Binomial Distribution

Dr. Tom Ilvento
FREC 408

Binomial Random Variables

- In many cases the responses to an experiment are dichotomous
  - Yes/No
  - Alive/Dead
  - Support/Don't Support

Binomial Random Variables

- When our focus is conducting an experiment **n times independently** and observing the number **x of times** that **one** of the **two outcomes occurs**
- This **x** is a Binomial Random Variable
- We can exploit this by using known formulas for a probability distribution

Examples of Binomial Random Variables

- 1,000 people are polled in a telephone survey and asked if they support George W. Bush
  - The responses are Yes (1) or No (0)
  - The proportion saying yes is designated as
    - p
    - (1-p) is the proportion saying No

Binomial Random Variable

- Yes it is a binomial random variable
- Conduct an experiment 1,000 times and observe the number x of times that Yes occurs

Characteristics of a Binomial Random Variable

- The experiment consists of *n* identical trials
- There are only two outcomes on each trial. Outcomes can be denoted as
  - **S for Success**
  - **F for Failure**

Characteristics of a Binomial Random Variable (cont.)

- The probability of S (success) remains the same from trial to trail
  - **Denoted as p   the proportion**
- The probability of F (failure)
  - **Denoted as q   q=(1-p)**
- The trials are independent of each other
- **The binomial random variable x is the number of Successes in *n* trials**
- **Also refer to *Conditions Required for a Binomial Experiment* on P246 of book.**

Example 1: Marketing example

- Marketing survey of 100 randomly chosen consumers
- Record their preferences for a new and an old diet soda – ask them to choose their preference
- Let x be number of 100 who choose the new brand
- **This is a binomial random variable**

Conduct an experiment 100 times and observe the number x of times that Yes occurs

Fitness Example

- Heart Association says only **10%** of adults over 30 can pass the fitness test
- Suppose **4 people** over 30 are selected at random
- Let x be the number who pass the minimum requirements
- Find the probability distribution for x

Conduct an experiment 4 times and observe the number x of times that pass occurs

How to solve the fitness problem – the way we used with discrete random variables

1. List the events
2. List the sample points that refer to that event
3. Calculate the probabilities
   - **p = .1   and q = (1.0 - .1) = .9**

Event x	**Sample Points**	**Probability**
All Fail	FFFF	(.9)(.9)(.9)(.9) = .6561

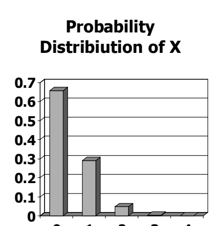
I multiply through on the probabilities because Each trial is independent of the others

Solve for Each Event

Event x	Notation	Probability
0 *All Fail*	FFFF	(.9)(.9)(.9)(.9) **= .6561**
1 *1 Pass 3 Fail*	SFFF FSFF FFSF FFFS	$4[(.1)(.9)^3]$ **=.2916**
2 *2 Pass 2 Fail*	SSFF SFSF SFFS FSSF FSFS FFSS	$6[(.1)^2(.9)^2]$ **=.0486**
3 *3 Pass 1 Fail*	SSSF SFSS SSFS FSSS	$4[(.1)^3(.9)]$ **=.0036**
4 *4 Pass*	SSSS	(.1)(.1)(.1)(.1) **= .0001**

Fitness Example

- When x = 0 All Fail P = .6561
- When x = 1 One Pass P = .2916
- When x =2 Two pass P =.0486
- When x=3 Three pass P = .0036
- When x=4 Four pass P = .0001



Fitness Example

- Find the probability that none of the adults pass the test
  - P(x=0) = .6561
- Find the probability that 3 of 4 adults pass the test
  - P(x=3) = .0036

When we have many trials the formulas get complicated

- We can also use the binomial probability distribution formula
- Using factorial notation
  - n! = n(n-1)(n-2)…(n-(n-1))
  - 5! = 5x4x3x2x1 = 120
  - 0! = 1, 1!=1, 2!=2x1=2, …
- The formula for any x in n trials is:

$$p(x) = \frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

Binomial Probability Distribution Formula (P248)

$$p(x) = \frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

aka

$$p(x) = \binom{n}{x} p^x q^{n-x}$$

Note: it uses the Combinatorial Rule as the first part of the formula

Most calculators will do all or part of this – become familiar before trying it out

What defines a binomial probability distribution?

