1.2 Sample Spaces

Probability theory is concerned with situations in which the outcomes occur randomly. Generically, such situations are called *experiments*, and the set of all possible outcomes is the **sample space** corresponding to an experiment. The sample space is denoted by Ω, and an element of Ω is denoted by ω. The following are some examples.

EXAMPLE **A** Driving to work, a commuter passes through a sequence of three intersections with traffic lights. At each light, she either stops, s, or continues, c. The sample space is the set of all possible outcomes:

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

where csc, for example, denotes the outcome that the commuter continues through the first light, stops at the second light, and continues through the third light. ■

EXAMPLE **B** The number of jobs in a print queue of a mainframe computer may be modeled as random. Here the sample space can be taken as

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

that is, all the nonnegative integers. In practice, there is probably an upper limit, N, on how large the print queue can be, so instead the sample space might be defined as

$$\Omega = \{0, 1, 2, \ldots, N\}$$ ■

EXAMPLE **C** Earthquakes exhibit very erratic behavior, which is sometimes modeled as random. For example, the length of time between successive earthquakes in a particular region that are greater in magnitude than a given threshold may be regarded as an experiment. Here Ω is the set of all nonnegative real numbers:

$$\Omega = \{t \mid t \geq 0\}$$ ■

We are often interested in particular subsets of Ω, which in probability language are called **events.** In Example A, the event that the commuter stops at the first light is the subset of Ω denoted by

$$A = \{sss, ssc, scc, scs\}$$

Events : Subsets of the sample space
measurable

Union (OR) : $A \cup B$ ← Why not "A and B"

Because "or" refers to the elements of A and B

"$x \in A \cup B$" if $x \in A$ OR $x \in B$ ... or both.

Intersection (AND) : $A \cap B$

Complement (NOT) A^c

Empty set : $\emptyset$

Disjoint : A & B disjoint if $A \cap B = \emptyset$

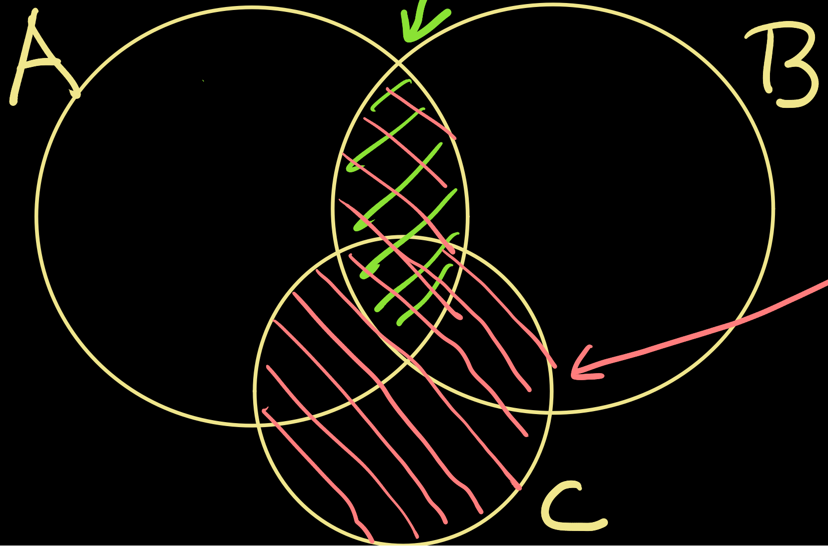
Commutative Laws: $A \cup B = B \cup A$

$A \cap B = B \cap A$

Associative Laws: $(A \cup B) \cup C = A \cup (B \cup C)$

Distributive Laws: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$

$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$



Definition of probability measure:

P is a function : Subsets of $\Omega \to [0,1]$

measurable

P is a probability measure if

1) $P(\Omega) = 1$

2) $A \subset \Omega$ then $P(A) \geq 0$

3a) If $A_1 \cap A_2 = \phi$ then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

3b) If $A_1, A_2, \ldots$ are mutually disjoint the

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

Definition probability measure:

P is a function : Subsets of $\Omega \to [0,1]$

measurable

P is a probability measure if

$\boxed{\text{Axioms of } \sigma\text{-additive probability}}$

"normed" 1) $P(\Omega) = 1$

"positive" 2) $A \subset \Omega$ then $P(A) \geq 0$

"additive" 3a) If $A_1 \cap A_2 = \emptyset$ then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$

"σ-additive" 3b) If $A_1, A_2, \ldots$ are mutually disjoint the

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

Can prove:

