University of Washington

STATISTICS

# Elements of Statistical Methods

# Probability (Ch 3)

Fritz Scholz

Spring Quarter 2010

April 5, 2010

# Frequentist Interpretation of Probability

In common language usage probability can take on several different meanings.

E.g., there is the possibility, not necessarily certainty.

There is a $40\%$ chance that it will rain tomorrow.

Given the current atmospheric conditions, under similar conditions in the past it has rained the next day in about $40\%$ of the cases.

or: about $40\%$ of the rain gauges in the forcast area showed rain.

Both of these are examples of the frequentists' notion of probability, in the long run what proportion of instances will give the target result?

The $40\%$ area forecast may need some thinking in that regard to make it fit.

# Subjectivist Interpretation of Probability

You look out the window and say:

there is a $40\%$ chance that it rains while I walk to the train station.

It is based on gut feeling (subjective) and internal vague memory of that person.

Having received a high PSA value for a prostate check

a man is $95\%$ certain that the biopsy will show no cancer.

A woman has to decide whether to undergo an operation.

Based on medical experience (frequentist) it will be succesful in $40\%$ of the cases.

But she feels she has a better than $80\%$ chance of success.

It is my lucky day, I feel $70\%$ certain that I will win something in the lottery.

# Axiomatic Probability

Rather than deciding which interpretation is correct or to adopt, we stay on neutral grounds and use the axiomatic probability model proposed by Kolmogorov (1933).

Both frequentists and subjectivists appear to accept this model as a basis and it has evolved into a rich and useful theory.

It consists of three entities:

$S$: A sample space, a universe of all possible outcomes for an experiment.

$\mathcal{C}$: A designated collection of observable subsets (called events) of $S$.

$P$: A probability measure, a function that assigns numbers (called probabilities) to the events in $\mathcal{C}$.

# Examples

Flipping a coin twice we can distinguish the following four outcomes:

$$S = \left\{ \begin{array}{cc} \mathtt{HH} & \mathtt{HT} \\ \mathtt{TH} & \mathtt{TT} \end{array} \right\}$$

As observable events we can take the collection of all subsets of $S$, i.e.,

$$\mathcal{C} = \left\{ \begin{array}{c} S \;, \qquad \emptyset \\ \{\mathtt{HH},\mathtt{HT},\mathtt{TH}\}, \{\mathtt{HH},\mathtt{HT},\mathtt{TT}\}, \{\mathtt{HH},\mathtt{TH},\mathtt{TT}\}, \{\mathtt{TH},\mathtt{HT},\mathtt{TT}\} \\ \{\mathtt{HH},\mathtt{HT}\}, \{\mathtt{HH},\mathtt{TH}\}, \{\mathtt{HH},\mathtt{TT}\}, \{\mathtt{TH},\mathtt{HT}\}, \{\mathtt{TH},\mathtt{TT}\}, \{\mathtt{HT},\mathtt{TT}\} \\ \{\mathtt{HH}\}, \{\mathtt{HT}\}, \{\mathtt{TH}\}, \{\mathtt{TT}\} \end{array} \right\}$$

Measuring a person's height. While we have some notion about such heights, e.g., $1\ \mathtt{foot} \le$ height $\le 9\ \mathtt{feet}$, it is more convenient to us $S = R$, even though most of these outcome values are not possible.

Observable events could be the collection of all intervals, plus what can be obtained by set operations $\cup$, $\cap$ and $^c$.

# Observable Events

We said that $\mathcal{C}$ should contain the observable events.

What does observable mean?

When $S$ is finite we can take the collection $\mathcal{C}$ of all subsets of $S$.

The same route can be taken with denumerable sample spaces $S$.

When $S = R$, it is no longer so easy to take all possible subsets of $R$ as $\mathcal{C}$.

We need to impose some axiomatic assumptions about $\mathcal{C}$ and $P$.

# Collection of Events: Required Properties

Any collection $\mathcal{C}$ of events should satisfy the following properties:

1. The sample space $S$ is an event, i.e., $S \in \mathcal{C}$

2. If $A$ is an event, i.e., $A \in \mathcal{C}$, so is $A^c$, i.e., $A^c \in \mathcal{C}$

3. For any countable sequence of events, $A_1, A_2, A_3 \ldots \in \mathcal{C}$

   their union should also be an event, i.e., $\bigcup_{i=1}^{\infty} A_i \in \mathcal{C}$.

Such a collection $\mathcal{C}$ with properties 1-3 is also called a sigma-field or sigma-algebra.

By 1. and 2.: $S \in \mathcal{C} \Longrightarrow S^c = \emptyset \in \mathcal{C}$. $\quad \mathcal{C} = \{\emptyset, S\}$ is the simplest sigma-field.

# Coin Flips Revisited

Suppose we cannot distinguish $\texttt{HT}$ and $\texttt{TH}$. Then we have as sample space

$$S = \{\{\texttt{H},\texttt{H}\},\{\texttt{H},\texttt{T}\},\{\texttt{T},\texttt{T}\}\}$$

We used set notation to describe the three elements,

order within a set is immaterial.

As sigma-field of all subsets we get

$$\mathcal{C} = \left\{ \begin{array}{c} S, \quad \emptyset, \quad \{\,\{\texttt{H},\texttt{H}\}\,\}, \quad \{\,\{\texttt{H},\texttt{T}\}\,\}, \quad \{\,\{\texttt{TT}\}\,\}, \\ \{\,\{\texttt{H},\texttt{H}\},\{\texttt{H},\texttt{T}\}\,\}, \quad \{\,\{\texttt{H},\texttt{H}\},\{\texttt{T},\texttt{T}\}\,\}, \quad \{\,\{\texttt{T},\texttt{T}\},\{\texttt{H},\texttt{T}\}\,\} \end{array} \right\}$$

$$\{\,\{\texttt{H},\texttt{H}\},\{\texttt{H},\texttt{T}\}\,\}^c = \{\,\{\texttt{TT}\}\,\}$$

The text treats this example within the context of the original sample space,

but introduces the indistinguishability of $\texttt{HT}$ and $\texttt{TH}$ by imposing the condition

that any event containing $\texttt{HT}$ also contains $\texttt{TH}$.    Compare the two models!

# The Probability Measure $P$

A probability measure on $\mathcal{C}$ assigns a number $P(A)$ to each event $A \in \mathcal{C}$ and satisfies the following properties (axioms):

1. For any $A \in \mathcal{C}$ we have $0 \leq P(A) \leq 1$.

2. $P(S) = 1$ The probability of some outcome in $S$ happening is $1$.

3. For any sequence $A_1, A_2, A_3, \ldots$ of pairwise disjoint events we have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{i.e.} \quad P\left(\lim_{n\to\infty} \bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} P(A_i)$$

The third property is referred to as countable additivity.

# $P(\emptyset) = 0$

This obvious property could have been added to the axioms, but it follows from 2-3.

Consider the specific sequence of pairwise disjoint events $S, \emptyset, \emptyset, \emptyset, \ldots$

and note that their infinite union is just $S$, i.e.,

$$S \cup \emptyset \cup \emptyset \cup \emptyset \cup \cdots = S$$

From properties 2 and 3 (axioms 2 and 3) we have

$$1 = P(S) = P(S \cup \emptyset \cup \emptyset \cup \emptyset \cup \cdots) = P(S) + \sum_{i=2}^{\infty} P(\emptyset) = 1 + \sum_{i=2}^{\infty} P(\emptyset)$$

it follows that $\sum_{i=2}^{\infty} P(\emptyset) = 0$ and thus $P(\emptyset) = 0$.

