Chapter 2

Probability

2.1 Definitions

- Experiment: Some action that takes place in the world
- Outcomes = Sample Space = Ω = the universe, every *basic outcome* that could happen
- Event = $A \subseteq \Omega$ = something that happened (could be more than one basic outcome
- Probability Distribution = $P : \Omega \to [0, 1], \sum_{x\in\Omega} P(x) = 1$, i.e. values sum to 1 and no value is negative

2.2 Example

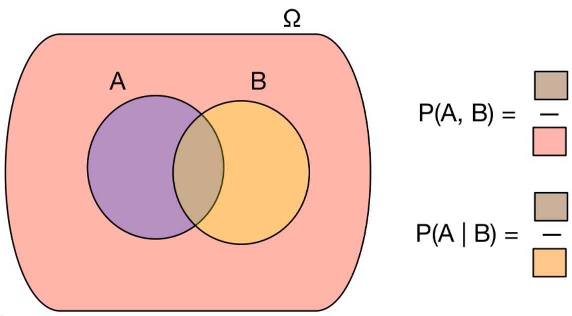
- Experiment = "toss a coin three times"
- $\Omega = \{HHH, HHT, HTT, HTH, THH, THT, TTT, TTH\}$
- Event A = "exactly two heads" = $\{HHT, HTH, THH\}$
- Event B = "first one was heads" = $\{HHH, HHT, HTT, HTH\}$
- Distribution: assign a number between 0 and 1 ('probability') to each basic outcome[1]; sum of all such numbers = 1
- Uniform Distribution: define $P(x) = c\forall c \in \Omega$...in this case?
- Probability of an event = sum of the probability of its basic outcomes
- So, $P(A) =?$ and $P(B) =?$

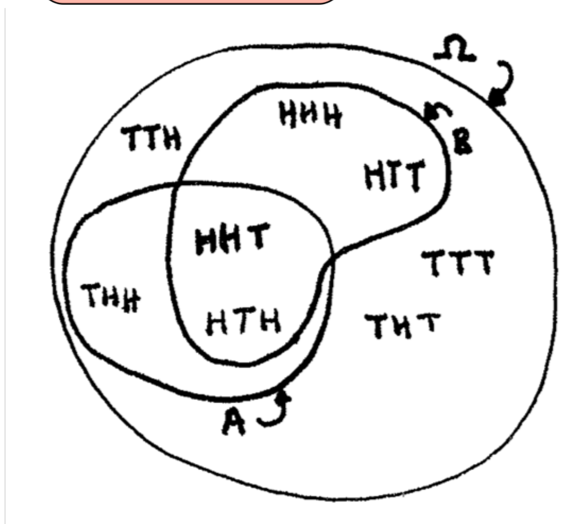
[1] actually to each event in a partition but we'll get back to that in a minute

2.3 Joint and Conditional Probability

$P(A, B) = P(A \cap B) =$ 'joint probability of A and B', i.e. probability of the event formed by the intersection operation (can think of it as probability of 'the joint event'

$P(A|B) = \frac{P(A \cap B)}{P(B)} =$ 'conditional probability of A given B', i.e. the joint event above, *assuming that* B *is* Ω



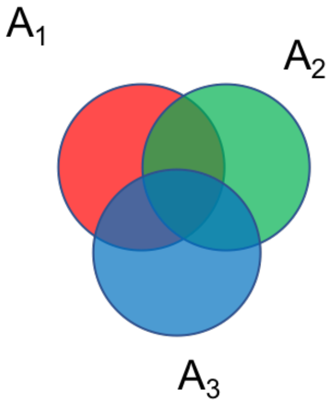


So $P(A) = 3/8$, and $P(B) = 1/2$
$P(A, B) = P(A) + P(B)$? (No. What is it?)
$P(B|A) =$ "if you've got two heads what's the chance your first was heads" = ?
$P(A|B) =$ "if your first is heads what's the chance you've got two" = ?

2.4 Chain Rule of Probability

(Not to be confused with the chain rule of Calculus)

Since $P(A|B) = \frac{P(A,B)}{P(B)}$ (by definition), we can rewrite terms to get $P(A, B) = P(A|B)P(B)$.



Now consider three events, A_1, A_2, A_3. How can we define $P(A_1, A_2, A_3) = P(A_1 \cap A_2 \cap A_3)$ in terms of conditional probabilities?

Recall that an event is just a set of basic outcomes. So let's define a new event

$$A_{23} = A_2 \cap A_3$$

.

