The second-ace puzzle

Alice gets two cards from a deck with four cards: A♠, 2♠, A♡, 2♡.

A♠ A♡	A♠ 2♠	A♠ 2♡
A♡ 2♠	A♡ 2♡	2♠ 2♡

Alice then tells Bob "I have an ace".

She then says "I have the ace of spades".

The situation is similar if if Alice says "I have the ace of hearts".

Puzzle: Why should finding out which particular ace it is raise the conditional probability of Alice having two aces?

The Monty Hall Puzzle

- You're on a game show and given a choice of three doors.
  - Behind one is a car; behind the others are goats.
- You pick door 1.
- Monty Hall opens door 2, which has a goat.
- He then asks you if you still want to take what's behind door 1, or to take what's behind door 3 instead.

Should you switch?

Random Variables

To deal with expectation, we formally associate with every element of a sample space a real number.

Definition: A *random variable* on sample space S is a function from S to the real numbers.

Example: Suppose we toss a biased coin ($\Pr(h) = 2/3$) twice. The sample space is:

- hh - Probability 4/9
- ht - Probability 2/9
- th - Probability 2/9
- tt - Probability 1/9

If we're interested in the number of heads, we would consider a random variable $\#H$ that counts the number of heads in each sequence:

$$\#H(hh) = 2; \quad \#H(ht) = \#H(th) = 1; \quad \#H(tt) = 0$$

Example: If we're interested in weights of people in the class, the sample space is people in the class, and we could have a random variable that associates with each person his or her weight.

Probability Distributions

If X is a random variable on sample space S, then the probablity that X takes on the value c, is:

$$\Pr(X = c) = \Pr(\{s \in S | X(s) = c\}$$

Similarly,

$$\Pr(X \leq c) = \Pr(\{s \in S | X(s) \leq c\}$$

This makes sense since the range of X is the real numbers.

Example: In the coin example,

$$\Pr(\#H = 2) = 4/9 \text{ and } \Pr(\#H \leq 1) = 5/9$$

Given a probability measure Pr on a sample space S and a random variable X, the *probability distribution* associated with X is $f_X(x) = \Pr(X = x)$.

- f_X is a probability measure on the real numbers.

The *cumulative distribution* associated with X is $F_X(x) = \Pr(X \leq x)$.

An Example With Dice

Suppose S is the sample space corresponding to tossing a pair of fair dice: $\{(i,j) \mid 1 \leq i,j \leq 6\}$.

Let X be the random variable that gives the sum:

- $X(i,j) = i + j$

$$
\begin{aligned}
&f_X(2) = \Pr(X = 2) = \Pr(\{(1,1)\}) = 1/36\\
&f_X(3) = \Pr(X = 3) = \Pr(\{(1,2),(2,1)\}) = 2/36\\
&\vdots\\
&f_X(7) = \Pr(X = 7) = \Pr(\{(1,6),(2,5),\ldots,(6,1)\}) = 6/36\\
&\vdots\\
&f_X(12) = \Pr(X = 12) = \Pr(\{(6,6)\}) = 1/36
\end{aligned}
$$

Can similarly compute the cumulative distribution:

$$
\begin{aligned}
&F_X(2) = f_X(2) = 1/36\\
&F_X(3) = f_X(2) + f_X(3) = 3/36\\
&\vdots\\
&F_X(12) = 1
\end{aligned}
$$

The Finite Uniform Distribution

The finite uniform distribution is an equiprobable distribution. If $S = \{x_1, \ldots, x_n\}$, where $x_1 < x_2 < \ldots < x_n$, then:

$$f(x_k) = 1/n$$
$$F(x_k) = k/n$$

The Binomial Distribution

Suppose there is an experiment with probability p of success and thus probability $q = 1 - p$ of failure.

- For example, consider tossing a biased coin, where $\Pr(h) = p$. Getting "heads" is success, and getting tails is failure.

Suppose the experiment is repeated independently n times.

- For example, the coin is tossed n times.

This is called a sequence of *Bernoulli trials*.

Key features:

- Only two possibilities: success or failure.
- Probability of success does not change from trial to trial.
- The trials are independent.

What is the probability of k successes in n trials?

Suppose $n = 5$ and $k = 3$. How many sequences of 5 coin tosses have exactly three heads?

- $hhhtt$
- $hhtht$
- $hhtth$

$\vdots$

$C(5, 3)$ such sequences!

What is the probability of each one?

$$p^3(1-p)^2$$

Therefore, probability is $C(5, 3)p^3(1-p)^2$.

