Review of Basic Probability Theory

2.1 PROBABILITY SPACE AND AXIOMS

Probability theory provides a set of mathematical rules to assign probabilities to outcomes of random experiments, e.g., coin flips, packet arrivals, stock prices, neural spikes, noise voltages, and so on. Given a random experiment, its sample space \( \Omega \) is the set of all outcomes. An event is a subset of the sample space and we say that an event \( A \subseteq \Omega \) occurs if the outcome \( \omega \) of the random experiment is an element of \( A \). Let \( \mathcal{F} \) be a set of events. A probability measure \( P : \mathcal{F} \rightarrow [0, 1] \) is a function that assigns probabilities to the events in \( \mathcal{F} \). We refer to the triple \((\Omega, \mathcal{F}, P)\) as the probability space of the random experiment.

The probability measure \( P \) must satisfy the following.

Axioms of probability.
1. \( P(A) \geq 0 \) for every event \( A \) in \( \mathcal{F} \).
2. \( P(\Omega) = 1 \).
3. Countable additivity. If \( A_1, A_2, \ldots \) are disjoint, i.e., \( A_i \cap A_j = \emptyset, i \neq j \), then
   \[
P\left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i).
   \]

For the probability measure \( P \) to be well-defined over all events of interest, the set of events \( \mathcal{F} \) must satisfy:
1. \( \emptyset \in \mathcal{F} \).
2. If \( A \in \mathcal{F} \), then \( A^c \in \mathcal{F} \).
3. If \( A_1, A_2, \ldots \in \mathcal{F} \), then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \).

Due to these defining properties, \( \mathcal{F} \) is often referred to as a \( \sigma \)-algebra or \( \sigma \)-field.
2.2 DISCRETE PROBABILITY SPACES

A probability space \((\Omega, \mathcal{F}, P)\) is said to be discrete if the sample space \(\Omega\) is countable, i.e., finite or countably infinite.

Example 2.1 (Flipping a coin). \(\Omega = \{H, T\}, \mathcal{F} = \{\emptyset, \{H\}, \{T\}, \Omega\},\) and
\[
P(\emptyset) = 0, \quad P(\{H\}) = p, \quad P(\{T\}) = 1 - p, \quad P(\Omega) = 1,
\]
where \(p \in [0, 1]\) is the bias of the coin. A fair coin has a bias of 1/2.

For discrete sample spaces, \(\mathcal{F}\) is often the set of all subsets of \(\Omega\), namely, the power set \(2^\Omega\) of \(\Omega\). (Recall that \(|2^\Omega| = 2^{|\Omega|}|.\) In this case, the probability measure \(P\) can be fully specified by assigning probabilities to individual outcomes (or singletons \(\{\omega\}\)) so that
\[
P(\{\omega\}) \geq 0, \quad \omega \in \Omega,
\]
and
\[
\sum_{\omega \in \Omega} P(\{\omega\}) = 1.
\]
Then it follows by the third axiom of probability that for any event \(A \subseteq \Omega\),
\[
P(A) = \sum_{\omega \in A} P(\{\omega\}).
\]

Example 2.2 (Rolling a fair die). \(\Omega = \{1, 2, 3, 4, 5, 6\}, \mathcal{F} = 2^\Omega = \{\emptyset, \{1\}, \{2\}, \ldots, \Omega\},\) and
\[
P(\{i\}) = \frac{1}{6}, \quad i = 1, 2, \ldots, 6.
\]
The probability of the event \(A\) “the outcome is even,” i.e., \(A = \{2, 4, 6\}\), is
\[
P(A) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6} = \frac{1}{2}.
\]

Example 2.3 (Flipping a coin \(n\) times). A coin with bias \(p\) is flipped \(n\) times. Then
\[
\Omega = \{H, T\}^n = \text{sequences of heads/tails of length } n,
\]
\[
\mathcal{F} = 2^\Omega,
\]
\[
P(\{\omega\}) = p^i(1-p)^{n-i},
\]
where \(i\) is the number of heads in \(\omega\). The probability of the event \(A_k\) “the outcome consists of \(k\) heads and \(n - k\) tails” is
\[
P(A_k) = \sum_{\omega \text{ has } k \text{ heads}} P(\{\omega\}) = \binom{n}{k} p^k (1-p)^{n-k}.
\]
We can verify that
\[
P(\Omega) = \sum_{k=0}^{n} P(A_k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.
\]
Example 2.4 (Flipping a coin until the first head). \( \Omega = \{ H, \, TH, \, TTH, \, TTTT, \ldots \} \), \( \mathcal{F} = 2^\Omega \), and
\[
P(\{ \omega \}) = (1 - p)^i p,
\]
where \( i \) is the number of tails in \( \omega \). Again we can verify that
\[
P(\Omega) = \sum_{\omega \in \Omega} P(\{ \omega \}) = \sum_{i=0}^{\infty} (1 - p)^i p = 1.
\]

