Probability (Multiplication Rule)

Ryan Miller

Estimating Probabilities

- As we've previously discussed, probabilities are defined as *long-run proportions*
  - The probability of a coin landing heads is 0.5 because it's what we'd expect after a large number of tosses

Estimating Probabilities

- As we've previously discussed, probabilities are defined as *long-run proportions*
  - The probability of a coin landing heads is 0.5 because it's what we'd expect after a large number of tosses
- So it makes sense to use proportions from a finite sample to *estimate* probabilities
  - For example, the probability of Stephen Curry making a free throw is 0.90
  - Obviously Steph hasn't shot infinitely many free throws, but he's taken enough for us to get a good estimate (also known as an *empirical probability*)

Conditional Probability

- Most our examples involving categorical variables centered around contingency tables and conditional proportions
  - Analogous to *conditional proportions* (row and column percentages), is the idea of a **conditional probability**

Conditional Probability

- Most our examples involving categorical variables centered around contingency tables and conditional proportions
  - Analogous to *conditional proportions* (row and column percentages), is the idea of a **conditional probability**
- Conditional probability is used in scenarios involving *dependent events*
  - For example, the probability of getting the flu (event A) might depend on if you've received a flu shot (event B)
- We denote conditional probability with the following notation: $P(A|B)$
  - Conditional probabilities can be estimated from a contingency table

Example

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX?

Example

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete?

Example

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete? $46/52 = 0.885$
3) An athlete has the XX genotype *and* is an endurance athlete?

Example

ACTN3 is known as the fast twitch gene, everyone has one of three genotypes (XX, RR, or RX). The table below summarizes a sample of 301 elite athletes:

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

From this table, let's estimate a few different probabilities:

1) An endurance athlete has genotype XX? $46/194 = 0.237$
2) An athlete with the XX genotype is an endurance athlete? $46/52 = 0.885$
3) An athlete has the XX genotype *and* is an endurance athlete? $46/301 = 0.153$

The Multiplication Rule

The relationship between these probabilities motivates the **multiplication rule**, which states:

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

In our previous example, notice:

1) $P(\text{XX}|\text{End}) = 46/194 = 0.237$
2) $P(\text{End}) = 194/301 = 0.645$
3) $P(\text{XX and End}) = 46/301 = 0.153$

	RR	RX	XX	Total
Sprint/power	53	48	6	107
Endurance	60	88	46	194
Total	113	136	52	301

It's easy verify the multiplication rule: $46/194 = \frac{46/301}{194/301}$

The Multiplication Rule

- The multiplication rule can be used to calculate the probability of *intersections* (and statements)
  - Some simple algebra shows:

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

- Notice it's also true that:

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

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$
  - Notice that independence does not mean the events are *disjoint*
  - $P(A \text{ and } B) = 0$ for disjoint events

Independence

- Two events are considered *independent* when $P(A \text{ and } B) = P(A) * P(B)$
  - Notice that independence does not mean the events are *disjoint*
  - $P(A \text{ and } B) = 0$ for disjoint events
- Independence tends to greatly simplify probability calculations
  - Consider 3 coin flips:

$$P(H_1 \text{ and } H_2 \text{ and } H_3) = P(H_1) * P(H_2) * P(H_3) = (1/2)^3 = 1/8$$

- This is a much easier calculation to think about compared to:

$$P(H_1 \text{ and } H_2 \text{ and } H_3) = P(H_3) * P(H_2|H_1) * P(H_3|H_1 \text{ and } H_2)$$

Conclusion

We've now covered 3 probability rules:

1) The addition rule, $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$, allows us to calculate the probability of *unions* of events
2) The multiplication rule, $P(A \text{ and } B) = P(A|B) * P(B)$, allows us to calculate the probability of *intersections* of events
3) The complement rule, $P(A) + P(A^C) = 1$, allows simpler calculations for large sample spaces

The final video on probability will go through a few examples illustrating how to determine which rules to apply in different situations