Notes 18

1 Probability

For the rest of the semester, we'll be talking about our final topic of the course — probability. This is a subject that comes up, in one form or another, in most areas of computer science. It's important in the design of algorithms, coding theory, cryptography, fault tolerance ... it also comes up in everyday life, in things like political polls and weather prediction.

Unfortunately, probability is one of the least understood and most abused subjects in all of mathematics — which is funny, since it's really very easy, provided you follow a few basic rules when analyzing a problem. People get into trouble when they rely too heavily on their intuition at the expense of these rules.

2 Let's Make a Deal

We'll start with a problem, the "Monty Hall Problem," that recently became sort of notorious. Just for a little background, Monty Hall was the host of a game show called *Let's Make a Deal*, which, depending on whom you ask, was either a great TV show or one of the all-time low points in the history of televsion. A typical scenario on *Let's Make a Deal* was the following: on stage are three doors, numbered 1, 2, and 3; and Monty asks you to pick one of the doors. Behind one of these doors is a new car, and behind the other two is something terrible, like a donkey, or a goat.

So you pick one door. Now Monty, who knows what's behind all the doors, selects a door you didn't pick that has a goat behind it, and opens it. Then he asks you,

> *Do you want to switch doors?*

That is, do you want to stay with the original door you picked, or switch to the other unopened door? We'll spend the rest of this lecture trying to answer this question.

This was the subject of an "Ask Marilyn" column in *Parade* Magazine a few years ago. Marylin wrote that your chances of winning were 2/3 if you switched — because if you switch, then you win if the car was originally behind either of the two doors you didn't pick.

Now, Marylin is listed in the *Guiness Book of World Records* as having the world's highest IQ, but for this answer she got a tidal wave of angry mail, much of it from people with Ph.D.'s in mathematics, telling her she was wrong, and that the answer was 1/2 — because after Monty opens a door, the car is now equally likely to be behind one of the two remaining doors.

It turned out that Marilyn was right. But both of the arguments above are based on intuition, and the answer is actually pretty subtle. So rather than try to come up with our own explanation in words, let's introduce some of the definitions of probability.

3 The Sample Space

The key to this problem, and almost all probability problems, is to look carefully at the space of all possible outcomes.

Definition 3.1 *The* sample space *for a problem (or experiment) is the set consisting of all possible outcomes.*

Definition 3.2 *An* outcome *(a.k.a* sample point *or* atomic event*) of an experiment consists of the total information about the experiment after it has been performed. (This includwes the value of all random choices that were made.)*

Let's re-phrase our main question a little.

> *What is the probability of winning if you switch doors?*

Is it 2/3 as Marylin said, or 1/2 as all those outraged readers said, or neither?

So let's work out the sample space for the problem. An outcome of the experiment consists of

1. The number of the door containing the car.
2. The number of the door chosen by the contestant.
3. The number of the door opened by Monty.

So a sample point is a 3-tuple of numbers, containing these three quantities. For example, in the sample point $(2, 1, 3)$, the contestant won, because the car was behind door 2, she picked 1, Monty opened 3, and therefore she switched to 2.

Not every 3-tuple is a sample point. For example, $(1, 2, 1)$ is not a sample point because Monty never opens the door containing the prize. Similarly, $(1, 2, 2)$ is not a sample point, because he never opens the door you chose, either.

A good way to draw the sample space S is in the form of a tree — first a branch for the first component of the 3-tuple, then a branch for the second, and then a branch for the third. Here's a picture of S.

