Lecture 7: Probability

Chapter 2.x

Outcomes

Probability theory forms the mathematical foundation of statistics. We use probability to build tools to describe and understand randomness.

We often frame probability in terms of a random process giving rise to an outcome.

Typical examples of random processes

- Dice: 6 possible outcomes
- Coin Flip: 2 possible outcomes

Disjoint AKA Mutually Exclusive Outcomes

Two outcomes are disjoint (also called mutually exclusive) if they cannot both occur at the same time.

Back to examples:

- Dice: Rolling a 1 and a 2 are disjoint.
- Coin Flip: Getting a head and a tail are disjoint.

But rolling a 1 and rolling "an odd number" are not disjoint, since they can both occur simultaneously.

Addition Rule of Probability

If A_1 and A_2 represent two disjoint outcomes, then the probability that one of them occurs is

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2)$$

Back to dice, rolling a 1 and a 2 are disjoint, so:

$$P(\text{rolling a 1 or a 2}) = P(\text{rolling a 1}) + P(\text{rolling a 2}) = \frac{1}{6} + \frac{1}{6}$$

General Addition Rule of Probability

If A_1 and A_2 represent two outcomes (not necessarily disjoint), then the probability that one of them occurs is

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) - P(A_1 \text{ and } A_2)$$

Venn diagram:

General Addition Rule of Probability

Events are just combinations of outcomes. So for example let

- A_1 be the event that we draw a diamond
- A_2 be the event that we draw a face card

These two events are not disjoint, as there are 3 diamond face cards. Venn diagram:

General Addition Rule of Probability

$$
\begin{aligned}
P(A_1 \text{ or } A_2) &= P(\text{diamond or a face card}) \\
&= P(\text{diamond}) + P(\text{face card}) - \\
&\quad P(\text{diamond AND face card}) \\
&= \frac{13}{52} + \frac{3 \times 4}{52} - \frac{3}{52} = \frac{22}{52} = 42.3\%
\end{aligned}
$$

Previously... Probability

The general addition rule of probability: If A_1 and A_2 represent two outcomes (not necessarily disjoint), then the probability that one of them occurs is

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) - P(A_1 \text{ and } A_2)$$

If A_1 and A_2 are disjoint, then they cannot both occur, so

$$P(A_1 \text{ and } A_2) = 0$$

and hence

$$P(A_1 \text{ or } A_2) = P(A_1) + P(A_2)$$

Sample Space and the Complement of Events

Rolling a die has 6 possible outcomes. The sample space is the set of all possible outcomes $S = \{1, 2, \ldots, 6\}$.

Say event D is the event of rolling an even number i.e $D = \{2, 4, 6\}$. The complement of event D is $D^c = \{1, 3, 5\}$ i.e. getting an odd number. A and A^c are disjoint.

So then for any event A and its complement A^c

$$P(A) + P(A^c) = 1$$

Independence

Two processes are independent if knowing the outcome of one provides no useful information about the outcome of the other. Otherwise we say they are dependent.

Consider:

1. You roll a die once, and then you roll it again.
2. You get a movie recommendation from your friend Robin, but then their significant other Sam also recommends it.
3. You compare test scores from two Grade 9 students in the same class. Then same school. Then same school district. Then same city. Then same state.

Independence

We say that events A and B are independent if

$$P(A \text{ and } B) = P(A) \times P(B)$$

Ex: Dice rolls are independent, so say you roll twice:

$$\begin{aligned} P(\text{rolling a 1 and then a 6}) &= P(\text{rolling a 1}) \times P(\text{rolling a 6}) \\ &= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36} \end{aligned}$$

Example Demonstrating You Already Know Cond. Prob.

Let's suppose I take a random sample of 100 Reed students to study their smoking habits.

	Smoker	Not Smoker	Total
Male	19	41	60
Female	12	28	40
Total	31	69	100

- What is the probability of a randomly selected male smoking?
- What is the probability that a randomly selected smoker is female?

Conditional Probability

The conditional probability of an event A given the event B, is defined by

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

This is read as "the probability of A given B" or "the probability of A conditional on B."