# Countable Additivity $\Longrightarrow$ Finite Additivity

Let $A_1, \ldots, A_n$ be a finite sequence of pairwise disjoint events. Then

$$P\left(A_1 \cup \ldots \cup A_n\right) = P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) = P(A_1) + \ldots + P(A_n)$$

Proof:

Augment the sequence $A_1, \ldots, A_n$ with an infinite number of $\emptyset$'s, then the infinite sequence $A_1, \ldots, A_n, \emptyset, \emptyset, \ldots$ is pairwise disjoint and their union is

$$A_1 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \cup \ldots = \bigcup_{i=1}^{n} A_i$$

From axiom 3 together with $P(\emptyset) = 0$ we get

$$\begin{aligned} P(A_1 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \cup \ldots) &= P(A_1) + \ldots + P(A_n) + P(\emptyset) + P(\emptyset) + \ldots \\ P(A_1 \cup \ldots \cup A_n) &= P(A_1) + \ldots + P(A_n) \end{aligned}$$

# $P(A^c) = 1 - P(A)$ & $A \subset B \Longrightarrow P(A) \leq P(B)$

In both proofs below we use the established finite additivity property.

$$S = A \cup A^c \qquad 1 = P(S) = P(A) + P(A^c) \implies P(A^c) = 1 - P(A)$$

Assume $A \subset B$, then $A$ and $B \cap A^c$ are mutually exclusive and their union is $B$

$$B = A \cup (B \cap A^c) \implies P(B) = P(A) + P(B \cap A^c) \geq P(A)$$

since $P(B \cap A^c) \geq 0$ by axiom 1.

# $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

For any $A, B \in \mathcal{C}$ we have the pairwise disjoint decompositions (see Venn diagram)

$$A \cup B = (A \cap B^c) \;\cup\; (A \cap B) \;\cup\; (B \cap A^c)$$

$$\Longrightarrow \quad P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(B \cap A^c)$$

and

$$A = (A \cap B^c) \;\cup\; (A \cap B) \quad \text{and} \quad B = (A \cap B) \;\cup\; (B \cap A^c)$$

$$\Longrightarrow \quad P(A) = P(A \cap B^c) + P(A \cap B) \quad \text{and} \quad P(B) = P(A \cap B) + P(B \cap A^c)$$

$$\begin{aligned} \Longrightarrow \quad P(A) + P(B) &= P(A \cap B^c) + 2P(A \cap B) + P(B \cap A^c) \\ &= P(A \cup B) + P(A \cap B) \\ \Longrightarrow \quad P(A) + P(B) - P(A \cap B) &= P(A \cup B) \end{aligned}$$

# Finite Sample Spaces

Suppose $S = \{s_1, \ldots, s_N\}$ is a finite sample space with outcomes $s_1, \ldots, s_N$.

Assume that $\mathcal{C}$ is the collection of all subsets (events) of $S$.

and that we have a probability measure $P$ defined for all $A \in \mathcal{C}$.

Denote by $p_i = P(\{s_i\})$ the probability of the event $\{s_i\}$.

Then for any event $A$ consisting of outcomes $s_{i_1}, \ldots, s_{i_k}$ we have

$$P(A) = P\left(\bigcup_{j=1}^{k} \{s_{i_j}\}\right) = P\left(\bigcup_{s_i \in A} \{s_i\}\right) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} p_i \qquad (1)$$

The probabilities of the individual outcomes determine the probability of any event.

To specify $P$ on $\mathcal{C}$, we only need to specify $p_1, \ldots, p_N$ with $0 \le p_i, i = 1, \ldots, N$

and $p_1 + \ldots + p_N = 1$.

This together with (1) defines a probability measure on $\mathcal{C}$, satisfying axioms 1-3.

This also works for denumerable $S$ with $0 \le p_i, i = 1, 2, \ldots$ and $\sum_{i=1}^{\infty} p_i = 1$.

# Equally Likely Outcomes

Many useful probability models concern $N$ equally likely outcomes, i.e.,

$$p_i = \frac{1}{N}\,, \quad i = 1,\ldots,N \qquad \text{Note } p_i \geq 0 \text{ and } \sum_{i=1}^{N} p_i = \frac{N}{N} = 1$$

For such models the above event probability (1) becomes

$$P(A) = \sum_{s_i \in A} p_i = \sum_{s_i \in A} \frac{1}{N} = \frac{\#(A)}{\#(S)} = \frac{\text{\# of cases favorable to } A}{\text{\# of possible cases}}$$

Thus the calculation of probabilities is simply a matter of counting.

# Dice Example

For a pair of symmetric dice it seems reasonable to assume that all 36 outcomes in a proper roll of a pair of dice are equally likely.

$$S = \left\{ \begin{array}{cccccc} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{array} \right\}$$

What is the chance of coming up with a 7 or 11?

$A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(6,5),(5,6)\}$

with $\#(A) = 8$ and thus $P(A) = 8/36 = 2/9$.

Counting by full enumeration can become tedious and shortcuts are desirable.

# Combinatorial Counts

40 patients in a preliminary drug trial are equally split between men and women.

We randomly split these 40 patients so that half get the new drug
while the other half get a look alike placebo.

What is the chance that the "new drug" patients are equally split between genders?

"randomly split" means: any group of 20 out of 40 is equally likely to be selected.

$$P(A) = \frac{\binom{20}{10} \times \binom{20}{10}}{\binom{40}{20}} = \texttt{choose}(20, 10)\hat{}2/\texttt{choose}(40, 20)$$

$$= \texttt{dhyper}(10, 20, 20, 20) = 0.2476289$$

See documentation in R on using `choose` and `dhyper`.

Enumerating $\binom{40}{20} = 137,846,528,820$ and $\binom{20}{10}^2 = 184,756^2$ cases is impractical.

# Rolling 5 Fair Dice

a) What is the probability that all 5 top faces show the same number?

$$P(A) = \frac{6}{6^5} = \frac{1}{6^4} = \frac{1}{1296}$$

b) What is the probability that the top faces show exactly 4 different numbers?

The duplicated number can occur on any one of the possible $\binom{5}{2}$ pairs of dice (order does not matter) & these two identical numbers can be any one of 6 values.

For the remaining 3 numbers we could have $5 \times 4 \times 3$ possibilities. Thus

$$P(A) = \frac{\#(A)}{\#(S)} = \frac{6 \cdot \binom{5}{2} \cdot 5 \cdot 4 \cdot 3}{6^5} = \frac{25}{54} = 0.462963$$

We repeatedly made use of the multiplication principle in counting the combinations of the various choices with each other.

# Rolling 5 Fair Dice (continued)

c) What is the chance that the top faces show exactly three 6s or exactly two 5s?

Let $A =$ event of seeing exactly three 6s and $B =$ event of seeing exactly two 5s?

$$P(A) = \frac{\binom{5}{3} \cdot 5^2}{6^5}, \qquad P(B) = \frac{\binom{5}{2} \cdot 5^3}{6^5}, \qquad P(A \cap B) = \frac{\binom{5}{3}}{6^5}$$

$$\begin{aligned} P(A \cup B) = P(A) + P(B) - P(A \cap B) &= \frac{\binom{5}{3} \cdot 5^2 + \binom{5}{2} \cdot 5^3 - \binom{5}{3}}{6^5} \\ &= \frac{250 + 1250 - 10}{6^5} = \frac{1490}{6^5} \approx 0.1916 \end{aligned}$$

Sometimes you have to organize your counting in manageable chunks.

# The Birthday Problem

Assuming a 365 day year and dealing with a group of $k$ students,

what is the probability of having at least one birthday match among them?

The basic operating assumption is that all $365^k$ birthday $k$-tuples $(d_1, \ldots, d_k)$

are equally likely to have appeared in this class of $k$.

