8.1 Binomial Situations

"Dumb Dora" didn't study for her final exam. It was a true/false test, so she decided to flip a coin for the answers. The statistics professor watched her the entire two hours as she was flipping the coin ... writing the answer ... flipping the coin ... writing the answer. At the end of the two hours, everyone else had finished the exam and left the room except for "Dumb Dora". The professor walked over and said, "Listen, I see that you did not study for this statistics test, you didn't even look at the exam questions. You flipped a coin over and over. You sat there doing nothing for a while and then just started flipping the coin again. If you are just flipping a coin for your answers, what in the world is taking you so long? Are you finished?" Still flipping the coin, she replied "Oh, I finished the test half an hour ago, but now I am checking my answers!"

X = the # of girls in a family with 3 kids
(assume non-multiple births)

B: How many possible outcomes are there for each birth?
I: Are the various "observations" (children) independent?
N: Are we told the # of “observations” (children) in the family?
S: Is the probability of each “success” (having a girl) the same for each “observation” (child)?

Traits of the **binomial setting**:

1. Each observation has just 2 outcomes: success or failure.
2. The n observations are all independent.
3. The variable counts the number of successes in a fixed number of n observations. (This will be different in 8.2.)
4. The probability, p, of a success is the same for each observation.

Binomial mnemonic: BINS

Binary-- just 2 outcomes: success or failure

Independent trials (success probability doesn't change)

(close enough if we pick <10% of population)

Number of trials fixed in advance

Success probability is = for each trial before we begin

P(defective light bulb)= 0.1 & a box contains 2 bulbs.

What's the random variable X?

Just 2 possible outcomes for each bulb?

Can we treat 1 bulb as independent of another?

Fixed number of bulbs/box?

Is P(success)= P(defective bulb) the same for each bulb?

So, is X binomial?

ex.
MC test
5 options per item
We randomly guess on each

- B: guess right or guess wrong
- I: knowledge of 1 question tells us nothing about others
- N: test has a set # of questions
- S: p=P(guess right) = 1/5

I but not S:
ex.
MC test
3 to 5 options per question
randomly guess on each

ex.
2 men & 1 woman at a travel agency. Suppose
P(woman choses cruise) =.3 &
P(man choses cruise) =.35.

S but not I:
ex.
Spread card deck face down.
Pick 1.
P(pointing to red)= 26/52.
Start turning over one at a time.
After 1st card revealed,
P(next is red) $\neq$ 26/52.

Weights of the 15 wrestlers on a team are recorded as part of the team statistics.

Binomial or not?

What condition(s) fail(s)?

The number of people who pass the driver's exam is recorded each day for a year.

Binomial or not?

What condition(s) fail(s)?

We draw 4 cards from a deck of cards with replacement and record the number of red cards.

Binomial or not?

What condition(s) fail(s)?

We draw 4 cards from a deck of cards without replacement and record the number of red cards.

Binomial or not?

What condition(s) fail(s)?

2 men & a woman (all single) enter a travel agency. There's a 30% chance a woman books a cruise & a 35% chance a man will. We are interested in the number of people in the trio that book a cruise.

Binomial or not?

What condition(s) fail(s)?

Now try p. 441 8.1

IF X is a discrete random variable &
counts # of successes &
has binomial setting

THEN
X has **binomial distribution**
with parameters
n= the number of observations &
p= P(success on one observation)
where $0 \leq X \leq n$

X~B(n,p)
(similar to notation: N(μ,σ).

example:

P(any light bulb is defective)= 0.1.