$+3-3+1$

$$P(A^c) = 1 - P(A)$$

$$P(\emptyset) = 0$$

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

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

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

$$P(A \cup B \cup C \cup D) = ?$$

Counting methods:

When Ω is finite and all elements have the same probability, then

$$P(A) = \frac{\#(A)}{\#(\Omega)} = \frac{\text{\# of ways } A \text{ can occur}}{\#(\Omega)}$$

EXAMPLE **B** *Simpson's Paradox*

A black urn contains 5 red and 6 green balls, and a white urn contains 3 red and 4 green balls. You are allowed to choose an urn and then choose a ball at random from the urn. If you choose a red ball, you get a prize. Which urn should you choose to draw from? If you draw from the black urn, the probability of choosing a red ball is $\frac{5}{11} = .455$ (the number of ways you can draw a red ball divided by the total number of outcomes). If you choose to draw from the white urn, the probability of choosing

... skipping for now — will be back!

SP has MUCH larger implications in the right context.

Multiplication Principle

If one experiment has m outcomes

and

for each outcome a second experiment has n outcomes,

then there are $m \times n$ outcomes altogether.

If experiment 1 has n_1 outcomes, and

for each, experiment 2 has n_2 outcomes, and

for each combination of prior expts,

experiment 3 has n_3 outcomes

...

for each combination of prior expts,

experiment p has n_p outcomes

Then there are $n_1 \times n_2 \times n_3 \times \cdots n_p$ outcomes altogether.

Permutations and Combinations

n distinct objects : $n!$ possible orderings.

Choosing r objects from n :

\# of ways : n^r with replacement

$n \times (n-1) \times \cdots (n-(r-1))$ without replacement

ambiguous

alters

$$= \frac{n!}{(n-r)!}$$

Note: "# of ways of choosing" implies order matters.

EXAMPLE **B** Suppose that from ten children, five are to be chosen and lined up. How many different lines are possible?

From Proposition A, there are $10 \times 9 \times 8 \times 7 \times 6 = 30{,}240$ different lines. ■

P. 10

$$\frac{10!}{5!}$$

BUT "how many 'sets' of children?

Then order does not matter.

There are $5!$ line-ups with the same set of children.

So # of sets is $\frac{10!}{5!} \Big/ 5! = \frac{10!}{5!\,5!}$

EXAMPLE E *Birthday Problem*

Suppose that a room contains n people. What is the probability that at least two of them have a common birthday?

This is a famous problem with a counterintuitive answer. Assume that every day of the year is equally likely to be a birthday, disregard leap years, and denote by A the event that at least two people have a common birthday. As is sometimes the case, finding $P(A^c)$ is easier than finding $P(A)$. This is because A can happen in many ways, whereas A^c is much simpler. There are 365^n possible outcomes, and A^c can happen in $365 \times 364 \times \cdots \times (365 - n + 1)$ ways. Thus,

$$P(A^c) = \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

and

$$P(A) = 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n}$$

The following table exhibits the latter probabilities for various values of n:

n	$P(A)$
4	.016
16	.284
23	.507
32	.753
40	.891
56	.988

From the table, we see that if there are only 23 people, the probability of at least one match exceeds .5. The probabilities in the table are larger than one might intuitively guess, showing that the coincidence is not unlikely. Try it in your class. ■

See R script:
BirthdayParadox.R

Combinations

How many **subsets** instead of **sequences**.

Choosing r from n:

$\frac{n!}{(n-r)!}$ sequences

in

For each subset of r there are $r!$ sequences

So there are $\frac{n!}{(n-r)!\,r!}$ distinct subsets.

$= \binom{n}{r}$ "n choose r"

Binomial coefficient:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Lottery: Probability of winning the jackpot at 6-49.

$$\frac{1}{\binom{49}{6}} = \frac{1}{13{,}983{,}816}$$

But this is misleading. Why?

Read Example I on p. 13 Capture-Recapture

(for interest - not on quiz).

Multinomial coefficients.

n distinct objects:

How many of grouping into

r classes of $n_1, n_2, \dots, n_r$

where $\sum_{i=1}^{r} n_i = n$

not multiplied

$$\binom{n}{n_1 \; n_2 \; n_3 \; \cdots \; n_r} = \frac{n!}{n_1! \, n_2! \cdots n_r!}$$

$$(a+b+c)^n = \sum_{\substack{n_1+n_2+n_3=n \\ n_1, n_2, n_3 \geq 0}} \binom{n}{n_1 \;\; n_2 \;\; n_3} a^{n_1} b^{n_2} c^{n_3}$$

1.5 Conditional Probability

Example:

$C+$: Covid

$C-$: No Covid.

