Announcements:

- We will start using R Commander again, so see one of us if you are still having problems downloading it.
- Next Friday I will be out of town, Shandong Min will give the lecture.

Homework: (Due *Wed,* Nov 3)

Chapter 8:

#33 (Note that problem continues on top of p. 324)

#39 (will want to use computer, or lots of computation!)

#87 (will need to use computer)

Section 8.4: Binomial Random Variables

What do the following random variables have in common?

Ex 1: A fair coin is flipped 10 times,
X = number of heads.

Ex 2: Ten births are observed at a hospital,
X = number of boys, assume P(B)=.5

Ex 3: A student takes a true/false test with 10 questions,
just guessing, X = number correct

Ex 4: Suppose *half* of all adults think genetically modified food is unsafe. Take a random sample of 10 adults, X = number (out of the 10 polled) who think this.

What do these have in common?

Each of these random variables has the exact same probability distribution function!

For instance, in each case, $P(X = 0) = (½)^{10}$
$P(X = 1)$ is same for all of them, and so on.
Note that X can be 0, 1, 2, ... , 10

In each case, X is called a *binomial random variable with* *n=10 and p=½.*

It is the outcome of a *binomial experiment*.

Properties of a Binomial Experiment

1. There are *n* "trials" where *n* is determined in advance. (10 Coin flips, births, T/F questions, adults polled)

2. There are *two possible outcomes* on each trial, called "success" and "failure" and denoted S and F. (Heads/tails; Boy/girl; Right/Wrong, Unsafe/not unsafe)

3. The *outcomes are independent* from one trial to the next. Knowledge of one does not help predict the next one. (True for all 4 examples.)

4. The probability of a "success" *remains the same* from one trial to the next, and this probability is denoted by p. The probability of a "failure" is $1-p$ for every trial.

Note that $n = 10$ and $p = ½$ for each example given.

NOTE: $p = ½$ is not always the case!
For example, multiple choice test with 5 choices, student is just guessing, p = 1/5.

A **binomial random variable** is defined as X = number of successes in the n trials of a binomial experiment.

Two examples (one binomial, one not):
Weekly quiz has 5 questions with 4 choices per question, worth 2 points each. Suppose someone is just guessing.
X = *Number correct*
X **is** a binomial random variable, n = 5 and p = 1/4
Y = *Points earned* = 2X
Y **is not** a binomial random variable (but Y/2 is).

Examples that are *not* binomial experiments:

1. A team plays 12 games in the season, X = number won.
   p = Probability of win does *not* stay the same
   Condition #4 does not hold.

2. Woman decides to have children until she has one girl or 4 children, whichever comes first.
Number of "trials" is not fixed in advance (Condition #1).

3. Deal a poker hand of 5 cards, X = number of aces.
Cards are drawn *without replacement* so outcomes are NOT independent (also, p changes). (Conditions #3, #4)

Once you recognize a binomial random variable, the pdf is always given by this formula (so you don't have to rely on Chapter 7 rules each time!):

Probability of exactly *k* successes:

$$\Pr(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n.$$

Factorial notation: n! = 1 × 2 × 3 × ... × (n-1) × (n)
0! = 1, by convention.

EX: n = 5, p = .25, k = 2 (e.g. 2 quiz Qs right, out of 5)

$$\Pr(X = 2) = \frac{5!}{2!(5-2)!}(.25)^2(1-.25)^{5-2} = 10(.0625)(.4219) = .2637$$

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

How this formula is found; Use example of n = 3, p = .25, k = 2:

$$\Pr(X = 2) = \frac{3!}{2!(3-2)!}(.25)^2(1-.25)^{3-2} = 3(.0625)(.75) = .14$$

- Individual string of *k* successes and (*n* – *k*) failures has probability $p^k(1\text{-}p)^{n\text{-}k}$ <u>Example</u>: P(SSF) = (.25)(.25)(.75)

- There are $\frac{n!}{k!(n-k)!}$ possible ways to get *k* successes

Example: *n* = 3, *k* = 2, could be {SSF, SFS, FSS}

$$\frac{n!}{k!(n-k)!} = \frac{3!}{2!(3-2)!} = \frac{3\times2\times1}{(2\times1)(1)} = 3$$

Finding binomial probabilities using a computer (to find pdf and cdf):

Excel – See page 297
=BINOMDIST(k,n,p,false) for the pdf
=BINOMDIST(k,n,p,true) for the cdf
(You type the equal sign then the command in any cell and it will put the requested probability in that cell.)

EX: BINOMDIST(2,3,.25,false) would give .14.

