Lecture 2 - Laws of Probability

6.042 - February 6, 2003

Let's play *Strange Dice*! The rules are simple. There are three dice, A, B, and C. Not surprisingly, the dice are numbered *strangely*, as shown in Figure 1.

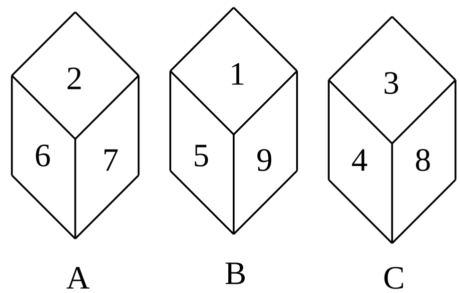

Figure 1: These are the three dice used in Strange Dice. The number on each concealed face is the same as the number on the opposite, exposed face.

The rules are simple. You pick one of the three dice, and then I pick one of the two remainders. We both roll and the player with the higher number wins.

Which of the dice should you choose to maximize your chances of winning? Die B is appealling, because it has a 9, the highest number overall. Then again, die A has two relatively large numbers, 6 and 7. But die C has an 8 and no very small numbers at all. Intuition gives no clear answer!

1 Analysis of Strange Dice

We can analyze Strange Dice using the standard, four-step method for solving probability problems. To fully understand the game, we need to consider three different experiments, corresponding to the three pairs of dice that could be pitted against one another.

1.1 Die A versus Die B

First, let's determine what happens when die A is played against die B.

Step 1: Find the sample space. The sample space for this experiment is worked out in the tree diagram show in Figure 2. It consists of nine ordered pairs:

$$\{\ (2,1),\ (2,5),\ (2,9),\ (6,1),\ (6,5),\ (6,9),\ (7,1),\ (7,5),\ (7,9)\ \}$$

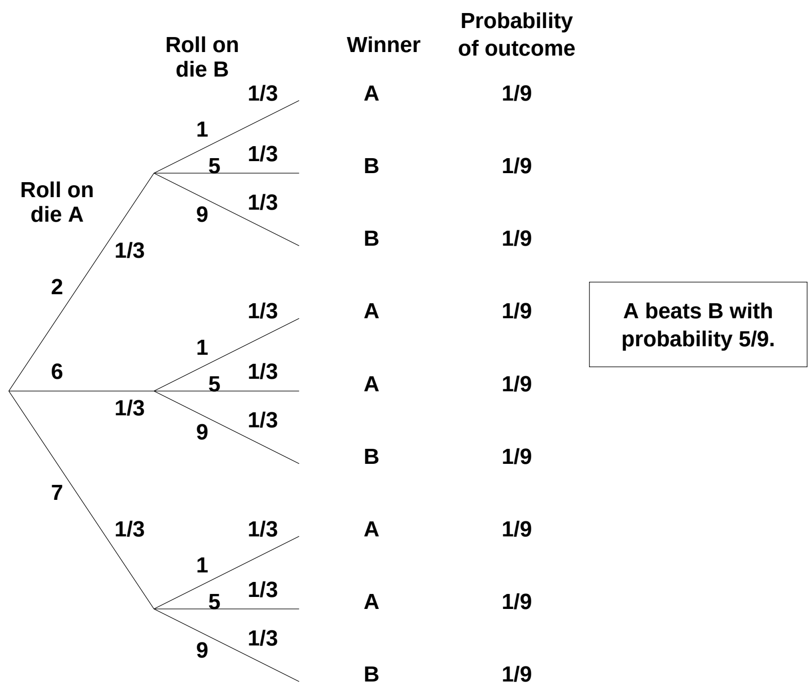

Figure 2: This is the tree diagram for the experiment of playing die A against die B. Die A beats die B with probability $5/9$.

Step 2: Define events of interest. We are interested in the event that the number on die A is greater than the number on die B. This event is a set of five outcomes:

$$\{\ (2,1),\ (6,1),\ (6,5),\ (7,1),\ (7,5)\ \}$$

These outcomes are marked A in the figure.