We will employ a useful trick. Sometimes it is easier to get your counting

arms around $A^c$ and then employ $P(A) = 1 - P(A^c)$.

$A^c$ means that all $k$ birthdays are different.

$$P(A^c) = \frac{365 \cdot 364 \cdots (365 - (k-1))}{365^k} \quad \& \quad P(A) = 1 - \frac{365 \cdot 364 \cdots (366 - k)}{365^k}$$

It takes just 23 students to get $P(A) > .5$.

# Matching or Adjacent Birthdays

What is the chance of having at least one matching or adjacent pair of birthdays? Again, going to the complement is easier. View Dec. 31 and Jan. 1 as adjacent.

Let $A$ be the event of getting $n$ birthdays at least one day apart. Then we have

$$P(A) = \binom{365-n-1}{n-1}\frac{(n-1)!365}{365^n}$$

$$= \frac{(365-2n+1)(365-2n+2)\cdots(365-2n+n-1)}{365^{n-1}}$$

365 ways to pick a birthday for person 1.  There are $365-n$ non-birthdays (NB).

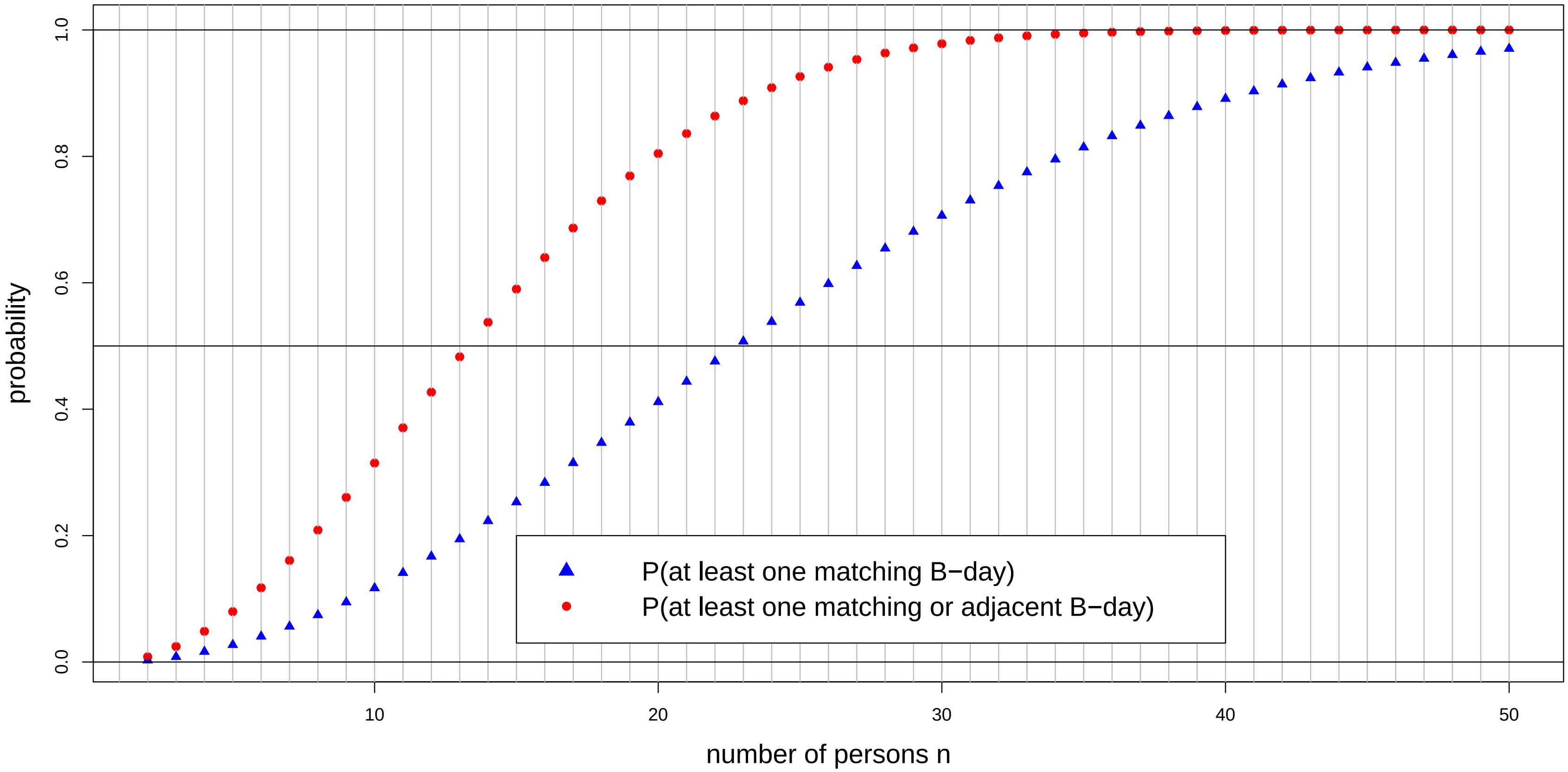
Use the remaining $n-1$ birthdays (BD) to each fill one of the remaining $365-n-1$ gaps between the non-birthdays, $\binom{365-n-1}{n-1}$ ways.

That fixes the circular NB–BD pattern, anchored on the BD of person 1.

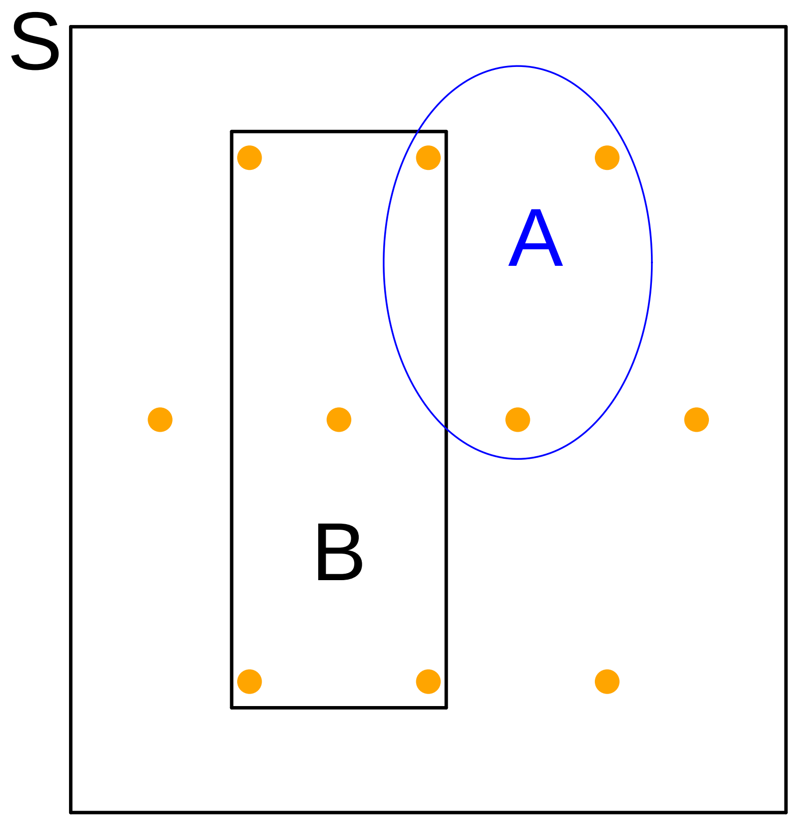
$(n-1)!$ ways to assigns these birthdays to the remaining $(n-1)$ persons.

# $P(M)$ and $P(A^c)$

$n = 14$ gives the smallest $n$ for which $P(A^c) \geq .5$, in fact $P(A^c) = .5375$.



# Conditional Probability




Conditional probabilities are a useful tool for breaking down probability calculations into manageable segments.

