ICS 6A
Notes on Discrete Probability

Sandy Irani *

February 1, 2000

1 Probability Spaces

When one performs an **experiment**, there is a set of possible outcomes. We will call this set the **sample space** and will denote it by $\mathcal{S}$. A given experiment can have a variety of different sample spaces associated with it, so it is important to define the sample space one has in mind. For example, consider the experiment of rolling a red die and a blue die. One can choose the sample space to be the set of outcomes which denote the number on each dice:

$$\{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6, \text{ where } i, j \text{ are integers }\}.$$

Here the i represents the number on the red die and the j represents the number on the blue die.

The fact that we are using parentheses instead of curly braces $\{\}$ to denote the pairs indicates that the order of the number matters. This means that $(1, 2)$ is different from $(2, 1)$. This results from the fact that the dice have different colors and we can determine which die shows which value.

An alternative sample space for this experiment is the set of possible values for the sum of the values of the dice:

$$\{i \mid 2 \leq i \leq 12\}.$$

An **event** is a subset of the sample space. Each individual point in the sample space is called an **elementary event**. Consider the sample space $\{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6, \text{ where } i, j \text{ are integers }\}$. An example for an event in this sample space is

$$A = \{(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6)\}.$$

*The last section on variance was written by Pierre Baldi and the section on Bayes Inversion rule was written by Rina Dechter.

Sometimes an event is defined by an occurrence. For example, we might say that A is the event that the red die is a three. What is really meant by this is that A denotes the set of all outcomes where the red die is a three.

The **probability** (or likelihood) that is assigned to each elementary event is called the **distribution** over the sample space. The probability of each elementary event is a real number in the interval from 0 to 1 (inclusive). Also, the sum of the probabilities of all of the elementary events is 1. For the first few examples, we will consider sample spaces where each outcome is equally likely to occur. In this case, we say that there is a **uniform distribution** over the sample space. In cases where some elementary events are more likely than others, there is a **non-uniform distribution** over the sample space.

Note that if the dice are not weighted, there is a uniform distribution over the first sample space but not for the second. For the second sample space, it is more likely that the sum of the values on the dice is 7 than 2 since there are several combinations of dice throws that sum to 7 and only one that sums to 2.

1.1 Sample Spaces with Uniform Distributions

For a given event, we will be interested in the **probability** that the event occurs.

Fact 1 *Let $\mathcal{S}$ be a sample space and let A be an event in the sample space (i.e. $A \subseteq \mathcal{S}$). If $\mathcal{S}$ has the property that each elementary event is equally likely, then*

$$Prob[A] = \frac{|A|}{|\mathcal{S}|}.$$

Thus, in the example above where

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

we have that

$$Prob[A] = \frac{|A|}{|\mathcal{S}|} = \frac{6}{36} = \frac{1}{6}.$$

Let's look at some more examples in the same sample space

Example 2

Experiment: rolling a red and a blue die

Sample Space: $\mathcal{S} = \{(i,j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6,$ where i,j are integer s $\}$, where the i represents the number on the red die and the j represents the number on the blue die.

Event: B is the event that the sum of the values on the two dice is three. $B = \{(1,2),(2,1)\}$, and

$$Prob[B] = \frac{|B|}{|\mathcal{S}|} = \frac{2}{36} = \frac{1}{18}.$$

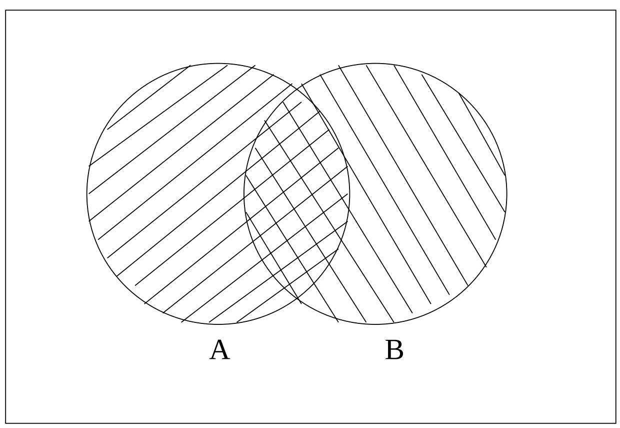
Example 3

Same experiment and sample space from Example 2. C is the event that either the red die or the blue die has a 3.

In order to determine $|C|$, we will define C as the union of the following two sets:

$$\begin{aligned} D &= \text{the red die has a 3} \\ E &= \text{the blue die has a 3} \end{aligned}$$

Note that $C = D \cup E$. As we have argued earlier, $|D| = |E| = 6$. Now we observe that $|C| = |D \cup E| = |D| + |E| - |D \cap E|$. The reason for this equality can be observed in the Venn diagram picture below.



If we sum the sizes of D and E, we have almost got the size of $D \cup E$, except that we have counted the elements in the intersection twice. In order to get $|D \cup E|$, we must subtract off $|D \cap E|$.

$$D \cap E = \text{the red die and the blue die have a 3} = \{(3,3)\}.$$

Thus, we know that $|D \cap E| = 1$ which gives that $|C| = |D \cup E| = |D| + |E| - |D \cap E| = 6 + 6 - 1 = 11$. We are now ready to determine the probability of event C:

$$Prob[C] = \frac{|C|}{|\mathcal{S}|} = \frac{11}{36}.$$

Example 4

Experiment: dealing a hand of five cards from a perfectly shuffled deck of 52 playing cards

Sample Space: the set of all possible five card hands from a deck of playing cards. Since the deck is perfectly shuffled, we can assume that each hand of five cards is equally likely.

G is the event that the hand is a full house.

