(Probabilistic) Experiment: $(\Omega, \mathcal{B}, P)$

Ω is the sample space $\equiv$ set of all possible outcomes.

(Often denoted S.)

ω denotes a particular outcome.

$\Omega = \{$ all possible ω $\}$

$\mathcal{B}$ is the class of "events" for which probabilities are defined.

(We mainly ignore $\mathcal{B}$ in this class. Assume all events of interest have well-defined probabilities.)

P is a "Probability function".

$P(A) =$ probability of the event A.

An event A is a subset of Ω.

Experiments and events are often depicted by Venn diagrams.

Example: Roll Two Fair Dice

$\Omega = \{(i, j) : 1 \leq i \leq 6, 1 \leq j \leq 6\}$

$\#(\Omega) = 36$

$\omega = (i, j)$

Example: Poker (5 card draw)

$\Omega =$ set of all poker hands

$\#(\Omega) = \binom{52}{5} = \frac{52!}{5!47!}$

a particular outcome is $\omega = \{A\heartsuit, 5\clubsuit, 5\spadesuit, K\heartsuit, 3\diamondsuit\}$

These are examples of experiments which are

discrete,
have finite Ω,
have equally likely outcomes ω.

In these situations:

$$P(A) = \frac{\#(A)}{\#(\Omega)}$$

Example: Toss a biased coin with $P(\text{Heads}) = 2/3$ three times.

$\Omega = \{$HHH, HHT, HTH, . . . , TTH, TTT$\}$

$\#(\Omega) = 8$

For $\omega =$ HTH , $P(\omega) = (2/3) \times (1/3) \times (2/3)$, etc.

This experiment is discrete, has finite Ω, has outcomes which are *not* equally likely.

Example: Turn on a Geiger counter for one minute and count the number of clicks. (Assume an average of λ clicks per minute.)

$\Omega = \{0, 1, 2, 3, \ldots\}$

A typical outcome might be $\omega = 3$.

$P(\omega)$ is given by Poisson distribution:

$$P(\omega) = \frac{\lambda^\omega e^{-\lambda}}{\omega!}$$

This experiment is discrete, has infinite (but countable) Ω, has outcomes which are *not* equally likely.

In these situations:

$$P(A) = \sum_{\omega \in A} P(\omega)$$

Example: Turn on a Geiger counter. Measure the length of time until you hear the first click. (Assume an average of λ clicks per minute.)

$\Omega = (0, \infty)$

$\#(\Omega) = \infty$ (and even worse, Ω is uncountable.)

For all outcomes ω, $P(\omega) = 0$.

This is an example of a continuous experiment where P is described in terms of a density function (pdf).

The time has an exponential distribution and

$$P([a,b]) = \int_a^b \lambda e^{-\lambda x}\, dx\,.$$

$$P(A) = \int_A \lambda e^{-\lambda x}\, dx\,.$$

Example: Toss a biased coin with $P(\text{Heads}) = 2/3$ *infinitely* many times. Record the sequence of heads and tails.

$\Omega = \{$ all possible sequences of H and T $\}$.

A typical $\omega = (H, H, T, H, H, H, T, T, \ldots)$

$\#(\Omega) = \infty$.

$P(\omega) = 0$ for all ω.

The experiment has an infinite (and uncountable) Ω.

Is this experiment discrete or continuous?

How to compute probabilities $P(A)$?

Example: Toss a dart at a square target (1 ft. by 1 ft.). Dart is tossed "at random" (uniformly).

$\Omega = \{(x, y) : 0 \le x \le 1,\, 0 \le y \le 1\}$

$\#(\Omega) = \infty$.

This is continuous experiment with P given by

$$P(A) = \frac{\text{Area}(A)}{\text{Area}(\Omega)}\,.$$

Example: Now suppose the dart is tossed according to a joint density $f(x, y)$ on the plane. Then (by definition)

$$P(A) = \int\int_A f(x, y)\,dx\,dy\,.$$

Comment: More complicated experiments lead to higher-dimensional sample spaces Ω and probability functions P described by higher-dimensional joint density or mass functions.

Properties of a Probability Function P

For any experiment (Ω, P):

$P(\Omega) = 1 \quad \star$

$P(\emptyset) = 0$

$0 \overset{\star}{\leq} P(A) \leq 1$

$P(A^c) = 1 - P(A) \qquad (\text{where } A^c = \Omega - A)$

$A \cap B = \emptyset \;\Rightarrow\; P(A \cup B) = P(A) + P(B)$

$$A_1, A_2, A_3, \ldots \text{ disjoint } \;\Rightarrow\; P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \star$$

$A \subset B \;\Rightarrow\; P(A) \leq P(B)$

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

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$$

etc.

[$\star$ denotes an axiom.]

[Can change ∞ to finite n above.]

