CHEM 116 January 23, 2020

Unit 2, Lecture 3

Numerical Methods and Statistics

1 Equations with Random Variables

Companion Reading

Bulmer Chapter 3

1.1 Conditional Probability Distribution

A conditional probability distribution allows us to fix one sample, event, or rv in order to calculate another. It is written as $\mathrm{P}(X = x|Y = y)$. For example, what is the probability of having the flu given I'm showing cold/flu symptomps. Conditionals are generally much easier specify, understand and measure than joints or marginals.

- The probability I visited node C given that I started in node A and ended in node B
- The probability a test shows I have influenza given that I do not have influenza (false negative)
- The probability that the sum of two dice is 7 given that one die shows 4 when rolling two dice.

The definition of a conditional probability distribution is

$$\mathrm{P}(x|y) = \frac{\mathrm{P}(x,y)}{\mathrm{P}(y)} \tag{1}$$

A CPD is a full-fledged PMF, so $\sum_{\mathcal{X}} \mathrm{P}(x|y) = 1$ due to normalization, sometimes called the law of total probability. If we forget what goes in the denominator on the right-hand-side we can quickly see that $\sum_{\mathcal{X}} \mathrm{P}(x,y)/\mathrm{P}(y) = 1$ whereas $\sum_{\mathcal{X}} \mathrm{P}(x,y)/\mathrm{P}(x) \neq 1$.

The definition is the same for continuous distributions.

This leads to an alternative way to marginalize:

$$\sum_{\mathcal{Y}} \mathrm{P}(x|y)\mathrm{P}(y) = \sum_{\mathcal{Y}} \mathrm{P}(x,y) = \mathrm{P}(x)$$

1.2 Viewing Conditionals as Sample Space Reduction

Consider guessing binary numbers at random between 0 and 7:

000
001
010
011
100
101
110
111

The probability of sampling 4 (100),

$$P(x = 100) = \frac{1}{8}$$

Now, consider the rv Y, the number of non-zero bits. What is

$$P(x = 100|Y = 1)$$

We could rewrite this in terms of the joint and marginal, as

$$P(x = 100|Y = 1) = \frac{(x = 100, Y = 1)}{Y = 1} = \frac{1/8}{3/8} = \frac{1}{3}$$

Or we could recognize that the condition of $Y = 1$ reduces the sample space to 3, because there are only 3 samples that are consistent with $Y = 1$. Thus, the probability of $x = 100$ is $1/3$, since $x = 100$ has a single permutation and Q_c, the conditional sample space is 3.

2 Tricky Concepts

Product Spaces A product space is for joining two possibly dependent sample spaces. It can also be used to join sequential trials.

Event vs Sample on Product Spaces Things which were samples on the components of a product space are now events due to permutations

Random Variables They assign a numerical value at each sample in a sample space, but we typically care about the probability of those numerical values (PMF). So X goes from sample to number and $P(x)$ goes from number to probability.

Continuous PDF A pdf is a tool for computing things, not something meaningful by itself.

Marginal Probability Distribution A marginal 'integrates/sums' out other samples/random variables/events we are not interested in.

Joint vs Conditional People almost never think in terms of joints. Conditionals are usually easier to think about, specify, and be a way to attack problems.

3 Working with Marginals, Condtionals, and Joints

3.0.1 Seasons Example

The sample space is a product space of the seasons T (Winter (0), Spring (1), Summer (2), Fall (3)) and W if the weather is nice or not (N=nice, S=not nice). We know that

$$\begin{array}{ll} P(W = N|T = 0) = 0 & P(W = N|T = 1) = 0.4 \\ P(W = N|T = 2) = 0.8 & P(W = N|T = 3) = 0.7 \end{array}$$

$$P(T = t) = \frac{1}{4}$$

What is the probability that the weather is not nice? Use marginalization of conditional:

$$P(W = S) = \sum_{\mathcal{T}} P(W = S|T)P(T) = 0.25(1 - 0) + 0.25(1 - 0.4) + 0.25(1 - 0.8) + 0.25(1 - 0.7)$$

$$\mathrm{P}(W=S)=0.525$$

What is the probability that it is fall given that it is nice? Start with definition of conditional :

$$\mathrm{P}(T=3|W=N)=\frac{\mathrm{P}(T=3,W=N)}{\mathrm{P}(W=N)}$$

We know that $\mathrm{P}(T=3,W=N)=\mathrm{P}(W=N|T=3)\mathrm{P}(T=3)$ due to definition of conditional:

$$\mathrm{P}(T=3|W=N)=\frac{\mathrm{P}(W=N|T=3)\mathrm{P}(T=3)}{\mathrm{P}(W=N)}$$

Finally, we can use $\mathrm{P}(W=N)=1-\mathrm{P}(W=S)$, the **NOT** rule:

$$\mathrm{P}(T=3|W=N)=\frac{0.7\times 0.25}{1-0.525}=0.368$$

3.1 Bayes' Theorem

We derived a well-known equation in that example called Bayes' Theorem:

$$\mathrm{P}(A|B)=\frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} \tag{2}$$

This is useful to swap the order of conditionals

3.2 Independence

Finally, we are ready to define independence. If the random variables X and Y are independent, then

$$\mathrm{P}(x|y)=\mathrm{P}(x) \tag{3}$$

This implies via Bayes' Theorem

$$\mathrm{P}(y|x)=\mathrm{P}(y) \tag{4}$$

And also implies via CPF definition

$$\mathrm{P}(x,y)=\mathrm{P}(x)\mathrm{P}(y) \tag{5}$$

which was our **AND** rule from last unit.

