Basics of Probability Theory

Andreas Klappenecker

Texas A&M University
The **probability space** or **sample space** $\Omega$ is the set of all possible outcomes of an experiment. For example, the sample space of the coin tossing experiment is $\Omega = \{\text{head, tail}\}$.

Certain subsets of the sample space are called **events**, and the probability of these events is determined by a **probability measure**.
If we roll a dice, then one of its six face values is the outcome of the experiment, so the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$.

An event is a subset of the sample space $\Omega$. The event $\{1, 2\}$ occurs when the dice shows a face value less than three.

The probability measures describes the odds that a certain event occurs, for instance $\Pr[\{1, 2\}] = 1/3$ means that the event $\{1, 2\}$ will occur with probability $1/3$. 
A probability measure is not necessarily defined on all subsets of the sample space \( \Omega \), but just on all subsets of \( \Omega \) that are considered events. Nevertheless, we want to have a uniform way to reason about the probability of events. This is accomplished by requiring that the collection of events form a \( \sigma \)-algebra.
A $\sigma$-algebra $\mathcal{F}$ is a collection of subsets of the sample space $\Omega$ such that the following requirements are satisfied:

**S1** The empty set is contained in $\mathcal{F}$.

**S2** If a set $E$ is contained in $\mathcal{F}$, then its complement $E^c$ is contained in $\mathcal{F}$.

**S3** The countable union of sets in $\mathcal{F}$ is contained in $\mathcal{F}$. 
The empty set $\emptyset$ is often called the **impossible event**.

The sample space $\Omega$ is the complement of the empty set, hence is contained in $\mathcal{F}$. The event $\Omega$ is called the **certain event**.

If $E$ is an event, then $E^c = \Omega \setminus E = \Omega - E$ is called the **complementary event**.
Let $\mathcal{F}$ be a $\sigma$-algebra.

**Exercise**

If $A$ and $B$ are events in $\mathcal{F}$, then $A \cap B$ in $\mathcal{F}$.

**Exercise**

The countable intersection of events in $\mathcal{F}$ is contained in $\mathcal{F}$.

**Exercise**

If $A$ and $B$ are events in $\mathcal{F}$, then $A - B = A \setminus B$ is contained in $\mathcal{F}$.
Let $A$ be a subset of $P(\Omega)$. Then the intersection of all $\sigma$-algebras containing $A$ is a $\sigma$-algebra, called the smallest $\sigma$-algebra generated by $A$. We denote the smallest $\sigma$-algebra generated by $A$ by $\sigma(A)$.

Example

Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $A = \{\{1, 2\}, \{2, 3\}\}$.

$$\sigma(A) = \emptyset, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{3, 4, 5, 6\}, \{2, 3\}, \{1, 4, 5, 6\}, \{2\}, \{1, 3, 4, 5, 6\}, \{1\}, \{2, 3, 4, 5, 6\}, \{3\}, \{1, 2, 4, 5, 6\}, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 3\}, \{2, 4, 5, 6\}\}$$
Exercise

Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{A} = \{\{2\}, \{1, 2, 3\}, \{4, 5\}\}$. Determine $\sigma(\mathcal{A})$. 
Exercise

Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{A} = \{\{2\}, \{1, 2, 3\}, \{4, 5\}\}$. Determine $\sigma(\mathcal{A})$.

Solution

We have

$$\mathcal{A} = \{\emptyset, \Omega, \{2\}, \{1, 3, 4, 5, 6\}, \{1, 2, 3\}, \{4, 5, 6\}, \{4, 5\}, \{1, 2, 3, 6\}, \{1, 3\}, \{2, 4, 5, 6\}, \{6\}, \{1, 2, 3, 4, 5\}, \{2, 6\}, \{1, 3, 4, 5\}, \{2, 4, 5\}, \{1, 3, 6\}\}$$
Probability Measure

Let $\mathcal{F}$ be a $\sigma$-algebra over the sample space $\Omega$. A **probability measure** on $\mathcal{F}$ is a function $\Pr : \mathcal{F} \rightarrow [0, 1]$ satisfying

**P1** The certain event satisfies $\Pr[\Omega] = 1$.

**P2** If the events $E_1, E_2, \ldots$ in $\mathcal{F}$ are mutually disjoint, then

$$\Pr[\bigcup_{k=1}^{\infty} E_k] = \sum_{k=1}^{\infty} \Pr[E_k].$$
Example

Probability Function Let $\Omega$ be a sample space and let $a \in \Omega$. Suppose that $\mathcal{F} = \mathcal{P}(\Omega)$ is the $\sigma$-algebra. Then $\Pr: \Omega \rightarrow [0, 1]$ given by

$$\Pr[A] = \begin{cases} 1 & \text{if } a \in A, \\ 0 & \text{otherwise.} \end{cases}$$

is a probability measure.