R Commander: See instructions linked to website.
Distributions →Discrete distributions → Binomial distribution → Binomial probabilities
(then fill in n and p in the popup box)

Mean and standard deviation *for binomial random variables* (only!):

Mean = expected value of X = E(X) = μ = ***np***

Variance = $\sigma^2 = np(1\text{-}p)$; standard deviation = $\sqrt{np(1-p)}$

Example:

n = 10, p = 0.2

mean = (10)(0.2) = **2**

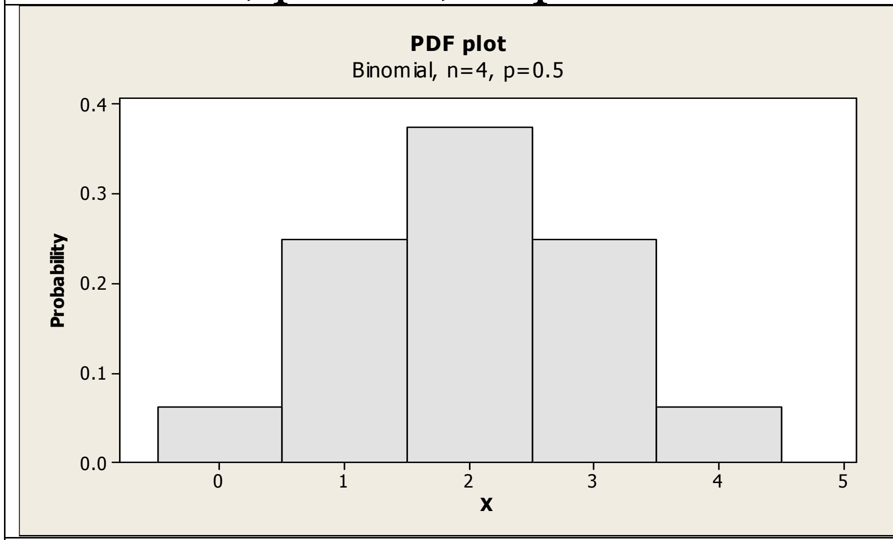
standard deviation = $\sqrt{10(.2)(.8)} = \sqrt{1.6}$ = 1.265

(not much use for now, but will be very useful soon)

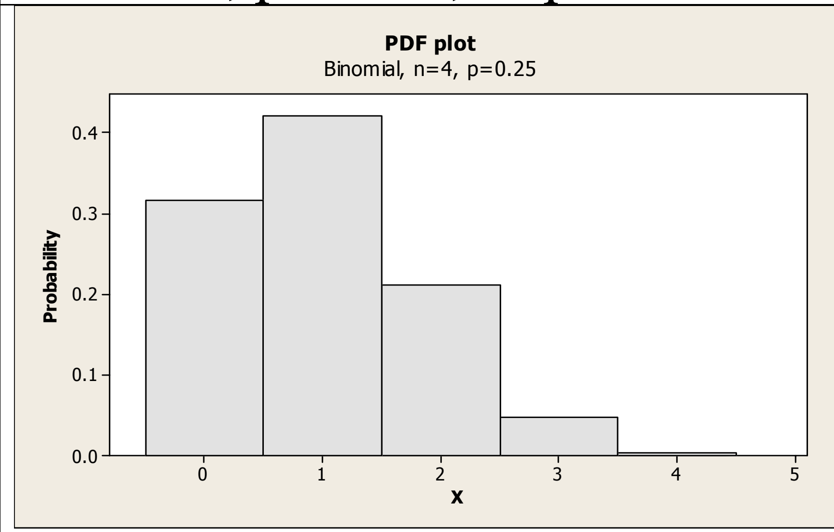
Let's look at some pictures of binomial pdfs with different n's and p's.

Binomial pdfs (Note change in n from 4 to 40):

