Probability Theory Dictionary I

Appart from the measure theoretical language, Probability Theory uses some special terms, and today we are going to get the first installment of the so-called Dictionary of Probability Theory.

**Definitions.** A *probability space* is a measure space \((\Omega, A, P)\), where the measure \(P\) has the property:

\[
P(X) = 1. \tag{1}
\]

The set \(X\) is referred to as the *sample space*. The \(\sigma\)-algebra \(A\) is referred to as the *event field*. The sets in \(A\) are called *events*. The measure \(P\) on \(A\), which by (1) takes values in \([0, 1]\), is called the *probability measure*. Given an event \(A \in A\), the number \(P(A) \in [0, 1]\) is interpreted as the *probability that, picking “randomly” an \(x \in X\), the relation \(x \in A\) takes place.*

The term “randomly” is to be used here with some caution, since some points in \(X\) may be “favoured” against others.

**Example 1.** The simplest example of a probability space is the case when \(X\) is a two-point set, say \(X = \{0, 1\}\), and \(A = \mathcal{P}\{0, 1\}\). In other words, if one defines \(A_0 = \{0\}\) and \(A_1 = \{1\}\), then \(A = \{\emptyset, A_0, A_1, X\}\). Any probability measure \(P : A \to [0, 1]\) is then completely determined by the numbers \(p_k = P(A_k), k = 0, 1\), that are subject to the restrictions \(p_0, p_1 \geq 0\) and \(p_0 + p_1 = 1\). This probability space is the model for coin flips, where 0 corresponds to heads, and 1 corresponds to tails, so \(p_0\) is the probability that in one toss we get a head, while \(p_1\) is the probability that we get a tail. In real life coins are “biased,” so \(p_0 \neq p_1\). A *fair* coin is one for which \(p_0 = p_1 = \frac{1}{2}\).

The above example is a particular case of a general type.

**Definition.** A probability space \((X, A, P)\) is said to be *finite*, if the event field \(A\) is finite.

In the above definition it is not required that the sample space \(X\) be finite. The advantages of working with finite probability spaces are explained in the following result.

**Proposition 1.** Let \(A\) be a finite algebra on some non-empty set \(X\).

(i) \(A\) is a \(\sigma\)-algebra.

(ii) Any finitely additive map \(\mu : A \to [0, \infty]\) is a measure on \(A\).
(iii) There exist and integer \( n \geq 1 \), and non-empty disjoint sets \( X_1, \ldots, X_n \in \mathcal{A} \) with \( \bigcup_{k=1}^{n} X_k = X \), such that the correspondence

\[
\Phi : \mathcal{P}\{1, \ldots, n\} \ni K \mapsto \bigcup_{k \in K} X_k \in \mathcal{A}
\]

is a bijection.

**Proof.** (i) It is obvious that \( \mathcal{A} \) is a monotone class, since any increasing sequence \((A_n)_{n=1}^{\infty}\) in \( \mathcal{A} \) is eventually constant. (The condition \( A_1 \subset A_2 \subset \ldots \) combined with the fact that \( \mathcal{A} \) is finite, force the existence of some \( p \) such that \( A_n = A_p, \forall n \geq p \), so \( \bigcup_{n=1}^{\infty} A_n = A_p \in \mathcal{A} \).)

(ii) Assume \( \mu \) is finitely additive, and let us prove that it is a measure. If one starts with some disjoint sequence \((A_n)_{n=1}^{\infty}\) in \( \mathcal{A} \), then the finitess of \( \mathcal{A} \) will force the existence of some \( p \) such that \( A_n = \emptyset, \forall n > p \), so by finite additivity (which also gives \( \mu(A_n) = 0, \forall n > p \)) we get

\[
\mu\left( \bigcup_{n=1}^{\infty} A_n \right) = \mu\left( \bigcup_{n=1}^{p} A_n \right) = \sum_{n=1}^{p} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A_n).
\]

(iii) Let us define, for every point \( x \in X \), the collection

\[
\mathcal{A}_x = \{ A \in \mathcal{A} : A \ni x \}.
\]

Since \( \mathcal{A} \) is an algebra, the collection \( \mathcal{A}_x \) is always non-empty, because it contains at least the total set \( X \). Since \( \mathcal{A} \) is finite, the intersection

\[
S(x) = \bigcap_{A \in \mathcal{A}_x} A
\]

belongs to \( \mathcal{A} \), for every \( x \in X \).