Example 2.5 (Counting the number of packets). Consider the number of packets arriving at a node in a communication network in time interval \((0, T]\) at rate \( \lambda \in (0, \infty) \). Then, \( \Omega = \{ 0, 1, 2, 3, \ldots \} \), \( \mathcal{F} = 2^\Omega \), and
\[
P(\{ k \}) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}, \quad k = 0, 1, 2, \ldots,
\]
provided that the number of packets are Poisson distributed. Note that
\[
P(\Omega) = \sum_{k=0}^{\infty} \frac{(\lambda T)^k}{k!} e^{-\lambda T} = 1.
\]

In all examples so far, \( \mathcal{F} = 2^\Omega \). This is not necessarily the case.

Example 2.6. (Rolling a colored die). Suppose that each face of a die is colored, say, 1 and 2 are red, and 3 through 6 are blue. Further suppose that the observer of a die roll can only note the color of the face, not the actual number. Then \( \Omega = \{ 1, 2, 3, 4, 5, 6 \} \) as before, but
\[
\mathcal{F} = \{ 0, \{ 1, 2 \}, \{ 3, 4, 5, 6 \}, \Omega \}.
\]
This is a valid \( \sigma \)-algebra (check!), but it is much smaller in size than the previous case and the probability measure is fully specified by \( P(\{ 1, 2 \}) \) alone. As an extreme, if all six faces are of the same color, then we have the trivial \( \sigma \)-algebra \( \mathcal{F} = \{ 0, \Omega \} \), which is still valid but hardly interesting. Thus, the choice of \( \mathcal{F} \) controls the level of granularity at which one can assign probabilities.

2.3 Continuous Probability Spaces

A continuous probability space has an uncountable number of elements in \( \Omega \). Unlike the discrete case, the choice of \( \mathcal{F} = 2^\Omega \), albeit valid, is too rich to admit an interesting probability measure under the standard axioms of probability. At the same time, specifying probabilities to singletons is not sufficient to extrapolate probabilities for other events. Hence, \( \mathcal{F} \) should be chosen more carefully, which is the main reason behind the intricate definitions of probability measure and \( \sigma \)-algebra.

Suppose that \( \Omega \) is the real line \( \mathbb{R} \) or its subinterval, e.g., \([0, 1] = \{ x \in \mathbb{R} : 0 \leq x \leq 1 \} \).
Then the set of events is typically taken to contain all open subintervals of \( \Omega \), i.e., all intervals of the form \((a, b)\), \(a, b \in \Omega\). More formally, let \( \mathcal{F} \) be the smallest \( \sigma \)-algebra that contains all open subintervals in \( \Omega \). This \( \sigma \)-algebra is commonly referred to as the Borel \( \sigma \)-algebra \( \mathcal{B} \) and accordingly each event in \( \mathcal{B} \) is called a Borel set.

Since \( \mathcal{B} \) is a \( \sigma \)-algebra, it is closed under complement, countable unions, and countable intersections (cf. Problem 2.1), and contains many subsets other than open intervals. For example, since the half-open interval \((a, b]\) can be represented by a countable intersection of open intervals (Borel sets) as

\[
(a, b] = \bigcap_{c \in \mathbb{Q} : c > b} (a, c),
\]

it is also Borel. As a matter of fact, \( \mathcal{B} \) contains all open subsets and thus is the smallest \( \sigma \)-algebra that contains all open subsets of \( \Omega \). The probability of any Borel set can be fully specified by assigning probabilities to open intervals (or to closed intervals, half-closed intervals, half-intervals, etc.).