Recall that a full house is three of a kind and a pair. We know that

$$|\mathcal{S}| = \binom{52}{5},$$

and that

$$|G| = 13 \cdot 12 \cdot \binom{4}{3}\binom{4}{2}.$$

The latter equality comes from the fact that there are thirteen ways to pick the face/number for the three of a kind and there are twelve remaining choices for the face/number of the pair. Once the faces/names have been chosen, there are $\binom{4}{3}$ ways to pick the three of a kind and $\binom{4}{2}$ ways to pick the pair. Thus, we have that

$$Prob[G] = \frac{|G|}{|\mathcal{S}|} = \frac{13 \cdot 12 \cdot \binom{4}{3}\binom{4}{2}}{\binom{52}{5}} = \frac{3744}{2598960} \approx .00144$$

Example 5

Experiment: tossing a fair coin three times.

Sample Space: the set of all possible sequences of outcomes from the three tosses.

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

Since the coin is a fair coin, we can assume that each sequence is equally likely.

G is the event that the number of heads is two.

The number of elements in G is the number of ways to choose the two tosses which result in heads. Thus, $|G| = \binom{3}{2}$, and

$$Prob[G] = \frac{|G|}{|\mathcal{S}|} = \frac{\binom{3}{2}}{8} = \frac{3}{8}.$$

Let's try another event in the same sample space. H is the event that the number of heads is at least one. In this case, it is easier to determine $\bar{H}$ and use the fact that $|H| + |\bar{H}| = |\mathcal{S}|$. (You should verify for yourselves that this follows from the formula that we derived above for the size of union of two sets and the facts that $|H \cap \bar{H}| = 0$ and $H \cup \bar{H} = \mathcal{S}$.)

The event $\bar{H}$ denotes all those sequences of three coin tosses where the number of heads is less than 1 (i.e. the number of heads is 0). Thus, $\bar{H} = \{TTT\}$, and $|\bar{H}| = 1$.

Using our fact above, we have that $|H| = |\mathcal{S}| - |\bar{H}| = 8 - 1 = 7$, so

$$Prob[H] = \frac{|H|}{|\mathcal{S}|} = \frac{7}{8}.$$

1.2 Sample Spaces with Non-uniform Distributions

We now turn to an example of a sample space with a non-uniform distribution. When the distribution is non-uniform, it is important to specify exactly what is the probability of each point in the sample space. This is called the **distribution** over the sample space. That is for each elementary event, $x \in \mathcal{S}$, we must specify $Prob[x]$. These probabilities are always real numbers in the interval from 0 to 1 (inclusive) and must always sum to 1:

$$\sum_{x \in \mathcal{S}} Prob[x] = 1.$$

Example 6

Experiment: rolling two dice

Sample Space: the set of all possible sums of the values of each die.

$$\mathcal{S} = \{i \mid 2 \leq i \leq 12, i \text{ is an integer}\}.$$

If the dice are fair dice, then the probabilities are as follows:

$$\begin{array}{rrr}
 & Prob[5] = \frac{1}{9} & Prob[9] = \frac{1}{9} \\
Prob[2] = \frac{1}{36} & Prob[6] = \frac{5}{36} & Prob[10] = \frac{1}{12} \\
Prob[3] = \frac{1}{18} & Prob[7] = \frac{1}{6} & Prob[11] = \frac{1}{18} \\
Prob[4] = \frac{1}{12} & Prob[8] = \frac{5}{36} & Prob[12] = \frac{1}{36}
\end{array}$$

Now consider the event I which denotes the set of outcomes in which the sum is even. $I = \{2, 4, 6, 8, 10, 12\}$. In order to determine the probability of the event I, we sum up the probabilities of all the elementary events in I:

$$Prob[I] = \sum_{x \in I} Prob[x] = \frac{1}{36} + \frac{1}{12} + \frac{5}{36} + \frac{5}{36} + \frac{1}{12} + \frac{1}{36} = \frac{1}{2}.$$

In general we have the following fact:

Fact 7 *Let $\mathcal{S}$ be a sample space and let A be an event in the sample space (i.e. $A \subseteq \mathcal{S}$), then*

$$Prob[A] = \sum_{x \in A} Prob[x].$$

Note that this more general fact is completely consistent with Fact 1 which covers the case where the distribution over the sample space is uniform. In the uniform case, every single element in the sample space has a probability of $\frac{1}{|\mathcal{S}|}$. If we plug these probabilities into our new definition we get that

$$Prob[A] = \sum_{x \in A} Prob[x] = \sum_{x \in A} \frac{1}{|\mathcal{S}|} = |A| \cdot \frac{1}{|\mathcal{S}|} = \frac{|A|}{|\mathcal{S}|}.$$

In the next example, we will use the following fact:

Fact 8 *Let $\mathcal{S}$ be a sample space and let A and B be events in the sample space (i.e. $A \subseteq \mathcal{S}$ and $B \subseteq \mathcal{S}$), then*

$$Prob[A \cup B] = Prob[A] + Prob[B] - Prob[A \cap B].$$

This fact is the same as saying:

$$\sum_{x \in A \cup B} Prob[x] = \sum_{x \in A} Prob[x] + \sum_{x \in B} Prob[x] - \sum_{x \in A \cap B} Prob[x].$$

In summing the probabilities of the elements in A and the probabilities of the elements in B, we are counting the probability of each element in $A \cap B$ twice and must subtract it off to get the sum of the probabilities in $A \cup B$. Note that this is very similar to the argument we used in determining the size of $A \cup B$.

Example 9

Consider the experiment and sample space from Example 6. Let J be the event that the sum of the numbers is prime or less than six.