The Venn diagram shows 10 equally likely outcomes and two events $A$ and $B$.

$$P(A) = \frac{\#(A)}{\#(S)} = \frac{3}{10} = 0.3$$

Suppose that we can restrict attention to the outcomes in $B$ as our new sample space, then

$$P(A|B) = \frac{\#(A \cap B)}{\#(S \cap B)} = \frac{1}{5} = 0.2$$

the conditional probability of $A$ given $B$.

# Conditional Probability: Formal Definition

The previous example with equally likely outcomes can be rewritten as

$$P(A|B) = \frac{\#(A \cap B)}{\#(S \cap B)} = \frac{\#(A \cap B)/\#(S)}{\#(B)/\#(S)} = \frac{P(A \cap B)}{P(B)}$$

which motivates the following definition in the general case,

not restricted to equally likely outcomes

**Definition:** When $A$ and $B$ are any events with $P(B) > 0$, then we define the

conditional probability of $A$ given $B$ by

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

This can be converted to the multiplication or product rule

$$P(A \cap B) = P(A|B)P(B) \quad \text{and} \quad P(A \cap B) = P(B \cap A) = P(B|A)P(A)$$

provided that in the latter case we have $P(A) > 0$.

# Two Headed/Tailed Coins

Ask Marilyn: Suppose that we have three coins, one with two heads on it (`HH`), one with two tails on it (`TT`) and a fair coin with head and tail (`HT`).

One of the coins is selected at random and flipped. Suppose the face up is `Heads`. What is the chance that the other side is `Heads` as well?

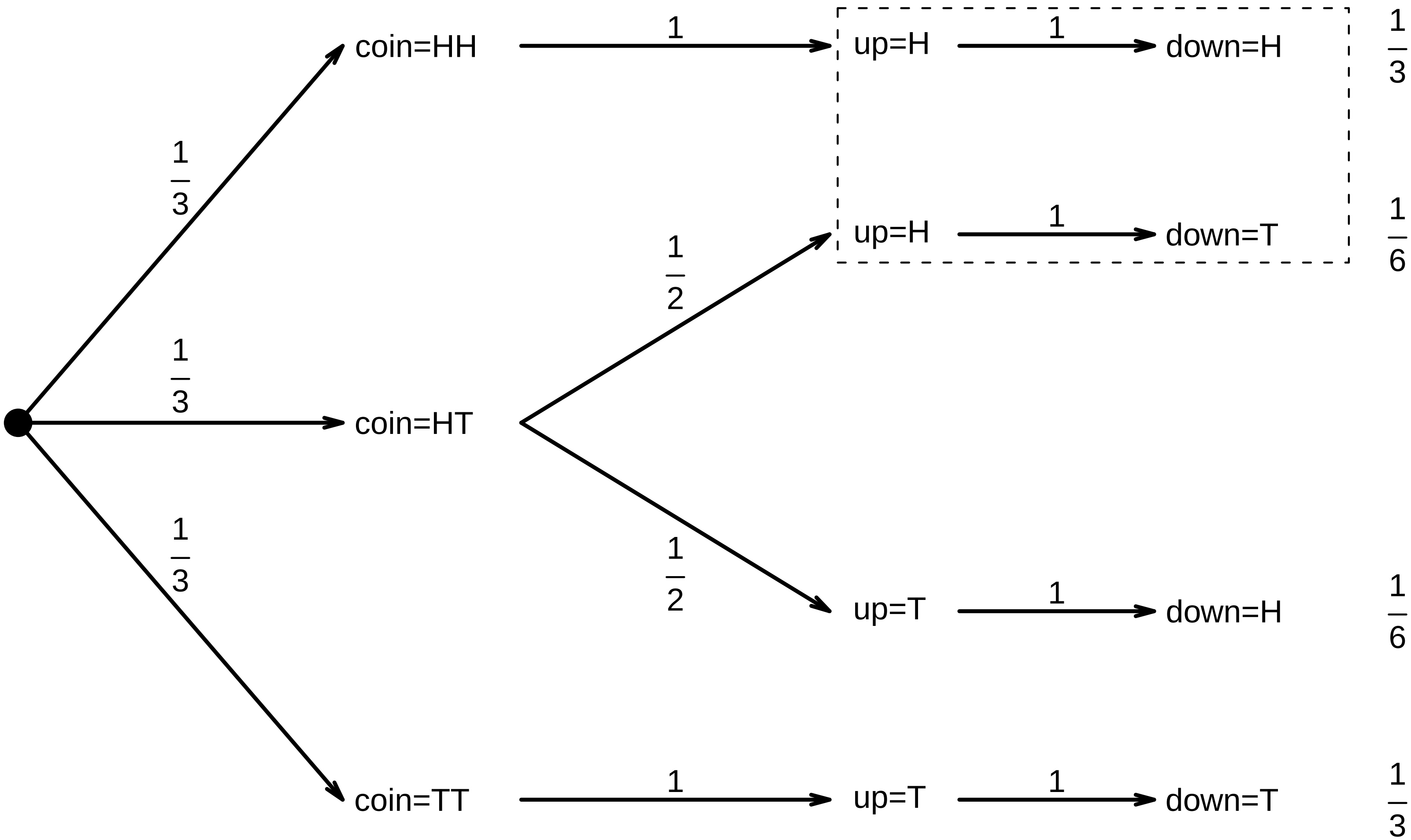
We could reason as follows:

1. Given the provided information, it can't be the `TT` coin. It must be `HH` or `HT`.

2. If `HH` was selected face down is `Heads`, if `HT` then face down is `Tails`

3. Thus the chance of having `Heads` as face down is 1/2.   Or is it?

# Tree Diagram



# Applying the Multiplication Rule

$$\begin{aligned}
P(\{\text{up} = H\}) &= P(\{\{\text{up} = H\} \cap \{\text{coin} = \mathtt{HH}\}\} \cup \{\{\text{up} = H\} \cap \{\text{coin} = \mathtt{HT}\}\}) \\
&= P(\{\text{up} = H\} \cap \{\text{coin} = \mathtt{HH}\}) + P(\{\text{up} = H\} \cap \{\text{coin} = \mathtt{HT}\}) \\
&= P(\{\text{up} = H\} | \{\text{coin} = \mathtt{HH}\}) \cdot P(\{\text{coin} = \mathtt{HH}\}) \\
&\quad + P(\{\text{up} = H\} | \{\text{coin} = \mathtt{HT}\}) \cdot P(\{\text{coin} = \mathtt{HT}\}) \\
&= 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}
\end{aligned}$$

$$\begin{aligned}
P(\{\text{down} &= \mathtt{H}\} | \{\text{up} = H\}) = \frac{P(\{\text{up} = H\} \cap \{\text{down} = \mathtt{H}\})}{P(\{\text{up} = H\})} \\
&= \frac{P(\{\text{coin} = \mathtt{HH}\})}{1/2} = \frac{1/3}{1/2} = \frac{2}{3} \neq \frac{1}{2}
\end{aligned}$$

# HIV Screening

A population can be divided into those that have HIV (denoted by $D$)

and those who do not (denoted by $D^c$).

A test either results in a positive, denoted by $T^+$, or in a negative, denoted by $T^-$.

The Venn diagram shows the possible outcomes for a randomly chosen person.

| | $T^+$ | $T^-$ |
|---|---|---|
| $D$ | $D \cap T^+$ | $D \cap T^-$ |
| $D^c$ | $D^c \cap T^+$ | $D^c \cap T^-$ |

Test is correct: $D \cap T^+$ or $D^c \cap T^-$; false positive: $D^c \cap T^+$; false negative: $D \cap T^-$.