Further comments on:

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

(2) $P\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} P(A_i)$

[There are proofs of both in text on pages 10–12.]

Proof of (2):

For $k = 2$, (2) becomes $P(A \cup B) \leq P(A) + P(B)$.

This follows immediately from (1) since $P(A \cap B) \geq 0$.

For $k = 3$, (2) is $P(A \cup B \cup C) \leq P(A) + P(B) + P(C)$.

This follows immediately from the result for $k = 2$:

$$\begin{aligned} P(A \cup B \cup C) &= P((A \cup B) \cup C) \\ &\leq P(A \cup B) + P(C) \\ &\leq P(A) + P(B) + P(C) \end{aligned}$$

Similarly,

$$\begin{aligned} P(A \cup B \cup C \cup D) &= P((A \cup B \cup C) \cup D) \\ &\leq P(A \cup B \cup C) + P(D) \end{aligned}$$

etc. (Use induction for a formal proof.)

Application of $P(\cup_i A_i) \le \sum_i P(A_i)$: Dunn's Multiple Comparison Procedure

Suppose a researcher (Ed) wishes to design an experiment to compare k treatments with a control (placebo).

(Take $k = 5$ for simplicity.)

After conducting the experiment, Ed will draw conclusions about the effectiveness of the treatments.

Suppose that **none** of the treatments are effective; they are all equivalent to the control. (Of course, Ed doesn't know this.)

Let $A_i = \{$Ed (falsely) claims treatment i is better than the control$\}$.

Define $B = A_1 \cup A_2 \cup \cdots \cup A_5 = \{$Ed (falsely) claims at least one of the treatments is better than the control$\}$.

B is the event that Ed makes an error. Suppose Ed wishes the probability of an error to be at most .05. How can he accomplish this?

One answer: If Ed designs his experiment so that $P(A_i) = .01$ for all i, then

$$P(B) = P(A_1 \cup A_2 \cup \cdots \cup A_5) \le \sum_{i=1}^{5} P(A_i) = 5 \times .01 = .05\,.$$

Property (1) : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
is the simplest case of the

Principle of Inclusion-Exclusion.

The next case is:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

The general case is:

$$\begin{aligned} P\left(\bigcup_{i=1}^{k} A_i\right) &= \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) \\ &\quad - \cdots + (-1)^{k-1} P(A_1 \cap A_2 \cap \cdots \cap A_k) \end{aligned}$$

There is a "picture proof" of the case with $k = 3$ where you keep track of how many times each region in the Venn diagram gets counted. (Do it!)

A rigorous formal argument can be given using the properties of probability we have covered.

What follows is a proof for $k = 3$ sets. The proof uses the property for $k = 2$ sets (which is property (1) above).

Proof of Principle of Inclusion-Exclusion for 3 sets

$$
\begin{aligned}
P(A \cup B \cup C) &= P((A \cup B) \cup C) \\
&\quad \text{Apply the case } k = 2. \\
&= P(A \cup B) + P(C) - P((A \cup B) \cap C) \quad (\ddagger)
\end{aligned}
$$

Now note that

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

and (using the distributive law for sets)

$$
\begin{aligned}
& P((A \cup B) \cap C) = P((A \cap C) \cup (B \cap C)) \\
& \quad \text{Apply the case } k = 2. \\
& = P(A \cap C) + P(B \cap C) - P((A \cap C) \cap (B \cap C)) \\
& \quad \text{Apply the associative and commutative laws} \\
& \quad \text{for } \cap \text{ to the event in the last term.} \\
& = P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)
\end{aligned}
$$

Plugging these facts back into ($\ddagger$) gives the final result

$$
\begin{aligned}
P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\
&\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\
&\quad + P(A \cap B \cap C)
\end{aligned}
$$

The proof for $k = 4$ sets is similar. Use induction to prove the general case.

“Applications” of the Principle of Inclusion-Exclusion (and more basic properties of probability)

Suppose a monkey types 5 letters “at random”. (The key strokes are independent with each letter having equal probability $= 1/26$. This is equivalent to saying that all 26^5 possibilities are equally likely.)

(#1) $P(\text{monkey types HELLO}) = \left(\frac{1}{26}\right)^5$ Why?

Solution:

$$\{\text{monkey types HELLO}\} = A_1 \cap A_2 \cap \cdots \cap A_5 \quad \text{where}$$

$$\begin{array}{lllll}
A_1 & = & \{\text{first letter is H}\} & = & \{\ell_1 = \text{H}\} \\
A_2 & = & \{\text{second letter is E}\} & = & \{\ell_2 = \text{E}\} \\
\vdots & & \vdots & & \vdots \\
A_5 & = & \{\text{fifth letter is O}\} & = & \{\ell_5 = \text{O}\}.
\end{array}$$

Since $A_1, A_2, \ldots, A_5$ are independent, we have

$$P(A_1 \cap \cdots \cap A_5) = P(A_1) \times \cdots \times P(A_5) = \left(\frac{1}{26}\right)^5.$$

(#2) $\quad P(\text{monkey types BURP}) = 2\left(\frac{1}{26}\right)^4 \quad$ Why?