We will define J as the union of two sets:

$K =$ outcomes which are prime $= \{2, 3, 5, 7, 11\}$
$L =$ outcomes which are less than $6 = \{2, 3, 4, 5\}$
$K \cap L = \{2, 3, 5\}$

We have that

$$Prob[K] = \sum_{x \in K} Prob[x] = \frac{1}{36} + \frac{1}{18} + \frac{1}{9} + \frac{1}{6} + \frac{1}{18} = \frac{15}{36}.$$

$$Prob[L] = \sum_{x \in L} Prob[x] = \frac{1}{36} + \frac{1}{18} + \frac{1}{12} + \frac{1}{9} = \frac{5}{18}.$$

$$Prob[L \cap K] = \sum_{x \in L \cap K} Prob[x] = \frac{1}{36} + \frac{1}{18} + \frac{1}{9} = \frac{7}{36}.$$

Since $J = K \cup L$,

$$\begin{aligned} Prob[J] &= Prob[K \cup L] = Prob[K] + Prob[L] - Prob[L \cap K] \\ &= \frac{15}{36} + \frac{5}{18} - \frac{7}{36} = \frac{1}{2} \end{aligned}$$

You can verify this by determining the set J directly and summing the probabilities.

For the next example, we will use the following fact:

Fact 10 *Let $\mathcal{S}$ be a sample space and let A be an event in the sample space (i.e. $A \subseteq \mathcal{S}$), then*

$$Prob[A] + Prob[\bar{A}] = 1.$$

You can verify this fact using the following three facts which have already been established:

$$\begin{aligned} Prob[A \cup \bar{A}] &= Prob[A] + Prob[\bar{A}] - Prob[A \cap \bar{A}] \\ A \cup \bar{A} &= \mathcal{S} \\ A \cap \bar{A} &= \emptyset \end{aligned}$$

Example 11

Consider the experiment and sample space from Example 6. The event K is the event that the sum of the dice is greater than two. This means that $\bar{K}$ is the event that the sum of the dice is at most two which implies that $\bar{K} = \{2\}$. $Prob[\bar{K}] = Prob[2] = \frac{1}{36}$. Using Fact 10, we have that

$$Prob[K] = 1 - Prob[\bar{K}] = 1 - \frac{1}{36} = \frac{35}{36}.$$

2 Conditional Probability

Sometimes when we are determining the probability that an event occurs, knowing that another event has occurred gives us some information about whether the first event is more or less likely to occur.

Suppose, for example, that I am determining the probability that attendance is greater than 90% for a randomly chosen lecture in this class. Suppose I tell you that the randomly chosen lecture falls on a Friday, would that change the probability that attendance is greater than 90%? What if I tell you that the lecture falls on a Monday?

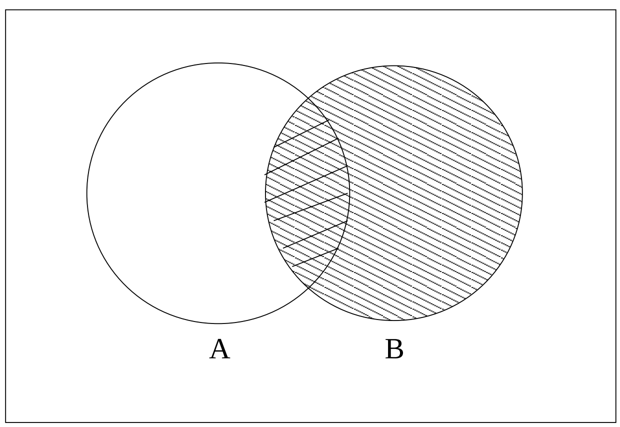
Alternatively, suppose that I am dealing you a five card hand from a deck of cards. Suppose we are trying to determine the probability that the fourth card you are dealt is an ace. If I tell you that there were two aces in the first three cards dealt, then it is less likely that the fourth card will be an ace since there are fewer aces in the deck.

In order to quantify this idea, we need the notion of **conditional probability**. Suppose we have a sample space $\mathcal{S}$ and two events A and B. The **probability of A given B** (also called the probability of event A conditional on event B), denoted $Prob[A \mid B]$, is the probability that A occurs if B occurs. The formula for the probability is:

$$Prob[A \mid B] = \frac{Prob[A \cap B]}{Prob[B]}.$$

This is probably best viewed using the Venn Diagram shown below. If I tell you that event B will definitely occur, then we are limiting our view to the portion of the sample space defined by B. That is, we are now defining B to be our new sample space. We need to divide by $Prob[B]$ in order to *re-normalize* so that the sum of the probabilities for the events in our new sample space sum to 1.

Once we have done this, we are interested in the probability of A, limited to the portion of the sample space defined by B. This is just $Prob[A \cap B]$.



Let's look at an example. There are thirty lectures in this course. There are ten that fall on a Friday. Of those nine have attendance 90% or greater and one has attendance less than 90%. There are ten that fall on a Wednesday. Of those five have attendance greater than 90%. There are ten that fall on a Monday. Of those, two have attendance greater than 90%. The sample space looks as follows:

$$\begin{aligned}
Prob[(M,\ \text{attendance } > 90\%)] &= \frac{2}{30} \\
Prob[(M,\ \text{attendance } <= 90\%)] &= \frac{8}{30} \\
Prob[(W,\ \text{attendance } > 90\%)] &= \frac{5}{30} \\
Prob[(W,\ \text{attendance } <= 90\%)] &= \frac{5}{30}
\end{aligned}$$

$$\begin{aligned} Prob[(F,\ \text{attendance } > 90\%)] &= \frac{9}{30} \\ Prob[(F,\ \text{attendance } <= 90\%)] &= \frac{1}{30} \end{aligned}$$

Let A be the event that a randomly chosen lecture has attendance $> 90\%$. Let B be the event that a randomly chosen lecture falls on a Friday. Let C be the probability that a randomly chosen lecture falls on a Monday.