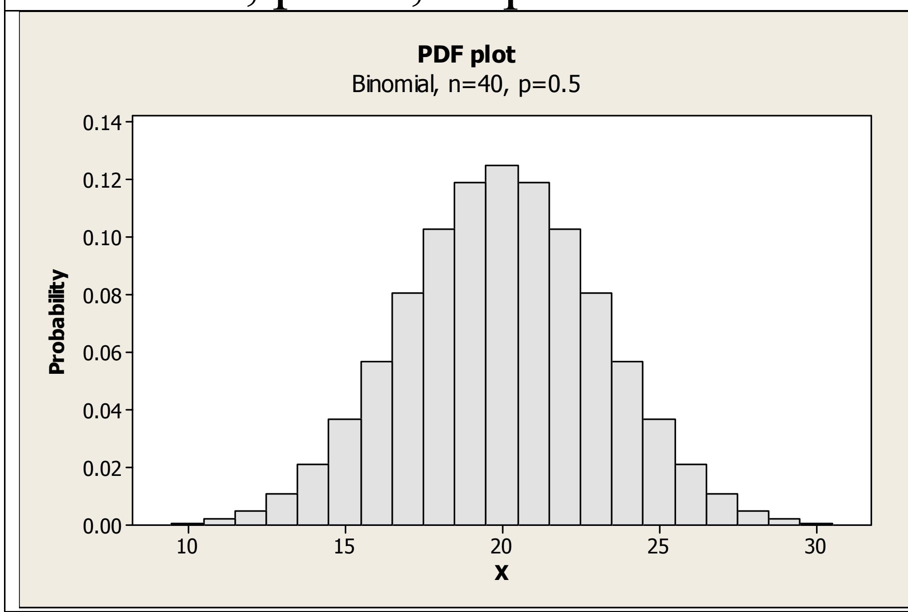
n = 4, p = 0.5, Exp. value = 2



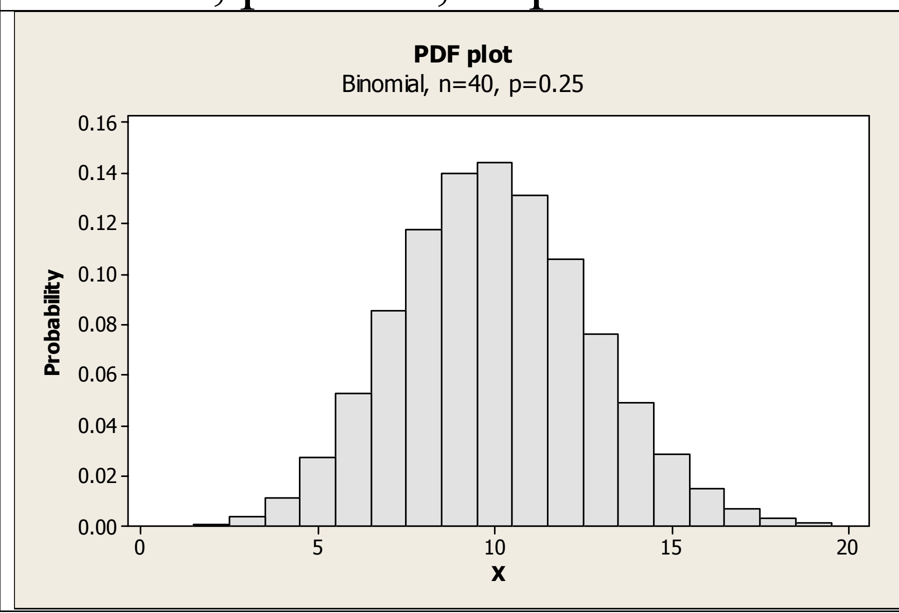
n = 4, p = 0.25, Exp. value = 1



n=40, p=0.5, Exp. value = 20



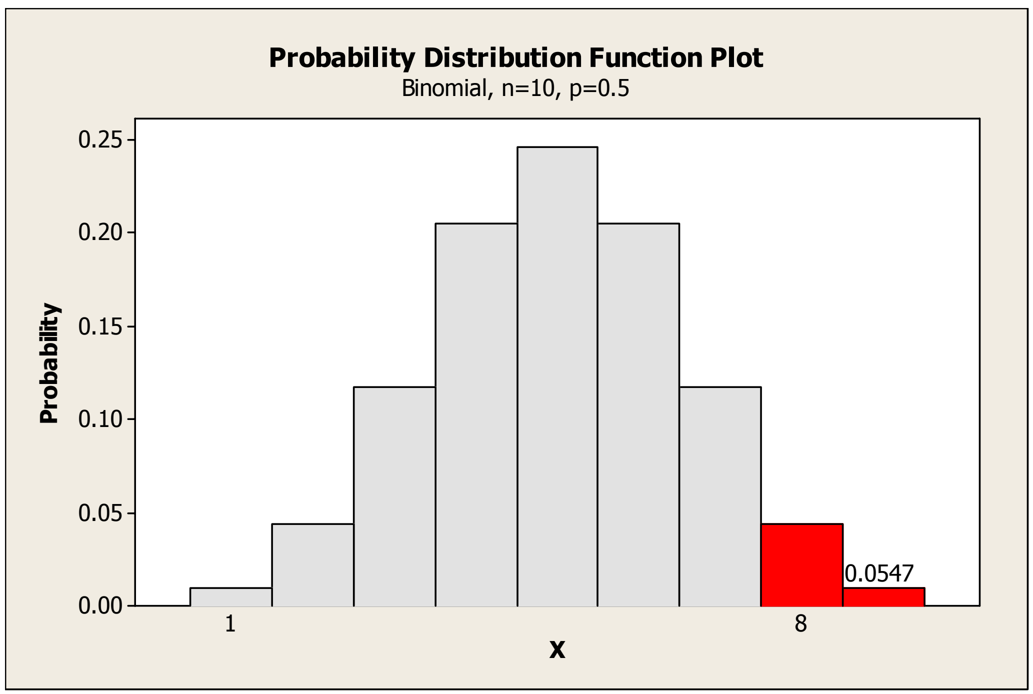
n=40, p = 0.25, Exp. value = 10



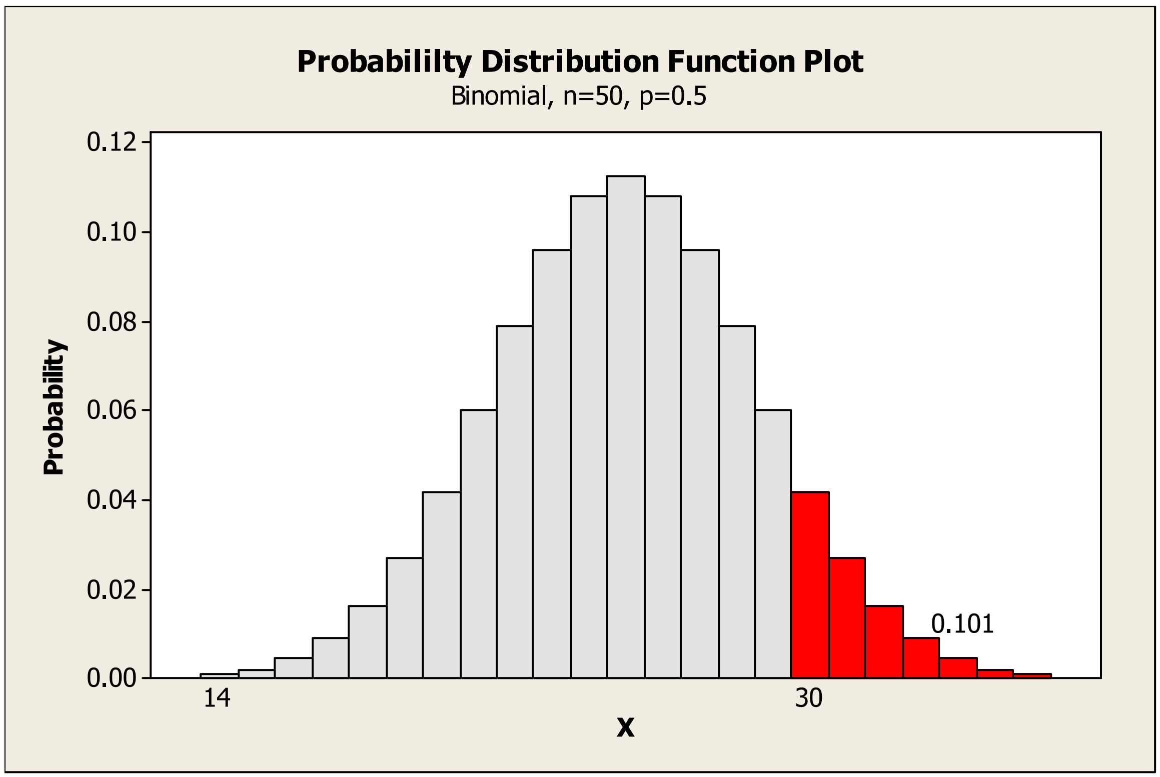
For binomial, the CDF often more interesting than PDF.

Ex: Test has 10 questions, pass if 80%, 8 or more, correct.

Find P(X = 8, 9, 10) = P(X ≥ 8) = 1 – P(X ≤ 7) = 1 − cdf for X = 7, which is **1 −.94531 = .0547** (if just guessing)
Probability = sum of areas of rectangles for those values!



Now suppose test has 50 questions, you need 60% correct to pass, so need 30 questions correct. If just guessing, $P(X \geq 30) = 1 - P(X \leq 29) = 1 - .899 = .101$ $= P(30)+P(31)+P(32)+\ldots.+P(50)$



Ex: Political poll with n = 1000. Suppose *true p = .48* in favor of a candidate.

X = number in poll who say they support the candidate.

X is a binomial random variable, n = 1000 and p = .48.

