Sample space: $S = \text{set of all potential outcomes of experiment}$

E.g., flip two coins: $S = \{(H,H), (H,T), (T,H), (T,T)\}$

Events: $E \subseteq S$ is an arbitrary subset of the sample space

$\geq 1$ head in 2 flips: $E = \{(H,H), (H,T), (T,H)\}$

Probability:
A function from subsets of $S$ to real numbers – $\Pr: 2^S \rightarrow [0,1]$

Probability Axioms:
Axiom 1 (Non-negativity): $0 \leq \Pr(E)$

Axiom 2 (Normalization): $\Pr(S) = 1$

Axiom 3 (Additivity): $EF = \emptyset \Rightarrow \Pr(E \cup F) = \Pr(E) + \Pr(F)$

Equally likely outcomes

Simplest case: sample spaces with equally likely outcomes.

- Coin flips: $S = \{\text{Heads, Tails}\}$
- Flipping two coins: $S = \{(H,H), (H,T), (T,H), (T,T)\}$
- Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$

$\Pr(\text{each outcome}) = \frac{1}{|S|}$

In that case,

$\Pr(E) = \frac{|E|}{|S|}$

Conditional probability & chain rule

General definition: $P(E \mid F) = \frac{P(EF)}{P(F)}$ where $P(F) > 0$

Implies: $P(EF) = P(E|F)P(F)$ ("the chain rule")

General definition of Chain Rule:

$P(E_1E_2 \ldots E_n) = \frac{P(E_1|E_2 \ldots E_n) \cdot P(E_2|E_3 \ldots E_n) \ldots P(E_n|E_1, E_2, \ldots, E_{n-1})}{P(E_1P(E_2 \ldots P(E_n))}$

Bayes Theorem

Most common form:

$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E)}$

Expanded form (using law of total probability):

$P(F \mid E) = \frac{P(E \mid F)P(F)}{P(E \mid F)P(F) + P(E \mid F^c)P(F^c)}$

Proof:

$P(F \mid E) = \frac{P(\text{EF})}{P(E)} = \frac{P(E \mid F)P(F)}{P(E)}$

Law of total probability

More generally, if $F_1, F_2, \ldots, F_n$ partition $S$ (mutually exclusive, $\bigcup F_i = S$, $P(F_i) > 0$), then

$P(E) = \sum_i P(E \mid F_i)P(F_i)$

(Analogous to reasoning by cases; both are very handy)

Independence
Independence of events

Intuition: \( E \) is independent of \( F \) if the chance of \( E \) occurring is not affected by whether \( F \) occurs.

Formally:

\[
Pr(E|F) = Pr(E) \quad \text{or} \quad Pr(E \cap F) = Pr(E)Pr(F)
\]

These two definitions are equivalent.

Independence

Draw a card from a shuffled deck of 52 cards.

\( E \): card is a spade

\( F \): card is an Ace

Are \( E \) and \( F \) independent?

Independence

Toss a coin 3 times. Each of 8 outcomes equally likely.

Define

\( A = \{ \text{at most one T} \} = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}\} \)

\( B = \{ \text{both H and T occur} \} = \{\text{HHH, TTT}\} \)

Are \( A \) and \( B \) independent?

Independence as an assumption

It is often convenient to assume independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

\( E = \{ \text{main chute doesn’t open} \} \quad Pr(E) = 0.02 \)

\( F = \{ \text{backup doesn’t open} \} \quad Pr(F) = 0.1 \)

What is the chance that at least one opens assuming independence?

Note: Assuming independence doesn’t justify the assumption! Both chutes could fail because of the same rare event, e.g. freezing rain.

Independence as an assumption

It is often convenient to assume independence.

People often assume it without noticing.

Example: A sky diver has two chutes. Let

\( E = \{ \text{main chute doesn’t open} \} \quad Pr(E) = 0.02 \)

\( F = \{ \text{backup doesn’t open} \} \quad Pr(F) = 0.1 \)

What is the chance that at least one opens assuming independence?

Using independence to define a probabilistic model

We can define our probability model via independence.

Example: suppose a biased coin comes up heads with probability \( 2/3 \), independent of other flips.

Sample space: sequences of 3 coin tosses.

