). Therefore,

\[
p(Y) = \frac{p(E \cap Y)}{p(E)} = \frac{p(Y \cap E)}{p(E)} = p(Y \mid E) \quad \diamond
\]

**Example 5.2.8:** Roll two standard loaded dice. Let \( Y \) be the event that the first die is a one, and \( E \) the event that the sum is odd. Then

\[
p(E) = 2 \cdot \frac{3}{7} \cdot \frac{4}{7} = \frac{24}{49} \quad \text{and} \quad \frac{p(12, 14, 16)}{1/21}
\]

\[
p(E \mid Y) = \frac{p(E \cap Y)}{p(Y)} = \frac{12}{21} = \frac{4}{7}
\]

\[
= \frac{2/441 + 4/441 + 6/441}{1/21} = \frac{12}{21} = \frac{4}{7}
\]
Example 5.2.9: Roll two fair dice. Let $Y$ be the event that the first die is a one, and $E$ the event that the sum is odd. Then

$$p(E) = 2 \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

and

$$p(E | Y) = \frac{p(E \cap Y)}{p(Y)} = \frac{p(12, 14, 16)}{1/6} = \frac{3/36}{1/6} = \frac{3}{6} = \frac{1}{2}$$

Remark: Disjoint events are usually NOT probabilistically independent.
RANDOM VARIABLES

DEF: A probability space is a pair consisting of a sample space $U$, called its domain and a probability measure on $U$.

DEF: A random variable is a real-valued function on the domain of a probability space.

Remark: Since a random variable is a function, it is not a variable, and it is not random.

Example 5.2.10: Flip a coin three times. Let $X(t)$ be the number of heads that occurs. Then

\[
\begin{align*}
X(TTT) &= 0 \\
X(TTH) &= X(THT) = X(HTT) = 1 \\
X(THH) &= X(HHT) = X(HTH) = 2 \\
X(HHH) &= 3
\end{align*}
\]
5.3 EXPECTED VALUE AND VARIANCE

DEF: The expected value of a random variable $X$ on a probability space $(S, p)$ is the sum

$$E(X) = \sum_{s \in S} X(s)p(s)$$

Example 5.3.1: The expected outcome of a fair die is

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

Example 5.3.2: The expected outcome of the standard loaded die is

$$1 \cdot \frac{1}{21} + 2 \cdot \frac{2}{21} + \cdots + 6 \cdot \frac{6}{21} = \frac{91}{21} = \frac{13}{3}$$

Over a non-uniform probability space the expected value of a random variable is the weighted mean.
Example 5.3.3: Flip a fair coin three times. The expected number of heads is

\[
\frac{0 \cdot 1}{8} + \frac{1 \cdot 3}{8} + \frac{2 \cdot 3}{8} + \frac{3 \cdot 1}{8} = \frac{12}{8} = \frac{3}{2}
\]

This calculation is based on the binomial distribution.

Example 5.3.4: Flip a standard loaded coin three times. The expected number of heads is

\[
\frac{0 \cdot 1}{125} + \frac{1 \cdot 12}{125} + \frac{2 \cdot 48}{125} + \frac{3 \cdot 64}{125} = \frac{0 + 12 + 96 + 192}{125} = \frac{300}{125} = \frac{12}{5}
\]

This calculation too is based on the binomial distribution.
SUMMING RANDOM VARIABLES

**Theorem 5.3.1.** Let $X_1$ and $X_2$ be random variables on a probability space $(S, p)$. Then

$$E(X_1 + X_2) = E(X_1) + E(X_2)$$

**Proof:**

$$E(X_1 + X_2) = \sum_{s \in S} (X_1(s) + X_2(s))p(s)$$

$$= \sum_{s \in S} X_1(s)p(s) + X_2(s)p(s)$$

$$= \sum_{s \in S} X_1(s)p(s) + \sum_{s \in S} X_2(s)p(s)$$

$$= E(X_1) + E(X_2)$$

**Example 5.3.5:** When two fair dice are rolled, here are both calculations:

$$E(X_1) + E(X_2) = \frac{7}{2} + \frac{7}{2} = 7$$

and

$$E(X_1 + X_2) = \frac{1}{36} \sum_{j=1}^{6} \sum_{k=1}^{6} (j + k) = \frac{252}{36} = 7$$
Theorem 5.3.2. Let $X_1, \ldots, X_n$ be random variables on a probability space $(S, p)$. Then