Solution:

$\{\text{monkey types BURP}\} = \{\text{BURP?}\} \cup \{\text{?BURP}\} = B_1 \cup B_2\,.$
Here "?" stands for any letter whatsoever.
This means

$$\begin{aligned} B_1 &= \{\ell_1 = \mathsf{B}\} \cap \{\ell_2 = \mathsf{U}\} \cap \{\ell_3 = \mathsf{R}\} \cap \{\ell_4 = \mathsf{P}\} \\ &= \{\ell_1 = \mathsf{B}, \ell_2 = \mathsf{U}, \ell_3 = \mathsf{R}, \ell_4 = \mathsf{P}\}\,, \\ B_2 &= \{\ell_2 = \mathsf{B}, \ell_3 = \mathsf{U}, \ell_4 = \mathsf{R}, \ell_5 = \mathsf{P}\} \end{aligned}$$

Using independence as in Example #1, we see that

$$P(B_1) = P(B_2) = \left(\frac{1}{26}\right)^4 .$$

Clearly, B_1 and B_2 are disjoint (mutually exclusive). Thus

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) = 2\left(\frac{1}{26}\right)^4 .$$

(#3) $\quad P(\text{monkey types ZIT}) = 3\left(\frac{1}{26}\right)^3 \quad$ Why?

Solution: This is just like the previous example.

$$\begin{aligned} \{\text{monkey types ZIT}\} &= \{\text{ZIT??}\} \cup \{\text{?ZIT?}\} \cup \{\text{??ZIT}\} \\ &= C_1 \cup C_2 \cup C_3\,. \end{aligned}$$

Clearly $P(C_1) = P(C_2) = P(C_3) = (1/26)^3$ and the events are disjoint. Thus

$$P(C_1 \cup C_2 \cup C_3) = P(C_1) + P(C_2) + P(C_3) = 3\left(\frac{1}{26}\right)^3 .$$

(#4) P(monkey types AAAA) $= 2(1/26)^4 - (1/26)^5$.

Solution:

$$\begin{aligned}\{\text{monkey types AAAA}\} &= \{\text{AAAA?}\} \cup \{\text{?AAAA}\}\\ &= D_1 \cup D_2\,.\end{aligned}$$

D_1 and D_2 are **not** disjoint: $D_1 \cap D_2 = \{\text{AAAAA}\}$. Thus

$$\begin{aligned}P(D_1 \cup D_2) &= P(D_1) + P(D_2) - P(D_1 \cap D_2)\\ &= (1/26)^4 + (1/26)^4 - (1/26)^5\,.\end{aligned}$$

(#5) P(monkey types AAA) $= 3(1/26)^3 - 2(1/26)^4$.

Solution:

$$\begin{aligned}\{\text{monkey types AAA}\} &= \{\text{AAA??}\} \cup \{\text{?AAA?}\} \cup \{\text{??AAA}\}\\ &= E_1 \cup E_2 \cup E_3\,.\end{aligned}$$

Since

$$\begin{aligned}E_1 \cap E_2 &= \{\text{AAAA?}\}\\ E_2 \cap E_3 &= \{\text{?AAAA}\}\\ E_1 \cap E_3 &= \{\text{AAAAA}\}\\ E_1 \cap E_2 \cap E_3 &= \{\text{AAAAA}\}\end{aligned}$$

we have

$$\begin{aligned}P(E_1 \cup E_2 \cup E_3) &= P(E_1) + P(E_2) + P(E_3)\\ &\quad - P(E_1 \cap E_2) - P(E_2 \cap E_3) - P(E_1 \cap E_3)\\ &\quad + P(E_1 \cap E_2 \cap E_3)\\ &= (1/26)^3 + (1/26)^3 + (1/26)^3\\ &\quad - (1/26)^4 - (1/26)^4 - (1/26)^5\\ &\quad + (1/26)^5\,.\end{aligned}$$

(#6) $P(\text{monkey types AA}) = 4\left(\frac{1}{26}\right)^2 - 3\left(\frac{1}{26}\right)^3 - \left(\frac{1}{26}\right)^4 + \left(\frac{1}{26}\right)^5$.

