Basics of Probability

Fractions, Proportions, and Percents

When we talk about "probability" or "chance" we can represent those as fractions, proportions, or percents.

Fractions (e.g., 1/4) are the easiest to work with when doing math with a pencil.

Proportions (e.g., 0.25) are the easiest to work with when doing math on a calculator.

Percents (e.g., 25%) are the easiest to talk and write about.

The Indicator Trick

Suppose a list of items (maybe numbers, maybe something else). You want to compute the proportion of the list that has some desired property.

Example: {A, B, C, C, B, C}
What proportion is this list is a C?

Example: {1, 2, 3, 2, 1, 4, 3, 5}
What proportion is this list is an even number?

Indicator Trick: Flag or "indicate" all the items with the desired feature with a 1, and all the others with a 0. Then take the average. This gives you the proportion that have the desired feature.

Shorthand: **Indicate, then average.**

One Note of Caution

Make a special note, though, that "%" means "per 100" or "/100."
Because of this, **you cannot mindlessly multiply percents**.

Notice that 10% x 20% does not equal 200%.

Instead, it equals $10\% \times 20\% = 10/100 \times 20/100 = 200/10{,}000 = 2/100 = 2\%$

While you can safely add percents, here's a good rule:

Do calculations with fractions or proportions;
only convert to a percent once you have your final answer.

Chance Processes

Define a **chance process** as a process that one can repeat (independently and under the same conditions) to produce a result from defined set of possible outcomes.

We could imagine rolling a die or tossing a coin.

When we draw **with replacement**, we draw multiple times, but replace the draws before continuing. This means that same cards can be drawn multiple times.

When we draw **without replacement**, we draw multiple times, but <u>do not</u> replace the draws before continuing. This means that same card cannot be drawn again.

Numbered-Ticket Model

The numbered-ticket model: Fill a box with *k* tickets numbered $t_1, t_2, t_3, \ldots, t_k$. Draw *N* times with replacement from the box. Record the average of the draws.

1. You choose:
   A. *k*: the number of tickets in the box.
   B. $t_1, t_2, t_3, \ldots, t_k$: the numbers written on the tickets.
   C. *N*: the number of times to draw from the box.
2. You always:
   D. Draw with replacement.
   E. Average the draws.

Numbered-Ticket Model

We analyze chance processes using **frequentist theory**. Under frequentist theory, we're interested in the probability that a chance process produces some event. For compactness, we sometimes write "the probability that A happens" as Pr(A).

Define the **probability** of an event as the proportion of times the event occurs in the long-run.

Computing Probabilities

Counting the Equally-Likely Outcomes: list all the equally-likely outcomes of a chance process, then compute the proportion of outcomes that fall into the category of interest.

The Probability of the Opposite: If you know that Pr(A doesn't happen) is 0.4, then the Pr(A happens) is 0.6. Subtract the probability of an event from 1 to obtain the probability of its opposite.

Conditional Probability: For some chances processes, probabilities can change depending on what's happened before. If you suspect that the probabilities might have changed, you need to <u>recount the equally-likely outcomes</u>. *This is especially true when sampling without replacement.* We write the "probability of A *given that B happens*" and Pr(A | B).

Multiplication Rule

Suppose I'm interested in the probability that two things both happen.

I draw a card that is red and a king.

I toss a head and then another head.

I draw a red marble and then another red marble.

Notice the word "and"–this means there are two things and I'm interested in **both** happening.

When you want to compute the probability that two things both happen, you multiply the probability of the first times the probability of the second, given the first.

More compactly, $Pr(A \underline{\text{and}} B) = Pr(A) \times Pr(B \mid A)$.

Independence

We say that two events are **independent** if the probability of the second is the same, regardless of how the first turns out.

Else, the two events are **dependent**.

Importantly for our purposes, we can say the following:

When drawing <u>without</u> replacement, the draws are dependent.

When drawing <u>with</u> replacement, the draws are independent.