Step 3: Compute outcome probabilities. To find outcome probabilities, we first assign probabilities to edges in the tree diagram. On each die, each number comes up with probability $1/3$, regardless of the value of the other die. Therefore, we assign all edges probability $1/3$. The probability of an outcome is the product of probabilities on the corresponding root-to-leaf path; this means that every outcome has probability $1/9$. These probabilities are recorded on the right side of the figure.

Step 4: Compute event probabilities. The probability of an event is the sum of the probabilities of the outcomes in that event. Therefore, the probability that die A comes up greater than die B is:

$$\begin{aligned}
\Pr\{A > B\} &= \Pr\{(2,1)\} + \Pr\{(6,1)\} + \Pr\{(6,5)\} + \Pr\{(7,1)\} + \Pr\{(7,5)\} \\
&= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \\
&= \frac{5}{9}
\end{aligned}$$

We conclude that die A beats die B more than half of the time. Therefore, you had better not choose die B or else I'll pick die A and have a better-than-even chance of winning!

1.2 Die B versus Die C

Now suppose that die B is played against die C. The tree diagram for this experiment is shown in Figure 3. The analysis is the same as before, with the conclusion that die B beats die C with probability 5/9 as well. Therefore, you had beter not choose die C; if you do, I'll pick die B and most likely win!

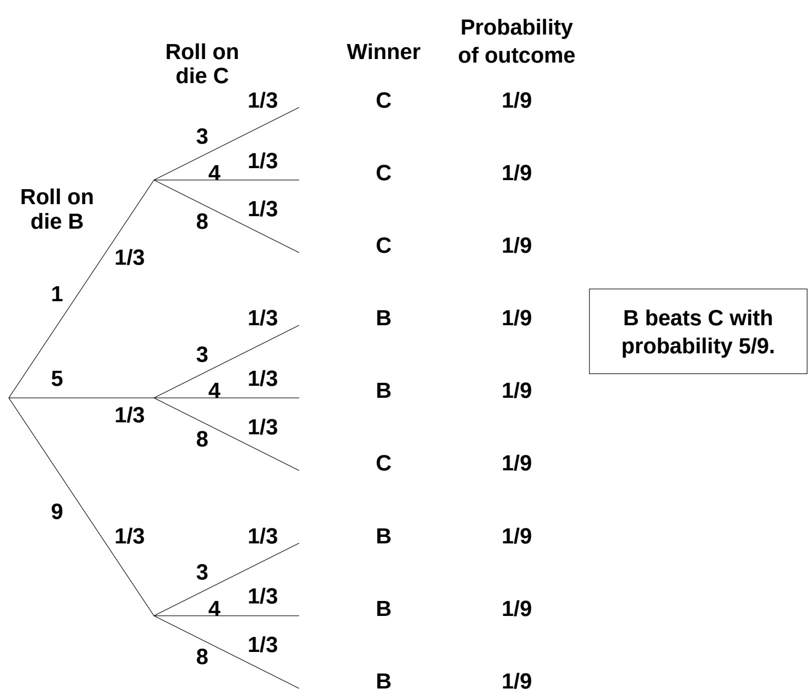

Figure 3: Die B beats die C with probability 5/9.

1.3 Die C versus Die A

We've seen that A beats B and B beats C. Apparently, die A is the best and die C is the worst. The result of a confrontation between A and C seems a forgone conclusion. A tree diagram for this final experiment is worked out in Figure 4. Surprisingly, die C beats die A with probability 5/9!