$T+$: Test Positive

$T-$: Test Negative

Suppose we have a way of knowing for sure whether someone had Covid or not.

Suppose, out of 1,000 people tested

Frequencies

	T+	T−	Total
C+	180	20	200
C−	10	790	800
Total	190	810	1,000

Relative frequencies.

	T+	T−	
C+	.180	.020	.200
C−	.010	.790	.800
	.190	.810	1

Suppose, out of 1,000 people tested

Frequencies

	T+	T−	Total
C+	180	20	200
C−	10	790	800
Total	190	810	1,000

Relative frequencies.

	T+	T−	
C+	.180	.020	.200
C−	.010	.790	.800
	.190	.810	1

joint relative frequencies

marginal relative frequencies

If we draw a person at random from the 1,000 Tested then we have probabilities:

	T+	T−	
C+	.180	.020	.200
C−	.010	.790	.800
	.190	.810	1

$P(C+ \cap T+) = 0.180$

$P(C+) = 0.2 \qquad P(T+) = 0.19$

Sensitivity of test - conditional probability

$$P(T+ \mid D+) = \frac{0.180}{0.200} = 0.9$$

Defn: A, B two events, $P(B) \neq 0$

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

$$P(T+ \mid D+) = \frac{P(T+ \cap D+)}{P(D+)} = \frac{0.180}{0.200} = 0.9$$

Multiplication Laws

If $P(B) \neq 0$

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

Law of Total Probability

Let $\{B_1, B_2, \ldots, B_n\}$ be a partition of Ω. i.e. $\bigcup_{i=1}^{n} B_i = \Omega$ and $B_i \cap B_j = \emptyset$ if $i \neq j$

with $P(B_i) > 0$

Then $$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

works for $n = \infty$

Proof: $A = \bigcup_{i=1}^{n} (A \cap B_i)$ ← disjoint

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

$\boxed{\text{BAYES RULE I}}$ If $P(B), P(A) > 0$

Reversing direction of conditionality

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

$$= P(A \mid B) \times \frac{P(B)}{P(A)}$$

BAYES RULE II

Let $\{B_1, \ldots, B_n\}$ be a partition of Ω with $P(B_i) > 0$. Let $P(A) > 0$, then

$$P(B_j \mid A) = \frac{P(A \mid B_j)\,P(B_j)}{P(A)}$$

$$= \frac{P(A \mid B_j)\,P(B_j)}{\sum_{k=1}^{n} P(A \mid B_k)\,P(B_k)}$$

Back to Covid

	T+	T−	
C+	.180	.020	.200
C−	.010	.790	.800
	.190	.810	1

Pretend this is all we know:

$$P(T+ \mid C+) = 0.9$$ "Sensitivity"

$$P(T- \mid C-) = \frac{.79}{.80} = 0.9875$$ "Specificity"

$$P(C+) = 0.2$$ "Prevalence"

If you test positive what is the probability that you have Covid?

Want $P(C+ \mid T+)$

$$= \frac{P(C+ \cap T+)}{P(T+)}$$

$$= \frac{P(T+ \mid C+)P(C+)}{P(T+ \cap C+) + P(T+ \cap C-)}$$

If you test positive what is the probability that you have Covid?

Want $P(C+ \mid T+)$

$$= \frac{P(C+ \cap T+)}{P(T+)}$$

$$= \frac{P(T+ \mid C+)\,P(C+)}{P(T+ \cap C+) + P(T+ \cap C-)}$$

Same

$$= \frac{.9 \times .2}{.9 \times .2 + \underbrace{P(T+|C-)}_{\substack{1-P(T-|C-) \\ 1-0.9875 \\ =0.0125}} \underbrace{P(C-)}_{\substack{1-P(C+) \\ 1-.2 \\ .8}}}$$

$$= \frac{.18}{.18 + 0.0125 \times .8} = 0.9474$$

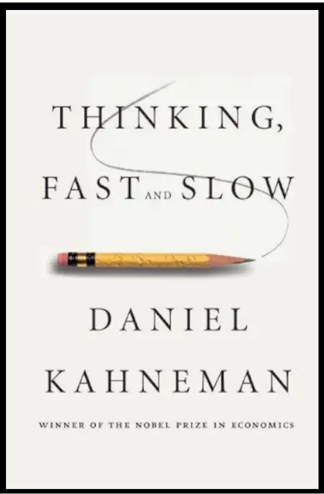
Read bottom of p. 24

As appealing as this formulation might be, a long line of research has demonstrated that humans are actually not very good at doing probability calculations in evaluating evidence. For example, Tversky and Kahneman (1974) presented subjects with the following question: “If Linda is a 31-year-old single woman who is outspoken on social issues such as disarmament and equal rights, which of the following statements is more likely to be true?

