Joint Probability

Many interesting problems involve not one random variable, but rather several interacting with one another. In order to create interesting probabilistic *models* and to reason in real world situations, we are going to need to learn how to consider several random variables jointly.

In this section we are going to use disease prediction as a working example to introduce you to the concepts involved in probabilistic models. The general question is: a person has a set of observed symptoms. Given the symptoms what is the probability over each possible disease?

We have already considered events that co-occur and covered concepts such as independence and conditional probability. What is new about this section is (1) we are going to cover how to handle *random variables* which co-occur and (2) we are going to talk about how computers can reason under large probabilistic models.

Joint Probability Functions

For single random variables, the most important information was the PMF or, if the variable was continuous, the PDF. When dealing with two or more variables, the equivalent function is called the Joint function. For discrete random variables, it is a function which takes in a value for each variable and returns the probability (or probability density for continuous variables) that each variable takes on its value. For example if you had two *discrete* variables the Joint function is:

$$P(X = x, Y = y) \qquad \text{Joint function for } X \text{ and } Y$$

You should read the comma as an "and" and as such this is saying the probability that $X = x$ and $Y = y$. Again like for single variables, as shorthand, we often write just the values and it implies that we are talking about the probability of the random variables taking on those values. This notation is convenient because it is shorter, and it makes it explicit that the function is operating over two parameters. It requires to recall that the event is a *random variable* taking on the given value.

$$P(x, y) \qquad \text{Shorthand for } P(X = x, Y = y)$$

If any of the variables are continuous we use different notation to make it clear that we need a probability density function, something we can integrate over to get a probability. We will cover this in detail:

$$f(X = x, Y = y) \qquad \text{Joint density function if } X \text{ or } Y \text{ are continuous}$$

The same idea extends to as many variables as you have in your model. For example if you had three discrete random variables X, Y, and Z, the joint probability function would state the likelihood of an assignment to all three: $P(X = x, Y = y, Z = z)$.

Joint Probability Tables

> **Definition**: Joint Probability Table
> A joint probability table is a way of specifying the "joint" distribution between multiple random variables. It does so by keeping a multi-dimensional lookup table (one dimension per variable) so that the probability mass of any assignment, eg $P(X = x, Y = y, \ldots)$, can be directly looked up.

Let us start with an example. In 2020 the Covid-19 pandemic disrupted lives around the world. Many people were unable to get tested and had to determine whether or not they were sick based on home diagnosis. Let's build a very simple probabilistic model to enable us to make a tool which can predict the probability of having the illness given observed symptoms. To make it clear that this is a pedagogical example, let's consider a made up illness called Determinitis. The two main symptoms are fever and loss of smell.

Variable	Symbol	Type
Has Determinitis	D	Bernoulli (1 indicates has Determinitis)
Fever	F	Categorical (none, low, high)
Can Smell	S	Bernoulli (1 indicates can smell)

A joint probability table is a brute force way to store the probability mass of a particular assignment of values to our variables. Here is a probabilistic model for our three random variables (aside: the values in this joint are realistic and based on reasearch, but are primarily for teaching. Consult a doctor before making medical decisions).

$$D = 0$$

	$S = 0$	$S = 1$
$F = \text{none}$	0.024	0.783
$F = \text{low}$	0.003	0.092
$F = \text{high}$	0.001	0.046

$$D = 1$$

	$S = 0$	$S = 1$
$F = \text{none}$	0.006	0.014
$F = \text{low}$	0.005	0.011
$F = \text{high}$	0.004	0.011

A few key observations:

- Each cell in this table represents the probability of one assignment of variables. For example the probability that someone can't smell, $S = 0$, has a low fever, $F = \text{low}$, and has the illness, $D = 1$, can be directly read off the table: $P(D = 1, S = 0, F = \text{low}) = 0.005$.
- These are joint probabilities *not* conditional probabilities. The value 0.005 is the value of illness, no smell and low fever. It is not the probability of no smell and low fever given illness. A table which stored conditional probabilities would be called a conditional probability table, this is a joint probability table.
- If you sum over all cells, the total will be 1. Each cell is a mutually exclusive combination of events and the cells are meant to span the entire space of possible outcomes.
- This table is large! We can count the number of cells using the step rule of counting. If n_i is the number of different values that random variable i can take on, the number of cells in the joint table is $\prod_i n_i$.

Properties of Joint Distributions

There are many properties of a random variable of any joint distribution some of which we will dive into extensively. Here is a brief summary. Each random variable has:

Property	Notation Example	Description
Distribution Function (PMF or PDF)	$P(X = x, Y = y, \ldots)$ or $f(X = x, Y = y, \ldots)$	A function which maps values the RV can take on to likelihood.
Cumulative Distribution Function (CDF)	$F(X < x, Y < y, \ldots)$	Probability that each variable is less than its corresponding parameter
Covariance	$\sigma_{X,Y}$	A measure of how much two random variables vary together.
Correlation	$\rho_{X,Y}$	Normalized co-variance