$$A = \{(M,\ \text{attendance } > 90\%), (W,\ \text{attendance } > 90\%), (F,\ \text{attendance } > 90\%)\}.$$

The probability of A is $Prob[A] = 2/30 + 5/30 + 9/30 = 16/30$. What is the probability of A given B? To determine this, we will need to know:

$$Prob[A \cap B] = Prob[(F,\ \text{attendance } > 90\%)] = 9/30.$$

We will also need to know:

$$\begin{aligned} Prob[B] = Prob[(F,\ \text{attendance } > 90\%), (F,\ \text{attendance } <= 90\%)] \\ = \frac{9}{30} + \frac{1}{30} = \frac{1}{3} \end{aligned}$$

The probability that a randomly chosen lecture has attendance greater than 90% given that the lecture falls on a Friday is:

$$Prob[A \mid B] = \frac{Prob[A \cap B]}{Prob[B]} = \frac{9/30}{1/3} = \frac{9}{10}$$

What about if the lecture falls on a Monday? We will need to know

$$Prob[A \cap C] = Prob[(M,\ \text{attendance } > 90\%)] = 2/30.$$

We will also need to know:

$$Prob[C] = Prob[(M,\ \text{attendance } > 90\%), (M,\ \text{attendance } <= 90\%)] \quad = \frac{2}{30} + \frac{8}{30} = \frac{1}{3}$$

The probability that a randomly chosen lecture has attendance greater than 90% given that the lecture falls on a Monday is:

$$Prob[A \mid C] = \frac{Prob[A \cap C]}{Prob[C]} = \frac{2/30}{1/3} = \frac{1}{5}$$

Let's look at another example: Suppose you are dealt a hand of five cards. The sample space will be the set of all possible five card hands. Let A be the event that the hand has only clubs and spades. If we look at the probability of A, it is just:

$$Prob[A] = \frac{\binom{26}{5}}{\binom{52}{5}} \approx .0253.$$

Now consider the event B that there are no hearts in the hand. If I tell you that event B has occurred, how will we revise our determination of the probability of A? In other words, what is the probability of A conditional on the event B?

$$Prob[A \mid B] = \frac{Prob[A \cap B]}{Prob[B]} = \frac{\binom{26}{5}}{\binom{39}{5}} \approx .114.$$

Now let the event A be the event that there are exactly two aces in the hand. Let the event B be the event that there are exactly two kings. What is the probability of A?

$$Prob[A] = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \approx .040.$$

What is the probability of A given B?

$$Prob[A \mid B] = \frac{Prob[A \cap B]}{Prob[B]} = \frac{\binom{4}{2}\binom{4}{2}44}{\binom{4}{2}\binom{48}{3}} \approx .015.$$

3 Independent Events

An important special case of conditional probabilities is when the information about whether event B has occurred gives you no information about whether A has occurred. This happens when the probability of A given B is the same as the probability of A. In this case, we say that the two events are **independent**.

$$Prob[A \mid B] = \frac{Prob[A \cap B]}{Prob[B]} = Prob[A].$$

This means that $Prob[A \cap B] = Prob[A] \cdot Prob[B]$ which also implies that

$$Prob[B \mid A] = \frac{Prob[B \cap A]}{Prob[A]} = Prob[B].$$

For example, suppose I toss a fair coin twice. If the first coin turns up heads, then it is still the case, that the probability that the next coin turns up heads is 1/2. This is why we can use the reasoning that the probability that both tosses come up heads is $(1/2) \cdot (1/2) = 1/4$. We can multiply the probability that the first toss comes up heads by the probability that the next toss comes up heads to get the probability that they both come up heads.

Note that this does **not** work if the two events are not independent. If we look at the example above for class attendance, suppose we want to determine the probability that a randomly chosen class falls on a Friday and has more than 90% attendance. The probability that a randomly chosen class falls on a Friday is 1/3. The probability that a randomly chosen class has more than 90% attendances is:

$$Prob[(M, > 90)] + Prob[(W, > 90)] + Prob[(F, > 90)] = \frac{9+5+2}{30} = \frac{8}{15}.$$

We have that:

$$Prob[M] \cdot Prob[> 90] = \left(\frac{1}{3}\right)\left(\frac{8}{15}\right) = \frac{8}{45}.$$

This is not equal to the true probability that a randomly chosen class falls on a Friday and has more than 90% attendance which is $Prob[(M, > 90)] = 9/30$.

If we have a whole series of events which are all mutually independent, we can multiply the probability that each event occurs to get the probability that they all occur. For example, suppose we roll a dice ten times. What is the probability that the number shown on the dice is odd every time? We will call this event A. Let A_j be the event that the number shown on the dice is odd on the j^{th} roll for $1 \leq j \leq 10$. We know that $Prob[A_j] = 1/2$. Since all ten events are independent, $Prob[A]$ is the product of all the $Prob[A_j]$'s. This tells us that $Prob[A] = (1/2)^{10}$.

Note: In the above example, we are assuming that all of the coin tosses are mutually independent. This is a standard assumption for experiments like multiple tosses of a coin or rolls of a die. However, in order to mathematically prove that two events A and B are independent, one has to determine the probabilities of A, B and $A \cap B$ and then show that

$$Prob[A] \cdot Prob[B] = Prob[A \cap B].$$

Try this with the following example: suppose you are using an ATM to take out some money from your account. Let E be the event that the machine eats your card. Let O be the event that the machine does not have enough money for your request. We have the following distribution on the combination of these events:

$$\begin{aligned} Prob[O \text{ and } E] &= \frac{1}{200} \\ Prob[O \text{ and } \bar{E}] &= \frac{9}{200} \\ Prob[\bar{O} \text{ and } E] &= \frac{19}{200} \\ Prob[\bar{O} \text{ and } \bar{E}] &= \frac{171}{200} \end{aligned}$$

Are the events O and E independent?