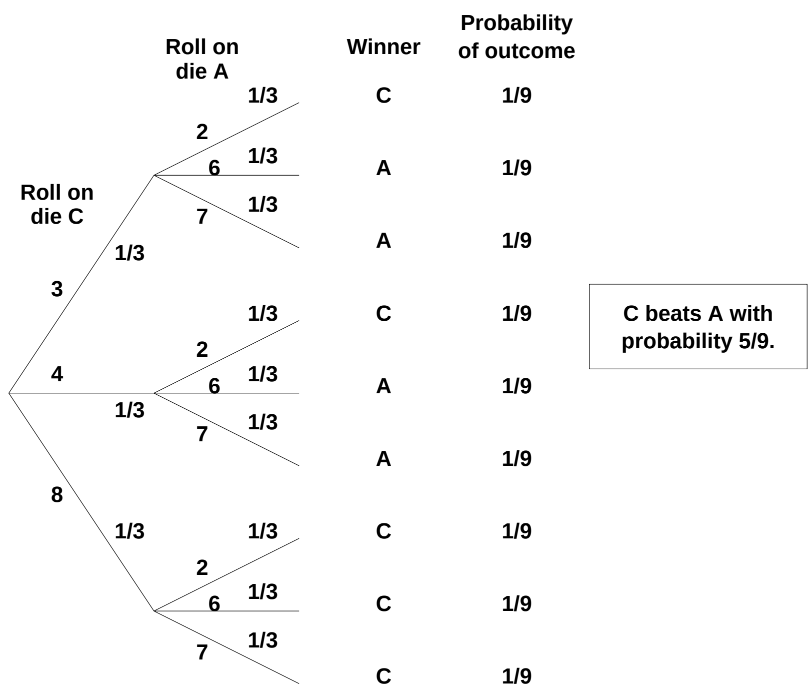

Figure 4: Die C beats die A with probability 5/9. Amazing!

In summary, die A beats B, B beats C, and C beats A! Evidently, there is a relation between the dice that is *not transitive*! This means that no matter what die the first player chooses, the second player can choose a die that beats it with probability 5/9. The player who picks first[1] is always at a disadvantage!

The dice can be renumbered so that A beats B and B beats C each with probability 2/3 and C still beats A with probability 5/9. Can you find such a renumbering?

2 Relations

In the preceding section, we found that there is a relation between the Strange Dice that is *not transitive*: A beats B and B beats C, but A does not beat C. Let's put that assertion on sound footing with a review of relations.

In general, a *relation* is a fundamental mathematical notion expressing a relationship between elements of a set. As one example, $<$ is a relation on the set of natural numbers, $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$. Here is the general definition:

[1] That would be... you. See note in Lecture 1 regarding complicated betting games and mathematicians.

Definition 1 *A* binary relation on a set *S is a subset $R \subseteq S \times S$.*

Thus, we can regard the relation $<$ as the set of all ordered pairs of natural numbers such that the first number is less than the second:

$$'<' \;=\; \{\ (0,1),\ (0,2),\ (0,3),\ \ldots,\ (1,2),\ (1,3),\ \ldots\ \}$$

(Regarding $<$ as the name of a set probably seems pretty strange, but strictly speaking that's what it is.) As another example, $=$ is a relation on the natural numbers as well:

$$'=' \;=\; \{\ (0,0),\ (1,1),\ (2,2),\ (3,3),\ \ldots\ \}$$

Ordinarily, one expresses the fact that relation R holds between elements $a, b \in S$ by writing the name of the relation between the two elements: aRb. (This is called infix notation.) For example, to say that the relation $<$ holds between $x, y \in \mathbb{N}$, we write $x < y$. Probably this comes as quite a shock. In principle, though, one could assert that a relation R holds between elements $a, b \in S$ using set notation by writing $(a, b) \in R$.

Some relations have special properties.

Definition 2 *A binary relation R on a set S is*

- reflexive *if aRa for all $a \in S$.*
- symmetric *if aRb implies bRa for all $a, b \in S$.*
- transitive *if aRb and bRc imply aRc for all $a, b, c \in S$.*

The $<$ relation is not reflexive, since 2 is not less than 2, for example. It is not symmetric, since 2 is less than 3, but 3 is not less than 2. But $<$ is transitive; if $a < b$ and $b < c$, then $a < c$.

The relation $=$ is reflexive ($a = a$), symmetric (if $a = b$, then $b = a$), and transitive (if $a = b$ and $b = c$, then $a = c$). Relations such as $=$ that possess all three properties are called *equivalence relations.*

Let's return to our original reason for discussing relations. In the preceding section, we implicitly established a relation on the set of Strange Dice, which might be called "usually beats". Since we showed that A usually beats B and B usually beats C, but it is not the case that A usually beats C, the relation is not transitive, as we asserted.