# $P(D|T^+)$?

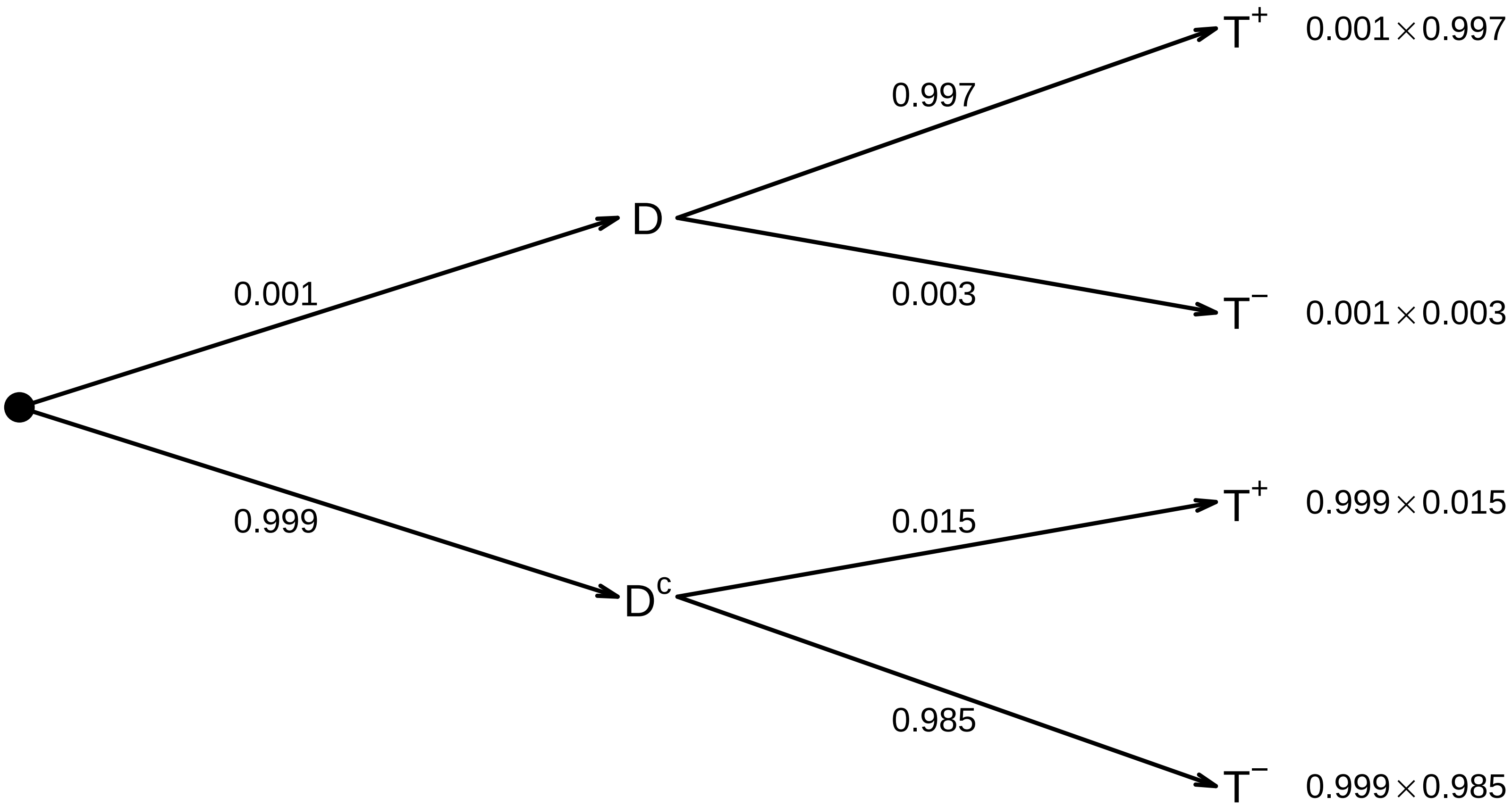
Typically something is known about the prevalence HIV, say $P(D) = 0.001$.

We may also know $P(T^+|D^c) = 0.015$ and $P(T^-|D) = .003$,

the respective probabilities of a false positive and a false negative.

$$
\begin{aligned}
P(D|T^+) &= \frac{P(D \cap T^+)}{P(T^+)} = \frac{P(D \cap T^+)}{P(\{T^+ \cap D\} \cup \{T^+ \cap D^c\})} = \frac{P(D \cap T^+)}{P(T^+ \cap D) + P(T^+ \cap D^c)} \\
&= \frac{P(T^+|D)P(D)}{P(T^+|D)P(D) + P(T^+|D^c)P(D^c)} \\
&= \frac{0.997 \cdot 0.001}{0.997 \cdot 0.001 + 0.015 \cdot 0.999} = 0.06238
\end{aligned}
$$

# HIV Test Tree Diagram




$$P(D \mid T^+) = \frac{P(D \cap T^+)}{P(T^+)} = \frac{0.001 \times 0.997}{0.001 \times 0.997 + 0.999 \times 0.015} = 0.06238268$$

# Independence

The concept of independence is of great importance in probability and statistics.

**Informally:** Two events are independent if the probability of occurrence of either is unaffected by the occurrence of the other.

The most natural way is to express this via conditional probabilities as follows:

$$P(A|B) = P(A) \qquad \text{and} \qquad P(B|A) = P(B)$$

$$\text{or} \qquad \frac{P(A \cap B)}{P(B)} = P(A) \qquad \text{and} \qquad \frac{P(A \cap B)}{P(A)} = P(B)$$

**Definition:** Two events $A$ and $B$ are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

Note that $P(A) > 0$ and $P(B) > 0$ are not required (as in $P(B|A)$ and $P(A|B)$).

# Comments on Independence

If $P(A) = 0$ or $P(B) = 0$ then $A$ and $B$ are independent.

Since $A \cap B \subset A$ and $A \cap B \subset B \implies 0 \leq P(A \cap B) \leq \min(P(A), P(B)) = 0$, thus

$$0 = P(A \cap B) = P(A) \cdot P(B) = 0$$

If $A \cap B = \emptyset$, i.e., $A$ and $B$ are mutually exclusive, and $P(A) > 0$ and $P(B) > 0$, then $A$ and $B$ cannot be independent.

The fact that $A$ and $B$ are spatially uncoupled in the Venn diagram does not mean independence, on the contrary there is strong dependence between $A$ and $B$

$$\text{because} \qquad P(A \cap B) = 0 < P(A) \cdot P(B)$$

or, knowing that $A$ occurred, leaves no chance for $B$ to occur (strong impact).

Thus $A$ and $A^c$ are not independent as long as $0 < P(A) < 1$.

# Implied Independence

If $A$ and $B$ are independent so are $A^c$ and $B$ and thus also $A^c$ and $B^c$.

Proof:

$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B) \cdot P(A) + P(B \cap A^c)$$

$$\implies P(B) \cdot (1 - P(A)) = P(B \cap A^c) \implies P(B \cap A^c) = P(B)P(A^c)$$

# Examples of Independence/Dependence

1. Given: $P(A) = 0.4$, $P(B) = 0.5$, and $P([A \cup B]^c) = 0.3$.

   Are $A$ and $B$ independent?

$$P(A \cup B) = 0.7 = P(A) + P(B) - P(A \cap B) = 0.4 + 0.5 - P(A \cap B)$$

$$\Longrightarrow\ P(A \cap B) = 0.2 = P(A) \cdot P(B)\ \Longrightarrow\ A \text{ and } B \text{ are independent!}$$

2. Given: $P(A \cap B^c) = 0.3$, $P(A^c \cap B) = 0.2$, and $P(A^c \cap B^c) = 0.1$.

   Are $A$ and $B$ independent?

$$0.1 = P(A^c \cap B^c) = P([A \cup B]^c) = 1 - P(A \cup B)\ \Longrightarrow\ P(A \cup B) = 0.9$$

$$0.9 = P(A \cup B) = P(A \cap B^c) + P(A^c \cap B) + P(A \cap B) = 0.3 + 0.2 + P(A \cap B)$$

$$\Longrightarrow\ P(A \cap B) = 0.4 \quad P(A) = 0.7 \quad P(B) = 0.6$$

and $P(A \cap B) = 0.4 \neq P(A) \cdot P(B) = 0.42$, i.e., $A$ and $B$ are dependent.