4 Random Variables

A random variable is a function which is defined on a sample space. For example, consider again the experiment of flipping three coins and define the sample space to be

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

Then one could define the random variable X to be the number of heads in the sequence. Note that in defining a random variable, we are defining a function which assigns a unique real number to ever elementary event in the sample space. For example, the number assigned to HHH is three. The number assigned to HTT is one.

Once the random variable X is defined, we can talk about events like $[X = 3]$ which denotes the subset of points in the sample space for which the value of the random variable is 3. Thus, the event $[X = 3]$ is just $\{HHH\}$. The event $[X \leq 1]$ is $\{TTT, TTH, THT, HTT\}$.

We can evaluate the probability of one of these events exactly as we did before. Suppose that each sequence of coin tosses is equally likely. Then

$$Prob[X \leq 1] = \frac{|\{TTT, TTH, THT, HTT\}|}{|\mathcal{S}|} = \frac{4}{8} = \frac{1}{2}.$$

4.1 Expectations of Random Variables

One of most fundamental properties one wants to determine about a random variable is its **expectation**. To give you some intuition about the expectation, if the distribution over the sample space is uniform, then the expectation is the average value of the random variable over the sample space.

Definition 12 *The expectation of random variable X is denoted by $E[X]$ and is defined to be*

$$E[X] = \sum_x (x \cdot Prob[X = x]),$$

where the sum ranges over all the possible values that X could be.

Let's consider the random variable X which is the number of heads in three consecutive coin tosses of a fair coin. We know that X can only be one of the following numbers $\{0, 1, 2, 3\}$.

We have the following set of probabilities for the four possible out comes:

$$\begin{aligned}
Prob[X = 0] &= \frac{|\{TTT\}|}{|\mathcal{S}|} = \frac{1}{8} \\
Prob[X = 1] &= \frac{|\{TTH, THT, HTT\}|}{|\mathcal{S}|} = \frac{3}{8} \\
Prob[X = 2] &= \frac{|\{THH, HTH, HHT\}|}{|\mathcal{S}|} = \frac{3}{8} \\
Prob[X = 3] &= \frac{|\{HHH\}|}{|\mathcal{S}|} = \frac{1}{8}
\end{aligned}$$

The set of probabilities associated with each of the values which X can be is called the **distribution** over X.

Now to determine the expectation of X:

$$\begin{aligned} E[X] &= \sum_{i=0}^{3} i \cdot Prob[X = i] \\ &= \left(0 \cdot \frac{1}{8}\right) + \left(1 \cdot \frac{3}{8}\right) + \left(2 \cdot \frac{3}{8}\right) + \left(3 \cdot \frac{1}{8}\right) = \frac{12}{8} = 1.5 \end{aligned}$$

Example 13

Suppose I offer you the following gamble. You must pay me \$10 up-front to play this game. I will deal you a hand of five cards from a perfectly shuffled deck of cards so that each hand of five cards is equally likely. You reveal your hand and I will give you \$20 for every ace in your hand. What are your expected winnings?

The sample space is the set of all five card hands which are all equally likely. The random variable which we define is your earnings for that hand. That is, if a hand has two aces, your total earnings would be $2 \cdot \$20 - \$10 = \$30$. (The $-\$10$ is for the price you paid to play the game). Since the number of aces in the hand can be either 0,1,2,3, or 4. This means that the amount you earn will be one of the following five values $\{-\$10, \$10, \$30, \$50, \$70\}$. We can assign the following probabilities to these amounts:

$$\begin{aligned} Prob[\text{Earnings } = -\$10] &= Prob[0 \text{ Aces}] = \frac{\binom{4}{0}\binom{48}{5}}{\binom{52}{5}} \approx .659 \\ Prob[\text{Earnings } = \$10] &= Prob[1 \text{ Aces}] = \frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}} \approx .299 \\ Prob[\text{Earnings } = \$30] &= Prob[2 \text{ Aces}] = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \approx .040 \\ Prob[\text{Earnings } = \$50] &= Prob[3 \text{ Aces}] = \frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} \approx .0017 \\ Prob[\text{Earnings } = \$70] &= Prob[4 \text{ Aces}] = \frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} \approx= .00002 \end{aligned}$$

Now to determine the expectation of your winnings:

$$\begin{aligned} E[X] &= \sum_{i=0}^{4} (20i - 10) \frac{\binom{4}{i}\binom{48}{5-i}}{\binom{52}{5}} \\ &\approx (-10 \cdot .659) + (10 \cdot .299) + (30 \cdot .04) + (50 \cdot .0017) + (70 \cdot .00002) \\ &\approx -\$2.30 \end{aligned}$$

Your expected winnings are $-\$2.30$. The error comes from the fact that we rounded the probabilities. We will see later that the expected winnings are in fact exactly $-\$2.30$.

Example 14

Now suppose I offer you another gamble. Again, I will deal you a five card hand from a perfectly shuffled deck, and again you must pay me \$10 to play. But this time, if you get no aces, I will pay you \$1, if you get one ace, I will pay you \$10, two aces, I pay you \$100, three aces and I pay you \$1000 and four aces and I will give you \$10000. What would your expected earnings be in this game? That is, if you have i aces, I will give you 10^i dollars.

