(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

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

$\#(\Omega) = 36$

$\omega = (i,j)$

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.

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

$\#(\Omega) = 8$

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 probabilites $P(A)$?

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

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

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

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)

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

Example: Break a stick (of length 1) into 5 pieces in the following way:

Break the stick "at random". Call the lengths of the left and right hand pieces X_1 and Y_1 ($X_1 + Y_1 = 1$) respectively. Now break the right piece "at random". Call the lengths of the left and right pieces X_2 and Y_2 ($X_2 + Y_2 = Y_1$). Now break the right piece of this and then again the right piece of that. In the end the stick is broken into 5 pieces of lengths X_1, X_2, X_3, X_4, Y_4.

$\Omega = \{(x_1, x_2, x_3, x_4) : x_i > 0$ for all i, $\sum_{i=1}^{4} x_i < 1\}$

Probabilities are described by
a joint density $f(x_1, x_2, x_3, x_4)$. (What is it?)

$$P(A) = \int\int\int\int_A f(x_1, x_2, x_3, x_4)\, dx_1\, dx_2\, dx_3\, dx_4\,.$$

Properties of a Probability Function P

For any experiment (Ω, P):

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

$P(\emptyset) = 0$

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

$P(A^c) = 1 - P(A) \quad$ (where $A^c = \Omega - A$)

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

$A_1, A_2, A_3, \ldots$ 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 $P(\cup_i A_i) \le \Sigma_i P(A_i)$:
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) $\qquad P(\text{monkey types HELLO}) = \left(\frac{1}{26}\right)^5 \qquad$ Why?

Solution:

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

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

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

Solution:

$\{\text{monkey types BURP}\} = \{\text{BURP?}\} \cup \{\text{?BURP}\} = B_1 \cup B_2\,.$

Here "?" stands for any letter whatsoever.
This means

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

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) $P(\text{monkey types ZIT}) = 3\left(\frac{1}{26}\right)^3$ 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.