**Claim:** For two points \( x, y \in X \), the sets \( S(x) \) and \( S(y) \) are either disjoint, or they coincide.

Assume \( S(x) \cap S(y) \neq \emptyset \), and let us prove the equality \( S(x) = S(y) \). Let us consider the disjoint sets \( A = S(x) \setminus S(y) \) and \( B = S(x) \cap S(y) \), which are both \( \mathcal{A} \), and satisfy \( A \cup B = S(x) \). Since \( x \in S(x) \), either \( x \in A \), or \( x \in B \), or both \( A \in \mathcal{A}_x \), or \( B \in \mathcal{A}_x \). This will force either \( A \supset S(x) \) or \( B \supset S(x) \), which in turn forces either \( A = S(x) \), or \( B = S(x) \). Since \( B \neq \emptyset \), we cannot have the equality \( A = S(x) \), so we must have the equality \( B = S(x) \), i.e. the inclusion \( S(x) \subset S(y) \). By symmetry we also have the inclusion \( S(y) \subset S(x) \), so indeed \( S(x) = S(y) \).

Having proven the Claim, let us list all the sets \( S(x) \), \( x \in X \) as a finite sequence \( S(x_1), \ldots, S(x_n) \). This means that

\[\begin{itemize}
\item[(a)] \( S(x_j) \neq S(x_k) \), which by the Claim forces \( S(x_j) \cap S(x_k) = \emptyset, \forall j \neq k \);
\item[(b)] for every \( x \in X \), there exists some \( j \) such that \( S(x) = S(x_j) \).
\end{itemize}\]
Let us remark that, since \( x \in S(x) \), \( \forall x \in X \), using (b) we have

\[
\bigcup_{k=1}^{n} S(x_k) = \bigcup_{x \in X} S(x) = X.
\]

Define the sets \( X_k = S(x_k), \ k = 1, \ldots, n \), and let us prove that the correspondence (2) is bijective. Let us first observe that, using (a), for every \( K \subset \{1, \ldots, n\} \), and every \( k \in \{1, \ldots, n\} \), one has

\[
X_k \cap \Phi(K) = \begin{cases} 
X_k & \text{if } k \in K \\
\emptyset & \text{if } k \notin K
\end{cases}
\]

Since the sets \( X_1, \ldots, X_n \) are non-empty, this gives the equality

\[
K = \{ k \in \{1, \ldots, n\} : \Phi(K) \cap X_k \neq \emptyset \},
\]

so \( \Phi \) is clearly injective. (If \( \Phi(K) = \Phi(L) \), then the above equality clearly forces \( K = L \).) To prove that \( \Phi \) is surjective, we start with an arbitrary set \( A \in \mathcal{A} \), we define

\[
K_A = \{ k \in \{1, \ldots, n\} : A \cap X_k \neq \emptyset \},
\]

so that we clearly have the equality \( A = \bigcup_{k \in K_A} (A \cap X_k) \), and all we need to prove is the fact that

\[
A \cap X_k = X_k, \ \forall k \in K_A.
\]  \hfill (3)

(In particular this will give the equality \( \Phi(K_A) = A \).) Fix some \( k \in K_A \), and let \( x \) be some point in \( A \cap X_k \). Since \( A \cap X_k \) belongs to \( \mathcal{A} \), it follows that \( A \cap X_k \in \mathcal{A}_x \), so \( A \cap X_k \supset S(x) \). We now have

\[
S(x_k) = X_k \supset A \cap X_k \supset S(x),
\]  \hfill (4)

so \( S(x) \cap S(x_k) = S(x) \neq \emptyset \), and then the above Claim forces the equality \( S(x) = S(x_k) \), which in turn forces all the inclusions in (4) to be equalities. Obviously this gives (3), and we are done. \( \square \)