Let $B_{n,p}(k)$ be the probability of getting k successes in n Bernoulli trials with probability p of success.

$$B_{n,p}(k) = C(n, k)p^k(1-p)^{n-k}$$

Not surprisingly, $B_{n,p}$ is called the *Binomal Distribution*.

Combining Distributions

If X and Y are random variables on a sample space S, so is $X+Y$, $X+2Y$, XY, etc.

For example, $(X+Y)(s) = X(s) + Y(s)$.

Example: If two dice are tossed, let X be the number that comes up on the first dice, and Y the number that comes up on the second.

- Formally, $X((i,j)) = i$, $Y((i,j)) = j$.

The random variable $X+Y$ gives the total number showing.

Example: Suppose we toss a biased coin n times (more generally, we perform n Bernoulli trials). Let X_k describe the outcome of the kth coin toss: $X_k = 1$ if the kth coin toss is heads, and 0 otherwise.

How do we formalize this?

- What's the sample space?

Notice that $\Sigma_{k=1}^{n} X_k$ describes the number of successes of n Bernoulli trials.

The Sum of Binomials

Suppose X has distribution $B_{n,p}$, Y has distribution $B_{m,p}$, and X and Y are independent.

$$\begin{array}{rll} & \Pr(X+Y=k) & \\ = & \Sigma_{j=0}^{k} \Pr(X=j \& Y=k-j) & \text{sum rule} \\ = & \Sigma_{j=0}^{k} \Pr(X=j)\Pr(Y=k-j) & \text{independence} \\ = & \Sigma_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j} \binom{m}{k-j} p^{k-j}(1-p)^{m-k+j} & \\ = & \Sigma_{j=0}^{k} \binom{n}{j}\binom{m}{k-j} p^k (1-p)^{n+m-k} & \\ = & (\Sigma_{j=0}^{k} \binom{n}{j}\binom{m}{k-j}) p^k (1-p)^{n+m-k} & \\ = & \binom{n+m}{k} p^k (1-p)^{n+m-k} & \end{array}$$

Thus, $X+Y$ has distribution $B_{n+m,p}$.

An easier argument: Perform $n+m$ Bernoulli trials. Let X be the number of successes in the first n and let Y be the number of successes in the last m. X has distribution $B_{n,p}$, Y has distribution $B_{m,p}$, X and Y are independent, and $X+Y$ is the number of successes in all $n+m$ trials, and so has distribution $B_{n+m,p}$.

Expected Value

Suppose we toss a biased coin, with $\Pr(h) = 2/3$. If the coin lands heads, you get \$1; if the coin lands tails, you get \$3. What are your expected winnings?

- 2/3 of the time you get \$1; 1/3 of the time you get \$3:
- $(2/3 \times 1) + (1/3 \times 3) = 5/3$

What's a good way to think about this? We have a random variable W (for winnings):

- $W(h) = 1$
- $W(t) = 3$

The expectation of W is

$$\begin{aligned} E(W) &= \Pr(h)W(h) + \Pr(t)W(t) \\ &= \Pr(W = 1) \times 1 + \Pr(W = 3) \times 3 \end{aligned}$$

More generally, the *expected value* of random variable X on sample space S is

$$E(X) = \Sigma_x x \Pr(X = x)$$

An equivalent definition:

$$E(X) = \Sigma_{s \in S} X(s) \Pr(s)$$

Example: What is the expected count when two dice are tossed?

Let X be the count:

$$
\begin{aligned}
& E(X) \\
= \ & \Sigma_{i=2}^{12} i \Pr(X = i) \\
= \ & 2\tfrac{1}{36} + 3\tfrac{2}{36} + 4\tfrac{3}{36} + \cdots + 7\tfrac{6}{36} + \cdots + 12\tfrac{1}{36} \\
= \ & \tfrac{252}{36} \\
= \ & 7
\end{aligned}
$$

Expectation of Binomials

What is $E(B_{n,p})$, the expectation for the binomial distribution $B_{n,p}$

- How many heads do you expect to get after n tosses of a biased coin with $\Pr(h) = p$?

Method 1: Use the definition and crank it out:

$$E(B_{n,p}) = \Sigma_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

This looks awful, but it can be calculated ...

Method 2: Use Induction; break it up into what happens on the first toss and on the later tosses.

- On the first toss you get heads with probability p and tails with probability $1 - p$. On the last $n - 1$ tosses, you expect $E(B_{n-1,p})$ heads. Thus, the expected number of heads is:

$$\begin{aligned} E(B_{n,p}) &= p(1 + E(B_{n-1,p})) + (1-p)(E(B_{n-1,p})) \\ &= p + E(B_{n-1,p}) \\ E(B_{1,p}) &= p \end{aligned}$$

Now an easy induction shows that $E(B_{n,p}) = np$.