# Postulated Independence

In practical applications independence is usually based on our understanding of physical independence, i.e., $A$ relates to one aspect of an experiment while $B$ relates to another aspect that is physically independent from the former.

In such cases we postulate probability models which reflect this independence.

**Example:** First flip a penny, then spin it, with apparent physical independence. The sample space is $S = \{\mathtt{HH}, \mathtt{HT}, \mathtt{TH}, \mathtt{TT}\}$, with respective probabilities

$$p_1 \cdot p_2, \quad p_1 \cdot (1 - p_2), \quad (1 - p_1) \cdot p_2, \quad (1 - p_1) \cdot (1 - p_2)$$

$$\begin{aligned} \text{where} \quad P(\mathtt{H}\ \text{on flip}) &= P(\{\mathtt{HT}\} \cup \{\mathtt{HH}\}) = p_1 \cdot p_2 + p_1 \cdot (1 - p_2) = p_1 \\ \text{and} \quad P(\mathtt{H}\ \text{on spin}) &= P(\{\mathtt{TH}\} \cup \{\mathtt{HH}\}) = (1 - p_1) \cdot p_2 + p_1 \cdot p_2 = p_2 \end{aligned}$$

and $P(\{\mathtt{H}\ \text{on flip}\} \cap \{\mathtt{H}\ \text{on spin}\}) = P(\{\mathtt{HH}\}) = p_1 p_2 = P(\mathtt{H}\ \text{on flip}) \cdot P(\mathtt{H}\ \text{on spin})$

# Common Dependence Situations

1. Consider the population of undergraduates at William & Mary, from which a student is selected at random. Let $A$ be the event that the student is female, $B$ be the event that the student is heading for elementary education.

   Being told $P(A) \approx .6$ and $P(A|B) \geq .9 \quad \Longrightarrow \quad A$ and $B$ are not independent.

2. Select a person at random from a population of registered voters.

   Let $A$ be the event that the person belongs to a country club,

   $B$ be the event that the person is a Republican. We probably would expect

   $$P(B|A) \gg P(B)$$

   i.e., $A$ and $B$ are not independent.

# Mutual Independence of a Collection $\{A_\alpha\}$ of Events

A collection $\{A_\alpha\}$ of events is said to consist of mutually independent events

if for any finite choice of events $A_{\alpha_1},\ldots,A_{\alpha_k}$ in $\{A_\alpha\}$ we have

$$P(A_{\alpha_1}\cap\ldots\cap A_{\alpha_k}) = P(A_{\alpha_1})\cdot\ldots\cdot P(A_{\alpha_k})$$

For example, for 3 events $A$, $B$, $C$, we not only require

$$P(A\cap B) = P(A)\cdot P(B), \qquad P(A\cap C) = P(A)\cdot P(C), \qquad P(B\cap C) = P(B)\cdot P(C)$$

but also $\qquad P(A\cap B\cap C) = P(A)\cdot P(B)\cdot P(C) \qquad$ (2)

Pairwise independence does not necessarily imply (2).

Counterexample: Flip 2 fair coins. Let $A = \{\texttt{H}$ on $1^{\text{st}}$ flip$\}$, $B = \{\texttt{H}$ on $2^{\text{nd}}$ flip$\}$,

$C = \{$same result on both flips$\}$ with $\quad P(A) = P(B) = P(C) = \frac{1}{2}\quad$ and

$P(A\cap C) = P(\texttt{HH}) = \frac{1}{4}$, etc., but $\quad P(A\cap B\cap C) = P(\texttt{HH}) = \frac{1}{4} \neq \frac{1}{8}$.

$\longrightarrow$ text example on "independence" of 3 blood markers (O.J. Simpson trial).

# Random Variables

In many experiments the focus is on numbers assigned to the various outcomes.

Numbers $\longrightarrow$ arithmetic and common arena for understanding experimental results.

The simplest and nontrivial example is illustrated by a coin toss: $S = \{\mathtt{H}, \mathtt{T}\}$,

where we assign the number $1$ to the outcome $\{\mathtt{H}\}$ and $0$ to $\{\mathtt{T}\}$.

Such an assignment can be viewed as a function $X : S \longrightarrow R$

$$\boxed{\begin{array}{c}\mathtt{H}\\ \mathtt{T}\end{array}} \quad \xrightarrow{X} \quad \boxed{\begin{array}{c}1\\ 0\end{array}} \qquad \text{with} \quad X(\mathtt{H}) = 1 \ \text{ and } \ X(\mathtt{T}) = 0$$

Such a function is called a random variable. We use capital letters from the end of the alphabet to denote such random variables (r.v.'s), e.g., $U, V, W, X, Y, Z$.

Using the word "variable" to denote a function is somewhat unfortunate, but it is customary. It emphasizes the varying values that $X$ can take on as a result of the (random) experiment.

It seems that $X$ only relabels the experimental outcomes, but there is more.

# Random Variables for Two Coin Tosses

Toss a coin twice.

Assign the number of heads to each outcome in $S = \{\mathtt{HH}, \mathtt{HT}, \mathtt{TH}, \mathtt{TT}\}$. $\quad Y : S \rightarrow R$

$$\boxed{\begin{matrix} \mathtt{HH} & \mathtt{HT} \\ \mathtt{TH} & \mathtt{TT} \end{matrix}} \quad \xrightarrow{Y} \quad \boxed{\begin{matrix} 2 & 1 \\ 1 & 0 \end{matrix}}$$

with $Y(\mathtt{HH}) = 2,\ Y(\mathtt{HT}) = Y(\mathtt{TH}) = 1,$ and $Y(\mathtt{TT}) = 0$

We may also assign a pair of numbers $(X_1, X_2)$ to each of the outcomes as follows

$$X_1(\mathtt{HH}) = 1, \quad X_1(\mathtt{HT}) = 1, \quad X_1(\mathtt{TH}) = 0, \quad X_1(\mathtt{TT}) = 0$$

$$X_2(\mathtt{HH}) = 1, \quad X_2(\mathtt{HT}) = 0, \quad X_2(\mathtt{TH}) = 1, \quad X_2(\mathtt{TT}) = 0$$

$X_1 =$ # of heads on the first toss and $X_2 =$ # of heads on the second toss.

$X = (X_1, X_2)$ is called a random vector (of length 2).

We can express $Y$ also as $Y = X_1 + X_2 = g(X_1, X_2)$ with $g(x_1, x_2) = x_1 + x_2$.

# Borel Sets

A random variable $X$ induces a probability measure on a sigma field $\mathcal{B}$ of certain subsets of $R = (-\infty, \infty)$.

This sigma field, the Borel sets, is the smallest sigma field containing all intervals $(-\infty, y]$ for $y \in R$, i.e., it contains all sets that can be obtained by complementation, countable unions and intersections of such intervals, e.g., it contains intervals like

$$[a,b], \quad [a,b), \quad (a,b], \quad (a,b) \quad \text{for any } a,b \in R \qquad \text{(why?)}$$

It takes a lot gyrations to construct a set that is not a Borel set.

We won't see any in this course.

How do we assign probabilities to such Borel sets $B \in \mathcal{B}$?