**Comments.** Let \( X \) and \( \mathcal{A} \) be as above. The sets \( X_1, \ldots, X_n \), that appear in part (iii) of the above result, are called the **atoms** of \( \mathcal{A} \). They are uniquely determined (up to a re-indexing), and they are intrinsically defined as those sets in \( \mathcal{A} \), that are minimal. (Call a set \( M \in \mathcal{A} \) **minimal** in \( \mathcal{A} \), if the only sets \( A \in \mathcal{A} \), with \( A \subset M \), are \( A = \emptyset \) and \( A = M \).) A probability measure \( \mathcal{P} \) on \( \mathcal{A} \) is then completely determined by the numbers \( p_k = \mathcal{P}(X_k), \ k = 1, \ldots, n \), namely, for every set \( A \in \mathcal{A} \), one has

\[
\mathcal{P}(A) = \sum_{k \in K_A} p_k,
\]

where \( K_A \subset \{1, \ldots, n\} \) is the (unique) set satisfying \( \bigcup_{k \in K_A} X_k = A \). The numbers \( p_1, \ldots, p_n \geq 0 \) are called the **weights** of the atoms. They are of course subject to the constraint \( \sum_{k=1}^{n} p_k = 1 \).
Although the sample space $X$ may be infinite, Proposition 1 states that the pair $(X, A)$ is “the same” as the pair $(\{1, \ldots, n\}, P(\{1, \ldots, n\}))$, in which the atoms are the singleton sets $\{1\}, \ldots, \{n\}$. The phrase “the same” is meant to suggest that probabilistic computations are identical in the two contexts.

The above discussion then legitimizes the following notion.

**Definition.** A *standard finite probability space* is a probability space of the form $(X, P(X), P)$, where $X$ is a finite set. The atoms for this spaces will be the singleton sets $\{x\}, x \in X$. The probability measure $P$ is completely described by the system of weights $(p_x)_{x \in X} \subset [0, 1]$, given by $p_x = P(\{x\})$, namely:

$$P(A) = \sum_{x \in A} p_x, \quad \forall A \in P(X). \quad (5)$$

The constraint is of course

$$\sum_{x \in X} p_x = 1. \quad (6)$$

**Definition.** A standard finite probability space $(X, P(X), P)$ is said to be *fair* (or unbiassed), if all atoms have equal weights. Using (6) this forces

$$p_x = \frac{1}{|X|}, \quad \forall x \in X,$$

where $|.|$ stands for the number of elements. In this case (5) reads:

$$P(A) = \frac{|A|}{|X|}, \quad \forall A \in P(X).$$

**Comment.** Since we do not know yet how to construct interesting *infinite* probability spaces, all the examples here will deal with finite ones (most of them standard and fair). The various notions that we are going to discuss however, make sense in arbitrary probability spaces.

The example below (which is discussed in several parts) will be used to illustrate the various important notions.

**Example 2.** Consider the experiment of rolling two *fair* dice. To make the distinction between the two dice we assume that one of them is red, and the other one is blue. Below are ten questions about this game.

1. What is the probability space that models this experiment?
2. What is the probability that the red die shows a six?
3. What is the probability that at least one die shows a six?
4. What is the probability that both dice show sixes?
5. What is the probability that no six is shown?

6. What is the probability that the sum of the two numbers shown is even?

7. If the red die shows a six, what is the probability that the sum is even?

8. If at least one die shows a six, what is the probability that the sum is even?

9. If the red die shows a six, what is the probability that the blue die shows a six?

10. If one die shows a six, what is the probability that both dice show sixes?

The answer of question 1 is of course the standard finite fair probability space whose sample space is the product

\[ X = \{1, \ldots, 6\} \times \{1, \ldots, 6\}, \]

which has \( |X| = 36 \). The key events that are used in answering questions

To answer questions 2-5, we are going to consider the events

- \( R \): the red die shows a six,
- \( B \): the blue die shows a six,

Notice that we have only described these two events in words. As sets, the two events are:

\[ R = \{6\} \times \{1, \ldots, 6\} \text{ and } B = \{1, \ldots, 6\} \times \{6\}. \]

Since \( |R| = 6 \), the answer to question 2 is:

\[ P(R) = \frac{|R|}{|X|} = \frac{6}{36} = \frac{1}{6}. \]

Note that we also have \( P(B) = \frac{1}{6} \).

