Marginalization

An important insight regarding probabilistic models with many random variables is that "the joint distribution is complete information." From the joint distribution you can compute all probability questions involving those random variables in the model. This chapter is an example of that insight.

The central question of this chapter is: Given a joint distribution, how can you compute the probability of random variables on their own?

Marginalization From Two Random Variables

To start, consider two random variables X and Y. If you are given the joint how can you compute $P(X = x)$? Recall that if you have the joint you have a way to know the probability $P(X = x, Y = y)$ for any value x and y. We already have a technique for computing $P(X = x)$ from the joint. We can use the [Law of Total Probability](#) (LOTP)! In this case the events $Y = y$ make up the "background events":

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

Note that to apply the LOTP it must be the case that the different events $Y = i$ must be mutually exclusvie and it must be the case that $\sum_y P(Y = y) = 1$. Both are true.

If we wanted $P(Y = y)$ we could again use the Law of Total Probability, this time with X taking on each of its possible values as the background events:

$$P(Y = y) = \sum_x P(X = x, Y = y)$$

Example: Favorite Number

Consider the following joint distribution for X and Y where X is a person's favorite binary digit and Y is their year at Stanford. Here is a real joint distribution form a past class:

Variable	Symbol	Type
Favorite Digit	X	Discrete number $\{0, 1\}$
Year in School	Y	Categorical $\{$Frosh, Soph, Junior, Senior, 5+$\}$

	$X = 0$	$X = 1$
$Y = \text{Frosh}$	0.01	0.13
$Y = \text{Soph}$	0.05	0.33
$Y = \text{Junior}$	0.04	0.21
$Y = \text{Senior}$	0.03	0.12
$Y = 5+$	0.02	0.06

What is the probability that a student's favorite digit is 0, $P(X = 0)$? We can use the LOTP to compute this probability:

$$P(X = 0) = \sum_y P(X = 0, Y = y)$$

$$
\begin{aligned}
&= + P(X = 0, Y = \text{Frosh}) \\
&\quad + P(X = 0, Y = \text{Soph}) \\
&\quad + P(X = 0, Y = \text{Junior}) \\
&\quad + P(X = 0, Y = \text{Senior}) \\
&\quad + P(X = 0, Y = 5+) \\
&= 0.01 + 0.05 + 0.04 + 0.03 + 0.02 \\
&= 0.15
\end{aligned}
$$

Marginalization with More Variables

The idea of marginalization can be extended to joint distributions with more than two random variables. Consider having three random variables X, Y, and Z, we could marginalize out any of the variables:

$$P(X = x) = \sum_{y,z} P(X = x, Y = y, Z = z)$$

$$P(Y = y) = \sum_{x,z} P(X = x, Y = y, Z = z)$$

$$P(Z = z) = \sum_{x,y} P(X = x, Y = y, Z = z)$$

Notation: **Double Sum**

In this case the double sum notation:

$$\sum_{y,z}$$

Also written equivalently as:

$$\sum_y \sum_z$$

means that we are summing over all possible values of y and z. For example, if Y is a random variable with 3 possible values and Z is a random variable with 4 possible values, then $\sum_{y,z}$ means that we are summing over all 12 possible combinations of y and z.

Here is an example in code. Assume we have a function `joint(x, y, z)` and that all X, Y and Z can take on values in the set $\{0, 1, 2, 3, 4\}$