- p = Probability of a success on a single trial
- q= (1-p) probability of failure
- n= number of trials
- x = number of successes in n trials

$$p(x) = \frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

$$p(x) = \frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

For x=3 in the fitness example

$$p(3) = \frac{4!}{3!(4-3)!}(.1)^3(.9)^{4-3}$$

$$= \frac{4\cdot 3\cdot 2\cdot 1}{(3\cdot 2\cdot 1)(1)}(.001)(.9)$$

$$= \frac{24}{6}(.0009) = .0036$$

This matches the number we generated the other way

Fitness Example Table

Event x	Notation	Probability
0 *All Fail*	FFFF	.6561
1 *1 Pass 3 Fail*	SFFF FSFF FFSF FFFS	.2916
2 *2 Pass 2 Fail*	SSFF SFSF SFFS FSSF FSFS FFSS	.0486
3 *3 Pass 1 Fail*	SSSF SFSS SSFS FSSS	**.0036**
4 *4 Pass*	SSSS	.0001

$$p(x)=\frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

For x=2 in fitness example

$$p(2)=\frac{4!}{2!(4-2)!}(.1)^2(.9)^{4-2}$$

$$=\frac{4\cdot3\cdot2\cdot1}{(2\cdot1)(2\cdot1)}(.01)(.81)$$

$$=\frac{24}{4}(.0081)=.0486$$

This matches the number we generated the other way

Fitness Example Table

Event x	Notation	Probability
0 *All Fail*	FFFF	.6561
1 *1 Pass 3 Fail*	SFFF FSFF FFSF FFFS	.2916
2 *2 Pass 2 Fail*	SSFF SFSF SFFS FSSF FSFS FFSS	**.0486**
3 *3 Pass 1 Fail*	SSSF SFSS SSFS FSSS	.0036
4 *4 Pass*	SSSS	.0001

Mean of a Binomial Random Variable

- Since a binomial is only a dichotomy, the formulas for the mean and the standard deviation will simplify
- From $\mu = \sum x \cdot p(x)$
  - To **$\mu = np$** (P252)

Variance and Standard Deviation of a Binomial Random Variable

- From $\sigma^2 = \sum(x-\mu)^2 p(x)$
  - To **$\sigma^2 = npq$** (P252)
- The standard deviation is then

$$\sigma = \sqrt{npq}$$

Mean and Standard Deviation for Fitness Example

- Heart Association says only 10% of adults over 30 can pass the fitness test
- Thus the proportion passing was estimated at .1, and n for the problem was 4 people
- $\mu = np = 4(.1) = .40$
- $\sigma^2 = npq = 4(.1)(.9) = .36$
- $\sigma = .60$

Fitness Example Table for mean and variance

Event x	Notation	Probability
0 *All Fail*	FFFF	.6561
1 *1 Pass 3 Fail*	SFFF FSFF FFSF FFFS	.2916
2 *2 Pass 2 Fail*	SSFF SFSF SFFS FSSF FSFS FFSS	.0486
3 *3 Pass 1 Fail*	SSSF SFSS SSFS FSSS	.0036
4 *4 Pass*	SSSS	.0001

I could have solved for the mean using the formula for discrete random variables

- To solve for the mean I would have:

$$E(x)=\sum_{i=1}^{n} x_i \cdot p(x_i)=\mu$$

  - E(x) = (0)(.6561) + (1)(.2916) + (2)(.0486) + (3)(.0036) + (4)(.0001)
  - E(x) = .4
- Binomial approach
  - E(x) = n·p = 4·(.1) = .4

I could have solved for the Variance using the formula for discrete random variables

- To solve for the variance I would have:

$$E[(x-\mu)^2]=\sum_{i=1}^{n}(x_i-\mu)^2 p(x_i)=\sigma^2$$

  - $E(x-\mu)^2$ = $(0-.4)^2(.6561)$ + $(1-.4)^2(.2916)$ + $(2-.4)^2(.0486)$ + $(3-.4)^2(.0036)$ + $(4-.4)^2(.0001)$
  - $E(x-\mu)^2$ = .36
- Binomial approach
  - E(x) = n·p·q = 4·(.1)(.9) = .36

Nitrous Oxide Example

- Suppose we were recording the number of dentists that use nitrous oxide (laughing gas) in their practice
- We know that 60% of dentists use the gas.
  - **p = .6 and q = .4**
- Let X = number of dentists in a random sample of **five** dentists use use laughing gas.
  - **n = 5**

Nitrous Oxide Example

- We said the probability that a dentist uses nitrous oxide is .6
- How would you assign probabilities to the values x could take when we randomly select five dentists?

