which in turn implies

$$P(Y = 2 \text{ and } X = 6) = P(Y = 2)\ P(X = 6) \tag{3.2}$$

In other words, the events $\{X = 6\}$ and $\{Y = 2\}$ are independent, and similarly the events $\{X = i\}$ and $\{Y = j\}$ are independent for any i and j. This leads to our formal definition of independence:

Definition 4 *Random variables X and Y are said to be independent if for any sets I and J, the corresponding events $\{X \text{ is in } I\}$ and $\{Y \text{ is in } J\}$ are independent, i.e.*

$$P(X \textit{ is in } I \textit{ and } Y \textit{ is in } J) = P(X \textit{ is in } I) \cdot P(Y \textit{ is in } J) \tag{3.3}$$

So the concept simply means that X doesn't affect Y and vice versa, in the sense that knowledge of one does not affect probabilities involving the other. The definition extends in the obvious way to sets of more than two random variables.

The notion of independent random variables is absolutely central to the field of probability and statistics, and will pervade this entire book.

3.4 Example: The Monty Hall Problem

This problem, while quite simply stated, has a reputation as being extremely confusing and difficult to solve [37]. Yet it is actually an example of how the use of random variables in "translating" the English statement of a probability problem to mathematical terms can simplify and clarify one's thinking, making the problem easier to solve. This "translation" process consists simply of naming the quantities. You'll see that here with the Monty Hall Problem.

Imagine, this simple device of introducing named random variables into our analysis makes a problem that has vexed famous mathematicians quite easy to solve!

The Monty Hall Problem, which gets its name from a popular TV game show host, involves a contestant choosing one of three doors. Behind one door is a new automobile, while the other two doors lead to goats. The contestant chooses a door and receives the prize behind the door.

The host knows which door leads to the car. To make things interesting, after the contestant chooses, the host will open one of the other doors not chosen, showing that it leads to a goat. Should the contestant now change her choice to the remaining door, i.e. the one that she didn't choose and the host didn't open?

Many people answer No, reasoning that the two doors not opened yet each have probability 1/2 of leading to the car. But the correct answer is actually that the remaining door (not chosen by the contestant and not opened by the host) has probability 2/3, and thus the contestant should switch to it. Let's see why.

Again, **the key is to name some random variables**. Let

- C = contestant's choice of door (1, 2 or 3)
- H = host's choice of door (1, 2 or 3), after contestant chooses
- A = door that leads to the automobile

We can make things more concrete by considering the case $C = 1,\ H = 2$. The mathematical formulation of the problem is then to find the probability that the contestant should change her mind, i.e. the probability that the car is actually behind door 3:

$$P(A = 3 \mid C = 1,\ H = 2) = \frac{P(A = 3,\ C = 1,\ H = 2)}{P(C = 1,\ H = 2)} \tag{3.4}$$

You may be amazed to learn that, really, we are already done with the hard part of the problem. Writing down (3.4) was the core of the solution, and all that remains is to calculate the various quantities above. This will take a while, but it is pretty mechanical from here on, simply going through steps like those we took so often in earlier chapters.

Write the numerator as

$$P(A = 3,\ C = 1)\ P(H = 2 \mid A = 3,\ C = 1) \tag{3.5}$$

Since C and A are independent random variables, the value of the first factor in (3.5) is

$$\frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \tag{3.6}$$

What about the second factor? Remember, in calculating $P(H = 2 \mid A = 3,\ C = 1)$, we are given in that case that the host knows that $A = 3$, and since the contestant has chosen door 1, the host will open the only remaining door that conceals a goat, i.e. door 2. In other words,

$$P(H = 2 \mid A = 3,\ C = 1) = 1 \tag{3.7}$$

Now consider the denominator in (3.4). We can, as usual, "break big events down into small events." For the breakdown variable, it seems natural to use A, so let's try that one:

$$P(C = 1,\ H = 2) = P(A = 3,\ C = 1,\ H = 2) + P(A = 1,\ C = 1,\ H = 2) \tag{3.8}$$

(There is no $A = 2$ case, as the host, knowing the car is behind door 2, wouldn't choose it.)

We already calculated the first term. Let's look at the second, which is equal to

$$P(A = 1,\ C = 1)\ P(H = 2 \mid A = 1,\ C = 1) \tag{3.9}$$

If the host knows the car is behind door 1 and the contestant chooses that door, the host would randomly choose between doors 2 and 3, so

$$P(H = 2 \mid A = 1,\ C = 1) = \frac{1}{2} \tag{3.10}$$

Meanwhile, similar to before,

$$P(A = 1,\ C = 1) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} \tag{3.11}$$

So, altogether we have

$$P(A = 3 \mid C = 1,\ H = 2) = \frac{\frac{1}{9} \cdot 1}{\frac{1}{9} \cdot 1 + \frac{1}{9} \cdot \frac{1}{2}} = \frac{2}{3} \tag{3.12}$$

Even Paul Erdös, one of the most famous mathematicians in history, is said to have given the wrong answer to this problem. Presumably he would have avoided this by writing out his analysis in terms of random variables, as above, rather than say, a wordy, imprecise and ultimately wrong solution.

3.5 Expected Value

3.5.1 Generality — Not Just for Discrete Random Variables

The concepts and properties introduced in this section form the very core of probability and statistics. **Except for some specific calculations, these apply to both discrete and continuous random variables, and even the exceptions will be analogous.**

The properties developed for *variance*, defined later, also hold for both discrete and continuous random variables.

3.5.2 Misnomer

The term "expected value" is one of the many misnomers one encounters in tech circles. The expected value is actually not something we "expect" to occur. On the contrary, it's often pretty unlikely or even impossible.

For instance, let H denote the number of heads we get in tossing a coin 1000 times. The expected value, you'll see later, is 500. This is not surprising, given the symmetry of the situation and the fact (to be brought in shortly) that the expected value is the mean. But $P(H = 500)$ turns out to be about 0.025. In other words, we certainly should not "expect" H to be 500.

Of course, even worse is the example of the number of dots that come up when we roll a fair die. The expected value will turn out to be 3.5, a value which not only rarely comes up, but in fact never does.

In spite of being misnamed, expected value plays an absolutely central role in probability and statistics.

3.5.3 Definition and Notebook View

Definition 5 *Consider a repeatable experiment with random variable X. We say that the expected value of X is the long-run average value of X, as*