**Example 2.7 (Picking a random number between 0 and 1).** \( \Omega = [0, 1] \), \( \mathcal{F} = \mathcal{B} \), and

\[
P((a, b)) = b - a, \quad 0 \leq a < b \leq 1.
\]

This is the *uniform* distribution over \( \Omega \). By (2.1) and the axioms of probability,

\[
P((a, b]) = \lim_{c \to b} (c - a) = b - a, \quad 0 \leq a < b \leq 1,
\]

It can be similarly checked that

\[
P([a, b]) = b - a, \quad 0 \leq a < b \leq 1.
\]

In particular, \( P([a]) = 0, a \in [0, 1] \), and the probability of picking any specific number is zero.

For any reasonable \( \Omega \) (such as a finite set or the \( d \)-dimensional Euclidean space \( \mathbb{R}^d \), but sometimes even a space of time series or functions), the Borel \( \sigma \)-algebra can be defined as the smallest \( \sigma \)-algebra that contains all open subsets. When \( \Omega \) is countable, the Borel \( \sigma \)-algebra is \( 2^\Omega \). Henceforth, we assume that \( \mathcal{F} \) is the Borel \( \sigma \)-algebra of \( \Omega \) and any event of our interest is Borel unless specified otherwise. Note, however, that for an uncountable \( \Omega \), there are many subsets of \( \Omega \) that are not Borel (if interested in these sets, refer to any graduate-level course on *measure theory*).

### 2.4 BASIC PROBABILITY LAWS

We can establish the following as simple corollaries of the axioms of probability.

1. \( P(A^c) = 1 - P(A) \).
2. If $A \subseteq B$, then $P(A) \leq P(B)$.

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

4. $P(A \cup B) \leq P(A) + P(B)$.

More generally, we have the following inequality, also known as *Boole's inequality*, that can be generalized to a countably infinite number of events.

**Union of events bound.** For any events $A_1, A_2, \ldots, A_n$,

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)$$

The following identity is very useful in finding the probability of a complicated event.

**Law of total probability.** Let $A_1, A_2, \ldots$ be events that partition $\Omega$, that is, $A_1, A_2, \ldots$ are disjoint ($A_i \cap A_j = \emptyset, i \neq j$) and $\bigcup_i A_i = \Omega$. Then for any event $B$,

$$P(B) = \sum_i P(A_i \cap B).$$

### 2.5 CONDITIONAL PROBABILITY AND THE BAYES RULE

So far probability measures are similar to other common *measures*, such as length, area, volume, and weight, all of which are nonnegative and countably additive. The notion of conditioning is a unique feature of probability theory that is not found in the general measure theory.

Let $B$ be an event such that $P(B) \neq 0$. The *conditional probability* of the event $A$ given $B$ is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

The function $P(\cdot \mid B) : \mathcal{F} \rightarrow [0, 1]$ is a probability measure in itself, that is, it satisfies the three axioms of probability:

1. $P(A \mid B) \geq 0$ for every $A$ in $\mathcal{F}$.
2. $P(\Omega \mid B) = 1$.
3. If $A_1, A_2, \ldots$ are disjoint, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i \mid B).$$
Assume that $P(A) \neq 0$ and $P(B) \neq 0$. Then the conditional probability of $A$ given $B$—the *a posteriori* probability (or *posterior* in short) of $A$—can be related to the unconditional probability of $A$—the *a priori* probability (or *prior* in short) of $A$. Using the definition of conditional probability twice, we have

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)}{P(B)} P(A). \quad (2.2)$$

By multiplying $P(B)$ on both sides, we establish the following useful identity.

**Chain rule.** For any pair of events $A$ and $B$,

$$P(A \cap B) = P(A) P(B \mid A) = P(B) P(A \mid B).$$

Note that the chain rule holds even when $P(A) = 0$ or $P(B) = 0$ if we interpret the product of zero and an undefined number to be zero. By induction, the chain rule can be generalized to more than two events. For example,

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) P(A_2 \mid A_1) P(A_3 \mid A_1 \cap A_2).$$

Let $A_1, A_2, \ldots, A_n$ be nonzero probability events that partition $\Omega$ and let $B$ be a nonzero probability event. By (2.2),

$$P(A_j \mid B) = \frac{P(B \mid A_j)}{P(B)} P(A_j). \quad (2.3)$$

By the law of total probability,

$$P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) P(B \mid A_i) \quad (2.4)$$

Substituting (2.4) into (2.3) yields the famous relationship between the priors $P(A_i)$, $i = 1, 2, \ldots, n$, and the posteriors $P(A_j \mid B)$, $j = 1, 2, \ldots, n$.