X	0	1	2	3	4	5

Solve for Each Event

Event x	Notation	Probability
0 *All Fail*	FFFFF	(.4)(.4)(.4)(.4)(.4) = .01024
1 *1 Pass 4 Fail*	SFFFF FSFFF FFSFF FFFSF FFFFS	[(.6)(.4)(.4)(.4)(.4)] x5 = .0768
2 *2 Pass 3 Fail*	SSFFF SFSFF SFFSF SFFFS FSSFF FSFSF FSFFS FFSSF FFSFS FFFSS	[(.6)(.6)(.4)(.4)(.4)] x10 = .2304
There	**Is**	**More!!**

Nitrous Oxide Example

- So what other way can we get to the probabilities?

X	0	1	2	3	4	5
P(X)	.0102	.0768	.2304	.3456	.2592	.0778

$$p(x)=\frac{n!}{x!(n-x)!}(p)^x(q)^{n-x}$$

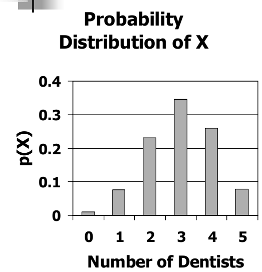
Nitrous Oxide Example

Solve for p(x=3) $p(x) = \frac{n!}{x!(n-x)!}(p)^x (q)^{n-x}$

$$p(3) = \frac{5!}{3!(5-3)!}(.6)^3(.4)^2$$

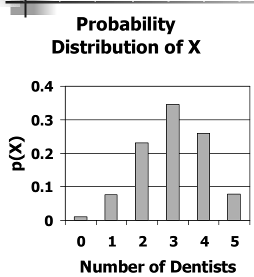
$$p(3) = \frac{120}{6(2)}(.216)(.16) =$$ **.3456**

Probability Distribution of the Discrete Variable X



- **E(X) = 3**
- **σ^2 = 1.1998**
- **σ = 1.09535**

Probability Distribution of the Discrete Variable X



- E(X) = 3
- σ^2 = 1.1998
- σ = 1.09535

- **E(X) = np=5(.6) = 3**
- **σ^2 = npq = 5(.6)(.4) = 1.2**
- **σ = 1.0954**

Example: Seedling Survival

- An agronomist knows from past experience that 80% of a citrus variety seedling will survive being transplanted.
- If we take a random sample of **6 seedlings** from current stock, what is the probability that **exactly 2 seedlings will survive**?

Example: Seedling Survival

- For the problem we can calculate
  - p = .8
  - q = .2
  - μ = np = 6(.8) = 4.8
  - σ^2 = npq = 6(.8)(.2) = .96
  - σ = .98

Example: Seedling Survival

- **Probability that exactly 2 survive is**

$$p(2) = \frac{6!}{2!(6-2)!}(.8)^2(.2)^{6-2}$$

$$= \frac{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(2\cdot 1)(4\cdot 3\cdot 2\cdot 1)}(.64)(.0016)$$

$$= \frac{720}{48}(.001024) = .01536$$

Example: Seedling Survival

- **Probability that exactly 3 survive is**

$$p(3) = \frac{6!}{3!(6-3)!}(.8)^3(.2)^{6-3}$$

$$= \frac{6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1}{(3\cdot 2\cdot 1)(3\cdot 2\cdot 1)}(.512)(.008)$$

$$= \frac{720}{36}(.0041) = .0820$$

Solve for Each Event

Event	Probability
x = 0 All fail	.0001
x = 1 One pass	.0015
x = 2 Two pass	.0154
x = 3 Three pass	.0820
x = 4 Four pass	.2460
x = 5 Five pass	.3932
x = 6 Six pass	.2621

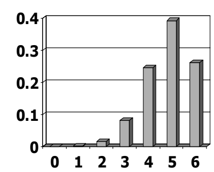
Look at the Cumulative Probabilities

Event	Probability	Cumulative p
x = 0 All fail	.0001	.0001
x = 1 One pass	.0015	.0016
x = 2 Two pass	.0154	.0170
x = 3 Three pass	.0820	.0990
x = 4 Four pass	.2460	.3450
x = 5 Five pass	.3932	.7382
x = 6 Six pass	.2621	1.0000

Citrus Example

- Mean = 6(.8) = 4.8
- Std Dev = .98
- It Makes Sense! Our expectation is that most seedlings will survive (i.e. 4.8 of 6)
- Look at the cumulative probability

Probability Distribiution for Citrus Example



Move to the Binomial Table

- We can also use a table to help
- **Appendix B, Table B2** (Page 734) contains cumulative probabilities for n= 5, 6, 7, 8, 9, 10, 15, 20, and 25
- Each table lists values of p across the top
  - P = .01, .05, .1, .2, .3, …, .95, .99
  - k = # of successes

Binomial Table n=6

P k	.01	.05		.8
0	.941	.735		.000
1	.999	.967		.002
2	1.000	.998		.017
3	1.000	1.000		.099
4	1.000	1.000		.345
5	1.000	1.000		.738

Binomial Table

- NOTE: The table is **cumulative** binomial probabilities, cumulative up to an including the value for k
- This means to find exact probabilities you might have to subtract two table values

Binomial Probabilities Using the Table for Citrus Example

- We said the probability that 4 survive is .2460
- From the Table
  - Cumulative up to 4 is .345
  - Subtract the probabilities for up to 3 (.099)
  - **.345 - .099 = .2460**

You have to be careful using the Table!