$$E(X_1 + \cdots + X_n) = E(X_1) + \cdots + E(X_n)$$

Proof: By induction on $n$, using Thm 5.3.1. ♦

Example 5.3.6: When 100 fair coins are tossed, the expected number of heads is

$$\frac{1}{2} \cdot 100 = 50$$

Example 5.3.7: When 100 standard loaded coins are tossed, the expected number of heads is

$$0.8 \cdot 100 = 80$$
GEOMETRIC DISTRIBUTION

DEF: The geometric distribution on the positive integers is

\[ pr(k) = (1 - p)^{k-1} p \]

Example 5.3.8: A coin with \( p(H) = p \) is tossed until the first occurrence of heads. Then the probability of requiring exactly \( k \) tosses is \( (1 - p)^{k-1} p \). We observe that

\[
\sum_{k=1}^{\infty} (1-p)^{k-1} p = p \sum_{k=1}^{\infty} (1-p)^{k-1} = \frac{p}{1 - (1-p)} = 1
\]

It is proved in the text that

\[
E(X) = \sum_{k=1}^{\infty} (1 - p)^{k-1} pk
= p \frac{d}{dx} (1 - x)^{-1}\bigg|_{x=1-p} = \frac{1}{p}
\]
INDEPENDENT RANDOM VARIABLES

DEF: The random variables $X$ and $Y$ on the probability space $(S, p)$ are independent if for all real numbers $r_1$ and $r_2$

$$p(X = r_1 \land Y = r_2) = p(X = r_1) \cdot p(Y = r_2)$$

Example 5.3.9: Suppose that $X$ is the sum of two fair dice and $Y$ is the product. Then

$$p(X = 2) = \frac{1}{36} \quad \text{and} \quad p(Y = 5) = \frac{1}{18}$$

However,

$$p(X = 2 \land Y = 5) = 0 \neq \frac{1}{36} \cdot \frac{1}{18}$$

Thus $X$ and $Y$ are not independent.
VARIANCE and STANDARD DEVIATION

DEF: The **variance** of a random variable $X$ on a probability space $(S, p)$ is the sum

$$\sigma^2(X) = V(X) = \sum_{s \in U} (X(s) - E(X))^2 p(s)$$

DEF: The **standard deviation** of a random variable $X$ on a probability space $(U, p)$ is

$$\sigma(X) = \sqrt{V(X)}$$

**Example 5.3.10:** Flip a fair coin three times. The variance of the number of heads is

$$\left(0 - \frac{3}{2}\right)^2 \cdot \frac{1}{8} + \left(1 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(2 - \frac{3}{2}\right)^2 \cdot \frac{3}{8} + \left(3 - \frac{3}{2}\right)^2 \cdot \frac{1}{8}$$

$$= \frac{9}{4} \cdot \frac{1}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{1}{4} \cdot \frac{3}{8} + \frac{3}{4} \cdot \frac{1}{8} = \frac{24}{32} = \frac{3}{4}$$

The standard deviation is

$$\sqrt{\frac{3}{4}} = \frac{\sqrt{3}}{2}$$
Example 5.3.11: Flip a standard loaded coin three times. The variance of the number of heads is

\[
\begin{align*}
&= (0 - \frac{12}{5})^2 \cdot \frac{1}{125} + (1 - \frac{12}{5})^2 \cdot \frac{12}{125} \\
&\quad + (2 - \frac{12}{5})^2 \cdot \frac{48}{125} + (3 - \frac{12}{5})^2 \cdot \frac{64}{125} \\
&= \frac{144 + 49 \cdot 12 + 4 \cdot 48 + 9 \cdot 64}{5^5} = \frac{1500}{5^5} = \frac{12}{25}
\end{align*}
\]

The standard deviation is

\[
\sqrt{\frac{12}{25}} = \frac{2\sqrt{3}}{5}
\]

**CHEBYSHEV INEQUALITY**

Chebyshev Inequality. Let \( X \) be a random variable on any probability space that has a mean and a variance. Then

\[
p(|X(s) - E(X)| \geq k\sigma(X)) \leq \frac{1}{k^2}
\]