**Rules for Logical Operations.** We have reached the point where we are going to explain how logical operations are to be interpreted in terms of set operations. The rules are contained in the table below:

<table>
<thead>
<tr>
<th>Operation Type</th>
<th>Logical</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>Union</td>
<td></td>
</tr>
<tr>
<td>AND</td>
<td>Intersection</td>
<td></td>
</tr>
<tr>
<td>NOT</td>
<td>Complement</td>
<td></td>
</tr>
</tbody>
</table>

The complement operation is particularly nice, because of the following.

**Proposition 1.** If \((X, \mathcal{A}, P)\) is a probability space, and \(A \in \mathcal{A}\) is some event, then the probability of its complement \(A' = X \setminus A\) is given by the formula

\[ P(A') = 1 - P(A). \]
Proof. This is pretty obvious from additivity (note that $A$ and $A'$ are disjoint and $A \cup A' = X$), which gives $P(A) + P(A') = P(X) = 1$. □

Example 2. (continued) With these rules in mind, the event described in question 3 can be described in words as “$R$ or $B$,” so in terms of sets, we have to consider the event $S = R \cup B$. To compute the probability $P(S)$ it is helpful to look at the complement $S' = X \setminus S$, which corresponds to the event described in words as: no sixes are shown. (This is, by the way, the event to consider when we answer question 5.) The point is that $S'$ is much easier to count, since one obviously has the equality

$$S' = \{1, \ldots, 5\} \times \{1, \ldots, 5\},$$

so $|S'| = 25$. The answer to question 5 is then

$$P(S') = \frac{|S'|}{|X|} = \frac{25}{36},$$

and the answer to question 3 is

$$P(S) = 1 - P(S') = 1 - \frac{25}{36} = \frac{11}{36}.$$

The event that describes question 4 is obviously the singleton (atom) set $R \cap B = \{(6, 6)\}$, so the answer to that question is

$$P(R \cap B) = \frac{1}{36}.$$

The event $E$ from question 6 can be described as a union $E = E_1 \cup E_2$, where

$$E_1 = \{1, 3, 5\} \times \{1, 3, 5\} \text{ and } E_2 = \{2, 4, 6\} \times \{2, 4, 6\}.$$

Since $E_1$ and $E_2$ are disjoint, and $|E_1| = |E_2| = 9$, the answer to question 6 is

$$P(E) = P(E_1) + P(E_2) = \frac{|E_1|}{|X|} + \frac{|E_2|}{|X|} = \frac{9}{36} + \frac{9}{36} = \frac{1}{2}.$$

To answer questions 7-10, we need to understand how to operate with the word “if.” Its meaning in the probability dictionary is quite interesting, and is described as follows.

Definition. Let $(X, \mathcal{A}, P)$ be a probability space, and let $M \in \mathcal{A}$ be an event with non-zero probability, i.e. $P(M) > 0$. For any event $A \in \mathcal{A}$ one defines the conditional probability of $A$ under $M$ as the number

$$P(A|M) = \frac{P(A \cap M)}{P(M)}.$$
Alternative notations for this number are $P(A \mid M)$ or $P_M(A)$.

**Comment.** Given a probability space $(X, \mathcal{A}, P)$, in order to define conditional probability with respect with an event $M \in \mathcal{A}$, we need the condition $P(M) > 0$. The events $M \in \mathcal{A}$, with $P(M) = 0$, will be called *essentially impossible*. The *absolutely impossible* event is the empty set $\emptyset$. It is possible however, to have essentially impossible events other than $\emptyset$, if for example $\mathcal{A}$ is infinite, or in the finite case when the probability measure $P$ vanishes on some atoms. In the same spirit, an event $M \in \mathcal{A}$ is said to be *essentially certain*, if $P(M) = 1$. One such example is the total set $X$, which is referred to as the *absolutely certain* event. An event is essentially certain, if and only if its complement is essentially impossible.

