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 . . .

Expectation is Linear

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

Proof: Recall that

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

Thus,

$$\begin{aligned} E(X+Y) &= \Sigma_{s\in S} \Pr(s)(X+Y)(s) \\ &= \Sigma_{s\in S} \Pr(s)X(s) + \Sigma_{s\in S} \Pr(s)Y(s) \\ &= E(X) + E(Y). \end{aligned}$$

Theorem: $E(aX) = aE(X)$

Proof:

$$E(aX) = \Sigma_{s\in S} \Pr(s)(aX)(s) = a\Sigma_{s\in S} X(s) = aE(X).$$

Example 1: Back to the expected value of tossing two dice:
Let X_1 be the count on the first die, X_2 the count on the second die, and let X be the total count.

Notice that

$$E(X_1) = E(X_2) = (1 + 2 + 3 + 4 + 5 + 6)/6 = 3.5$$

$$E(X) = E(X_1 + X_2) = E(X_1) + E(X_2) = 3.5 + 3.5 = 7$$

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, $k = 1, \dots, n$:

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

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

Therefore

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

Conditional Expectation

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

$$\begin{aligned} E(X \mid A) &= \Sigma_x x \Pr(X = x \mid A) \\ &= \Sigma_x x \Pr(X = x \cap A) / \Pr(A) \end{aligned}$$

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

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

Proof:

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

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.

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

$$\begin{aligned} E(\#H) &= E(\#H \mid A) \Pr(A) + E(\#H \mid \overline{A}) \Pr(\overline{A}) \\ &= p_1 \frac{2}{3} + p_2 \frac{1}{3} \end{aligned}$$

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

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

The *standard deviation* of X is

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

Why not use $|X(s) - E(X)|$ as the measure of distance instead of variance?

- $(X(s) - E(X))^2$ 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

$$\frac{1}{3}(100 - 50)^2 + \frac{1}{3}(50 - 50)^2 + \frac{1}{3}(0 - 50)^2 = \frac{5000}{3}$$

Expectation and variance are two ways of compactly describing a distribution.

- They don't completely describe the distribution
- But they're still useful!

Variance: Examples

Let X be Bernoulli, with probability p of success. Recall that $E(X) = p$.

$$\begin{aligned} \mathrm{Var(X)} &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= p(1-p)[p + (1-p)] \\ &= p(1-p) \end{aligned}$$

Theorem: $\mathrm{Var(X)} = \mathrm{E(X^2)} - \mathrm{E(X)^2}$.

Proof:

$$\begin{aligned} E((X - E(X))^2) &= E(X^2 - 2E(X)X + E(X)^2) \\ &= E(X^2) - 2E(X)E(X) + E(E(X)^2) \\ &= E(X^2) - 2E(X)^2 + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Example: Suppose X is the outcome of a roll of a fair die.

- Recall $E(X) = 7/2$.
- $E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \ldots + 6^2 \cdot \frac{1}{6} = \frac{91}{6}$
- So $\mathrm{Var(X)} = \frac{91}{6} - (\frac{7}{2})^2 = \frac{35}{12}$.

Some Examples

Example 1: A fair die is rolled. Let X denote the number that shows up. What is the probability distribution of $Y = X^2$?

$$\begin{aligned} \{s : Y(s) = k\} &= \{s : X^2(s) = k\} \\ &= \{s : X(s) = -\sqrt{k}\} \cup \{s : X(s) = \sqrt{k}\}. \end{aligned}$$

Conclusion: $f_Y(k) = f_X(\sqrt{k}) + f_X(-\sqrt{k})$.
So $f_Y(1) = f_Y(4) = f_Y(9) = \cdots f_Y(36) = 1/6$.
$f_Y(k) = 0$ if $k \notin \{1, 4, 9, 16, 25, 36\}$.

Example 2: A coin is flipped. Let X be 1 if the coin shows H and -1 if T. Let $Y = X^2$.

- In this case $Y \equiv 1$, so $\Pr(Y = 1) = 1$.

Example 3: If two dice are rolled, 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 $X + Y$ is the total number showing.

Example 4: 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.

- If the probability of a single success is p, then $\Sigma_{k=1}^n X_k$ has distribution $B_{n,p}$
  - The binomial distribution is the sum of Bernoullis

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

Propositional 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 brillig
- 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. $P \land Q$
3. $P \lor Q$
4. $P \land \neg Q$ (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:

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

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 ($\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$) 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

Truth Table for Implication

For $\Rightarrow$:

P	Q	$P \Rightarrow Q$
T	T	
T	F	
F	T	
F	F	

Why is this right? What should the truth value of $P \Rightarrow Q$ be when P is false?

- This choice is mathematically convenient
- As long as Q is true when P is true, then $P \Rightarrow Q$ will be true no matter what.

For $\Leftrightarrow$:

P	Q	$P \Leftrightarrow Q$
T	T	T
T	F	F
F	T	F
F	F	T

How many possible truth tables are there with two primitive propositions?

P	Q	?
T	T	
T	F	
F	T	
F	F	

By the product rule, there are 16.

We've defined connectives corresponding to 4 of them: $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$.

- Why didn't we bother with the rest?
- They're definable!

Other Equivalences

It's not hard to see that $P \oplus Q$ is also equivalent to $\neg(P \Leftrightarrow Q)$

Thus, $P \Leftrightarrow Q$ is equivalent to $\neg(P \oplus Q)$, which is equivalent to

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

Thus, we don't need $\Leftrightarrow$ either.

We also don't need $\Rightarrow$:

$P \Rightarrow Q$ is equivalent to $\neg P \vee Q$

We also don't need $\vee$:

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

Each of the sixteen possible connectives can be expressed using $\neg$ and $\wedge$ (or $\vee$)