3 Laws of Probability

In this section, we list some useful facts about probabilities. You'll use these all the time when you're proving theorems or doing calculations. These rules hold for *all* finite sample spaces S and *all* probability funtions Pr. Many of the proofs rely on set identities, which you can find in Rosen.

Sum Rule If $A_1, A_2, \ldots, A_n \subseteq S$ are disjoint events, then:

$$\Pr\left\{\bigcup_{i=1}^{n} A_i\right\} = \sum_{i=1}^{n} \Pr\left\{A_i\right\}$$

(Recall that events are just sets. So *disjoint events* are disjoint sets, meaning that they have no outcomes in common. As a result, disjoint events are mutually exclusive; if one happens, the other does not.)

Intuitively, the two sides of the equation above are equal because both sum the probabilities of all outcomes in the event $A_1 \cup A_2 \cup \ldots \cup A_n$. The fact that the events A_i are disjoint is critical; if an outcome were in two events, then its probability would be counted twice on the right side.

But this is an appeal to intuition, not a proof. For now, we only prove the sum rule for the case of two disjoint sets, A and B.

$$\begin{aligned}
\Pr\{A \cup B\} &= \sum_{x \in A \cup B} \Pr\{x\} \\
&= \sum_{x \in A} \Pr\{x\} + \sum_{x \in B} \Pr\{x\} \\
&= \Pr\{A\} + \Pr\{B\}
\end{aligned}$$

The first step uses the definition of the probability of an event; namely, the probability of an event is the sum of the probabilities of the outcomes in that event. In the second step, we split the sum into two parts. Since the events A and B are disjoint, each term in the original sum appears in exactly one of the two partial sums. The final step again uses the definition of the probability of an event.

Complement Rule For every event A:

$$\Pr\{A\} + \Pr\{\overline{A}\} = 1$$

Here $\overline{A}$ denotes the *complement* of A, which is the set of all outcomes in the sample space that are not in A. The complement rule follows from the sum rule:

$$\begin{aligned}
\Pr\{A\} + \Pr\{\overline{A}\} &= \Pr\{A \cup \overline{A}\} \\
&= \Pr\{S\} \\
&= 1
\end{aligned}$$

The first step uses the sum rule, which is applicable because A and $\overline{A}$ are disjoint. The second step uses the definition of complement, which implies that the union of an event and its complement is the entire sample space. In the final step, we use the fact that the sum of the probabilities of all outcomes in a sample space is 1. This was one of the two conditions on every probabililty function Pr that we imposed when we defined the notion of a probability space.

Difference Rule For all events A and B:

$$\Pr\{B - A\} = \Pr\{B\} - \Pr\{A \cap B\}$$

This again follows from the sum rule:

$$\begin{aligned}
\Pr\{B - A\} + \Pr\{A \cap B\} &= \Pr\{(B - A) \cup (A \cap B)\} \\
&= \Pr\{B\}
\end{aligned}$$

The first equation uses the sum rule, which is applicable because events $B - A$ and $A \cap B$ are disjoint. In the second step, we use the set identity $(B - A) \cup (A \cap B) = B$. The difference rule is obtained by subtracting $\Pr\{A \cap B\}$ from the first expression and the last one.

Inclusion-Exclusion For all events A and B:

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}$$

The inclusion-exclusion rule can be regarded as a generalization of the sum rule where the sets A and B are not required to be disjoint. Intuitively, the probability of an outcome that lies in both events is counted just once on the left. On the right side, that outcome's probability is *included* by the first term, *included* again by the second term, and then *excluded* by the third term. Thus, overall, its probability is counted just once on the right side as well.

The proof uses the sum rule and set identities:

$$\begin{aligned}
\Pr\{A \cup B\} &= \Pr\{A\} + \Pr\{B - A\} \\
&= \Pr\{A\} + (\Pr\{B - A\} + \Pr\{A \cap B\}) - \Pr\{A \cap B\} \\
&= \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}
\end{aligned}$$

The first step is a valid application of the sum rule, since A and $B - A$ are disjoint and $A \cup (B - A) = A \cup B$. In the second step, we add and subtract $\Pr\{A \cap B\}$ on the right side. The final step uses the sum rule again.