Again, we will define a random variable which defines the amount of earnings you get for each hand. If a hand has i aces, then the value of the random variable is $10^i - 10$. We can use the same probabilities from the previous example:

$$\begin{aligned}
E[X] &= \sum_{i=0}^{4}(10^i - 10)\frac{\binom{13}{i}\binom{48}{5-i}}{\binom{52}{5}} \\
&\approx (-9 \cdot .659) + (0 \cdot .299) + (90 \cdot .04) + (990 \cdot .0017) + (9990 \cdot .00002) \\
&\approx -\$2.702
\end{aligned}$$

4.2 Linearity of Expectations

A very useful property of expectation is that in order to get the expectation of the sum of two random variables, you simply have to sum their expectations:

Fact 15 *Let X and Y be any two random variables, then*

$$E[X+Y] = E[X] + E[Y].$$

This actually applies to any number of random variables which you may wish to sum:

Fact 16 *Let $X_1, X_2, \ldots, X_n$ be any n random variables. Then*

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

This fact can be very helpful in reducing the amount of calculations necessary in determining expectations of random variables.

For example, in the beginning of Section 4.1, we were determining the expected number of heads in a sequence of three coin tosses. We could have done this by defining three random variables:

$$\begin{aligned} X_1 &= \text{the expected number of heads in the first coin toss} \\ X_2 &= \text{the expected number of heads in the second coin toss} \\ X_3 &= \text{the expected number of heads in the third coin toss} \end{aligned}$$

Notice that if X is the total number of heads in the sequence, then $X = X_1 + X_2 + X_3$. This means that $E[X] = E[X_1] + E[X_2] + E[X_3]$. X_i will be easier to determine than X.

The value of X_1 is either 1 or 0. The probability that $X_1 = 1$ is the probability the first coin toss turns up heads which is $\frac{1}{2}$. The probability that $X_1 = 0$ is the probability that the first coin toss ends up tails. Thus, we have that

$$E[X_1] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}.$$

Since X_2 and X_3 are distributed exactly the same as X_1, we have that $E[X_1] = E[X_2] = E[X_3] = \frac{1}{2}$ which gives that $E[X] = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} = \frac{3}{2}$.

(If you are not convinced that this method is easier, try determining the expected number of heads in a sequence of 20 coin tosses).

Here is another useful fact which would have saved us some work in the previous section:

Fact 17 *Let X be a random variable and let c_1 and c_2 be two real numbers. Then*

$$E[c_1 \cdot X + c_2] = c_1 \cdot E[X] + c_2.$$

Let's revisit the example where you pay me \$10 to play the game where I deal you five cards and give you \$20 for each ace in your hand. Let Y denote the random variable which is the number of aces in your hand. Let Z be the random variable which denotes your earnings for a given hand. Then $Z = 20Y - 10$. We now know that $E[Z] = 20 \cdot E[Y] - 10$, so all we have to do is to determine $E[Y]$.

To determine $E[Y]$ we will use the additive properties of expectations. Define Y_j to be the number of aces in the j^{th} card which you are dealt for $1 \leq j \leq 5$. Y_j is 1 if the j^{th} card is an ace and Y_j is 0 if it is not an ace. Note that $Y = Y_1 + Y_2 + Y_3 + Y_4 + Y_5$.

What is the distribution over Y_j? A given random card is an ace with probability $\frac{4}{52} = \frac{1}{13}$. This means that $Y_j = 1$ with probability $\frac{1}{13}$ and $Y_j = 0$ with probability $\frac{12}{13}$. For $1 \leq j \leq 5$,

$$E[Y_j] = \left(1 \cdot \frac{1}{13}\right) + \left(0 \cdot \frac{12}{13}\right) = \frac{1}{13}.$$

So we have that

$$E[Y] = E[Y_1] + E[Y_2] + E[Y_3] + E[Y_4] + E[Y_5] = \frac{5}{13}.$$

Putting it all together our expected earnings

$$E[Z] = 20 \cdot E[Y] - 10 = 20\left(\frac{5}{13}\right) - 10 = -2.30.$$

5 Conditional Expectations

The idea of conditional probabilities can be extended to random variables as well. Suppose a football team tends to perform better at their home games than their away games. Let's consider the random variable P which denotes the number of points the team scores in a given game. We will base our probabilities on last year's record, so the expectation of P is the average number of points the team scored over all of last season's games. Now consider the event H that a game is at home. If we look at the random variable P given that the game is at home, we limit our attention to those games where the team was at home. The expectation of random variable P is simply:

$$\sum_j j \cdot Prob[P = j].$$

The expectation of P **conditional on the event** H is

$$\sum_j j \cdot Prob[P = j \mid H].$$

The $Prob[P = j \mid H]$ is the same kind of conditional probability we saw in the previous section. It is the probability that the event $P = j$ occurs, conditional on the event H.

Let's make this example more concrete with numbers. Suppose there were ten games last season. We will denote each game by a pair in $\{A, H\} \times \mathcal{N}$. The first item in the pair says whether the game was home or away. The second item is a number which denotes the score of the team in that game. For last season, we have:

$$(H, 24), (A, 13), (H, 27), (A, 21), (H, 14), (A, 31), (H, 35), (A, 0), (H, 24), (A, 21).$$

Let P denote the random variable which is the number of points the team scores in a randomly chosen game from last season. $P \in \{0, 13, 14, 21, 24, 27, 31, 35\}$. We have the following distribution over P:

$$\begin{aligned}
Prob[P = 0] &= \frac{1}{10} \\
Prob[P = 13] &= \frac{1}{10} \\
Prob[P = 14] &= \frac{1}{10} \\
Prob[P = 21] &= \frac{1}{5}
\end{aligned}$$

$$
\begin{aligned}
Prob[P = 24] &= \frac{1}{5} \\
Prob[P = 27] &= \frac{1}{10} \\
Prob[P = 31] &= \frac{1}{10} \\
Prob[P = 35] &= \frac{1}{10}
\end{aligned}
$$