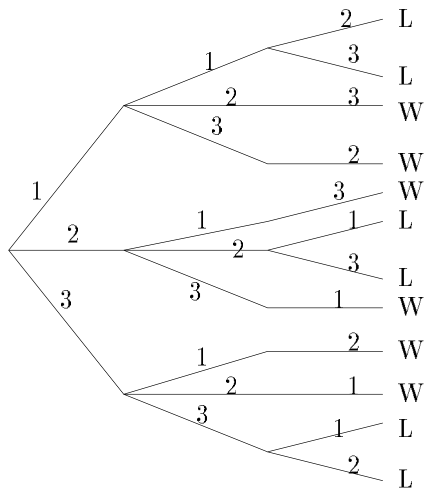

Figure 1: The sample space for our problem

Now, what we really care about is whether each outcome leads to a win (W) or a loss (L) — so we label each sample point accordingly.

Let $S_W \subseteq S$ denote the subset of winning outcomes. We say that S_W is the *event* that the contestant won.

Definition 3.3 *An* event *is a subset of the sample space.*

Any subset of the sample space is an event.

4 The Probability of Winning

What we really want is the probability of winning. In our space S, we have six winning outcomes and six losing outcomes. Does this mean that the probability of winning is 1/2? No! At least, not necessarily! Some outcomes may be more or less likely to occur than others, so we have to be careful. (Embedded in here is the mistake most of Marylin's readers were probably making — though of course we can't know for sure.)

In general, the only way to determine the probability of an event is to define the sample space associated with the problem.

Definition 4.1 *A* probability space *consists of a sample space S and a probability function $Pr : S \to \mathbb{R}$, such that*

1. *For all $s \in S$, $0 \le Pr(S) \le 1$.*
2. $\sum_{s \in S} Pr(s) = 1$.

So a probability space is a sample space with a probability for each outcome. Assigning the correct probabilities to outcomes is the most important part of analyzing one of these problems.

Definition 4.2 *For any event $S' \subseteq S$,*

$$Pr(S') = \sum_{s \in S'} Pr(s).$$

That is, to determine the probability of an event, we just have to add up the probabilities of the outcomes it contains.

To make up probabilities for our outcomes, we'll make some reasonable but arbitrary assumptions. Namely,

1. The prize is placed behind any door with equal probability ($= 1/3$).
2. The contestant picks any door with equal probability, regardless of where the prize was placed ($= 1/3$).
3. Now, if the contestant picks the door with the prize, the Monty has a choice of which of two doors to open, and he picks among them with equal probability ($= 1/2$). Otherwise, he picks the only door he is allowed to open under the rules.

Given this, we can label each outcomes with a probability, as in the next figure. Basically, a 3-tuple that results in Monty having a choice of doors to open has probability

$$\frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{18},$$

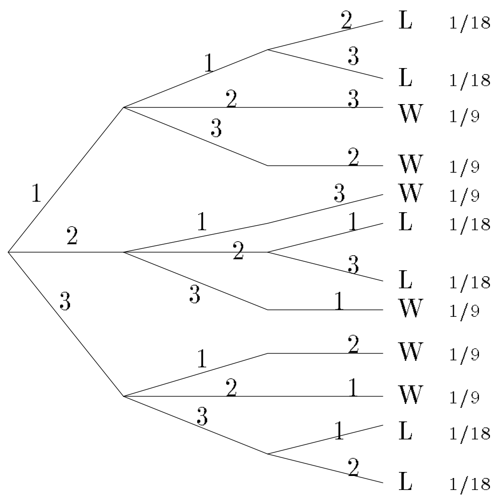

Figure 2: The probability space for our problem

while a 3-tuple that results in Monty having no choice of which door to open has probability

$$\frac{1}{3} \cdot \frac{1}{3} \cdot 1 = \frac{1}{9}.$$

Now it's (finally) easy to go through the space, add up the probabilities of the winning outcomes, and thereby determine the probability of the event S_W, that the contestant wins by switching doors.

The answer is

$$6 \cdot \frac{1}{9} = \frac{6}{9} = \frac{2}{3}.$$

So Marilyn was right, and all those math Ph.D.'s who reasoned with her, pleaded with her, and even threatened her, were wrong.

Tune in next week for another edition of *Let's Make a Deal* ...