We have a box of 2 light bulbs.   B(2, 0.1)

P(2 defective)=

P(0 defective)=

P(only 1^{st} defective)=

P(only 2^{nd} defective)=

+

P(1 defective)=

X	0	1	2
P(x)			

example:
We count the number of girls X in a family of 3 children.
Why is this binomial?   Which binomial? B(___, ___)

P(3 girls)=

P(0 girls)=

3 ways to have 2 girls and a boy:
P(GBG)=

P(BGG)=

P(GGB)=
+ ______________
P(2 girls and a boy)=

probability distribution function (pdf) gives P(X=x).

pdf for B(3,0.5) here:

X				
P(X)				

probability histogram (page 393) for a binomial pdf:

http://www.stat.wvu.edu/SRS/Modules/NormalApprox/normalapprox.html

On the TI-84, press 2nd VARS
binompdf(n,p,X)

example:
My great grandparents had 10 girls!
These 8 that lived to adulthood.
My grandma

A "Plinko board" (The Price is Right) can simulate binomial settings:

http://www.math.psu.edu/dlittle/java/probability/plinko/index.html

B(10, 0.5) is symmetric. Some binomials are <u>approximately</u> normal.

http://www.stat.wvu.edu/SRS/Modules/NormalApprox/normalapprox.html

<u>cumulative distribution function (cdf)</u> gives $P(X \leq x)$.
(up to and including x successes) or (at most x successes)

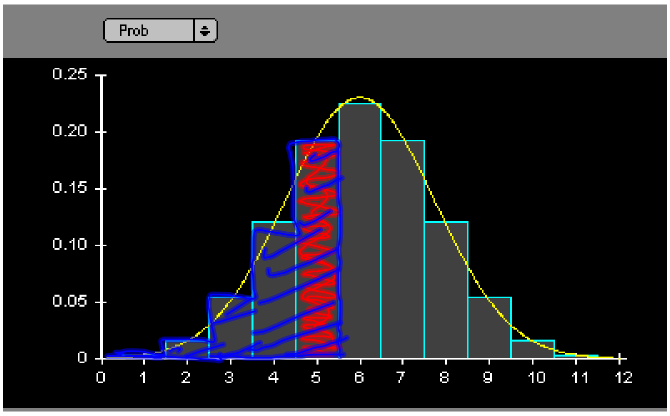
example:
X=# of girls in a family of 3 children. B(3, 0.5)
P(at most 1 girl)=$P(X \leq 1)$=

P(at most 2 girls)=$P(X \leq 2)$=

P(at most 3 girls)=$P(X \leq 3)$=

On the TI-84, 2nd VARS
binomcdf(n,p,X)

Now try p. 445 8.3, 5, 7



pdf

cdf

For the # of ways k successes out of n trials can happen, use the **binomial coefficient**:

$C(n,k) = {}_nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$

(Notation depends on book, calculator, or software used.)

The notation "n!" called "n factorial" tells you to multiply #s from 1 to n.

n! = n(n-1)(n-2)...(4)(3)(2)(1).

Note: 0!=1, by definition.

X~B(n,p), then X takes on values 0, 1, 2, ..., n.
The **binomial probability** is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

On the TI: binompdf(n,p,X)

Light bulb situation: X~B(2, 0.1).
Use the formula to find P(X=1).

NO
CALCULATOR
SYNTAX...
EVER!!

Use standard notation:

B(20, 0.5)   P(X=4) =

or use English:

"it's a binomial setting with n=20 and p=0.5 and I need to find P(X=4)."

example:
Suppose 25% of users of an allergy medication report headaches. Out of an SRS of 15 users, what's P(5 or more with headaches)?

Avoid "calculator speak" when using binomcdf.

This is correct, but too lengthy:

P(5 or more users with headaches)

$=P(X\geq 5)$

$=P(X=5)+P(X=6)+P(X=7)+P(X=8)+P(X=9)+P(X=10)+P(X=11)+P(X=12)+P(X=13)+P(X=14)+P(X=15)$

$=({}_{15}C_5)(.75^{10})(.25^5)+({}_{15}C_6)(.75^9)(.25^6)+({}_{15}C_7)(.75^8)(.25^7)+({}_{15}C_8)(.75^7)(.25^8)+({}_{15}C_9)(.75^6)(.25^9)+({}_{15}C_{10})(.75^5)(.25^{10})+({}_{15}C_{11})(.75^4)(.25^{11})+({}_{15}C_{12})(.75^3)(.25^{12})+({}_{15}C_{13})(.75^2)(.25^{13})+({}_{15}C_{14})(.75^1)(.25^{14})+({}_{15}C_{15})(.75^0)(.25^{15})$

$=.3135$

Instead, do the following:

1st indicate method verbally or preferably by formula, perhaps even using sigma notation.