Each r.v. $X$ induces its own probability measure on the Borel sets $B \in \mathcal{B}$.

# Induced Events and Probabilities

Suppose we have a r.v. $X : S \longrightarrow R$ with corresponding probability space $(S, \mathcal{C}, P)$.

For any Borel set $B \in \mathcal{B}$ we can determine the set $X^{-1}(B)$ of all outcomes in $S$ which get mapped into $B$, i.e.,

$$X^{-1}(B) = \{s \in S : X(s) \in B\}$$

How do we know that $X^{-1}(B) \in \mathcal{C}$ is an event? We don't.

Thus we require it in our definition of a random variable.

**Definition:** A function $X : S \longrightarrow R$ is a random variable if and only if the

induced event $\quad X^{-1}((-\infty, y]) \in \mathcal{C} \quad$ for any $y \in R$

and thus the induced probability $P_X$ (induced by $P$ and $X$)

$$P_X((-\infty, y]) = P(\{s \in S : X(s) \leq y\}) \qquad \text{exists for all } y \in R$$

# Cumulative Distribution Function (CDF)

A variety of ways of expressing the same probability (relaxed and fastidious):

$$\begin{aligned} P_X((-\infty, y]) &= P\left(X^{-1}((-\infty, y])\right) = P(\{s \in S : X(s) \in (-\infty, y]\}) \\ &= P(-\infty < X \leq y) = P(X \leq y) \text{ (most relaxed)} \end{aligned}$$

**Definition:** The cumulative distribution function (cdf) of a random variable $X$ is the function $F : R \longrightarrow [0, 1]$ defined by

$$F(y) = P(X \leq y)$$

# CDF for Single Coin Toss or Coin Spin

**Example** (coin toss $P(\mathtt{H}) = 0.5$):

Since $X$ takes only the values 1 and 0 for $\mathtt{H}$ and $\mathtt{T}$ we have

$$P(X \le y) = \begin{cases} 0.0 = P(\emptyset) & \text{for} \quad y < 0 \\ 0.5 = P(X = 0) = P(\mathtt{T}) & \text{for} \quad 0 \le y < 1 \\ 1.0 = P(X = 0 \cup X = 1) = P(\mathtt{T} \cup \mathtt{H}) & \text{for} \quad 1 \le y \end{cases}$$

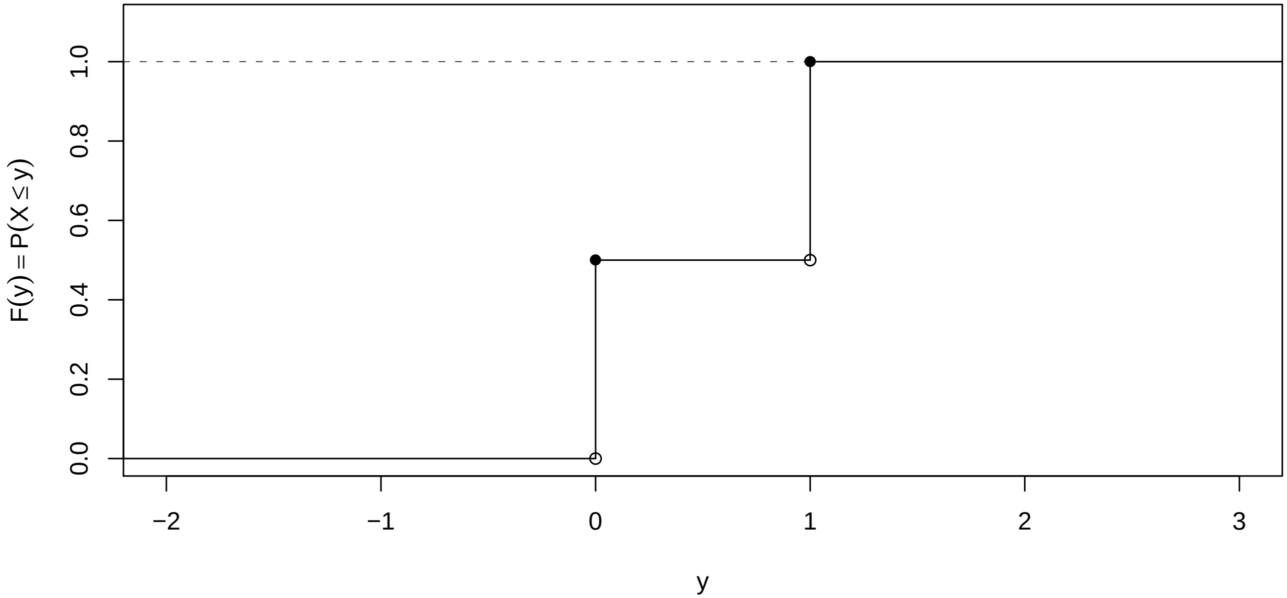
**Example** (coin spin $P(\mathtt{H}) = 0.3$):

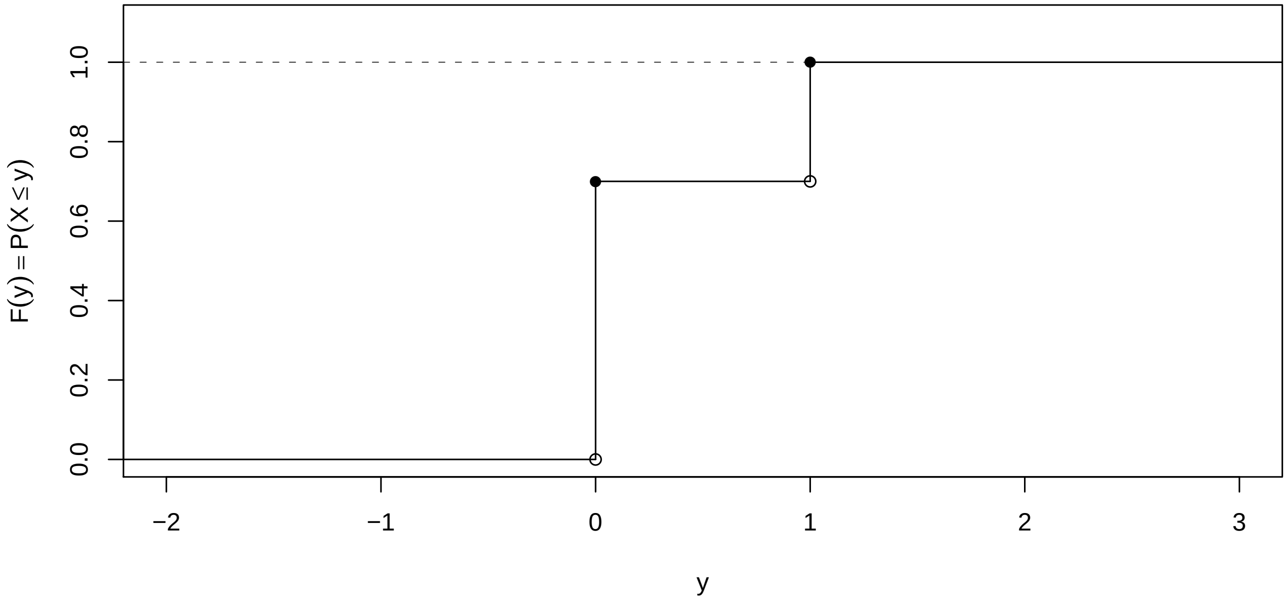
$$P(X \le y) = \begin{cases} 0.0 = P(\emptyset) & \text{for} \quad y < 0 \\ 0.7 = P(X = 0) = P(\mathtt{T}) & \text{for} \quad 0 \le y < 1 \\ 1.0 = P(X = 0 \cup X = 1) = P(\mathtt{T} \cup \mathtt{H}) & \text{for} \quad 1 \le y \end{cases}$$

The jump sizes at 0 and 1 represent $1 - P(\mathtt{H}) = P(\mathtt{T})$ and $P(\mathtt{H})$, respectively.