Then we would write

$$P(A_1, A_{23}) = P(A_1|A_{23})P(A_{23})$$

.

Now, substitute back in the joint event that A_{23} represents

$$P(A_1, A_2, A_3) = P(A_1|A_2, A_3)P(A_2, A_3)$$

.

Now substitute the definition of joint probabilities in terms of conditional probabilities again

$$P(A_1, A_2, A_3) = P(A_1|A_2, A_3)P(A_2|A_3)P(A_3)$$

.

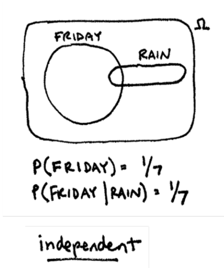
Of course $P(A_1, A_2, A_3) = P(A_3, A_2, A_1)$ (set intersection is commutative) so you could write this instead as $P(A_3|A_2, A_1)P(A_2|A_1)P(A_1)$

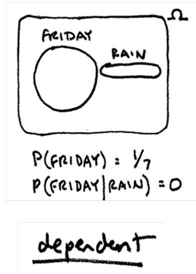
The general chain rule for probabilities is:

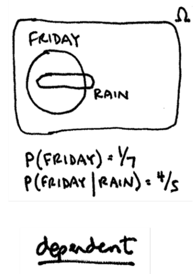
$$P(A_1, \ldots, A_N) = P(A_1|A_2, \ldots, A_N) \times \ldots \times P(A_{N-1}|A_N) \times P(A_N)$$

2.5 Independence

A and B are *independent* if the occurrence of one does not affect the occurrence of the other, i.e. if $P(A|B) = P(A)$. Corollary: $P(B|A) = P(B)$. Corollary: $P(A, B) = P(A)P(B)$. **Exercise:** Prove that these three statements are corollaries of each other.







2.6 Bayes' Rule/Theorem/Law

$P(A|B) = \frac{P(A,B)}{P(B)}$, by definition. Thus, $P(A, B) = P(A|B)P(B)$.

Because intersection is commutative (see above), $P(A, B) = P(B|A)P(A)$. This also explains the corollary noted in Section 2.5. This leads to Bayes' Rule/Theorem/Law:

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

This can be very helpful when you have information about one conditional direction but you want info about the other direction.

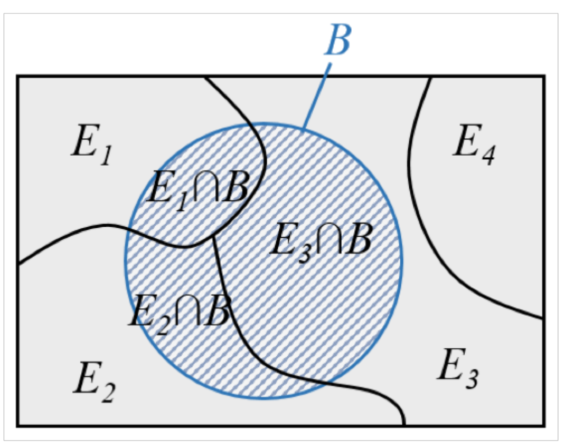
2.7 Law of Total Probability

We say events $E_1, \ldots, E_n$ *partition* Ω if:

$$\forall i, j \in [1, n], E_i \cap E_j = \varnothing$$

and

$$\sum_{i=1}^{n} P(E_i) = 1$$



The Law of Total Probability says, given partitioning events $E_1 \ldots E_n$ and event B:

$$P(B) = \sum_{i=1}^{n} P(B, E_i)$$

2.8 Example

Some people can read minds, but not many: $P(MR) = 1/100,000 = .00001$.

There is a test to read minds; if you are a mind reader I can detect this very well: $P(T|MR) = 0.95$ and if you're not I can detect this even better: $P(\neg T|\neg MR) = 0.995$. Note: $\{T, \neg T\}$ partition the event space, as do $\{MR, \neg MR\}$.

If Jill gets a positive result on the test, how likely is it she is the mind reader?

i.e., $P(MR|T) =?$

By Bayes' Law, $P(MR|T) = \frac{P(T|MR)P(MR)}{P(T)}$.

We need to get $P(T)$. By law of total probability, $P(T) = P(T, MR) + P(T, \neg MR)$. By definition of conditional probability, $P(T, MR) = P(T|MR)P(MR) = 0.95 \times .00001$; $P(T, \neg MR) = P(T|\neg MR)P(\neg MR) = 0.005 \times .99999$. $P(T) = .00500945$ and $P(MR|T) \approx .002$