**Remark 1.** If $(X, \mathcal{A}, P)$ is a fair standard finite probability space, then for an event $M \in \mathcal{A}$ the condition $P(M) > 0$ is equivalent to the fact that $M$ is non-empty. Moreover, in this case one has:

$$P(A|M) = \frac{|A \cap M|}{|M|}, \quad \forall A \in \mathcal{A}.$$  

**Example 2.** (continued) We are now in position to answer questions 7-10.

To answer question 7, we must compute the conditional probability $P(E|R)$, where $E$ is the event discussed in question 6. Since

$$E \cap R = \{ (6, 2), (6, 4), (6, 6) \},$$

the answer to question 7 is

$$P(E|R) = \frac{|E \cap R|}{|R|} = \frac{3}{6} = \frac{1}{2}.$$  

To answer question 8, we must compute the conditional probability $P(E|S)$, where $S$ is the event discussed in question 3. Since

$$E \cap S = \{ (6, 2), (6, 4), (6, 6), (6, 2), (6, 4) \},$$

the answer to question 8 is

$$P(E|S) = \frac{|E \cap S|}{|S|} = \frac{5}{11}.$$  

The answer to question 9 is

$$P(B|R) = \frac{|B \cap R|}{|R|} = \frac{1}{6}.$$  

The answer to question 10 is

$$P(R \cap B \mid S) = \frac{|R \cap B \cap S|}{|S|} = \frac{1}{11}.$$
The following two result contain some useful properties of conditional probability.

**Proposition 2.** Let \((X, \mathcal{A}, P)\) be a probability space, and let \(M \in \mathcal{A}\) be an event with \(P(M) > 0\). The map \(P_M : \mathcal{A} \to [0,1]\) is a probability measure on \(\mathcal{A}\) with \(P_M(M') = 0\).

**Proof.** It is obvious that \(P_M(\emptyset) = P_M(M) = 0\) and \(P_M(X) = 1\). To prove that \(P\) is a measure, we start with a disjoint sequence \((A_n)_{n=1}^\infty\), and we prove the equality

\[
P_M\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty P_M(A_n)
\]

The left hand side is

\[
P_M\left(\bigcup_{n=1}^\infty A_n\right) = \frac{1}{P(M)} \cdot P\left(\bigcup_{n=1}^\infty A_n \cap M\right) = \frac{1}{P(M)} \cdot P\left(\bigcup_{n=1}^\infty [A_n \cap M]\right).
\]

Since the sets \((A_n \cap M)_{n=1}^\infty \subset \mathcal{A}\) are disjoint, we have the equality

\[
P\left(\bigcup_{n=1}^\infty [A_n \cap M]\right) = \sum_{n=1}^\infty P(A_n \cap M),
\]

so the preceding computation yields

\[
P_M\left(\bigcup_{n=1}^\infty A_n\right) = \frac{1}{P(M)} \sum_{n=1}^\infty P(A_n \cap M) = \sum_{n=1}^\infty \frac{P(A_n \cap M)}{P(M)} = \sum_{n=1}^\infty P_M(A_n). \quad \Box
\]

Going back to Example 2, the answer question 9 is particularly interesting. It says that the chances of the blue die to show a six, given that the red die shows a six already, are unchanged if this condition is removed, namely one has the equality

\[P(B \text{ if } R) = P(B).\]

This situation deserves a closer look, and we designate it using the following terminology.

**Definition.** Let \((X, \mathcal{A}, P)\) be a probability space. Two events \(A, B \in \mathcal{A}\) are said to be independent, if they satisfy the equality

\[P(A \cap B) = P(A) \cdot P(B).\]

As a matter of language, we are going to state this by saying that \(\{A, B\}\) is an independent pair. (We use braces \(\{\}\), rather than brackets \((\)\) since this condition is clearly symmetric.)
Remark that one has the following equivalent reformulations

\[
P(A \text{ if } B) = P(A), \text{ if } P(B) > 0; \\
P(B \text{ if } A) = P(B), \text{ if } P(A) > 0.
\]

**Example 2.** (continued) With the above terminology, we see that the events \(R\) (red die shows a six) and \(B\) (blue die shows a six) are **independent**.

**Remark 3.** Let \((X, \mathcal{A}, P)\) be a probability space, and let \(A, B \in \mathcal{A}\) be two events.

(i) If \(\{A, B\}\) is an independent pair, then so are the pairs \(\{A', B\}\), \(\{A, B'\}\) and \(\{A', B'\}\).

(ii) If one of the two events is essentially impossible, or essentially certain, then the two events are independent.

