1 Independence

**Definition 1.** Events $E$ and $F$ are independent if $Pr(E|F) = Pr(E)$

The intuition is that the additional information that $F$ occurred gives us no new information about the probability of $E$ occurring.

**Lemma 2.** Events $E$ and $F$ are independent iff $Pr(E \cap F) = Pr(E)Pr(F)$

**Proof.** The proof is simple algebra using the definition:

If $Pr(E \cap F) = Pr(E)Pr(F)$ then $Pr(E|F) = \frac{Pr(E \cap F)}{Pr(F)} = \frac{Pr(E)Pr(F)}{Pr(F)} = Pr(E)$

The reverse direction also holds. 

If $Pr(E|F) > Pr(E)$ we say that $E$ and $F$ are positively correlated, if $Pr(E|F) < Pr(E)$ we say that $E$ and $F$ are negatively correlated.

**Example 3.** Suppose a family has 2 children. The probabilities of the four possible boy (B) / girl (G) children – namely BB, BG, GB, and GG—are all the same. Given that one of the children is a boy, what is the probability that the family has 2 boys?

We have that $E=\{BB\}$, $F=\{BG, GB, BB\}$, $Pr(E|F) = \frac{Pr(E \cap F)}{Pr(F)} = \frac{1/4}{3/4} = \frac{1}{3}$

We see in the 2-child example that the events ‘2 boys’ and ‘at least one boy’ are correlated.
Example 4. Consider a family with 3 children, again each combination of boys and girls being equally likely.

Consider the events

\( E = \text{‘there are children of both sexes’} \), and \( F = \text{‘there is at most one boy’} \).

Straightforward calculations (do them!) show that \( \Pr(E) = 6/8 = 3/4 \), \( \Pr(F) = 4/8 = 1/2 \), and \( \Pr(E \cap F) = 3/8 \).

So \( \Pr(E \cap F) = 3/8 = 3/4 \times 1/2 = \Pr(E)\Pr(F) \), which shows that \( E \) and \( F \) are independent.

This does not fit with our intuition! (Which shows that you should believe in Math.)

EXERCISE: Do the same problem for 4 children. Are the events still independent?

1.1 Independence for several events

If we have 3 events, \( A_1, A_2, \) and \( A_3 \) it is not necessarily the case that if they are pairwise independent (that is, for \( i \neq j \) \( \Pr(A_i \cap A_j) = \Pr(A_i)\Pr(A_j) \)) then they are independent (that is \( \Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1)\Pr(A_2)\Pr(A_3) \)).

EXERCISE: Give an example!

One can similarly define \( k \)-wise independence: independence is the special case of \( k = 2 \). Similarly, for \( k \) events, parities independence, ... \( (k-1) \)-wise independence are not enough to ensure total independence.

HARDER EXERCISE: Give an example for every \( k \)!

2 Birthday Paradox

We will make two assumptions:

1. a year has 366 days

2. all birthdays are equally likely (the pdf of birthdays is uniform over the days of the year. Birthdays of distinct people are uncorrelated

Neither is true, but this is a Math problem...

The question is, what is the smallest number \( n \) such that in a room with \( n \) people the probability of the event ‘two people have the same birthday’ is greater than 1/2?
We know that \( n = 367 \) is sufficient (why?) but we want a smaller number.

1/2 of 366 is a popular guess—it is wrong. Let us compute the probability \( p_k \) that no two people have the same birthday in a group of \( k \) people.

Clearly, \( p_2 = \frac{365}{366} \) (the second person has only one bad choice). Similarly, the 3rd person must avoid 2 days, so \( p_3 = \frac{364}{366} \times \frac{365}{366} \), and \( p_n = \frac{365 \times 364 \times (366-n+1)}{366^n} \).

We want \( n \) such that \( p_n < 1/2 \). There is no easy shortcut, but it is not hard to write code—and the answer is .... 23.

This is surprisingly small!

The same kind of analysis is useful in other contexts—for example analyzing the behavior of hash functions. See Rosen.

3 Random Variables

Definition 5. Given a sample space \( \Omega \), and a pdf on it, a random variable is a function \( \Omega \to \mathbb{R}^+ \).