In our weather example, is the season and weather independent?

$$\mathrm{P}(W=N|T=0)\neq P(W=N|T=1)$$

so no.

3.3 Compound Conditionals

When writing conditionals, this is a common short-hand:

$$\mathrm{P}(x=2\,|\,Y=A,Z=0)$$

for

$$\mathrm{P}([x=2]\,|\,[Y=A,Z=0])$$

The conditional is always evaluated last, for example:

$$\mathrm{P}(x=2,Y=A\,|\,Z=0)$$

is also possible and means the probability of the joint $(x=2,Y=A)$ given that $Z=0$

3.4 Conditional Independence

X_0 and X_1 are conditionally independent given Z if

$$\mathrm{P}(X_0|Z, X_1) = \mathrm{P}(X_0|Z) \tag{6}$$

which is equivalent to

$$\mathrm{P}(X_0, X_1|Z) = \mathrm{P}(X_0|Z)\mathrm{P}(X_1|Z) \tag{7}$$

To use conditional independence, you must condition on Z. For example, if I want to calculate $\mathrm{P}(X_0, X_1)$, I'll need to condition it. Marginalizing a conditional is one way to get that quantity, using a compound conditional:

$$\mathrm{P}(X_0, X_1) = \sum_{\mathcal{Z}} \mathrm{P}(X_0, X_1|Z)\mathrm{P}(Z)$$

Now it is in a form where the conditional independence applies:

$$\mathrm{P}(X_0, X_1) = \sum_{\mathcal{Z}} \mathrm{P}(X_0|Z)\mathrm{P}(X_1|Z)$$

This is common for sequential trials, where the trials are independent when conditioned on some underlying property, but dependent if we do not know the property. For example, let's say I have two dice, one that is fair and one that follows the biased die model we saw in class. Let's further assume that $P(D = 0) = 0.1$, where D indicates the chosen die.

$$\mathrm{P}(X = x|D = 0) = \frac{1}{6}$$
$$\mathrm{P}(X = x|D = 1) = \frac{x}{21}$$

If I know which die I'm rolling, then every roll is independent as expected:

$$\mathrm{P}(X_0 = 6, X_1 = 1|D) = \mathrm{P}(X_0 = 6|D)\mathrm{P}(X_1 = 1|D)$$

however, consider

$$\mathrm{P}(X_1 = 1|X_0 = 6) = \frac{\mathrm{P}(X_1 = 1, X_0 = 6)}{\mathrm{P}(X_0 = 6)}$$

As above, we'll try to condition it to exploit conditional independence.

$$\mathrm{P}(X_1 = 1, X_0 = 6) = \mathrm{P}(X_1 = 1|D = 0)\mathrm{P}(X_0 = 6|D = 0)\mathrm{P}(D = 0)+$$
$$\mathrm{P}(X_1 = 1|D = 1)\mathrm{P}(X_0 = 6|D = 1)\mathrm{P}(D = 1)$$
$$\mathrm{P}(X_1 = 1, X_0 = 6) = \frac{1}{6}\frac{1}{6}\frac{1}{10} + \frac{1}{21}\frac{6}{21}\frac{9}{10} = 0.0150$$
$$\mathrm{P}(X_0 = 6) = \sum_{X_1} \mathrm{P}(X_1, X_0 = 6) = \frac{1}{6}\frac{1}{10}\underbrace{\sum_x \frac{1}{6}}_{=1} + \frac{6}{21}\frac{9}{10}\underbrace{\sum_x \frac{x}{21}}_{=1}$$
$$\mathrm{P}(X_0 = 6) = 0.274$$
$$\mathrm{P}(X_1 = 1|X_0 = 6) = \frac{0.0150}{0.274} = 0.0549$$

Finally, to show they are not independent:

$$\mathrm{P}(X_1 = 1) = \sum_{X_1} \mathrm{P}(X_1 = 1, X_0) = \frac{1}{6}\frac{1}{10}\underbrace{\sum_x \frac{1}{6}}_{=1} + \frac{1}{21}\frac{9}{10}\underbrace{\sum_x \frac{x}{21}}_{=1} = 0.0595$$

The intuition here is that the marginal of $X_1 = 1$ considers both die according to the die marginals. However, knowing that $X_0 = 6$ shifts it so that the biased die is more likely. This can be seen with Bayes theorem to:

$$\mathrm{P}(D = 1|X_0 = 6) \neq \mathrm{P}(D = 1)$$

and thus $\mathrm{P}(X_1|X_0 = 6)$ changes from $\mathrm{P}(X_1)$.

3.5 Table of Equations

$P(x) = \sum_y P(x, y)$	Definition of Marginal
$P(x\|y) = \frac{P(x,y)}{P(y)}$	Definition of Conditional
$P(x, y) = P(x\|y)P(y)$	Definition of Conditional
$\sum_x P(x\|y) = 1$	Law of Total Probability/Normalization
$\sum_y P(x\|y)P(y) = P(x)$	Marginilzation of Conditional
$P(x\|y) = \frac{P(y\|x)P(x)}{P(y)}$	Bayes' Theorem
$P(x, y) = P(x)P(y)$	Definition of Independence
$P(x\|y) = P(x)$	Independence Property (x,y independent)
$P(x, y\|z) = P(x\|z)P(y\|z)$	Conditional Independence