The expectation of P is:

$$0 \cdot \left(\frac{1}{10}\right) + 13 \cdot \left(\frac{1}{10}\right) + 14 \cdot \left(\frac{1}{10}\right) + 21 \cdot \left(\frac{1}{5}\right) + 24 \cdot \left(\frac{1}{5}\right) + 27 \cdot \left(\frac{1}{10}\right) + 31 \cdot \left(\frac{1}{10}\right) + 35 \cdot \left(\frac{1}{10}\right) = 21.$$

Now suppose we consider the distribution of P conditioned on the event that the game is a home game. First note that the probability that a randomly chosen game from last season is a home game is $5/10 = 1/2$. Now consider the distribution of P conditional on event H:

$$
\begin{aligned}
Prob[P = 0 \mid H] &= \frac{Prob[P = 0 \ and \ H]}{Prob[H]} = 0 \\
Prob[P = 13 \mid H] &= \frac{Prob[P = 13 \ and \ H]}{Prob[H]} = 0 \\
Prob[P = 14 \mid H] &= \frac{Prob[P = 14 \ and \ H]}{Prob[H]} = \frac{1/10}{1/2} = \frac{1}{5} \\
Prob[P = 21 \mid H] &= \frac{Prob[P = 21 \ and \ H]}{Prob[H]} = 0 \\
Prob[P = 24 \mid H] &= \frac{Prob[P = 24 \ and \ H]}{Prob[H]} = \frac{2/10}{1/2} = \frac{2}{5} \\
Prob[P = 27 \mid H] &= \frac{Prob[P = 27 \ and \ H]}{Prob[H]} = \frac{1/10}{1/2} = \frac{1}{5} \\
Prob[P = 31 \mid H] &= \frac{Prob[P = 31 \ and \ H]}{Prob[H]} = 0 \\
Prob[P = 35 \mid H] &= \frac{Prob[P = 35 \ and \ H]}{Prob[H]} \frac{1/10}{1/2} = \frac{1}{5}
\end{aligned}
$$

Note that the sum of the probabilities is 1. The expectation of P conditional on H is:

$$14\left(\frac{1}{5}\right) + 24\left(\frac{2}{5}\right) + 27\left(\frac{1}{5}\right) + 35\left(\frac{1}{5}\right) = 24.8$$

The general rule is that when there is a random variable X defined over a sample space and an event E defined over the same sample space:

$$E[X \mid E] = \sum_j j \cdot Prob[X = j \mid E] = \sum_j \frac{Prob[X = j \ and \ E]}{Prob[E]}.$$

6 Independent Random Variables

Suppose that we have two random variables X and Y. We say that X and Y are **independent** if for every $x, y \in \mathcal{R}$, The events $X = x$ and $Y = y$ are independent. That is, $Prob[X = x \mid Y = y] = Prob[X = x]$ and $Prob[Y = y \mid X = x] = Prob[Y = y]$. In this case, no matter what the value of Y, this gives no information about the value of X and no matter what the value of X, this gives no information about the value of Y. When two random variables are independent, we can calculate the expectation of their product as follows:

$$
\begin{aligned}
E[X \cdot Y] &= \sum_x \sum_y xyProb[X = x \text{ and } Y = y] \\
&= \sum_x \sum_y xyProb[X = x]Prob[Y = y] \\
&= \sum_x \sum_y (x \cdot Prob[X = x])(y \cdot Prob[Y = y]) \\
&= \left(\sum_x x \cdot Prob[X = x]\right)\left(\sum_y y \cdot Prob[Y = y]\right) \\
&= E[X] \cdot E[Y]
\end{aligned}
$$

Warning: this identity

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

is not necessarily true if X and Y are not independent.

Suppose we consider two independent rolls of a fair die. Let X be the random variable which denotes the value of the dice on the first roll. Let Y denote the value of the die on the second roll. X and Y are independent in this case, because if you are told the value of X, the distribution over Y remains exactly the same. So we can use the fact that $E[X \cdot Y] = E[X] \cdot E[Y] = (3.5) \cdot (3.5) = 12.75$.

Note that in the above example, we are assuming that the outcomes of the two rolls are independent. This is generally a standard assumption with experiments like multiple tosses of a coin or rolls of a die. However, if you are asked to prove that two random variables X and Y, then you have to show that for every pair of values x and y,

$$Prob[X = x] \cdot Prob[Y = y] = Prob[X = x \text{ and } Y = y].$$

If you are proving that X and Y are *not* independent, then you only have to find two specific values for x and y such that

$$Prob[X = x] \cdot Prob[Y = y] \neq Prob[X = x \text{ and } Y = y].$$

Try this with the following two random variables: let Z be the random variable which denotes the sum of the values of the two tosses. Are X and Z independent? Why or why not?

7 Bayes Inversion Rule and other Useful Rules

We saw (Fact 10) that the probability of two complimentary events sum to 1. In particular, it is possible to show that

$$P(A) = P(A \cap \bar{B}) + P(A \cap B) \tag{1}$$

Because $A = (A \cap B) \cup (A \cap \bar{B})$ and Fact 8. More generally, if $B_i,\ i = 1, 2, ..., n$ is a set of exhaustive and mutually exclusive events, then $P(A)$ can be computed from $P(A \cap B_i),\ i = 1, 2, ..., n$ using the sum