Binomial Formula using Excel

- In Excel, the formula for the Binomial Distribution function is:
- BINOMDIST(X,N,P,*cumulative*)
  - X is the number of successes
  - N is the number of independent trials
  - P is the probability of success on each trial
  - *Cumulative* is an argument
    - Entering TRUE gives a cumulative probability up to and including X successes
    - Entering FALSE gives the exact probability of X successes in N trials

Binomial Formula using Excel

- For our example of citrus plants
  - BINOMDIST(2,6,.8,TRUE)
    - cumulative probability up to and including 2 successes
    - .01696
  - BINOMDIST(2,6,.8,FALSE)
    - the exact probability of X successes in N trials
    - = .01536

Look at the Citrus Seedling Table

Event	Probability	Cumulative p
x = 0 All fail	.0001	.0001
x = 1 One pass	.0015	.0016
x = 2 Two pass	**.0154**	**.0170**
x = 3 Three pass	.0820	.0990
x = 4 Four pass	.2460	.3450
x = 5 Five pass	.3932	.7382
x = 6 Six pass	.2621	1.0000

Excel Binomial Distribution File

				p =	0.80
				q =	0.20
X	p(X)	Cum p(X)		n =	6.00
0	0.00006	0.00006			
1	0.00154	0.00160		Mean	4.80
2	0.01536	0.01696		Variance	0.96
3	0.08192	0.09888		Std Dev	0.98
4	0.24576	0.34464			
5	0.39322	0.73786			
6	0.26214	1.00000			
7	#NUM!	#NUM!			
8	#NUM!	#NUM!			
9	#NUM!	#NUM!			
10	#NUM!	#NUM!			
Formula	Cum = FALSE	0.01536		X successes =	2
Formula	Cum= TRUE	0.01696			

The Rare Event Approach

- What if we had 6 seedlings selected randomly and all of them died?
- Given p=.8, this would be a very rare event
  - P(x=0) = .001
- **Was this just by chance????**

Example Problem

- A study in the American Journal of Public Health found that 80% of female Japanese students from heavy-smoking families showed signs of nasal allergies
- Consider a random sample of 25 female Japanese students exposed daily to heavy smoking
- **What is the probability that fewer than 20 of the students will have nasal allergies?**

Answer to Problem

- **What is the probability that more than 15 of the students will have nasal allergies?**

Let's revisit the psychic problem

- Remember that a crystal is randomly placed under one of ten boxes and the psychic is asked to guess where it is.
- This experiment is repeated seven times, and x is the number of correct decisions in seven tries. Thus it is a Binomial random variable.
- If the psychic is guessing, what is the value of p, the probability of a correct decision on each trial?
  - **p = .1**
  - **the probability of a "success" is .1**

Let's view this as a discrete random variable – a binomial random variable

X	0	1	2	3	4	5	6	7
p(x)	.4783					.0002	.0000	.0000

Can you fill in the rest of the table?

To solve

- **Use the table on page 734**
  - n = 7
  - p = .1
  - k = the values of our discrete random variable
- For p(x=1)
  - For k = 1, the probability is .850 which is the cumulative probability up and including 1
  - To find the exact p(x=1), subtract the value for k=0 from the value k = 1
  - **p(x=1) = .850 - .478 = .372**

To solve

- **Use the formula**

$$p(1) = \frac{7!}{1!(7-1)!}(.1)^1(.9)^{7-1}$$

$$= \frac{5040}{720}(.0531) = .3720$$

Solve for all

X	0	1	2	3	4	5	6	7
p(x)	.4783	.3720	.1240	.0230	.0026	.0002	.0000	.0000

Can you solve for the mean and standard deviation of this binomial random variable?

Solve for Expected Value and Variance

X	0	1	2	3	4	5	6	7
p(x)	.4783	.3720	.1240	.0230	.0026	.0002	.0000	.0000

Expected value = mean = n•p = 7*.1 = .7

Variance = n•p•q = 7*(.1)*(.9) = .63

Standard Deviation = (.63).5 = .794

Poisson Distribution

- Applies to situations where we describe the number of events occurring in a specific time period or in a specific area

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

Where $\lambda = \mu$
e = natural logarithm = 2.7183