\( Pr(3 \text{ heads}) =? \)

\( Pr(3 \text{ tails}) =? \)

\( Pr(2 \text{ heads}) =? \)
Suppose a biased coin comes up heads with probability $p$, independent of other flips.

- $P(n \text{ heads in } n \text{ flips}) = p^n$
- $P(n \text{ tails in } n \text{ flips}) = (1-p)^n$
- $P(\text{HHTHTTT}) = p^2(1-p)p(1-p)^3 = p^2(1-p)^{n-3}$
- $P(\text{exactly } k \text{ heads in } n \text{ flips}) = \binom{n}{k} p^k (1-p)^{n-k}$

Aside: note that the probability of some number of heads is as it should, by the binomial theorem.

Consider the following parallel network with $n$ routers, $i^{th}$ has probability $p_i$ of failing independently.

- $P(\text{there is functional path}) = 1 - P(\text{all routers fail})$
Contrast: a series network

\[ P(\text{there is functional path}) = (1 - p_1)(1 - p_2) \cdots (1 - p_n) \]

A data structure problem: fast access to small subset of data drawn from a large space.

A solution: hash function \( h: \mathbb{D} \rightarrow \{0, \ldots, n-1\} \) crunches/scrambles names from large space into small one.

E.g., if \( x \) is integer: \( h(x) = x \mod n \)

Everything that hashes to same location stored in linked list. Good hash functions approximately randomize placement.

Scenario: Hash \( m \leq n \) keys from \( \mathbb{D} \) into size \( n \) hash table.

How well does it work?

**Worst case:** All collide in one bucket. (Perhaps too pessimistic?)

**Best case:** No collisions. (Perhaps too optimistic?)

A middle ground: Probabilistic analysis.

Below, for simplicity, assume:

- Keys drawn from \( \mathbb{D} \) randomly, independently (with replacement)
- \( h \) maps equal numbers of domain points into each range bin, i.e., \( |\mathbb{D}| = k|R| \) for some integer \( k \), and \( |h^{-1}(i)| = k \) for all \( 0 \leq i \leq n-1 \)

Many possible questions; a few analyzed below.
m keys hashed (non-uniformly) to table w/ n buckets
Each string hashed is an independent trial, with probability $p_i$ of getting hashed to bucket $i$

$E = \text{At least 1 of first } k \text{ buckets gets } \geq 1 \text{ key}$

What is $P(E)$?

Solution:
$F_i = \text{at least one key hashed into } i\text{-th bucket}$
$P(E) = P(F_1 \cap \ldots \cap F_k) = 1 - P(F_1^c \cup \ldots \cup F_k^c)$
$= 1 - (1 - p_1 - p_2 - \ldots - p_k)^m$

Perfect hashing (i)

Let $|R| = n, D_0 \subseteq D, |D_0| = m$. A hash function $h: D \rightarrow R$ is perfect for $D_0$ if $h: D_0 \rightarrow R$ is injective (no collisions). How likely is that?

1) Fix $h$; pick $m$ elements of $D_0$ independently at random $\in D$

$P(h \text{ is perfect for } D_0) = \frac{n \cdot n - 1 \cdot \ldots \cdot n - m + 1}{n^m}$

Except for very empty tables, a "perfect" hash is improbable.

If $E$ and $F$ are independent, then so are $E$ and $F^c$
and so are $E^c$ and $F$
and so are $E^c$ and $F^c$

Proof:
$P(EF^c) = P(E) - P(EF)$
$= P(E) - P(E)P(F)$
$= P(E)(1 - P(F))$
$= P(E)P(F^c)$
Independence of several events

Three events E, F, G are mutually independent if

\[ Pr(E \cap F) = Pr(E)Pr(F) \]
\[ Pr(F \cap G) = Pr(F)Pr(G) \]
\[ Pr(E \cap G) = Pr(E)Pr(G) \]
\[ Pr(E \cap F \cap G) = Pr(E)Pr(F)Pr(G) \]

Example: Show that E is independent of F U G.

\[
Pr( F U G | E) = Pr( F | E) + Pr( G | E) - Pr( F G | E)
\]
\[
= Pr( F) + Pr( G) - Pr( F G)/Pr(E)
\]
\[
= Pr( F) + Pr( G) - Pr( F G) = Pr( F U G )
\]