- Linda is bank teller.
- Linda is a bank teller and active in the feminist movement.”

More than 80% of those questioned chose the second statement, despite Property C of Section 1.3.

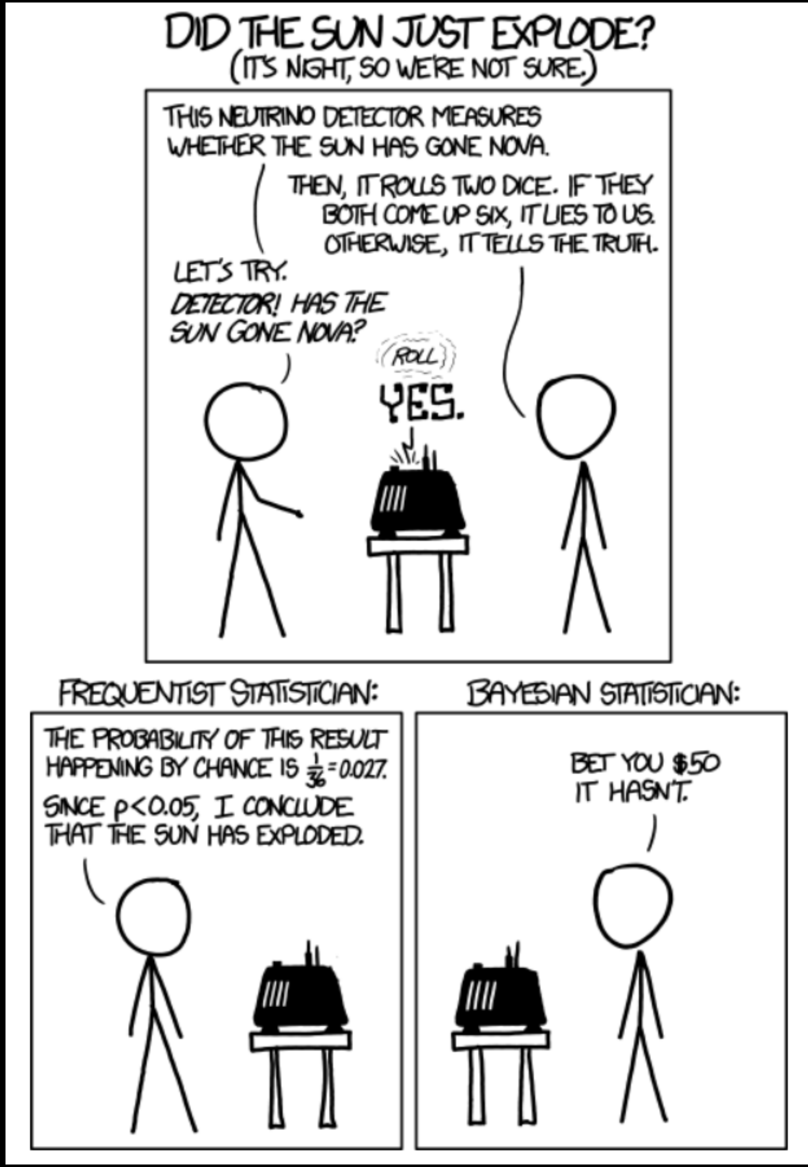
More on this later:



1.6 Independence

Defn A & B are independent iff $P(A \cap B) = P(A)P(B)$

Question: If $A \subset B$ is it possible for A & B to be independent?



QUIZ QUESTION.

How does the Bayesian statistician guess the probability that the Sun has gone nova?

Let N be "Sun went Nova"

Let Y be "Detector Say Yes"

We can agree that

$P(N) = \varepsilon$ a very very small number.

The Bayesian statistician believes that the correct way to use the information that the detector said "Yes" is to calculate $P(N|Y)$.

Pretend that $P(N) = 10^{-20}$

Find $P(N|Y)$.

Why does using only the p-value lead to a different conclusion than the Bayesian analysis?