There's an even easier way ...

The Expectation of $X + Y$

Theorem: $E(X+Y) = E(X) + E(Y)$

Proof: See the text.

Example 1: Back to the expected value of tossing two dice:
Let X be the count on the first die, Y the count on the second die.

Notice that

$$E(X) = E(Y) = (1+2+3+4+5+6)/6 = 3.5$$

$$E(X+Y) = E(X) + E(Y) = 3.5 + 3.5 = 7$$

Example 2: Back to the expected value of $B_{n,p}$.

Let X be the total number of successes and let X_k be the outcome of the kth experiment, $k = 1, \ldots, n$

$$E(X_k) = p \cdot 1 + (1-p) \cdot 0 = p$$

$$X = X_1 + \cdots + X_n$$

Therefore,

$$E(X) = E(X_1) + \cdots + E(X_n) = np$$

Conditional Expectation

$E(X|A)$ is the *conditional expectation* of X given A.

$E(X|A) = \Sigma_x x \Pr(X = x|A) = \Sigma_x x \Pr(X = x \cap A)/\Pr(A)$

Theorem: For all events A such that $\Pr(A), \Pr(\overline{A}) > 0$:

$$E(X) = E(X|A)\Pr(A) + E(X|\overline{A})\Pr(\overline{A})$$

Proof:

$$\begin{aligned}
& E(X) \\
= \; & \Sigma_x x \Pr(X = x) \\
= \; & \Sigma_x x (\Pr((X = x) \cap A) + \Pr((X = x) \cap \overline{A})) \\
= \; & \Sigma_x x (\Pr(X = x|A)\Pr(A) + \Pr(X = x|\overline{A})\Pr(\overline{A})) \\
= \; & \Sigma_x (x \Pr(X = x|A)\Pr(A)) + (x \Pr(X = x|\overline{A})\Pr(\overline{A})) \\
= \; & E(X|A)\Pr(A) + E(X|\overline{A})\Pr(\overline{A})
\end{aligned}$$

Example: I toss a fair die. If it lands with 3 or more, I toss a coin with bias p_1 (towards heads). If it lands with less than 3, I toss a coin with bias p_2. What is the expected number of heads?

Let A be the event that the die lands with 3 or more.

$$\Pr(A) = 2/3$$

$$E(\#H) = E(\#H|A)\Pr(A) + E(\#H|\overline{A})\Pr(\overline{A}) = p_1\frac{2}{3} + p_2\frac{1}{3}$$

Variance and Standard Deviation

Expectation summarizes a lot of information about a random variable as a single number. But no single number can tell it all.

Compare these two distributions:

- Distribution 1:

$$\Pr(49) = \Pr(51) = 1/4; \quad \Pr(50) = 1/2.$$

- Distribution 2: $\Pr(0) = \Pr(50) = \Pr(100) = 1/3$.

Both have the same expectation: 50. But the first is much less "dispersed" than the second. We want a measure of *dispersion*.

- One measure of dispersion is how far things are from the mean, on average.

Given a random variable X, $(X(s) - E(X))^2$ measures how far the value of s is from the mean value (the expectation) of X. Define the *variance* of X to be

$$Var(X) = E((X-E(X))^2) = \Sigma_{s\in S} \Pr(s)(X(s)-E(X))^2$$

The *standard deviation* of X is

$$\sigma_X = \sqrt{Var(X)} = \sqrt{\Sigma_{s\in S} \Pr(s)(X(s) - E(X))^2}$$

- Why not use $|X(s) - E(X)|$ as the measure of distance?
- $(X(s) - E(X))^2$ turns out to have nicer mathematical properties.
- In R^n, the distance between $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ is $\sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}$

Example:

- The variance of distribution 1 is
$$\frac{1}{4}(51-50)^2 + \frac{1}{2}(50-50)^2 + \frac{1}{4}(49-50)^2 = \frac{1}{2}$$
- The variance of distribution 2 is
$$\frac{1}{3}(100-50)^2 + \frac{1}{3}(50-50)^2 + \frac{1}{3}(0-50)^2 = \frac{5000}{3}$$

Note: We are not covering 6.6, 6.7, or the negative binomial and Poisson distributions discussed in 6.4.

Logic

Logic is a tool for formalizing reasoning. There are lots of different logics:

- probabilistic logic: for reasoning about probability
- temporal logic: for reasoning about time (and programs)
- epistemic logic: for reasoning about knowledge

The simplest logic (on which all the rest are based) is *propositional logic*. It is intended to capture features of arguments such as the following:

> Borogroves are mimsy whenever it is brillig. It is now brillig and this thing is a borogrove. Hence this thing is mimsy.

