Week 3 Cheat Sheet

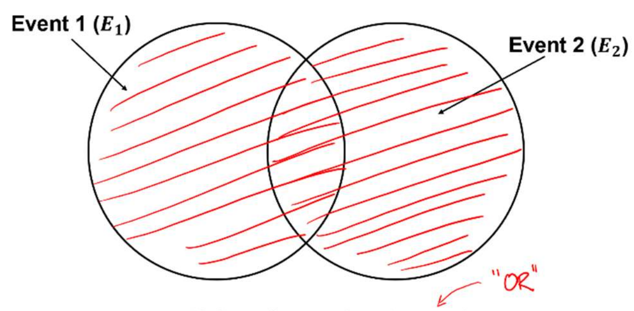
Statistics and Data Analysis with Excel, Part 1

Charlie Nuttelman

Here, I provide the mathematical equations and some of the important Excel functions required to perform various calculations in Week 3 of the course. The headings represent the screencasts in which you will find those calculations and concepts. Not all screencasts are referenced below – just the ones that have complex mathematical formulas or Excel formulas that are tricky to use.

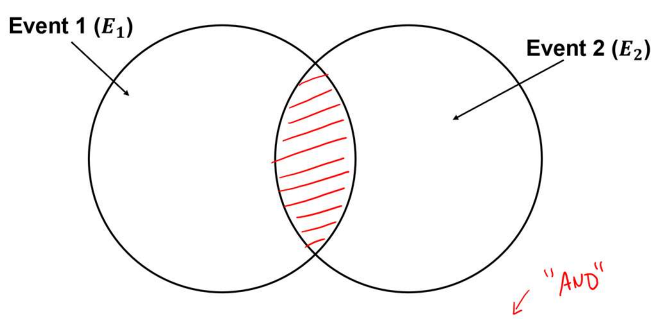
<u>Sample Spaces and Events</u>

Union of two events:



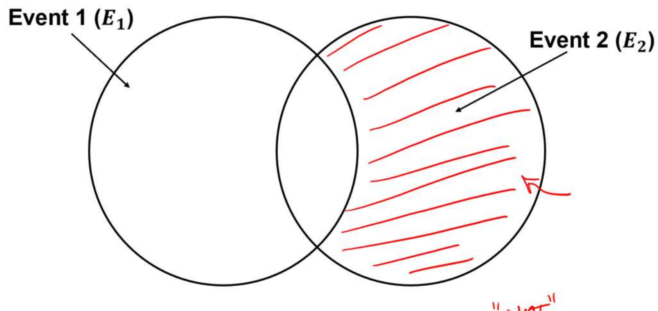
<u>Union</u> of E_1 and E_2 ($E_1 \cup E_2$)

Intersection of two events:



<u>Intersection</u> of E_1 and E_2 ($E_1 \cap E_2$)

Complement of an event:



<u>Complement</u> of E_1 (E_1')

Mutually exclusive events: Two events are mutually exclusive if their intersection is the null set (zero) ($E_1 \cap E_2 = \emptyset$).

<u>Permutations and Combinations</u>

Permutation: The number of ways we can put r items into n spots if order matters.

Denoted by: ${}_nP_r$

Calculation: ${}_nP_r = \frac{n!}{(n-r)!}$ (! denotes the factorial)

Easy to calculate in Excel: **PERMUT(n,r)**

Combination: The number of ways we can put r items into n spots if order does not matter.

Denoted by: ${}_nC_r$

Calculation: ${}_nC_r = \frac{n!}{(n-r)!r!}$ (! denotes the factorial)

Alternative notation: $\binom{n}{r}$ (This is especially important in some of the functions in Week 4)

In Excel: **COMBIN(n,r)**

Permutations of like objects: When members of subgroups are indistinguishable, the number of ways that we can arrange them if order matters is:

$$permutations\ of\ like\ objects = \frac{n!}{n_1! \cdot n_1! \cdot \ldots \cdot n_k!}$$

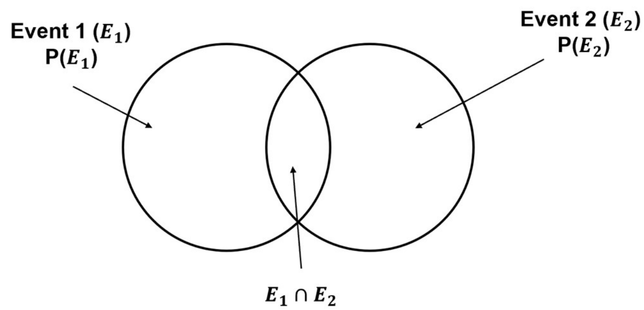
Here, the items are separated into k distinct groups, and $n_1, n_2, \ldots, n_k$ represent the number of items in each of those distinct groups.