2nd use the formula with the first few terms and last term inserted to show you know the pattern.

3rd use the calculator for the answer.

0, 1, 2, 3, 4, 5 ... 13, 14, 15

cdf

Like this:

X is binomially distributed with n=15 and P(headache)=p=.25
B(15, 0.25)

P(5 or more users with headaches)

$=P(X\geq 5)$

= sum of $({}_{15}C_x)(.75^{15-x})(.25^x)$ for x values 5 to 15

$=({}_{15}C_5)(.75^{10})(.25^5)+({}_{15}C_6)(.75^9)(.25^6)+\dots+({}_{15}C_{15})(.75^0)(.25^{15})$

$=.3135$

Now try p. 449 8.9, 11, 13

If X has a binomial distribution with parameters n and p, then:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\mu_X = np$$

$$\sigma_X = \sqrt{np(1 - p)}$$

X~B(n,p), then

the **mean** of X = $\mu_x = np$

standard deviation of X = $\sigma_x = \sqrt{np(1-p)}$

example:

X counts girls in a 10 child family, X~B(10, 0.5)

μ = avg. # of girls in such families

μ =

σ describes variability in population of all such families

σ =

example

defective light bulbs ~ B(2,0.1)

$\mu =$

$\sigma =$

example

Someone guesses on the whole MC part of the AP Stats exam (40 questions, 5 choices each)

The distribution is B(________,________)

$\mu =$

$\sigma =$

X~B(n,p) If we can expect 10 success and 10 failures, then X is approximately normal:
If np≥10 and n(1-p)≥10, then B(n,p) ≈ $N\left(np, \sqrt{np(1-p)}\right)$.

example: Should we use a normal approximation for...

- defective bulbs: B(2, 0.1)?

- girls in a family of 10 kids: B(10, 0.5)?

- correct guesses on a 500 question exam with 5 choices each B(500, 0.2)?

Simulating a binomial situation

01 - 44 success
45 -99, 00 failure

Let p=0.44 and n=20.
Choose integers 0 through 43 to be successes
and 44 through 99 to be failures.

Option 1:
Simulate a random integer from 0 through 99,
and let the calculator do that 20 times.
Then you have to count the successes:
MATH PRB RandInt(0,99,20)

Option 2:
Simulate one binomial trial at a time,
with p= .44, 20 times total.
Then let the calculator count the successes:
MATH PRB RandBin(1,.44,20)

An ice cream stand reports 12% of the cones they sell are "jumbo" size. You want to see what a "jumbo" cone looks like, so you stand and watch the sales for a while. Assuming independenct, what is the probability there is exactly 1 jumbo among the first 6 cones sold?

A)84% B) 6% C) 12% D)38% E)54%

Answer

Ex.
A die with s sides marked 1 through s is tossed n times and the number of times a 1 is tossed is recorded. After many repetitions of the experiment, it is found that the number of 1s tossed has a mean of 33 and a standard deviation of 5.5.

a) If the die is assumed to be fair, what is the most likely number of sides it has?

b) If the die is assumed to be fair, how many times was n tossed in each experiment?

Answer shown down here

(a) np = 33
np(1-p)= 5.5 squared = 30.25

33(1-p)= 30.25
(1-p) = .9167
p = 0.0833
0.0833 is the probability any one side will be rolled. So, 1/.0833 = 12

(b) np =33, but we know p = 1/12.
If 1/12 n =33, n = 396

Now try p. 454 15bc, 17, 19abc, and p. 459 21, 23, 25

In case you want another example regarding independence
and probability being the same for each trial:

"Suppose that we are counting the number of people who go up the hill next Saturday to fetch a pail of water. There are two trials: Jack and Jill.