We know that $\textbf{P1}$ holds, since $\Pr[\Omega] = 1$. $\textbf{P2}$ holds as well. Indeed, if $E_1, E_2, \ldots$ are mutually disjoint events in $\mathcal{P}(\Omega)$, then at most one of the events contains $a$.

$$\sum_{k=1}^{\infty} \Pr[E_k] = \begin{cases} 1 & \text{if some set } E_k \text{ contains } a, \\ 0 & \text{if none of the sets } E_k \text{ contains } a. \end{cases} = \Pr[\bigcup_{k=1}^{\infty} E_k]$$
These axioms have a number of familiar consequences. For example, it follows that the complementary event $E^c$ has probability

$$\Pr[E^c] = 1 - \Pr[E].$$

In particular, the impossible event has probability zero, $\Pr[\emptyset] = 0$. 
Another consequence is a simple form of the inclusion-exclusion principle:

\[ \Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F], \]

which is convenient when calculating probabilities.
Immediate Consequences

Another consequence is a simple form of the **inclusion-exclusion principle**:

\[
\Pr[E \cup F] = \Pr[E] + \Pr[F] - \Pr[E \cap F],
\]

which is convenient when calculating probabilities. Indeed,

\[
\Pr[E \cup F] = \Pr[E \setminus (E \cap F)] + \Pr[E \cap F] + \Pr[F \setminus (E \cap F)] \\
= \Pr[E] + \Pr[F \setminus (E \cap F)] + (\Pr[E \cap F] - \Pr[E \cap F]) \\
= \Pr[E] + \Pr[F] - \Pr[E \cap F].
\]
Exercises

Exercise

Let $E$ and $F$ be events such that $E \subseteq F$. Show that

$$\Pr[E] \leq \Pr[F].$$

Exercise

Let $E_1, \ldots, E_n$ be events that are not necessarily disjoint. Show that

$$\Pr[E_1 \cup \cdots \cup E_n] \leq \Pr[E_1] + \cdots + \Pr[E_n].$$
Conditional Probabilities
Let $E$ and $F$ be events over a sample space $\Omega$ such that $\Pr[F] > 0$. The **conditional probability** $\Pr[E | F]$ of the event $E$ given $F$ is defined by

$$\Pr[E | F] = \frac{\Pr[E \cap F]}{\Pr[F]}.$$  

The value $\Pr[E | F]$ is interpreted as the probability that the event $E$ occurs, assuming that the event $F$ occurs.

By definition, $\Pr[E \cap F] = \Pr[E | F] \Pr[F]$, and this simple multiplication formula often turns out to be useful.
Law of Total Probability (Simplest Version)

Let $\Omega$ be a sample space and $A$ and $E$ events. We have

$$
\Pr[A] = \Pr[A \cap E] + \Pr[A \cap E^c]
= \Pr[A | E] \Pr[E] + \Pr[A | E^c] \Pr[E^c].
$$

The events $E$ and $E^c$ are disjoint and satisfy $\Omega = E \cup E^c$. Therefore, we have

$$
\Pr[A] = \Pr[A \cap E] + \Pr[A \cap E^c].
$$

The second equality follows directly from the definition of conditional probability.
Bayes’ Theorem

\[
\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}.
\]

We have

\[
\Pr[A \mid B] \Pr[B] = \Pr[A \cap B] = \Pr[B \cap A] = \Pr[B \mid A] \Pr[A].
\]

Dividing by \(\Pr[B]\) yields the claim.
Bayes’ Theorem (Version 2)

\[
\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B \mid A] \Pr[A] + \Pr[B \mid A^c] \Pr[A^c]}
\]

By the first version of Bayes’ theorem, we have

\[
\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}
\]

Now apply the law of total probability with \( \Omega = A \cup A^c \) to the probability \( \Pr[B] \) denominator.
Polynomial Identities
Suppose that we use a library that is supposedly implementing a polynomial factorization. We would like to check whether the polynomials such as

\[ p(x) = (x + 1)(x - 2)(x + 3)(x - 4)(x + 5)(x - 6) \]
\[ q(x) = x^6 - 7x^3 + 25 \]

are the same.