To prove (i) it suffices to prove the implication

\[
\{A, B\} \text{ independent } \Rightarrow \{A', B\} \text{ independent.}
\]

But this is clear, since by additivity we have

\[
P(A' \cap B) = P(B \setminus [A \cap B]) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = \\
= [1 - P(A)]P(B) = P(A')P(B).
\]

To prove (ii), using (i) and symmetry, it suffices to consider the case when \(A\) is essentially impossible, i.e. \(P(A) = 0\). In this case we must prove the equality \(P(A \cap B) = 0\), which is however trivial, since the inclusion \(A \cap B \subset A\) will force \(0 \leq P(A \cap B) \leq P(A)\).

**Exercise 1.** Let \((X, \mathcal{A}, P)\) be a probability space, let \(A, B \in \mathcal{A}\) be two events with \(0 < P(A), P(B) < 1\). Prove that if either \(A \subset B\), or \(A \subset B'\), then \(A\) and \(B\) are not independent.

**Exercise 2.** Let \((X, \mathcal{A}, P)\) be a probability space, let \(B \in \mathcal{A}\) be some event, and let \((A_i)_{i \in I} \subset \mathcal{A}\) be a countable disjoint collection of events. Prove that, if \(\{A_i, B\}\) is an independent pair, for each \(i \in I\), then \(\bigcup_{i \in I} A_i, B\) is also an independent pair.

The following example has been discussed in class (with a slightly different language).

**Example 3.** Two brothers, Gabe and Steve, go to the County Fair and stop at the “Duck Pond” game. In this game there are 9 rubber duck floating in a tub. The brothers know that 6 of the nine duck are winners, the other 3 are
loosers. Each brother is supposed to pick one duck and keep it. It is assumed that one of them picks first, so the brother who picks second will pick one of 8 ducks. The brothers pick in the order decided by a coin flip (or by their mother, who is unbiased.) Below are several questions about this game.

1. What is the probability space that models this experiment?
2. What is the probability that Gabe picked first?
3. What is the probability that Gabe picks a winning duck?
4. If Gabe won, what is the probability that he picked first?
5. Is it in Gabe’s (winning) interest to pick first, or second?

The answer Question 1 is as follows. Let us number the nine ducks as 1, 2, . . . , 9, with the convention that the losing ones are 7, 8 and 9. One forms the set $D$ of all possible draws of two duck, that is

$$D = \{(x, y) \in \{1, \ldots, 9\} \times \{1, \ldots, 9\} : x \neq y\},$$

and one considers the set $T = \{g, s\}$ of all possible choices of who goes first. The sample space is then the product space $D \times T$. The probability space is then the fair standard finite one on $X$. Note that $|X| = |D| \cdot |T| = 9 \cdot 8 \cdot 2$. (We keep it written like this, for future simplifications.)

Let us consider the events $A$ (Gabe picks first), and $B$ (Gabe wins). As sets these events are described as

$$A = \{(x, y, g) : (x, y) \in D\} = D \times \{g\};$$
$$B = \{(x, y, z) \in D \times T : \text{either } [z = g \text{ and } x \leq 6], \text{ or } [z = s \text{ and } y \leq 6]\}.$$ 

The answer to question 2 is

$$P(A) = \frac{|A|}{|X|} = \frac{|D|}{|D| \cdot |T|} = \frac{1}{2}.$$ 

To answer question 3, we need to find $|B|$. Notice that if we define the sets

$$B_1 = \{(x, y, g) : (x, y) \in D, x \leq 6\},$$
$$B_2 = \{(x, y, s) : (x, y) \in D, y \leq 6\},$$

the $B = B_1 \cup B_2$, with $B_1 \cap B_2 = \emptyset$, and $|B_1| = |B_2| = 6 \cdot 8$. So the answer to question 3 is

$$P(B) = \frac{|B_1| + |B_2|}{|X|} = \frac{6 \cdot 8 \cdot 2}{9 \cdot 8 \cdot 2} = \frac{2}{3}.$$ 

The answer to question 4 is the conditional probability $P(A|B)$. Since $A \cap B = B_1$, we have

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|B_1|}{|B|} = \frac{6 \cdot 8}{6 \cdot 8 \cdot 2} = \frac{1}{2}.$$ 

10
At this point, we notice that since the above computation shows that
\[ P(A|B) = \frac{1}{2} = P(A), \]
it follows that \( A \) and \( B \) are independent events. So the answer to question 5 is that it does not matter if Gabe picks first or second.

The notion of independence can be expanded beyond pairs.