Does a fixed value of p imply that trials are independent?

No. For example, suppose Jack and Jill plan to go together up the hill to fetch a pail of water next Saturday, as long as it does not rain. If it rains, neither will go. The probability of rain is 0.2. Either both Jack and Jill will go up the hill (two successes) or neither will go (no successes). The probability of a success on each trial is fixed at 0.8. That is, the probability that Jack goes up the hill is 0.8, and the probability that Jill goes up the hill is 0.8. However, the trials are not independent. If Jack goes, so will Jill, and vice versa.

Does the fact that the trials are independent imply a fixed value of p?

Again, no. Suppose next Saturday Jack and Jill each plan to flip a penny. If Jack's penny lands heads, he goes up the hill. Same thing with Jill. They use different pennies, and Jill's is a two-headed penny. This means that the probability that she goes up the hill is 1 while the probability that Jack goes up the hill is 1/2. In this case, the trials are independentâ€"whether Jack goes up the hill or not does not change the probability of Jill going, and vice versa. But the probability of going up the hill is different."

Ann E. Watkins
Professor of Mathematics
California State University, Northridge

In case you want still more examples regarding independence and probability being the same for each trial:

Ex. A female frog lays a single fertilized egg in each of several different bromeliads, which are jungle plants that collect pools of water in their leaves. Let X be the number of those eggs that successfully grow to adult frogs. We might reasonably model each of these trials as independent of one another, since the bromeliads are all different, but the probabilities of success may be very different: one egg could be laid in a bromeliad full of hungry dragonfly nymphs ready to gobble up a frog egg, while another egg could be laid in a bromeliad with nothing in it but harmless and edible mosquito larvae.

Ex. A female frog lays a hundred eggs in a single pool of water, with limited resources. Let X be the number of those eggs that successfully grow to adult frogs. We might reasonably think of these tadpoles as identical at the outset, and hence they all have the same probability of succeeding. But with limited resources in the pool, their survivals are far from independent. The success of one diminishes the chance of success of all the others.

Ex. You and I buy raffle tickets for different raffles, me for my daughter's school's turkey raffle, you for your son's school's turkey raffle. One person will win each raffle--not necessarily you or I! Let X be the number of people among you and me who win a turkey. The event of me winning a turkey is independent of the event of you winning a turkey, since the raffles are different. But the probabilities are different, since (almost certainly) the number of people who entered my daughter's raffle isn't the same as the number of people who entered your son's raffle.

Ex. You and I each buy a single raffle ticket for the same turkey raffle. Now our chances of winning are the same, but clearly they are not independent: only one person can win.

--Floyd Bullard

Where's the "10 successes and 10 failures" rule from?

B(n,p) has from 0 to n successes possible.
It is discrete (think histogram).
Successes come in whole numbers.

N(μ,σ) extends forever in each direction
It is continuous (think curve).
To use N(μ,σ) to approximate B(n,p), we cut off the tails.
More than 3 σ from μ doesn't matter much.
Anything within 3 σ from μ does matter.

N(μ,σ) is a useful model for B(n,p) when the mean
of B(n,p) is more than 3 SDs from the "ends" at 0 and n.

Consider the lower end. (Symmetry provides the other half of the Success/Failure Condition.)

We need: $\mu - 3\sigma > 0$
That's: $np - 3\sqrt{(npq)} > 0$
Or: $np > 3\sqrt{(npq)}$
Squaring: $n^2 * p^2 > 9npq$
Divide by np: $np > 9q$
Since q must be at most 1, we can guarantee that
by requiring $np > 9$.

So why not say "at least 10"?
That's close enough for statisticians...

Some books are satisfied with 5.
That'll be okay if we cut off the Normal only 2 σ each way.
That won't matter much, as long as $p \approx 0.5$.