Solution:

$$\begin{aligned}
&\{\text{monkey types AA}\} \\
&\quad = \{\text{AA???}\} \cup \{\text{?AA??}\} \cup \{\text{??AA?}\} \cup \{\text{???AA}\} \\
&\quad = F_1 \cup F_2 \cup F_3 \cup F_4
\end{aligned}$$

and

$$P(F_1 \cup F_2 \cup F_3 \cup F_4) = \sum_i P(F_i) - \sum_{i<j} P(F_i \cap F_j) + \sum_{i<j<k} P(F_i \cap F_j \cap F_k) - P(F_1 \cap F_2 \cap F_3 \cap F_4)\,.$$

To calculate this, you must find all the intersections and their probabilities. For example,

$$F_1 \cap F_4 = \{\text{AA?AA}\} \quad \text{so that} \quad P(F_1 \cap F_4) = (1/26)^4\,.$$

(#7) $P(\text{monkey types A}) = 1 - (25/26)^5$.

Solution: It is possible (but very tedious) to do this by inclusion-exclusion using

$$\{\text{monkey types A}\} = \{\text{A????}\} \cup \{\text{?A???}\} \cup \cdots \cup \{\text{????A}\}\,.$$

But much better is to switch to the complement:

$$P(\text{monkey types A}) = 1 - P(\text{monkey does not type A})\,.$$

$$\{\text{does not type A}\} = \{\ell_1 \neq \text{A}\} \cap \{\ell_2 \neq \text{A}\} \cap \cdots \cap \{\ell_5 \neq \text{A}\}\,.$$

Now use the independence of the key strokes.

Coming up: More examples with equally likely outcomes.

But first . . .

Fundamental Theorem of Counting (FTC)

Suppose you want to calculate the number of **different** ways that some "task" can be performed.

If

1. The task can be broken down into k steps, and

2. Step i can be done in n_i different ways regardless of how the previous steps have been performed, and

3. Each different way of performing the steps gives a different way of performing the task,

Then the task can be done in $n_1 \times n_2 \times \cdots \times n_k$ different ways.

Poker Problems

Draw 5 cards from a well-shuffled deck of 52.

Problem #1: What is P(full house)?

> A "full house" means (Three of a kind) + (Pair).
> For example $\{$**K**$\heartsuit$**, K**$\spadesuit$**, K**$\diamondsuit$**, 5**$\clubsuit$**, 5**$\heartsuit\}$.

Let $A = \{$draw a full house$\}$.

$$P(A) = \frac{\#(A)}{\#(\Omega)} \text{ where } \#(\Omega) = \binom{52}{5} = 2{,}598{,}960\,.$$

Use FTC to calculate $\#(A)$.

Task:	Construct Full House	# of ways
Step 1:	Choose value to be repeated 3 times.	$\binom{13}{1} = 13$
Step 2:	Choose 3 cards of that value.	$\binom{4}{3} = 4$
Step 3:	Choose value for pair.	$\binom{12}{1} = 12$
Step 4:	Choose 2 cards of that value.	$\binom{4}{2} = 6$

Thus $\#(A) = 13 \times 4 \times 12 \times 6 = 3744$.

So $P(A) = \dfrac{3744}{2598960} = .00144\,.$

Problem #2: What is P(5 of a kind) with 2's wild?

Solution: Let $A = \{$draw 5 of a kind$\}$ with 2's wild.

"5 of a kind" can arise in four ways:

K K K K 2
K K K 2 2
K K 2 2 2
K 2 2 2 2

Define $A_i = \{$5 of a kind with i wild cards$\}$.

These events are **disjoint** and $A = A_1 \cup A_2 \cup A_3 \cup A_4$.

Thus

$$
\begin{aligned}
P(A) &= P(A_1) + P(A_2) + P(A_3) + P(A_4) \quad \text{(since disjoint)} \\
&= \frac{\#(A_1) + \#(A_2) + \#(A_3) + \#(A_4)}{\#(\Omega)} \quad \text{(since outcomes are equally likely)}
\end{aligned}
$$

Now use FTC to find $\#(A_i)$.

Task:	Construct 5 of a kind with i wild cards.	# of ways
Step 1:	Choose i wild cards.	$\binom{4}{i}$
Step 2:	Choose other value to be repeated.	$\binom{12}{1} = 12$
Step 3:	Choose $5 - i$ cards of that value.	$\binom{4}{5-i}$

Thus $\#(A_i) = 12\binom{4}{i}\binom{4}{5-i}$.

$$\begin{aligned} P(A) &= \frac{\displaystyle\sum_{i=1}^{4} 12\binom{4}{i}\binom{4}{5-i}}{2598960} = \frac{12\,(4\cdot 1 + 6\cdot 4 + 4\cdot 6 + 1\cdot 4)}{2598960} \\ &= \frac{12\cdot 56}{2598960} \approx .0002586 \end{aligned}$$

<u>Question:</u> Can we find $P(A)$ directly using the FTC (without first breaking down A into the four cases A_1, A_2, A_3, A_4) by inserting

Step 0: Choose i (the # of wild cards) between 1 and 4.

before Steps 1–3?

No!

