1 Probability Notation

Suppose we have 3 random variables A, B, and C. Consider the expression

$$P(+b, C) = \sum_{a \in \{a_1, a_2, a_3\}} P(a, +b, C)$$

In this course, we denote **discrete random variables** by capital letters and use them to represent all possible disjoint outcomes. In the above example, A, B, and C are random variables.

We use lower case letters to denote **outcomes**, i.e. possible values our variables can take on, such as $+b$ for the variable B, or a_1, a_2, and a_3 for the variable A in the above example.

We also have **variables for values** like a. Note that these variables are also represented by lower case letters and only represent a single outcome (as opposed to random variables).

2 Basic Rules

Definition of Conditional Probability:

$$P(X \mid Y) = \frac{P(X, Y)}{P(Y)}$$

Product Rule:

$$\begin{aligned} P(X, Y) &= P(X \mid Y)P(Y) \\ &= P(Y \mid X)P(X) \\ P(X_1, X_2, X_3) &= P(X_1, X_2 \mid X_3)P(X_3) \\ &= P(X_1 \mid X_2, X_3)P(X_2, X_3) \end{aligned}$$

Bayes' Theorem:

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)}$$

Normalization:

$$P(Y \mid X) = \frac{P(X, Y)}{P(X)} = \frac{P(X, Y)}{\sum_y P(X, y)}$$

$$P(Y \mid X) \propto P(X, Y)$$

$$P(Y \mid X) = \alpha P(X, Y) \qquad \textit{Note this difference between } \propto \textit{ and } \alpha$$

$$\alpha = \frac{1}{P(X)} = \frac{1}{\sum_y P(X, y)}$$

Chain Rule:

$$\begin{aligned} P(X_1, X_2, X_3) &= P(X_1 \mid X_2, X_3)P(X_2, X_3) \\ &= P(X_1 \mid X_2, X_3)P(X_2 \mid X_3)P(X_3) \end{aligned}$$

$$P(X_1, ..., X_N) = \prod_{n=1}^{N} P(X_n \mid X_1, ..., X_{n-1})$$

All of these basic probability rules hold when conditioning on a set of random variables or outcomes. To make this work, the conditioned variables need to be included in each term in the rule. For example, take Bayes' Theorem from above, but now conditioned upon variables A and B:

$$P(Y \mid X, A, B) = \frac{P(X \mid Y, A, B)P(Y \mid A, B)}{P(X \mid A, B)}$$

3 Marginalization

Marginalization uses the law of total probability to "sum out" variables from a joint distribution. This is useful when we are given the joint probability distribution and want to find the probability distribution over just a subset of the variables. Marginalization has the following forms:

To sum out a single variable:

$$P(X) = \sum_y P(X, y)$$

To sum out multiple variables:

$$P(X) = \sum_z \sum_y P(X, y, z)$$

This also works for conditional distributions when summing out a variable that is not conditioned upon, i.e. a variable to the left of the $|$:

$$P(A \mid C, d) = \sum_b P(A, b \mid C, d)$$

This does NOT work when summing over a variable that is conditioned upon, i.e. a variable to the right of the $|$:

$$P(A, b \mid C) \neq \sum_d P(A, b \mid C, d)$$

4 Independence

If two variables X and Y are **independent** ($X \perp\!\!\!\perp Y$), by definition the following are true:

- $P(X, Y) = P(X)P(Y)$
- $P(X) = P(X \mid Y)$
- $P(Y) = P(Y \mid X)$

If two variables X and Y are **conditionally independent given** Z ($X \perp\!\!\!\perp Y \mid Z$), by definition the following are true:

- $P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)$
- $P(X \mid Y, Z) = P(X \mid Z)$
- $P(Y \mid X, Z) = P(Y \mid Z)$

5 Probability Tables

When representing probabilities with capital letters, e.g. $P(A, B)$, we are referring to all the combinations of outcomes that the discrete random variables can have. Thus, we have a table of probabilities rather than a single value. This is also true for conditional probabilities, e.g. $P(A, B \mid C)$. When there is a mixture of capital letters and lower case letters, e.g. $P(A, b \mid C, d)$, the table contains all the combinations of outcomes for the random variables, A and C (while the discrete values b and d are fixed).

6 Important Note About Conditional Probability Tables

It is important to understand when a probability table contains the complete distribution, or in other words, when a probability table sums to one.

A probability table will sum to one when:

1. there is exactly one specific combination of outcomes that is conditioned upon and
2. we are considering all possible combinations of the other random variables.

Another way to phrase this: a probability table will sum to one, when:

1. there are no capital letters on the right-hand side of the $\mid$, and
2. there are only capital letters on the left-hand side.