Back to Example

	Smoker	Not Smoker	Total
Male	19	41	60
Female	12	28	40
Total	31	69	100

- What is the probability of a randomly selected male smoking?

$$P(S|M) = \frac{P(S \text{ and } M)}{P(M)} = \frac{19/100}{60/100} = \frac{19}{60}$$

- What is the probability that a randomly selected smoker is female?

$$P(F|S) = \frac{P(F \text{ and } S)}{P(S)} = \frac{12/100}{31/100} = \frac{12}{31}$$

Put It Together! Independence and Conditional Prob.

If A and B are independent events, then

$$P(A \text{ and } B) = P(A) \times P(B)$$

then

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A)$$

i.e. $P(A|B) = P(A)$: the event B occurring has no bearing on the probability of A

Independence Assumption

For almost all the statistical tools we will use in this class, we need the assumption of independence[1]. The issue is when we draw samples from a population.

Once we've sampled someone/something, we don't typically put them back into the pool of potential samples.
i.e. we cross them off the list.

This is the concept from probability theory of sampling with replacement vs sampling without replacement.

[1] proven in MATH 391/392

Independence Assumption

Say we are interested in the probability of sampling Wayne and Mario. For independence to hold we need

$$P(\text{ Wayne \& Mario }) = P(\text{ Wayne }) \times P(\text{ Mario })$$

Compare $N = 4$ & $N = 10000$ and assume we pick Mario first. By conditional probability:

$$\begin{aligned}
P(\text{ Wayne \& Mario }) &= P(\text{ Mario }) \times P(\text{ Wayne } \mid \text{ Mario }) \\
P(\text{ Wayne \& Mario }) &= \frac{1}{4} \times \frac{1}{3} \neq \frac{1}{4} \times \frac{1}{4} \\
P(\text{ Wayne \& Mario }) &= \frac{1}{10000} \times \frac{1}{9999} \approx \frac{1}{10000} \times \frac{1}{10000}
\end{aligned}$$

Independence Assumption

Moral of the story: when sampling without replacement, as the size of the sample n grows to be a larger and larger proportion of the total study population N, the independence assumption breaks down.

Gambler's Fallacy: Roulette

You can bet on individual numbers, sets of numbers, or red vs black. Let's assume no 0 or 00, so that $P(\text{red}) = P(\text{black}) = \frac{1}{2}$.

Gambler's Fallacy: Roulette

One of the biggest cons in casinos: spin history boards.

Let's ignore the numbers and just focus on what color occurred.
Note: the white values on the left are black spins.

Gambler's Fallacy: Roulette

Let's say you look at the board and see that the last 4 spins were red.

You will always hear people say "Black is due!"

Ex. on the 5th spin people think:

$$\begin{aligned} P(\text{black}_5 \quad &| \quad \text{red}_1 \text{ and } \text{red}_2 \text{ and } \text{red}_3 \text{ and } \text{red}_4) > \\ P(\text{red}_5 \quad &| \quad \text{red}_1 \text{ and } \text{red}_2 \text{ and } \text{red}_3 \text{ and } \text{red}_4) \end{aligned}$$

Gambler's Fallacy: Roulette

But assuming the wheel is not rigged, spins are independent i.e. $P(A|B) = P(A)$. So:

$$\begin{aligned} P(\text{black}_5|\text{red}_1 \text{ and } \text{red}_2 \text{ and } \text{red}_3 \text{ and } \text{red}_4) = P(\text{black}_5) &= \frac{1}{2} \\ P(\text{red}_5|\text{red}_1 \text{ and } \text{red}_2 \text{ and } \text{red}_3 \text{ and } \text{red}_4) = P(\text{red}_5) &= \frac{1}{2} \end{aligned}$$

Food for Thought: How People Perceive Randomness

Compare the following streaks on a non-rigged wheel:

$$RRRRRRR \text{ vs. } RBRRBBB$$

Both have the exact same probability $\left(\frac{1}{2}\right)^7$ of occurring, but people tend to somehow give different credence to the randomness of the pattern on the left vs right.

Next Week's Lab

Basketball players who make several baskets in succession are described as having a "hot hand." This refutes the assumption that each shot is independent of the next.

We are going to investigate this claim with data from a particular basketball player: Kobe Bryant of the Los Angeles Lakers in the 2009 NBA finals.

Next Time

Discuss the Normal Distribution