Propositional logic is good for reasoning about

- conjunction, negation, implication ("if … then …")

Amazingly enough, it is also useful for

- circuit design
- program verification

Proposition Logic: Syntax

To formalize the reasoning process, we need to restrict the kinds of things we can say. Propositional logic is particularly restrictive.
The *syntax* of propositional logic tells us what are legitimate formulas.

We start with *primitive propositions*. Think of these as statements like

- It is now brilling
- This thing is mimsy
- It's raining in San Francisco
- n is even

We can then form more complicated *compound propositions* using connectives like:

- $\neg$: not
- $\wedge$: and
- $\vee$: or
- $\Rightarrow$: implies
- $\Leftrightarrow$: equivalent (if and only if)

Examples:

- $\neg P$: it is not the case that P
- $P \wedge Q$: P and Q
- $P \vee Q$: P or Q
- $P \Rightarrow Q$: P implies Q (if P then Q)

Typical formula:

$$P \wedge (\neg P \Rightarrow (Q \Rightarrow (R \vee P)))$$

Wffs

Formally, we define *well-formed formulas* (*wffs* or just *formulas*) inductively (remember Chapter 2!):
The wffs consist of the least set of strings such that:

1. Every primitive proposition $P, Q, R, \ldots$ is a wff
2. If A is a wff, so is $\neg A$
3. If A and B are wffs, so are $A \wedge B$, $A \vee B$, $A \Rightarrow B$, and $A \Leftrightarrow B$

Disambiguating Wffs

We use parentheses to disambiguate wffs:

- $P \vee Q \wedge R$ can be either $(P \vee Q) \wedge R$ or $P \vee (Q \wedge R)$

Mathematicians are lazy, so there are standard rules to avoid putting in parentheses.

- In arithmetic expressions, $\times$ binds more tightly than $+$, so $3 + 2 \times 5$ means $3 + (2 \times 5)$
- In wffs, here is the precedence order:
  - $\neg$
  - $\wedge$
  - $\vee$
  - $\Rightarrow$
  - $\Leftrightarrow$

  Thus, $P \vee Q \wedge R$ is $P \vee (Q \wedge R)$;
  $P \vee \neg Q \wedge R$ is $P \vee ((\neg Q) \wedge R)$
  $P \vee \neg Q \Rightarrow R$ is $(P \vee (\neg Q)) \Rightarrow R$
- With two or more instances of the same binary connective, evaluate left to right:

  $P \Rightarrow Q \Rightarrow R$ is $(P \Rightarrow Q) \Rightarrow R$

Translating English to Wffs

To analyze reasoning, we have to be able to translate English to wffs.
Consider the following sentences:

1. Bob doesn't love Alice
2. Bob loves Alice and loves Ann
3. Bob loves Alice or Ann
4. Bob loves Alice but doesn't love Ann
5. If Bob loves Alice then he doesn't love Ann

First find appropriate primitive propositions:

- P: Bob loves Alice
- Q: Bob loves Ann

Then translate:

1. $\neg P$
2. $P \wedge Q$
3. $P \vee Q$
4. $P \wedge \neg Q$ (note: "but" becomes "and")
5. $P \Rightarrow \neg Q$

Evaluating Formulas

Given a formula, we want to decide if it is true or false.

How do we deal with a complicated formula like:

$$P \land (\neg P \Rightarrow (Q \Rightarrow (R \lor P)))$$

The truth or falsity of such a formula depends on the truth or falsity of the primitive propositions that appear in it. We use *truth tables* to describe how the basic connectives ($\neg$, $\land$, $\lor$, $\Rightarrow$, $\Leftrightarrow$) work.

Truth Tables

For $\neg$:

P	$\neg P$
T	F
F	T

For $\wedge$:

P	Q	$P \wedge Q$
T	T	
T	F	
F	T	
F	F	

For $\vee$:

P	Q	$P \vee Q$
T	T	
T	F	
F	T	
F	F	

This means $\vee$ is *inclusive* or, not *exclusive* or.

Exclusive Or

What's the truth table for "exclusive or"?

P	Q	$P \oplus Q$
T	T	F
T	F	T
F	T	T
F	F	F

$P \oplus Q$ is equivalent to $(P \wedge \neg Q) \vee (\neg P \wedge Q)$

P	Q	$\neg P$	$\neg Q$	$P \wedge \neg Q$	$Q \wedge \neg P$	$(P \wedge \neg Q) \vee (\neg P \wedge Q)$
T	T	F	F	F	F	F
T	F	F	T	T	F	T
F	T	T	F	F	T	T
F	F	T	T	F	F	F