We can multiply the terms both polynomials and simplify. This uses \( \Omega(d^2) \) multiplications for polynomials of degree \( d \).
If the polynomials $p(x)$ and $q(x)$ are the same, then we must have
\[ p(x) - q(x) \equiv 0. \]

If the polynomials $p(x)$ and $q(x)$ are not the same, then an integer $r \in \mathbb{Z}$ such that
\[ p(r) - q(r) \neq 0 \]
would be a witness to the difference of $p(x)$ and $q(x)$.

We can check whether $r \in \mathbb{Z}$ is a witness in $O(d)$ multiplications.
We get the following randomized algorithm for checking whether \( p(x) \) and \( q(x) \) are the same.

Input: Two polynomials \( p(x) \) and \( q(x) \) of degree \( d \).

\[
\text{for } i = 1 \text{ to } t \text{ do } \\
\quad r = \text{random}(1..100d); \\
\quad \text{return } '\text{different}' \text{ if } p(r) - q(r) \neq 0 \\
\text{end}
\]

return 'same'
Polynomial Identities

If $p(x) \equiv q(x)$, then every $r \in \mathbb{Z}$ is a non-witness.

If $p(x) \not\equiv q(x)$, then an integer $r$ in the range $1 \leq r \leq 100d$ is a witness if and only if it is not a root of $p(x) - q(x)$. The polynomial $p(x) - q(x)$ has at most $d$ roots.

The probability that the algorithm will return 'same' when the polynomials are different is at most

$$\Pr[\text{'same'}|p(x) \not\equiv q(x)] \leq \left( \frac{d}{100d} \right)^t = \frac{1}{100^t}.$$
Independent Events
Two events $E$ and $F$ are called independent if and only if

$$\Pr[E \cap F] = \Pr[E] \Pr[F].$$

Two events that are not independent are called dependent.
Example

Suppose that we flip a fair coin twice. Then the sample space is \( \{HH, HT, TH, TT\} \). The probability of each elementary event is given by 1/4. For instance, \( \Pr[\{HH\}] = 1/4 \).

The event \( E \) that the first coin is heads is given by \( \{HH, HT\} \). We have \( \Pr[E] = 1/2 \). The event \( F \) that the second coin is tails is given by \( \{HT, TT\} \). We have \( \Pr[F] = 1/2 \).

Then \( E \cap F \) models the event that the first coin is heads and the second coin is tails. The events \( E \) and \( F \) are independent, since

\[
\Pr[E \cap F] = \frac{1}{4} = \Pr[E] \Pr[F].
\]
If $E$ and $F$ are independent, then

$$
\Pr[E \mid F] = \frac{\Pr[E \cap F]}{\Pr[F]} = \frac{\Pr[E] \Pr[F]}{\Pr[F]} = \Pr[E].
$$

In this case, whether or not $F$ happened has no bearing on the probability of $E$. 
Suppose that \( E_1, E_2, \ldots, E_n \) are events. The events are called **mutually independent** if and only if for all subsets \( S \) of \( \{1, 2, \ldots, n\} \), we have

\[
\Pr\left[ \bigcap_{i \in S} E_i \right] = \prod_{i \in S} \Pr[E_i].
\]

Please note that it is not sufficient to show this condition for \( S = \{1, 2, \ldots, n\} \), but we really need to show this for all subsets.
We toss a fair coin three times. Consider the events:

- \( E_1 \) = the first two values are the same,
- \( E_2 \) = the first and last value are the same,
- \( E_3 \) = the last two values are the same.