**Bayes rule.** If $A_1, A_2, \ldots, A_n$ are nonzero probability events that partition $\Omega$, then for any nonzero probability event $B$,

$$P(A_j \mid B) = \frac{P(B \mid A_j)}{\sum_{i=1}^{n} P(A_i) P(B \mid A_i)} P(A_j), \quad j = 1, 2, \ldots, n.$$

The Bayes rule also applies to a countably infinite number of events.
Example 2.8 (Binary communication channel). Consider the *probability transition diagram* for a noisy binary channel in Figure 2.1. This is a random experiment with sample space

\[ \Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}, \]

where the first entry is the bit sent (the input of the channel) and the second is the bit received (the output of the channel). Define the two events

- \( A = \{0 \text{ is sent}\} = \{(0, 1), (0, 0)\}, \)
- \( B = \{0 \text{ is received}\} = \{(0, 0), (1, 0)\}. \)

The probability measure on \( \Omega \) is determined by \( P(A) \), \( P(B | A) \), and \( P(B^c | A^c) \), which are given on the probability transition diagram. To find \( P(A | B) \), we use Bayes rule:

\[
P(A | B) = \frac{P(B | A)}{P(A) P(B | A) + P(A^c) P(B | A^c)} P(A)
\]

to obtain

\[
P(A | B) = \frac{0.9}{0.2 \cdot 0.9 + 0.8 \cdot 0.025} \cdot 0.2 = \frac{0.9}{0.2} \cdot 0.2 = 0.9.
\]

Note that the posterior \( P(A | B) = 0.9 \) is much larger than the prior \( P(A) = 0.2 \); even though the observation is noisy, it still reveals some useful information about the input.

![Probability transition diagram](image)

*Figure 2.1.* The probability transition diagram for a binary communication channel. Here \( p(\cdot) \) denotes the probability of the input and \( p(\cdot | \cdot) \) denotes the conditional probability of the output given the input.

### 2.6 Independence

Two events \( A \) and \( B \) are said to be *statistically independent* (or *independent* in short) if

\[
P(A \cap B) = P(A) P(B).
\]
When $P(B) \neq 0$, this is equivalent to

$$P(A \mid B) = P(A).$$

In other words, knowing whether $B$ occurs provides no information about whether $A$ occurs.

**Example 2.9.** We revisit the binary channel discussed in Example 2.8. Assume that two independent bits are sent over the channel and we would like to find the probability that both bits are in error. Define the two events

- $E_1 = \{\text{First bit is in error}\}$
- $E_2 = \{\text{Second bit is in error}\}$

Since the bits are sent independently, the probability that both are in error is

$$P(E_1 \cap E_2) = P(E_1) P(E_2).$$

To find $P(E_1)$, we express $E_1$ in terms of the events $A_1$ (0 is sent in the first transmission) and $B_1$ (0 is received in the first transmission) as

$$E_1 = (A_1 \cap B_1^c) \cup (A_1^c \cap B_1).$$

Since $E_1$ has been expressed as the union of disjoint events,

$$P(E_1) = P(A_1 \cap B_1^c) + P(A_1^c \cap B_1)$$

$$= P(A_1) P(B_1^c \mid A_1) + P(A_1^c) P(B_1 \mid A_1^c)$$

$$= 0.2 \cdot 0.1 + 0.8 \cdot 0.025$$

$$= 0.04.$$  

The probability that the two bits are in error is

$$P(E_1 \cap E_2) = P(E_1) P(E_2) = (0.04)^2 = 1.6 \times 10^{-3}.$$  

In general, the events $A_1, A_2, \ldots, A_n$ are said to be *mutually independent* (or *independent* in short) if for every subset $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ of the events,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j}).$$

For example, $A, B,$ and $C$ are independent if all of the following hold:

$$P(A \cap B) = P(A) \ P(B), \quad \text{(2.5)}$$
$$P(A \cap C) = P(A) \ P(C), \quad \text{(2.6)}$$
$$P(B \cap C) = P(B) \ P(C), \quad \text{(2.7)}$$
$$P(A \cap B \cap C) = P(A) \ P(B) \ P(C). \quad \text{(2.8)}$$
2.6 Independence

Note that the last identity (2.3), or more generally,
\[ P(A_1 \cap A_2 \cap \cdots \cap A_n) = \prod_{j=1}^{n} P(A_j) \]
is not sufficient for mutual independence.