Because then the number of ways you can do Steps 1 and 3 depends on the choice made in Step 0.

Comment: Condition 3 of the FTC sometimes requires careful thought. The division of the task into steps must be such that any difference in the way the steps are performed must lead to a different outcome (a different way of performing the task).

Example: (A **Wrong** Solution of Exercise 1.20)

I have 12 friends. Each will call once on a randomly chosen day of the coming week (7 days). What is the probability that I get at least one call each day?

Solution: There are 7^{12} different ways that my friends could call. Let A be the event that I get at least one call each day. Then $P(A) = \#(A)/7^{12}$.

$\#(A)$ = the number of ways of assigning the 12 friends to the 7 days of the week with at least one friend each day.

Break down the task of assigning the friends to the days into the following steps:

1. Choose 7 friends. $\binom{12}{7}$ ways
2. Assign each of the 7 to a different day. $7!$ ways
3. Assign the remaining 5 friends to days (without any restrictions). 7^5 ways

Therefore $\#(A) = \binom{12}{7} 7!\, 7^5$.

What is wrong with this solution?

Refer to the friends as A, B, . . . , K, L.

Consider these two different ways to perform the tasks:

1. Choose friends A, B, C, D, E, F, G.
2. Assign A to day 1, B to day 2, . . . , G to day 7.
3. Assign H to day 1, I to day 2, . . . , L to day 5.

1. Choose friends H, I, J, K, L, F, G.
2. Assign H to day 1, I to day 2, . . . , L to day 5.
   Then assign F to day 6 and G to day 7.
3. Assign A to day 1, B to day 2, . . . , E to day 5.

They both lead to the **same** assignment of friends to days!!

So, this way of breaking down the task into steps is bogus.

Conditional Probability

Define $P(A \,|\, B) = \dfrac{P(A \cap B)}{P(B)}$ (so long as $P(B) > 0$).

If the sample space Ω consists of equally likely outcomes, this becomes

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

Immediate Consequences of Definition

$$P(A \cap B) = P(B)\, P(A \,|\, B) = P(A)\, P(B \,|\, A)$$

$$\begin{aligned} P(A \cap B \cap C) &= P(A)\, P(B \,|\, A)\, P(C \,|\, A \cap B) \\ &\ \text{since the RHS can be written as} \\ &= P(A)\, \frac{P(A \cap B)}{P(A)}\, \frac{P(A \cap B \cap C)}{P(A \cap B)} \\ &\ \text{and we can cancel terms.} \end{aligned}$$

And similarly

$P(A \cap B \cap C \cap D) = P(A)\, P(B \,|\, A)\, P(C \,|\, A \cap B)\, P(D \,|\, A \cap B \cap C)$,

etc.

Urn Model: An urn contains given numbers of colored balls. At each draw from the urn, all the balls in the urn are equally likely to be drawn.

Different types of urn models

Sampling with replacement: After a ball is drawn, it is put back in the urn.

Sampling withOUT replacement: After a ball is drawn, it is NOT put back.

A Pólya Urn: After drawing a ball, you replace it and add another ball of the same color.

A Simple Example

An urn contains 7 red balls and 3 white balls. Draw 3 balls (in sequence) withOUT replacement. What is the probability all 3 balls are red?

Let $A_i = \{i\text{-th ball is red}\}$.

$$\begin{aligned} P(\text{all 3 are red}) &= P(A_1 \cap A_2 \cap A_3) \\ &= P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \\ &= \frac{7}{10} \cdot \frac{6}{9} \cdot \frac{5}{8} \end{aligned}$$

The same problem with a Pólya Urn:

$$\begin{aligned} P(\text{all 3 are red}) &= P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \\ &= \frac{7}{10} \cdot \frac{8}{11} \cdot \frac{9}{12} \end{aligned}$$

Return to Poker

You can view the cards in the Poker hand as being ordered or **un**ordered.

If your outcomes ω are all the **un**ordered Poker hands (order does not matter), then $\#(\Omega) = \binom{52}{5}$.

If your outcomes ω are all the ordered Poker hands (order does matter), then $\#(\Omega) = 52 \cdot 51 \cdot 50 \cdot 49 \cdot 48$.

In both points of view the outcomes are equally likely.

Many problems can be solved using either approach. Both methods are correct and will lead to the same answer **if they are consistently followed**.

With the **ordered** point of view, we can think of the deck as an urn with 52 balls. Dealing a hand is drawing 5 balls (in sequence) withOUT replacement from the urn .

Finding the Probability of a Full House Viewing the Deck as an Urn Model

$P(\text{full house}) = ?$

A full house can occur in 10 different orders:

$$\{\text{full house}\} = \{XXXYY\} \cup \{XXYXY\} \cup \{XYXXY\} \cup \cdots \cup \{YYXXX\}$$

$$= B_1 \cup B_2 \cup \cdots \cup B_{10}$$

where $B_1 = \{XXXYY\}$, etc.