Boole's Inequality For all events A and B:

$$\Pr\{A \cup B\} \leq \Pr\{A\} + \Pr\{B\}$$

Boole's inequality is also sometimes called the *union bound*. It is useful for upper bounding the probability of a union of events. For example, the probability that you ace some quiz in 6.042 is at most the probability that you ace the first one plus the probability that you ace the second.

Boole's Inequality follows from the inclusion-exclusion rule by noting that $\Pr\{A \cap B\}$ is nonnegative.

Monotonicity For all events A and B such that $A \subseteq B$:

$$\Pr\{A\} \leq \Pr\{B\}$$

As usual, the proof uses the sum rule and a set identity:

$$\begin{aligned}
\Pr\{B\} &= \Pr\{A\} + \Pr\{B - A\} \\
&\leq \Pr\{A\}
\end{aligned}$$

The first step uses the sum rule, which applies since A and $B - A$ are disjoint and $A \cup (B - A) = B$. The second step uses the fact that $\Pr\{B - A\}$ is nonnegative.

4 An Upper Bound on Intransitivity

The Strange Dice have the unusual property that A beats B, B beats C, and C beats A, all with probability 5/9. The following theorem puts an upper bound on just how intransitive dice can be. In particular, it says that there is no way to relabel or weight the dice so that each die beats the next with probability greater than 2/3.

Theorem 3 *Suppose that we roll three dice, A, B, and C, with a finite number of sides, arbitrary numbering, and arbitrary weighting. Define the following events:*

$$\begin{aligned} A_1 &= \textit{event that } A \textit{ beats } B \\ A_2 &= \textit{event that } B \textit{ beats } C \\ A_3 &= \textit{event that } C \textit{ beats } A \end{aligned}$$

For these events, the following inequality holds:

$$\min(\ \Pr\{A_1\},\ \Pr\{A_2\},\ \Pr\{A_3\}\ ) \ \leq\ \frac{2}{3}$$

The proof uses many of the probability laws listed in the preceding section.

Proof. We reason as follows. Each step is explained below.

$$\begin{aligned}
\min(\Pr\{A_1\},\Pr\{A_2\},\Pr\{A_3\}) &\leq \frac{1}{3}(\Pr\{A_1\}+\Pr\{A_2\}+\Pr\{A_3\}) && (1)\\
&= \frac{1}{3}\left((1-\Pr\{\overline{A_1}\})+(1-\Pr\{\overline{A_2}\})+(1-\Pr\{\overline{A_3}\})\right) && (2)\\
&= 1-\frac{1}{3}\left(\Pr\{\overline{A_1}\}+\Pr\{\overline{A_2}\}+\Pr\{\overline{A_3}\}\right) && (3)\\
&\leq 1-\frac{1}{3}\left(\Pr\{\overline{A_1}\cup\overline{A_2}\cup\overline{A_3}\}\right) && (4)\\
&= 1-\frac{1}{3}\left(\Pr\{\overline{A_1\cap A_2\cap A_3}\}\right) && (5)\\
&= 1-\frac{1}{3}(1-\Pr\{A_1\cap A_2\cap A_3\}) && (6)\\
&= \frac{2}{3} && (7)
\end{aligned}$$

The first step uses the observation that the minimum of three numbers must be less than or equal to the average. In the second step, we apply the complement rule three times. The next step uses only algebraic simplification. The fourth line uses Boole's inequality for three events. The fifth line uses DeMorgan's Law, a set identity which says that:

$$\overline{X \cap Y \cap Z} \;=\; \overline{X} \cup \overline{Y} \cup \overline{Z}$$

The sixth line uses the complement rule once more. The final line use the fact that $A_1 \cap A_2 \cap A_3 = \emptyset$, since it is not possible that A beats B, B beats C, and C beats A *simultaneously*. □

The preceding theorem can be strengthened by a more complicated argument. In particular, the fraction 2/3 can be replaced by the *golden ratio*:

$$\begin{aligned} \phi &= \frac{\sqrt{5}-1}{2} \\ &\approx 0.618 \end{aligned}$$