Probability Rules

Probability can be thought of as the ratio of the number of ways of satisfying a particular constraint to the total number of possibilities:

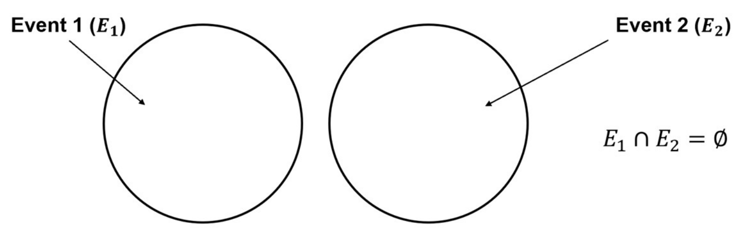
$$P = \frac{Total\ ways\ of\ satisfying\ a\ constraint}{Total\ possibilities}$$

Probability of a union:



$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

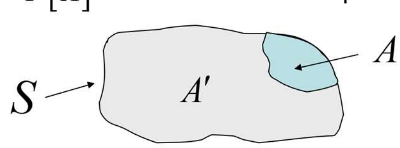
Mutually exclusive events (no intersection):



$$P[E_1 \cup E_2] = P[E_1] + P[E_2]$$

Probability of a complement:

$P[A'] = 1 - P[A]$ where A' is complement of A

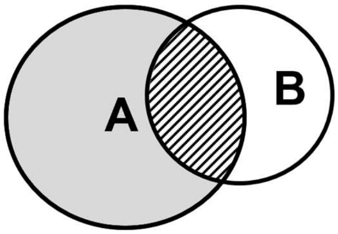


$$P[S] = 1$$

Conditional Probability

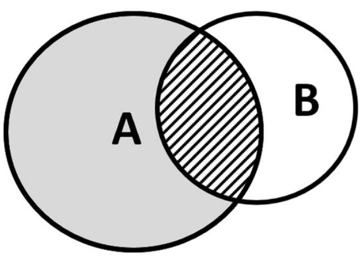
The conditional probability of B given A is defined by:

$$P\left[B|A\right]=\frac{P\left[A\cap B\right]}{P\left[A\right]}$$



Multiplication Rule

Starting with rearranged versions of conditional probabilities (see above), we can set the left-hand side of the following two equations equal to each other:



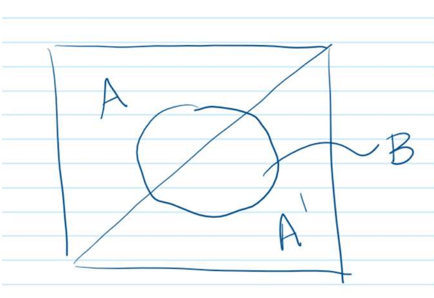
$$P\left[A\cap B\right]=P\left[B|A\right]P\left[A\right]$$
$$P\left[A\cap B\right]=P\left[A|B\right]P\left[B\right]$$

to arrive at the Multiplication Rule:

$$P[B|A]\cdot P[A]=P[A|B]\cdot P[B]$$

Total Probability Rule

Consider two events A and A' that are exhaustive (meaning that they take up the entire sample set) and each has an intersection with event B:



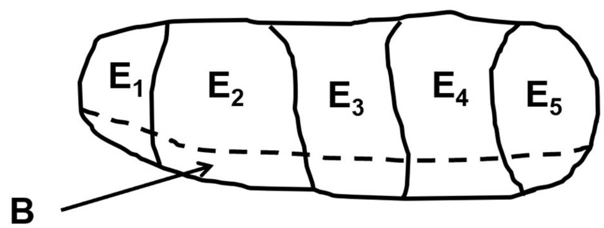
The total probability of B is given by

$$P[B] = P[B \cap A] + P[B \cap A']$$

Using the rules of conditional probability (see above), we arrive at the Total Probability Rule:

$$P[B] = P[B|A] \cdot P[A] + P[B|A'] \cdot P[A']$$

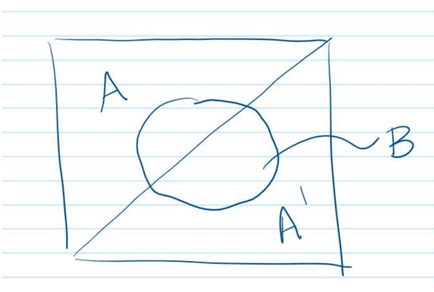
This can be extended to cases with multiple exhaustive events (E_1 through E_5, in the following example), where each of those events has an intersection with event B:



$$P[B] = P[B|E_1] \cdot P[E_1] + P[B|E_2] \cdot P[E_2] + P[B|E_3] \cdot P[E_3] + P[B|E_4] \cdot P[E_4] + P[B|E_5] \cdot P[E_5]$$

<u>Bayes' Theorem</u>

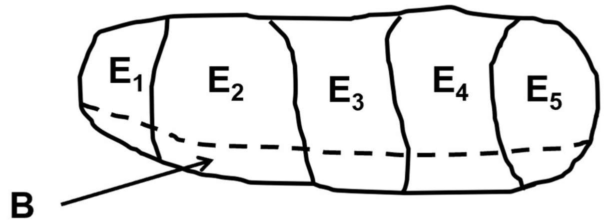
Again consider two events A and A' that are exhaustive (meaning that they take up the entire sample set) and each has an intersection with event B:



We can rearrange the Multiplication Rule and substitute in the Total Probability Rule to arrive at Bayes' Theorem:

$$P[A|B] = \frac{P[B|A]P[A]}{P[B|A]P[A] + P[B|A']P[A']}$$

Like the Total Probability Rule, this can be extended to cases with multiple exhaustive events (E_1 through E_5, in the following example), where each of those events has an intersection with event B:



$$P\left[E_j \middle| B\right] = \frac{P\left[B \middle| E_j\right] \cdot P\left[E_j\right]}{\sum_{i=1}^{k} P\left[B \middle| E_i\right] P\left[E_i\right]}$$