**Example 2.10.** Roll two fair dice independently. Define the events
\[
A = \{\text{The first die roll is } 1, 2, \text{ or } 3\},
B = \{\text{The first die roll is } 2, 3, \text{ or } 6\},
C = \{\text{The sum of the two rolls is } 9\} = \{(3, 6), (4, 5), (5, 4), (6, 3)\}.
\]
Since the dice are fair and the experiments are done independently, the probability of any pair of outcomes is 1/36. Therefore
\[ P(A) = \frac{1}{2}, \quad P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{9}. \]
Since \( A \cap B \cap C = \{(3, 6)\} \),
\[ P(A \cap B \cap C) = \frac{1}{36} = P(A)P(B)P(C). \]
But \( A, B, \) and \( C \) are not independent because
\[ P(A \cap B) = \frac{1}{3} \neq \frac{1}{4} = P(A)P(B). \]

Similarly, pairwise independence, e.g., (2.5)–(2.7), does not imply independence.

**Example 2.11.** Flip two fair coins independently. Define the events
\[
A = \{\text{The first coin is a head}\},
B = \{\text{The second coin is a head}\},
C = \{\text{Both coins are the same}\}.
\]
Since the flips are independent,
\[ P(A \cap B) = P(B \cap C) = P(C \cap A) = \frac{1}{4} \]
\[ = P(A)P(B) = P(B)P(C) = P(C)P(A), \]
and \( A, B, \) and \( C \) are pairwise independent. However, they are not independent since
\[ P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C). \]

Two events \( A \) and \( B \) are said to be conditionally independent give a third event \( C \) with \( P(C) > 0 \) if
\[ P(A \cap B \mid C) = P(A \mid C)P(B \mid C). \]
Example 2.12. We continue Example 2.10. Since $A \cap C = \{(3, 6)\}$, $B \cap C = \{(3, 6), (6, 3)\}$, and $A \cap B \cap C = \{(3, 6)\}$,

$$P(A \cap B \mid C) = \frac{1}{4} \neq \frac{1}{8} = P(A \mid C)P(B \mid C).$$

Hence, $A$ and $B$ are not conditionally independent given $C$. Now define the event

$$D = \{\text{The sum of the two rolls is 4}\} = \{(1, 3), (2, 2), (3, 1)\}.$$ 

Since $A \cap D = \{(1, 3), (2, 2), (3, 1)\}$, $B \cap D = \{(2, 2), (3, 1)\}$, and $A \cap B \cap D = \{(2, 2), (3, 1)\}$,

$$P(A \cap B \mid D) = \frac{2}{3} = P(A \mid D)P(B \mid D).$$

Hence, $A$ and $B$ are conditionally independent given $D$.

Conditional independence of more than two events given another event $C$ is defined similarly as independence with respect to the probability measure $P(\cdot \mid C)$. Conditional independence neither implies nor is implied by (unconditional) independence. In Example 2.12, $A$ and $B$ are conditionally independent given $D$, but are not independent unconditionally as shown in Example 2.10. In Example 2.11, $A$ and $B$ are independent, but they are not conditionally independent given $C$ as

$$P(A \cup B \mid C) = \frac{1}{2} \neq \frac{1}{4} = P(A \mid C)P(B \mid C).$$

PROBLEMS

2.1. $\sigma$-algebra. Show that if $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

2.2. Limits of probabilities. Show

(a) $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} P\left(\bigcup_{i=1}^{n} A_i\right)$.

(b) $P\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{n \to \infty} P\left(\bigcap_{i=1}^{n} A_i\right)$.

2.3. Extension of a probability measure. Consider a discrete probability space $(\Omega, 2^\Omega, P)$, where $\Omega$ is a subset of $\mathbb{R}$. Show that $P(\cdot \cap \Omega)$ is a valid probability measure for the sample space $\mathbb{R}$ and the set of events $2^\mathbb{R}$, that is, it satisfies the axioms of probability.

2.4. Independence. Show that the events $A$ and $B$ are independent if $P(A \mid B) = P(A \mid B^c)$.

2.5. Conditional independence. Let $A$ and $B$ be two events such that $P(A \cap B) > 0$. Show that $A$ and $B$ are conditionally independent given $A \cap B$.

