Stat 321 - Day 13
Binomial Distributions

Example: Multiple Choice Exam (cont.)
Suppose that a student takes a pop quiz with three multiple choice options on each question. Suppose further that he has to guess randomly on each question.

Predictions:

(a) Before doing any formal analysis, make some predictions:

- If a student needs to get at least 50% of the questions correct in order to pass, is that more likely to happen with a four-question quiz or a ten-question quiz?

- With a four-question quiz, what do you think is the most likely value for the number of correct answers? How about with a ten-question quiz?

- If several students took a four-question quiz and guessed randomly on each question, how many do you think they would answer correctly on average? How about with a ten-question quiz?

(b) Each of you will take a four-question quiz and a ten-question quiz on which you guess randomly at every answer. Once you are told the "answer key," determine how many questions you got correct on each quiz. Record the results for the class on the four-question quiz in the following table:

# correct	0	1	2	3	4	total
# students						
emp. prob.						

(c) How many and what proportion of students passed this quiz? How about for the ten-question quiz? Which is more likely, and how does that compare with your prediction in (a)?

Consider the quiz consisting of four questions. The following lists all possible outcomes in the sample space. For example, $S_1S_2S_3F_4$ means a student answered the first 3 questions correctly (Successes), but then missed the fourth (Failure).

$$\{S_1S_2S_3S_4 \;\; S_1S_2S_3F_4 \;\; S_1S_2F_3S_4 \;\; S_1S_2F_3F_4 \;\; S_1F_2S_3S_4 \;\; S_1F_2S_3F_4 \;\; S_1F_2F_3S_4 \;\; S_1F_2F_3F_4$$

$$F_1S_2S_3S_4 \;\; F_1S_2S_3F_4 \;\; F_1S_2F_3S_4 \;\; F_1S_2F_3F_4 \;\; F_1F_2S_3S_4 \;\; F_1F_2S_3F_4 \;\; F_1F_2F_3S_4 \;\; F_1F_2F_3F_4\}$$

(d) How many possible outcomes are there? Explain why this makes sense from the general product rule.

(e) Determine the value of the random variable X = "number of correct answers" for each outcome in the sample space. Is X a discrete or a continuous random variable?

(f) How many of these 16 outcomes contain exactly one success and three failures? Is the empirical estimate of the probability that X=1 close to 4/16? Explain why this is not the case- what assumption would this require that is not true here?

We will now derive the probability distribution of X through counting techniques and probability rules.

(g) To find the probability of any specific outcome, such as $F_1F_2S_3F_4$, we need to consider P(S)=1/3 and P(F) = 2/3. Under what conditions can we write: $P(F_1 \cap F_2 \cap S_3 \cap F_4) = P(F_1)P(F_2)P(S_3)P(F_4)$, and do you believe that this condition is met in this scenario? Explain, and then calculate this probability.

(h) What is true about the probability for each of the outcomes $\{F_1F_2F_3S_4, F_1F_2S_3F_4, F_1S_2F_3F_4, S_1F_2F_3F_4\}$ that correspond to X=1?

(i) To determine P(X=1), can we add the probabilities $P(F_1F_2F_3S_4) + P(F_1F_2S_3F_4) + P(F_1S_2F_3F_4) + P(S_1F_2F_3F_4)$? Explain why, and then determine this probability.

Thus, we can multiply the probability for any one of these outcomes by the number of outcomes corresponding to a particular value of X.

(j) How many outcomes are there that correspond to X=2? Express this as a combination (the number of ways of choosing two correct answers among the four questions).

(k) Use this information (and the probability of, say, $S_1S_2F_3F_4$) to determine P(X=2).

(l) Repeat this process to find the probabilities for X=0, X=3, and X=4. Complete the following table for the probability mass function (pmf) of X:

Number correct	0	1	2	3	4
Probability					

(m) How do these theoretical probabilities compare to the empirical estimates from our class results in (b)?

You should have found that the probability distribution of X = number of correct answers in a sample of four questions (subject to our assumptions that the success probability remains 1/3 on every question and that the questions are independent) is given by the expression

$$P(X=x)=\binom{4}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{4-x}.$$

(m) Now suppose that the student has probability p of answering any one question correctly, that this probability does not change from question to question, that the questions are independent, and that there are a total of n questions. Generalize from the above result to determine the formula for the probability distribution of the random variable X = number of successful answers in n attempts.

P(X=x) =

This probability distribution is known as the ***binomial distribution***. It applies to situations that consists of trials for which:

- Each trial has two possible outcomes (typically referred to as "success" and "failure").
- The probability of "success" remains constant on each trial (call it p).
- The trials are independent.
- The random variable of interest (call it X) is the number of successes in a *fixed number* (call it n) of trials.

You should have derived above the result that the probability distribution of a binomial random variable with parameters n and p is given by $P(X=x)=\binom{n}{x}p^{x}(1-p)^{n-x}$ for $x=0,1,\ldots,n$. It can be shown that E(X)=np and V(X)=np(1-p).

We can use Minitab (`Calc> Probability Distributions> Binomial`) to perform calculations involving the binomial distribution. Table A.1 of your text also provides binomial probabilities.

(n) Use Minitab to calculate P(X=1) that you found in (i) above. Choose `Calc> Probability Distributions> Binomial` from the Minitab menu. The first button should be set to "Probability." Enter 4 as the number of trials (the number of questions) and .3333 as the probability of success. Click the button next to "input constant" and enter the number 1 in the box. Click OK. Does this correspond to P(X=1) that you calculated in (i) above?

(o) Use Minitab to calculate the entire probability distribution of X. First enter the values 0, 1, 2, 3 and 4 into `c1`. Then choose `Calc> Probability Distributions> Binomial` as before. Leave the parameters where you set them, but click on "input column" and enter `c1` in the box. Enter `c2` as the storage column. Click OK. Do the probabilities correspomd to those that you recorded in (l) above?

Now consider the ten-question quiz, and let the random variable Y = number of questions answered correctly.

(p) Use Minitab to determine the probability of failing the quiz (answering nine or fewer questions correctly). Once again choose `Calc> Probability Distributions> Binomial`, but change the number of trials to 10. Even more importantly, click on "cumulative probability" in order to calculate the probability of nine or fewer successes rather than the probability of exactly nine successes. Finally, click on "input constant" and enter the value 9. Record the probability that Minitab produces:

$P(Y \leq 9) =$

(q) Use the complement rule and your answer to (p) to find the probability of passing the quiz. Is this probability lower than that of passing the four-question quiz?

$P(Y \geq 10) =$

Notice that when calculating binomial probabilities in Minitab, it is very handy to make use of the relationship $P(X \geq k) = 1 - P(X \leq k-1)$.

(r) Use Minitab to determine the entire probability distribution of Y:

```
MTB> set c4
DATA> 0:20
DATA> end
```

Choose `Calc> Probability Distributions> Binomial`. Click on the "Probability" button. Enter 20 as the number of trials and .3333 as the probability of success. Enter c4 as the input column and c5 as the output storage column. Alternatively, you could type:

```
MTB> pdf c4 c5;
SUBC> binomial 20 .3333.
```

Record the probabilities (to four decimal places) in the table:

y	0	1	2	3	4	5	6	7	8	9	10
$P(Y=y)$											

(s) Repeat (r) but clicking on the "Cumulative Probability" button and storing the results in `c6`. Record the results in the table below. Explain why these probabilities differ (with one exception) from the ones in (r).

y	0	1	2	3	4	5	6	7	8	9	10
$P(Y \leq y)$											