3.1 Bernouilli Trials

A Bernouilli trial is a sequence of \( n \) independent events, each with the same underlying structure:

- There are two possible outcomes: S (success) and F (failure), also denoted by 1 and 0
- \( Pr(S) = p \) for some \( p \in (0, 1) \); \( Pr(F) = q = 1 - p \)

We can view this as a (possibly biased) coin, where the probability of Heads is \( Pr(S) = p \). The Bernouilli trial is then a sequence of coin tosses, each toss independent of all others, but with the same underlying probability distribution.

We are usually interested in the number of successes. We denote the probability of \( k \) successes in \( n \) trials, where the probability of success in each trial is \( p \) by \( b(k; n, p) \).

We can think of such sequences as a sample space, and consider the random variable that associate to each sequence the number of successes.

Example 6. Consider a sequence of 3 coin tosses. Let \( i \) be the number of heads H. We have
| i = 0 | TTT |
| i = 1 | TTH, THT, HTT |
| i = 2 | THH, HTH, HHT |
| i = 3 | HHH |

The probabilities are 1/8, 3/8, 3/8, and 1/8, respectively.

**Theorem 7.** \( b(k; n, p) = \binom{n}{k} p^k (1 - p)^{n-k} \)

**Proof.** The formula looks much like a term of the Binomial Theorem. This is not a coincidence! A sequence of \( n \) terms with \( k \) successes can be obtained by considering a sequence of \( n \) letters, each chosen from \{S, F\}, with exactly \( k \) places where we chose S (and, implicitly, chose F in the remaining \( n - k \) places). The probability of a particular sequence is \( p^k (1 - p)^{n-k} \) (by the product rule). There are exactly \( \binom{n}{k} \) such sequences. \( \square \)

### 3.2 Expectation

**Definition 8.** The expectation \( E(X) \) of a random variable \( X \) is

\[
E(X) = \sum_{\omega \in \Omega} X(\omega) Pr(X = \omega)
\]

**Example 9.** Consider the random variable that is the value of the upward-looking face of an honest die. Its expectation is \( 1/6(1 + 2 + 3 + 4 + 5 + 6) = 21/6 = 7/2. \)

If, instead the die was dishonest, with 1 having probability 1/2, and the others having equal probabilities of 1/10, the expectation would be \( 1/2 + 1/10(2 + 3 + 4 + 5 + 6) = 2.5 \)

**Lemma 10.** For any random variable \( X \), \( \min_{\omega \in \Omega} X(\omega) \leq E(X) \leq \max_{\omega \in \Omega} X(\omega) \)

The proof is left to the reader.

The next theorem is quite important.

**Theorem 11.** Let \( X_1, X_2, \ldots, X_k \) be random variables on the same sample space, and define the random variable \( X \) by \( X = \sum_{i=1}^{k} X_i \). Then

\[
E(X) = E(\sum_{i=1}^{k} X_i) = \sum_{i=1}^{k} E(X_i)
\]
Note that there were no conditions imposed on the $X_i$—in particular they could be correlated, there could be repetitions, etc.

**Proof.** We simply use the definition of expectation, and some symbol pushing.

$$E(X) = \sum_{\omega \in \Omega} \left( \sum_{i=1}^{k} X_i(\omega) \right) Pr(\omega)$$

$$= \sum_{\omega \in \Omega} \sum_{i=1}^{k} X_i(\omega) Pr(\omega)$$

$$= \sum_{i=1}^{k} \sum_{\omega \in \Omega} X_i(\omega) Pr(\omega)$$

$$= \sum_{i=1}^{k} \left( \sum_{\omega \in \Omega} X_i(\omega) Pr(\omega) \right) = \sum_{i=1}^{k} E(X_i) \quad \square$$

It is easy to see that, for any $\alpha \in R$ and any random variable $X$, $E(\alpha X) = \alpha E(X)$

**Corollary 12** (Linearity of Expectation). *Let $X_1, X_2, \ldots, X_k$ be random variables on the same sample space, $\alpha_i \in R$, and define the random variable $X$ by $X = \sum_{i=1}^{k} \alpha_i X_i$. Then*

$$E(X) = E\left( \sum_{i=1}^{k} \alpha_i X_i \right) = \sum_{i=1}^{k} \alpha_i E(X_i)$$

This is a very useful fact. For example we could compute the expected sum of throwing three fair dice as follows:

Let $X_i$ be the random variable that represents the face of the $i$-th die. We already computed $E(X_i) = 7/2$. So $E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 3 \times 7/2 = 21/2$