- *n* trials = 1000 people (without replacement, but for large population treat as if with replacement)
- *"success"* = support, *"failure"* = doesn't support
- Trials are *independent*, knowing how one person answered doesn't change others probabilities
- *p* remains fixed at .48 for each random draw of a person to ask

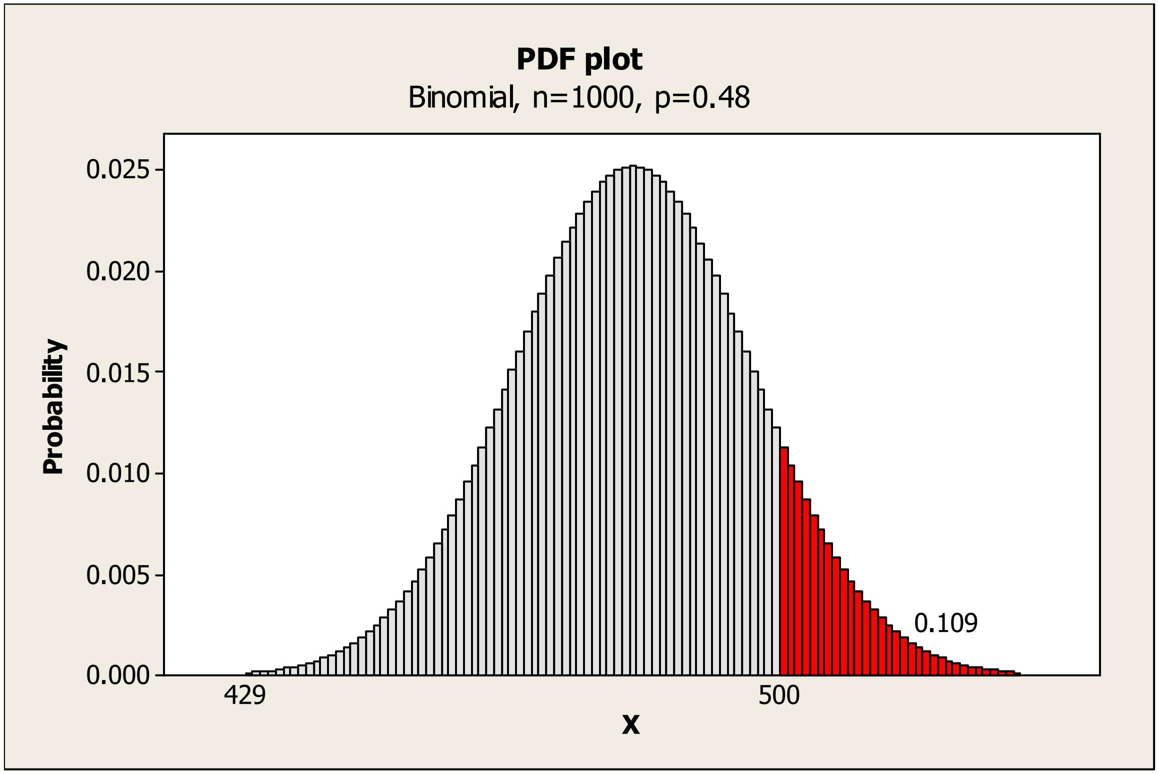
Mean = np = (1000)(.48) = **480**.

Standard dev, = $\sqrt{np(1-p)} = \sqrt{1000(.48)(.52)}$ = **15.8**

What is the probability that *at least half* of the *sample* support the candidate? (Remember only 48% of population supports him or her.)

$P(X \geq 500) = P(X = 500) + P(X = 501) + \ldots + P(X = 1000)$.

Using Excel: $1 - P(X \leq 499) = 1 - .891 = .109$.



Note what this says:
In polls of 1000 people in which 48% favor something, the poll will say *at least half favor it* with probability of just over .10 or in just over 10% of polls.

In Section 8.7, will learn how to *approximate* this using normal curve.

Now turn to power point for example of using binomial random variable in testing for psychic abilities.

Binomial example you can try: Online ESP test:
http://www.gotpsi.org

Try doing 5 guesses where there are 5 choices each time.

Assuming no ESP, n = 5 and p = 1/5 or .2.

What should be expected by chance?

X = number correct, E(X) = np = (5)(1/5) = 1.

PDF is P(X = k), CDF is P(X ≤ k)

Also interesting to find P(X ≥ k)

k	**pdf**	**cdf**	**P(X ≥ *k*)**
0	0.32768	0.32768	1.00000
1	0.40960	0.73728	0.67232
2	0.20480	0.94208	0.26272
3	0.05120	0.99328	0.05792
4	0.00640	0.99968	0.00672
5	0.00032	1.00000	0.00032