CHEM 116 January 27, 2019

Unit 2, Lecture 2

Numerical Methods and Statistics

1 Equations with Random Variables

Companion Reading

Bulmer Chapter 3

1.1 Joint Probability Distribution

We will indicate the probability that two rvs, X and Y, adopt a particular value x and y *simultaneously* as $\mathrm{P}(x, y)$. This is called a joint probability distribution. Joints indicate simultaneous occurance, unlike our **AND** from last lecture.

To treat successive observations, we simply rearrange our current definitions. Take flipping a coin. We redefine our sample space to be the product space of trials flip 1 and flip 2, so (H, H), (H, T), (T, H) and (T, T) where H =heads, T =tails. Next, we say X is the observation of trial 1 and Y is the observation of trial 2.

The continuous analogue is $p(x, y)$, which is not meaningful unless integrated over an area. Example: observing particle in fixed area.

1.2 Marginal Probability Distribution

The marginal probability distribution function is defined as

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

The marginal means the probability of $X = x, Y = y_0$ **OR** $X = x, Y = y_1$ **OR** $X = x, Y = y_2$, etc. For example, if X is the weather and Y is the day of the week, it is the probability of the weather being a particular value (e.g., nice) 'averaged' over all possible weekdays.

The marginal allows us to remove a rv or sample space dimension. That process is called **marginalizing** and the resulting $\mathrm{P}(x)$ is called the marginal. Marginalization is especially important if the two pieces of the joint are not independent.

1.3 Revisiting Independence

Recall that we defined independence in terms of multiple samples in Unit 1. Now, we'll discuss independence of *random variables*. Conceptually, independence of two random variables means that if I know the value of one variable, I have no additional evidence for what the value of another random variable is. For example, if I roll two dice I can have a random variable indicating the roll of die 1, R_1, and die 2, R_2. If I know that $R_1 = 4$, that doesn't help me determine the value of R_2 therefore these two random variables are independent. Now consider R_1 and the sum of the two dice S. If I know that $R_1 = 4$, I know that the sum of the two dice has to be greater than 4. That's because you cannot have a sum of two dice be less than 5 if one of the rolls is 4. Therefore, R_1 and S are *not independent*. One way to assess independence is the following theorem which hold for 2 indenpendent random variables:

$$P(X = x, Y = y) = P(X = x)P(Y = y), \text{ If } X \text{ and } Y \text{ are independent.} \tag{2}$$

We can prove independence if that expression holds for all values of X and Y. We can disprove independence by showing it doesn't hold for a single value.

2 Concept Review

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.

Independence of Random Variables Random variables can be independent and can be numerically evaluated by seeing if the produce of their marginals are equal to their joints.

2.1 Table of Equations

$P(x) = \sum_y P(x,y)$	Definition of Marginal
$\sum_x P(x) = 1$	Law of Total Probability/Normalization
$P(x,y) = P(x)P(y)$	Definition of Independence