Note: $\binom{5}{3} = 10$ different choices of 3 positions for the X's.

These events are disjoint so that

$$P(\text{full house}) = P(B_1) + \cdots + P(B_{10}).$$

$B_1 = C_1 \cap C_2 \cap C_3 \cap C_4$ where

$C_1 = \{\text{second card same as first}\}$, $\rightarrow X$

$C_2 = \{\text{third card same as first}\}$,

$C_3 = \{\text{fourth card different from first}\}$,

$C_4 = \{\text{fifth card same as fourth}\}$.

Y

Thus

$$P(B_1) = P(C_1)\,P(C_2 | C_1)\,P(C_3 | C_1 \cap C_2)\,P(C_4 | C_1 \cap C_2 \cap C_3)$$

$$= \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{48}{49} \cdot \frac{3}{48}$$

$$= .000144058$$

Can similarly evaluate $P(B_2), \ldots, P(B_{10})$.
You find $P(B_1) = P(B_2) = \cdots = P(B_{10})$.

Thus

$$P(\text{full house}) = 10\, P(B_1)$$

$$= .00144058$$

Definition: Events $B_1, B_2, \ldots, B_k$ form a **partition** (of Ω) if $B_1, B_2, \ldots, B_k$ are disjoint and $\cup_{i=1}^k B_i = \Omega$.

(Note: $k = \infty$ is allowed here.)

If $B_1, B_2, \ldots, B_k$ is a partition (of Ω), then for any event A we have . . .

(1) Law of Total Probability:

$$P(A) = \sum_{j=1}^{k} P(B_j)P(A|B_j)$$

(2) Bayes Rule: For all i,

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{P(A)} = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^k P(B_j)P(A|B_j)}$$

Proof of (1):

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_k)$$

(Draw a picture.)

The events in this union are disjoint. Thus

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k)$$

Now (1) follows since $P(A \cap B_i) = P(B_i)P(A|B_i)$ for all i.

Proof of (2): By the definition of conditional probability

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

Now substitute $P(A \cap B_i) = P(B_i)P(A|B_i)$ and $P(A) = \sum_{j=1}^k P(B_j)P(A|B_j)$ and we are done.

Examples for Law of Total Probability and Bayes Theorem

Situation: 4 coins in a hat.

Coins are numbered 1, 2, 3, 4.

Tosses of coin i are independent with probability π_i of heads.

i	π_i
1	.2
2	.4
3	.6
4	.8

Choose a coin $\underbrace{\text{"at random"}}_{\text{equally likely}}$.

Toss it 6 times. (The <u>same</u> coin is tossed 6 times.)

Observe the results.

We will consider various problems.

Preliminaries

$$\Omega = \{ (\underset{\substack{1 \le i \le 4 \\ \text{coin number selected}}}{i}, \underbrace{r_1, r_2, \ldots, r_6}_{r_i \in \{H, T\}}) \}$$

Typical outcome $\omega = (3, H, H, T, T, H, H)$.

$\#(\Omega) = 4 \times 2^6$, but outcomes are not equally likely.

Problem 1:

What is $P\{\underbrace{\text{observe 4 heads}}_{\text{call this } A}\}$?

Solution: Use Law of Total Probability, conditioning on the chosen coin.

Define $B_i = \{\text{choose coin } i\}$.

B_1, B_2, B_3, B_4 form a partition.

Therefore

$$P(A) = \sum_{i=1}^{4} P(B_i)\, P(A|B_i)$$

$\frac{1}{4}$

coin is chosen at random

$\binom{6}{4} \pi_i^4 (1-\pi_i)^2$

Let X = # of heads in 6 tosses of a coin with prob. π_i of heads.

$X \sim \text{Binomial}(n=6, \pi_i)$

$P(A|B_i) = P(X=4)$

$$= \frac{15}{4} \sum_{i=1}^{4} \pi_i^4 (1-\pi_i)^2$$

$$= \frac{15}{4}\left[(.2)^4(.8)^2 + (.4)^4(.6)^2 + (.6)^4(.4)^2 + (.8)^4(.2)^2\right]$$

$$= \frac{15}{4}(.04736) = .1776$$

Problem 2: Suppose we observe 4 heads (in 6 tosses).

What is the probability we have chosen coin i?

Solution: Bayes Theorem

$$\underset{\substack{\swarrow \quad\quad \searrow \\ \text{chose coin } i \quad \text{observe 4 H's}}}{P(B_i \mid A)} = \frac{P(B_i)\,P(A \mid B_i)}{\underbrace{\sum_j P(B_j)\,P(A \mid B_j)}_{P(A) = .1776}}$$

$$= \frac{\frac{1}{4}\binom{6}{4}\pi_i^4(1-\pi_i)^2}{.1776}$$

We evaluate this for $i = 1, 2, 3, 4$.