**Definition.** Let \((X, \mathcal{A}, P)\) be a probability space, and let \(C \subset \mathcal{A}\) be a collection of events. We say that \(C\) is independent, if

- for any finite sub-collection \(\{A_1, A_2, \ldots, A_n\} \subset C\) (it is assumed here that the sets \(A_1, A_2, \ldots, A_n\) are all distinct), one has the equality
  \[ P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n). \]

**Remark 3.** If \(C\) is an independent collection, then any two (distinct) sets in \(C\) form an independent pair. The converse however is not true, as indicated in the exercise below.

**Exercise 4.** Consider the standard fair probability space over the set \(X = \{1, 2, 3, 4\}\), and let \(A = \{1, 2\}\), \(B = \{1, 3\}\) and \(C = \{1, 4\}\). Analyze the independence of the pairs \(\{A, B\}\), \(\{A, C\}\), \(\{B, C\}\), and of the collection \(\{A, B, C\}\).

The exercise below provides a “model” for independent collections.

**Exercise 5.** Let \(n \geq 2\) be an integer, and let \((X_k, \mathcal{A}_k, P_k)_{k=1}^n\) be a system of standard finite probability spaces (not necessarily fair). Consider the product space \(X = \prod_{k=1}^n X_k\), and define the map \(\epsilon : X \to [0, 1]\) by
\[
\epsilon(x) = \prod_{k=1}^n P_k(\{x_k\}), \quad \forall x = (x_1, \ldots, x_n) \in X.
\]
Define the map \(P : \mathcal{P}(X) \to [0, \infty]\) by
\[
P(A) = \sum_{x \in A} \epsilon(x), \quad \forall A \subset X.
\]

(i) Prove that \(P\) is a probability measure on \(\mathcal{P}(X)\).

(ii) Prove that, given sets \(B_k \subset X_k, k = 1, \ldots, n\), one has the equality
\[
P(B_1 \times \cdots \times B_n) = P_1(B_1) \cdot P_2(B_2) \cdots P_n(B_n).
\]
(iii) Suppose some sets \( \emptyset \neq S_k \subseteq X_k, k = 1, \ldots, n \), are given. For each \( k \in \{1, \ldots, n\} \), define the sets \( S_1^k, \ldots, S_n^k \subset X_k \) by

\[
S_j^k = \begin{cases} 
S_k & \text{if } j = k \\
X_k & \text{if } j \neq k
\end{cases}
\]

and the sets \( A_j = S_1^j \times S_2^j \times \cdots \times S_n^j, j = 1, \ldots, n \). Prove that the collection \( \mathcal{C} = \{A_1, A_2, \ldots, A_n\} \) is independent in the probability space \((X, \mathcal{P}(X), P)\).

**Example 4.** Many interesting probability problems originate in lottery games. A simple lottery game is played as follows. One starts with a set of balls identified by numbers, say \( S = \{1, 2, \ldots, n\} \), and a positive integer \( k < n \). A draw is simply a set \( \alpha \subset S \) with \(|\alpha| = k\). The order in which the elements of \( \alpha \) are picked is irrelevant. The sample space for the lottery game is then the set \( X = \{\alpha \subset S : |\alpha| = k\} \).

It is well known that

\[
|X| = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}.
\]

Suppose now a player buys a lottery ticket, say \( \tau = \{1, 2, \ldots, k\} \). When the lottery draw \( \alpha \) is made, the player compares it with \( \tau \) and determines if the ticket is a winning one. The prize is dependent on the number of matches. If \( \tau = \alpha \), the player won the jackpot. If only \( k-1 \) numbers match, the player wins 1st prize. If only \( k-2 \) numbers match, the player wins the 2nd prize, and so on. If \( k-p \) numbers match, the player wins \( p \)th prize. (The jackpot corresponds to \( p = 0 \). Of course, \( p = k-1 \) should be the last prize that could be awarded.)

The following are natural questions that arise in connection with this game.