2.6. Conditional probabilities. Let $P(A) = 0.8$, $P(B^c) = 0.6$, and $P(A \cup B) = 0.8$. Find

(a) $P(A^c \mid B^c)$.

(b) $P(B^c \mid A)$. 
2.7. Let $A$ and $B$ be two events with $P(A) \geq 0.5$ and $P(B) \geq 0.75$. Show that $P(A \cap B) \geq 0.25$.

2.8. Monty Hall. Gold is placed behind one of three curtains. A contestant chooses one of the curtains, Monty Hall (the game host) opens one of the unselected empty curtains. The contestant has a choice either to switch his selection to the third curtain or not.

(a) What is the sample space for this random experiment? (Hint: An outcome consists of the curtain with gold, the curtain chosen by the contestant, and the curtain chosen by Monty.)

(b) Assume that placement of the gold behind the three curtains is random, the contestant choice of curtains is random and independent of the gold placement, and that Monty Hall’s choice of an empty curtain is random among the alternatives. Specify the probability measure for this random experiment and use it to compute the probability of winning the gold if the contestant decides to switch.

2.9. Negative evidence. Suppose that the evidence of an event $B$ increases the probability of a criminal’s guilt; that is, if $A$ is the event that the criminal is guilty, then $P(A|B) \geq P(A)$. Does the absence of the event $B$ decrease the criminal’s probability of being guilty? In other words, is $P(A|B^c) \leq P(A)$? Prove or provide a counterexample.

2.10. Random state transition. Consider the state diagram in Figure 2.2. The sample space is

$$\Omega = \{(\alpha, \alpha), (\alpha, \beta), \ldots, (y, y)\},$$

where the first entry is the initial state and the second entry is the next state. Define the events

$$A_1 = \{\text{the initial state is } \alpha\}, \quad A_2 = \{\text{the next state is } \alpha\},$$

$$B_1 = \{\text{the initial state is } \beta\}, \quad B_2 = \{\text{the next state is } \beta\},$$

$$C_1 = \{\text{the initial state is } \gamma\}, \quad C_2 = \{\text{the next state is } \gamma\}.$$

Assume that $P(A_1) = 0.5$, $P(B_1) = 0.2$, and $P(C_1) = 0.3$.

(a) Find $P(A_2)$, $P(B_2)$, and $P(C_2)$.

(b) Find $P(A_1 | A_2)$, $P(B_1 | B_2)$, and $P(C_1 | C_2)$.

(c) Find two events among $A_1, A_2, B_1, B_2, C_1, C_2$ that are pairwise independent.

2.11. Geometric pairs. Consider a probability space consisting of the sample space

$$\Omega = \{1, 2, 3, \ldots\}^2 = \{(i, j) : i, j \in \mathbb{N}\},$$

i.e., all pairs of positive integers, the set of events $2^\Omega$, and the probability measure specified by

$$P((i, j)) = p^i(1 - p)^{j-i-2}, \quad 0 < p < 1.$$
Figure 2.2. The state diagram for a three-state system. Here the label of each edge $i \to j$ denotes the transition probability from state $i$ to state $j$, that is, the conditional probability that the next state is $j$ given the initial state is $i$.

(a) Find $P(\{(i, j) : i \geq j\})$.
(b) Find $P(\{(i, j) : i + j = k\})$.
(c) Find $P(\{(i, j) : i \text{ is an odd number}\})$.
(d) Describe an experiment whose outcomes $(i, j), i, j \in \mathbb{N}$, have the probabilities $P((i, j))$.

2.12. Juror’s fallacy. Suppose that $P(A\mid B) \geq P(A)$ and $P(A\mid C) \geq P(A)$. Is it always true that $P(A\mid B \cap C) \geq P(A)$? Prove or provide a counterexample.

2.13. Polya’s urn. Suppose we have an urn containing one red ball and one blue ball. We draw a ball at random from the urn. If it is red, we put the drawn ball plus another red ball into the urn. If it is blue, we put the drawn ball plus another blue ball into the urn. We then repeat this process. At the $n$-th stage, we draw a ball at random from the urn with $n + 1$ balls, note its color, and put the drawn ball plus another ball of the same color into the urn.

(a) Find the probability that the first ball is red.
(b) Find the probability that the second ball is red.
(c) Find the probability that the first three balls are all red.
(d) Find the probability that two of the first three balls are red.