$(\pi_i = .2, .4, .6, .8)$

The results are:

i	$P(B_i \mid A)$
1	.02162
2	.19459
3	.43784
4	.34595
Total	1.0000

Situation: Gambler's Ruin

A casino runs a game which consists of a sequence of coin tosses.

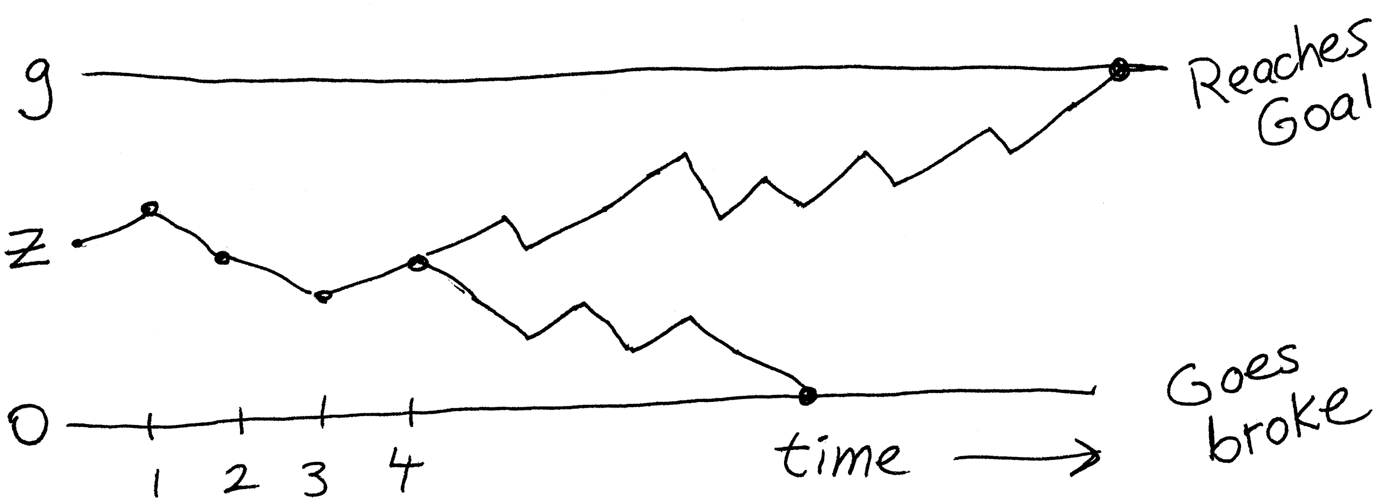
Toss	Gain	Prob	
H	+1	1/2	Not realistic!
T	−1	1/2	

A gambler starts with initial fortune z.

Goal: To increase total fortune to g.

(think of, say, $g = 100$)

The gambler decides to play until he reaches g or goes broke (hits zero).



$$\text{What is } P\underbrace{\left\{\begin{array}{l}\text{reaches goal of } g\\ \text{starting from}\\ \text{initial fortune } z\end{array}\right\}}_{\text{call this } A} \equiv \psi(z)\,?$$

Clearly $\psi(0) = 0$ and $\psi(g) = 1$.

We want $\psi(z)$ for $0 < z < g$.

<u>Solution</u>: Use Law of Total Probability, conditioning on the result of the first toss.

Define $B_1 = \{\text{first toss is } H\}$,

$B_2 = \{\text{first toss is } T\}$.

B_1, B_2 forms a partition.

Therefore

$$\underbrace{P(A)}_{\psi(z)} = \underbrace{P(B_1)}_{\frac{1}{2}}\,\underbrace{P(A|B_1)}_{\psi(z+1)} + \underbrace{P(B_2)}_{\frac{1}{2}}\,\underbrace{P(A|B_2)}_{\psi(z-1)}.$$

$$\psi(z) = \frac{1}{2}\left(\psi(z+1) + \psi(z-1)\right) \text{ for } 0 < z < g. \qquad (*)$$

This says $\psi(z)$ is obtained by linear interpolation from the values $\psi(z+1), \psi(z-1)$.

Since this is true for all z, the function ψ must be linear!

Fact: Any linear function $\psi(z) = az + b$ satisfies (*).

Proof:

$$\begin{aligned}
&\frac{1}{2}[\,\psi(z+1) + \psi(z-1)\,] \\
&= \frac{1}{2}\,[\,(a(z+1)+b) + (a(z-1)+b)\,] \\
&= \frac{1}{2}\,[\,az + a + b + az - a + b\,] \;=\; az + b \\
&= \psi(z)
\end{aligned}$$

Fact: The only linear function $\psi(z) = az + b$ which satisfies $\psi(0) = 0$ and $\psi(g) = 1$ is $\psi(z) = z/g$.