The probabilities are \( \Pr[E_1] = \Pr[E_2] = \Pr[E_3] = 1/2 \). We have

\[
\Pr[E_1 \cap E_2] = \Pr[E_2 \cap E_3] = \Pr[E_1 \cap E_3] = \Pr[\{HHH, TTT\}] = \frac{1}{4}.
\]

Thus, all three pairs of events are independent. But

\[
\Pr[E_1 \cap E_2 \cap E_3] = \frac{1}{4} \neq \Pr[E_1] \Pr[E_2] \Pr[E_3] = \frac{1}{8},
\]

so they are not mutually independent.
Example

A school offers as electives $A =$ athletics, $B =$ band, and $C =$ Mandarin Chinese.

\[
\begin{align*}
\Pr[A \cap B \cap C] & = 0.04 & \Pr[\overline{A} \cap B \cap C] & = 0.2 \\
\Pr[A \cap B \cap \overline{C}] & = 0.06 & \Pr[\overline{A} \cap B \cap \overline{C}] & = 0.1 \\
\Pr[A \cap \overline{B} \cap C] & = 0.1 & \Pr[\overline{A} \cap \overline{B} \cap C] & = 0.16 \\
\Pr[A \cap \overline{B} \cap \overline{C}] & = 0 & \Pr[\overline{A} \cap \overline{B} \cap \overline{C}] & = 0.34 \\
\end{align*}
\]

Then \( \Pr[A \cap B \cap C] = 0.04 = \Pr[A] \Pr[B] \Pr[C] = 0.2 \cdot 0.4 \cdot 0.5. \) But no two of the three events are pair-wise independent:

\[
\Pr[A \cap B] = 0.1 \neq \Pr[A] \Pr[B] = 0.2 \cdot 0.4 = 0.08
\]
Verifying Matrix Multiplication
## The Problem

Let $A$, $B$, and $C$ be $n \times n$ matrices over $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z}$. Is $AB = C$?

If we use traditional matrix multiplication, then forming the product of $A$ and $B$ requires $\Theta(n^3)$ scalar operations. Using the fastest known matrix multiplications takes about $\Theta(n^{2.37})$ scalar operations. Can we do better using a randomized algorithm?
A **witness** for $AB \neq C$ would be a vector $v$ such that

$$ABv \neq Cv.$$

We can check whether a vector is a witness in $O(n^2)$ time.
Verifying Matrix Multiplication

Theorem

If \( AB \neq C \), and we choose a vector \( v \) uniformly at random from \( \{0, 1\}^n \), then \( v \) is a witness for \( AB \neq C \) with probability \( \geq 1/2 \). In other words,

\[
\Pr_{v \in \mathbb{F}_2^n}[ABv = Cv \mid AB \neq C] \leq \frac{1}{2}.
\]
Simple Observation

Lemma

Choosing \( \mathbf{v} = (v_1, v_2, \ldots, v_n) \in \mathbb{F}_2^n \) uniformly at random is equivalent to choosing each \( v_k \) independently and uniformly at random from \( \mathbb{F}_2 \).

Proof.

If we choose each component \( v_k \) independently and uniformly at random from \( \mathbb{F}_2 \), then each vector \( \mathbf{v} \) in \( \mathbb{F}_2^n \) is created with probability \( 1/2^n \).

Conversely, if \( \mathbf{v} \in \mathbb{F}_2^n \) is chosen uniformly at random, then the components are independent and \( v_k = 1 \) with probability \( 1/2 \).
Proof of the Theorem

Let \( D = AB - C \neq 0 \). Then \( ABv = Cv \) if and only if \( Dv = 0 \).

Since \( D \neq 0 \), the matrix \( D \) must have a nonzero entry. Without loss of generality, suppose that \( d_{11} \neq 0 \).

If \( Dv = 0 \), then we must have

\[
\sum_{k=1}^{n} d_{1k} v_k = 0.
\]

Since \( d_{11} \neq 0 \), this is equivalent to

\[
v_1 = -\frac{\sum_{k=2}^{n} d_{1k} v_k}{d_{11}}.
\]
Idea (Principle of Deferred Decisions)

Rather than arguing with the vector $v \in \mathbb{F}_2^n$, we can choose each component of $v$ uniformly at random from $\mathbb{F}_2$ in order from $v_n$ down to $v_1$. 
Proof of the Theorem

Suppose that the components $\nu_n, \nu_{n-1}, \ldots, \nu_2$ have been chosen. This determines the right-hand side of

$$\nu_1 = -\frac{\sum_{k=2}^{n} d_{1k} \nu_k}{d_{11}}.$$

Now there is just one choice of $\nu_1$ that will make the equality true, so the probability that this equation is satisfied is at most $1/2$. In other words, the probability

$$\Pr[AB\nu = C\nu | AB \neq C] \leq 1/2.$$