See CDF plots on next slide.

# CDF for Coin Toss/Coin Spin



# 2 Fair Coin Tosses

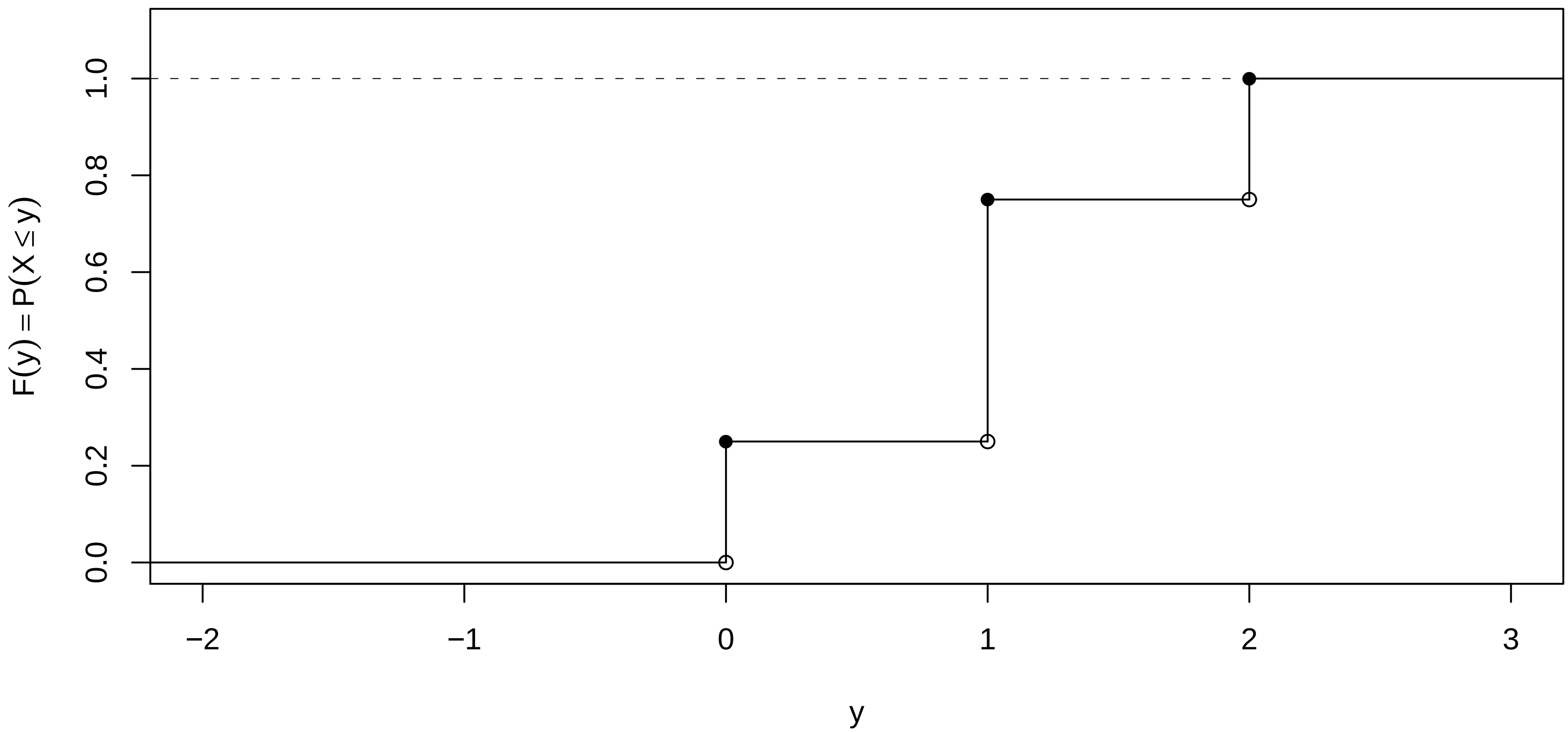
For two fair coin tosses the number $X$ of heads takes the values 0,1,2

for $s = \mathtt{TT}$, $s = \mathtt{HT}$ or $s = \mathtt{TH}$, and $s = \mathtt{HH}$ with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, respectively.

$$P(X \le y) = \begin{cases} 0.0 = P(\emptyset) & \text{for} \quad y < 0 \\ 0.25 = P(X = 0) = P(\mathtt{TT}) & \text{for} \quad 0 \le y < 1 \\ 0.75 = P(X = 0 \cup X = 1) = P(\mathtt{TT} \cup \mathtt{HT} \cup \mathtt{TH}) & \text{for} \quad 1 \le y < 2 \\ 1.0 = P(X = 0 \cup X = 1 \cup X = 2) = P(S) & \text{for} \quad 2 \le y \end{cases}$$

See CDF plot on next slide.

# CDF for 2 Fair Coin Tosses



# General CDF Properties

1. $0 \leq F(y) \leq 1$ for all $y \in R$ $\quad (F(y)$ is a probability)

2. $y_1 \leq y_2 \implies F(y_1) \leq F(y_2)$ (monotonicity property)

   This follows since $\{X \leq y_1\} \subset \{X \leq y_2\} \implies P(X \leq y_1) \leq P(X \leq y_2)$.

3. Limiting behavior as we approach $\pm\infty$

$$\lim_{y \to -\infty} F(y) = 0 \qquad \text{and} \qquad \lim_{y \to \infty} F(y) = 1$$

   This follows (with some more attention to technical detail) since

$$\lim_{y \to -\infty} \{X \leq y\} = \bigcap_{y \to -\infty} \{X \leq y\} = \emptyset \quad \text{and} \quad \lim_{y \to \infty} \{X \leq y\} = \bigcup_{y \to \infty} \{X \leq y\} = S$$

   Note that in our examples we had $F(y) = 0$ for sufficiently low $y$ ($y < 0$) and $F(y) = 1$ for sufficiently high $y$. $X$ had a finite and thus bounded value set.

# Two Independent Random Variables

Two random variable $X_1$ and $X_2$ are independent if any events defined in terms of $X_1$ is independent of any event defined in terms of $X_2$.

The following weaker but more practical definition is equivalent.

**Definition:** Let $X_1 : S \longrightarrow R$ and $X_2 : S \longrightarrow R$ be random variables defined on the same sample space $S$. $X_1$ and $X_2$ are independent if and only if for each $y_1 \in R$ and $y_2 \in R$

$$P(X_1 \leq y_1, X_2 \leq y_2) = P(X_1 \leq y_1) \cdot P(X_2 \leq y_2)$$

Note the shorthand notation

$$P(X_1 \leq y_1, X_2 \leq y_2) = P(\{X_1 \leq y_1\} \cap \{X_2 \leq y_2\})$$

You also often see $P(AB)$ for $P(A \cap B)$.

# Independent Random Variables

A collection of random variable $\{X_\alpha\}$ is mutually independent if the above product property holds for any finite subsets of these random variables, i.e., for any integer $k \geq 2$ and finite index subset $\alpha_1, \ldots, \alpha_k$ we have for all $y_1, \ldots, y_k \in R$

$$P(X_{\alpha_1} \leq y_1, \ldots, X_{\alpha_k} \leq y_k) = P(X_{\alpha_1} \leq y_1) \cdot \ldots \cdot P(X_{\alpha_k} \leq y_k)$$

Whether the independence assumption is appropriate in a given application is mainly a matter of judgment or common sense.

With independence we have access to many powerful and useful theorems in probability and statistics.