Proof:

$\psi(0) = a0 + b = 0$ implies $b = 0$.

$\psi(g) = ag + 0 = 1$ implies $a = 1/g$.

This almost proves that $\psi(z) = z/g$. But we have not yet shown that $\psi(z)$ **must** be a linear function.

A complete proof of $\psi(z) = z/g$ is given on the next page.

With a little algebra, we see that (*) is equivalent to

$$\psi(z+1)-\psi(z)=\psi(z)-\psi(z-1) \quad \text{for } 0<z<g. \qquad (\dagger)$$

By using (†) repeatedly we find

$$\psi(z+1)-\psi(z)=\psi(1)-\psi(0) \quad \text{for } 0<z<g.$$

Using this we see, for example, that

$$\begin{aligned}
&\psi(3)\\
&= (\psi(3)-\psi(2))+(\psi(2)-\psi(1))+(\psi(1)-\psi(0))+\psi(0)\\
&\qquad \text{since everything but } \psi(3) \text{ cancels}\\
&= (\psi(3)-\psi(2))+(\psi(2)-\psi(1))+(\psi(1)-\psi(0))\\
&\qquad \text{since } \psi(0)=0\\
&= 3(\psi(1)-\psi(0)) \qquad \text{by using } (\dagger)\\
&= 3\psi(1) \qquad \text{by again using } \psi(0)=0\,.
\end{aligned}$$

More generally

$$\begin{aligned}
\psi(z) &= (\psi(z)-\psi(z-1))+(\psi(z-1)-\psi(z-2))+\cdots\\
&\qquad +(\psi(2)-\psi(1))+(\psi(1)-\psi(0))+\psi(0)\\
&= z(\psi(1)-\psi(0)) \qquad \text{by using } \psi(0)=0 \text{ and } (\dagger)\\
&= z\psi(1) \qquad \text{by using } \psi(0)=0 \text{ again}\,.
\end{aligned}$$

Finally, setting $z=g$ in the above gives

$$1=\psi(g)=g\psi(1) \quad \Longrightarrow \quad \psi(1)=1/g$$

Therefore $\quad \psi(z)=\dfrac{z}{g} \quad$ for $0\le z\le g$.

Example: Applications of Bayes Rule to Gambler's Ruin

Suppose you start with an initial fortune z and your goal is to reach g.

What is the probability the first toss was heads, given that you reach the goal?

$$P(\text{first toss H} \,|\, \text{reach goal}) = P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(A)}$$

$$= \frac{\frac{1}{2}\cdot\psi(z+1)}{\psi(z)} = \frac{\frac{1}{2}\cdot(z+1)/g}{z/g} = \frac{1}{2}\cdot\frac{z+1}{z} \quad \left(> \frac{1}{2}\right).$$

Remarks: This does not depend on g! What about if $z = 1$?

What is the probability the first two tosses were tails, given that you reach the goal? (Assume $z \geq 2$.)

Let $C = \{\text{first two tosses are tails}\}$.

$$P(C|A) = \frac{P(C)P(A|C)}{P(A)} = \frac{\frac{1}{4}\cdot\psi(z-2)}{\psi(z)}$$

$$= \frac{\frac{1}{4}\cdot(z-2)/g}{z/g} = \frac{1}{4}\cdot\frac{z-2}{z} \quad \left(< \frac{1}{4}\right).$$

Remark: What if $z = 2$?

It is often useful to think of probability as the long-run fraction of times that an event occurs when an experiment is repeated very many times.

Computer simulations can be helpful in understanding and illustrating probabilities.

Example: The Gambler's Ruin problem with goal $g = 15$ and initial fortune $z = 5$.

According to our results:

$$P(\text{reach goal}) = \frac{z}{g} = \frac{5}{15} = \frac{1}{3} \approx 0.3333$$

$$P(\text{first toss H} \,|\, \text{reach goal}) = \frac{1}{2} \cdot \frac{z+1}{z} = \frac{1}{2} \cdot \frac{6}{5} = \frac{3}{5} = 0.6$$

$$P(\text{start with TT} \,|\, \text{reach goal}) = \frac{1}{4} \cdot \frac{z-2}{z} = \frac{1}{4} \cdot \frac{3}{5} = \frac{3}{20} = 0.15.$$

It is easy to simulate a gambler tossing a fair coin until he achieves the goal or goes broke.

I did this 1,000,000 times.

The gambler achieved the goal 333,688 times.

Among these 333,688 times (in which the goal was achieved):

- 200,139 times started with H.
- 50,182 times started with TT.

Note that: $\dfrac{333,688}{1,000,000} = .333688$ $\qquad \dfrac{200,139}{333,688} = .5997788$

$\dfrac{50,182}{333,688} = .150386$ $\qquad$ (see code in mordor)