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) \tag{2}$$

For example, the probability of A= "the outcome of two dice are equal" can be computed by summing over the joint events (A and B_i), $i = 1, 2, 3, 4, 5, 6$, where B_i stands for the proposition "the outcome of the first die is i", yielding,

$$P(A) = \sum_{i=1}^{6} P(A \cap B_i) = 6 \cdot 1/36 = 1/6.$$

As we saw, if A and B are independent $P(A|B) = P(A)$. Since, by definition

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

We get that:

$$P(A \cap B) = P(A|B)P(B) \tag{3}$$

In fact, philosophers frequently stress that natural phenomena provide empirical observations in the form of conditional probabilities, from which the probability of the joint event can be assessed. From EQ. 2 and EQ. 3 we get:

$$P(A) = \sum_{i} P(A|B_i)P(B_i) \tag{4}$$

Namely, the probability of an event A can be computed by *conditioning* it on a set of exhaustive and mutually exclusive events $B_i,\ i = 1, 2, ..., n$. For example, the probability of Equality can be assessed by

$$P(Equality) = \sum_{i} P(Equality|B_i)P(B_i) = 6 \cdot 1/6 \cdot 1/6 = 1/6.$$

The above decomposition provides the basis for assumption-based reasoning. For example, if we wish to calculate the probability that the outcome X of the first die will be greater than the outcome Y of the second, we can condition the event $A : X > Y$ on all the possible values of X and obtain:

$$P(A) = \sum_{i=1}^{6} P(Y < X|X = i)P(X = i)$$

$$= \sum_{i=1}^{6} P(Y < i)\frac{1}{6} = \sum_{i=1}^{6}\sum_{j=1}^{i-1} P(Y = j)\frac{1}{6}$$

$$= \frac{1}{6} \sum_{i=2}^{6} \frac{(i-1)}{6} = \frac{5}{12}$$

Another useful generalization of the product rule (3) is the *chain formula.* It states that if we have a set of n events, $E_1, ..., E_n$, then the probability of the joint event:

$$P(E_1 \cap E_2 \cap ... \cap E_n) = P(E_1|E_{n-1}, ..., E_2, E_1)...P(E_2|E_1)P(E_1)$$

This product can be derived by repeated application of EQ. 3 in any convenient order.

Finally, a very useful formula, Bayes Inversion Rule, allows assessing the probability of hypothesis given evidence. Let H be an hypothesis and E be some observed evidence, then,

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

$P(H|E)$ is sometimes called the posterior probability, and $P(H)$ the prior probability. Notice that the denominator can be derived by an operation called normalization:

$$P(E) = P(E|H)P(H) + P(E|\bar{H})P(\bar{H})$$

Example 18 *Consider the case when you have two urns containing red and blue balls. Urn1 has 3 red balls and 2 blue balls, urn2 had has 3 red balls and 5 blue balls. You picked an urn randomly and selected one ball. Lets assume that the outcome is red. What is the probability that you selected urn 1.*

Solution: Let X be the random variable standing for the urn that is selected. Let Y, be the color of the ball selected. We want to compute: $P(X = urn1|Y = red)$. Using Bayes inversion rule we get:

$$P(X = urn1|Y = red) = \frac{P(Y = red|X = urn1)P(X = urn1)}{P(Y = red|X = urn1)P(X = urn1) + P(Y = red|X = urn2)P(X = urn2)}$$

$$P(X = urn1|Y = red) = \frac{3/5 \cdot 1/2}{3/5 \cdot 1/2 + 3/8 \cdot 1/2} = 48/78$$

8 The Variance of a Random Variable

In order to summarize the information contained in a random variable, we can use its expectation. We have seen that the expectation corresponds to the mean value. The expectation alone, however, does not provide us with any information about how concentrated or dispersed a random variable is. The variance gives us such a measure of dispersion.

Definition 19 *The variance of a random variable X is denoted by* $\mathrm{Var}[X]$ *and is defined to be*

$$\mathrm{Var}[X] = E\left([X - E(X)]^2\right) = \sum_x \left([x - E(X)]^2 \cdot Prob[X = x]\right),$$

where the sum ranges over all the possible values that X could be.

Thus the variance is the average or expected squared deviation from the expectation. Alternatively, we can also use the standard deviation.

Definition 20 *The standard deviation of a random variable X is the square root of its variance.*

Consider a fair die with the random variable X associated with the outcome of the toss. Then the expectation is

$$E(X) = \sum_{i=1}^{6} i \cdot \frac{1}{6} = 3.5$$

and the variance

$$\mathrm{Var}(X) = \sum_{i=1}^{6} (i - 3.5)^2 \cdot \frac{1}{6} = \frac{35}{12} \approx 2.92$$

The standard deviation of X is $\sqrt{35/12} \approx 1.71$.

In contrast, consider now a die that is *not* fair, with a random variable Y associated with only two possible outcomes $Y = 3$ and $Y = 4$. Assume that each outcome has probability 0.5. Then clearly $E(Y) = E(X) = 3.5$. The two random X and Y are indistinguishable from their expectation alone. But Y is more concentrated. Indeed, this is reflected in its much smaller variance:

$$\mathrm{Var}(Y) = \sum_{i=3}^{4} (i - 3.5)^2 \cdot \frac{1}{2} = \frac{1}{4} = 0.25$$

The standard deviation of Y is $\sqrt{1/4} = 0.5$.

Fact 21 *The variance satisfies*

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

To prove this fact, we expand the square in the definition of the variance

$$\mathrm{Var}(X) = \sum_x \left(x^2 + [E(X)]^2 - 2xE(X)\right) \cdot Prob[X = x].$$

We then use the distributivity and calculate each term separately. By definition of the expectation, the first term satisfies $\sum_x x^2 \cdot Prob[X = x] = E(X^2)$. By using the linearity of the expectation, the second term satisfies $\sum_x [E(X)]^2 \cdot Prob[X = x] = [E(X)]^2$. Using again the linearity of the expectation, the third term satisfies $\sum_x -2xE(X)) \cdot Prob[X = x] = -2E(X)E(X) = -2[E(X)]^2$. Collecting terms, we finally get $\mathrm{VarX} = E(X^2) + [E(X)]^2 - 2[E(X)]^2 = E(X^2) - [E(X)]^2$.

Exercise: Calculate the variance of the dice above using both the definition and this formula.