1. **What is the probability that \( p \)th prize is won?**
2. **What is the probability that the \( p \)th prize, or better, is won?**
3. **Given that \( p \)th prize, or better, is won, what is the probability that \( q \)th prize is won?**
4. **Given that \( p \)th prize, or better, is won, what is the probability that \( q \)th prize, or better, is won?**

To answer these questions, we are going to define the events \( A_0, A_1, \ldots, A_k \) by

\[
A_j = \{\alpha \in X : |\tau \triangle \alpha| = j\},
\]

so that \( A_j \) consists of all draws that match \( \tau \) at exactly \( k-j \) numbers. With these notations, the event considered in question 1 is \( A_p \). The event in question 2 is \( B_p = A_0 \cup A_1 \cup \cdots \cup A_p \). The answer to question 3 is the conditional probability \( P(A_q|B_p) \), and the answer to question 4 is the conditional probability \( P(B_q|B_p) \).
These considerations show that all we have to do is to compute the numbers \(|A_j|, j = 0, 1, \ldots, k\). (Notice that the \(A\)'s are disjoint, so for every \(\ell \leq k\), one has the equality \(|B_\ell| = \sum_{j=0}^\ell |A_j|\).) Fix now some number \(j \in \{0, 1, \ldots, k\}\), and let us compute \(|A_j|\). A draw \(\alpha \in A_j\) can be thought as being formed as a combination of two “mini-draws.” One first picks the \(k - j\) matching numbers, and then one picks the \(j\) bad numbers. This shows that the set \(A_j\) is in one-to-one correspondence with the set

\[ T_j = \{(\alpha_1, \alpha_2) \in \mathcal{P}\{1, \ldots, k\} \times \mathcal{P}\{k + 1, \ldots, n\} : |\alpha_1| = k - j, |\alpha_2| = j\} \]

This immediately gives the equality

\[ |A_j| = \binom{k}{k-j} \cdot \binom{n-k}{j} = \frac{k! \cdot (n-k)!}{(j!)^2 \cdot (k-j)! \cdot (n-k-j)!}. \]

This formula holds if \(j \leq \min\{k, n-k\}\), otherwise \(|A_j| = 0\).

**Exercise 6.** In the card game called Bridge, the 52 cards are dealt to 4 players (identified as North, South, East, and West), who form two teams: N-S and E-W. Answer the following questions.

1. If N-S have seven hearts, what is the probability that E-W have the remaining six hearts distributed 3-3? Under same hypothesis, what is the probability that one of the players East or West has 4 hearts? How about 5 hearts, and 6 hearts? Which is the most likely distribution?
2. If North has 2 Aces, what is the probability that E-W have the other two Aces?
3. If North has at least 2 cards from every suit, what if the probability that South has at least 2 cards from every suit?

Going back to the notion of conditional probability, the following result is very useful.

**Proposition 3.** (Partition Formula) Let \((X, \mathcal{A}, \mathcal{P})\) be a probability space, and let \((B_i)_{i \in I}\) be a countable \(\mathcal{A}\)-partition of \(X\), that is:

- \(I\) is countable;
- \(B_i \in \mathcal{A}, \forall i \in I;\)
- \(B_i \cap B_j = \emptyset, \forall i \neq j;\)
- \(\bigcup_{i \in I} B_i = X.\)

If we define the set \(I_0 = \{i \in I : \mathcal{P}(B_i) > 0\}\), then for any event \(A \in \mathcal{A}\), one has the equality

\[ \mathcal{P}(A) = \sum_{i \in I} \mathcal{P}(A|B_i)\mathcal{P}(B_i). \]
Proof. Start with some event $A \in \mathcal{A}$. Since one has the equality

$$A = \bigcup_{i \in I} [A \cap B_i],$$

by $\sigma$-additivity we get

$$P(A) = \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I_0} P(A \cap B_i) + \sum_{j \in I \setminus I_0} P(A \cap B_j).$$

Notice that, since $A \cap B_j \subset B_j$, one has $0 \leq P(A \cap B_j) \leq P(B_j)$, which in particular gives

$$P(A \cap B_j) = 0, \quad \forall j \in I \setminus I_0.$$

The preceding computation can be continued then as:

$$P(A) = \sum_{i \in I_0} P(A \cap B_i) = \sum_{i \in I_0} \frac{P(A \cap B_i)}{P(B_i)} \cdot P(B_i) = \sum_{i \in I_0} P(A|B_i)P(B_i). \quad \square$$

The Partition Formula is extremely useful, especially when we work in infinite probability spaces. In fact one can use it even when the underlying probability space is not explicitly defined, as suggested in the discussion tomorrow.