Discrete Probability Distribution part 2

Week 10

Discrete Probability Distribution

- Bernoulli Distribution
- Binomial Distribution
- Hypergeometric Distribution
- Poisson Distribution

1. Bernoulli Distribution

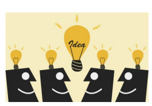
- Consider only two outcomes
- "success" or "failure"

- P : probability of success
- 1-P : probability of failure
- X = 1 if success, x = 0 if failure.

- Bernoulli Probability Function
- P(0) = 1-P
- P(1) = P

- Consider E(X) and Var(X) if X ~ Bernoulli
- $E(X) = \sum xP(x) = 0*(1-P) + 1*(P) = P$
- $Var(X)= E[(x-\mu)^2]=\sum(x-\mu)^2 * P(x) = \ldots$
- $=P(1-P)$. (Solved on board)

Practice



- P(X: making a sale) = 0.4
- 1-P(X) = 0.6.

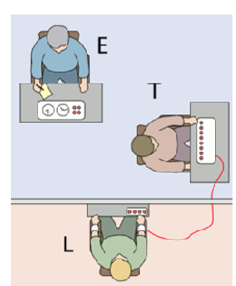
- Find E(X) and Var(X).

- E(X)=1*0.4 + 0*0.6 = 0.4 = P
- Var(X) = P(1-P) = 0.4*0.6=0.24.

Milgram experiment

Stanley Milgram, a Yale University psychologist, conducted a series of experiments on obedience to authority starting in 1963.

- Experimenter (E) orders the teacher (T), the subject of the experiment, to give severe electric shocks to a learner (L) each time the learner answers a question incorrectly.
- The learner is actually an actor, and the electric shocks are not real, but a pre-recorded sound is played each time the teacher administers an electric shock.



http://en.wikipedia.org/wiki/File:Milgram_Experiment_v2.png

Milgram experiment (cont.)

- These experiments measured the willingness of study participants to obey an authority figure who instructed them to perform acts that conflicted with their personal conscience.
- Milgram found that about 65% of people would obey authority and give such shocks.
- Over the years, additional research suggested this number is approximately consistent across communities and time.

Bernoulli random variables

- Each person in Milgram's experiment can be thought of as a *trial*.
- A person is labeled a *success* if she refuses to administer a severe shock, and *failure* if she administers such shock.
- Since only 35% of people refused to administer a shock, *probability of success* is p = 0.35.
- When an individual trial has only two possible outcomes, it is called a *Bernoulli random variable*.

- E(X) from this experiment = P = 0.35
- Var(X) = P(1-P) = 0.35*0.65 = 0.2275

Binomial Probability Distribution

<= n repeated Bernoulli experiments

- A fixed number of observations, n
- (e.g. 15 tosses of a coin)

- Two mutually exclusive and collectively exhaustive categories- "success (P)" and "failure (1-P)".
- (e.g. Head or Tail, Defective, not Defective..)

- Constant probability for each observation.
- (e.g. probability of getting a tail is the same each time we toss a coin)

- Observations are independent.
- : The outcome of one observation does not affect the outcome of the other.

Binomial Probability Distribution

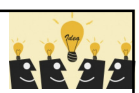
$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$

- P(X): probability of x successes in n trials, with P = probability of success.

- X: # of success in sample. (x= 0, 1, 2, …n)
- n: sample size
- P: probability of success.

Practice

$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$



- Flip a coin 4 times. X = # of heads.

- Q: What is the probability of 3 heads out of 4 flips?

- n = 4
- P=0.5
- X=0,1,2,3,4

$$(X = 3) = \frac{4!}{3!\,(4-3)!} 0.5^3 (1-0.5)^{(4-1)}$$

= 0.25

Why use Combination? $\frac{n!}{x!\,(n-x)!}$

- There are four possible scenarios for 3 Hs out of 4 flips. 4!/(3!1!) = 4.

	1st flip	2nd flip	3rd flip	4th flip
1st scenario	H	H	H	T
2nd scenario	H	H	T	H
3rd scenario	H	T	H	H
4th scenario	T	H	H	H

Binomial distribution
:Milgram's Experiment re-visited.

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$

P(X=1| n=4, p=0.35)
=[4!/(1!3!)]*0.35^1 *0.65^3
=4*0.35*0.65^3
=0.388

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) refuse}}{0.35} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) refuse}}{0.35} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) refuse}}{0.35} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) refuse}}{0.35} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) refuse}}{0.35} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 4: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) refuse}}{0.35} = 0.0961$

Suppose we randomly select four individuals to participate in this experiment. What is the probability that exactly 1 of them will refuse to administer the shock?

Let's call these people Allen (A), Brittany (B), Caroline (C), and Damian (D). Each one of the four scenarios below will satisfy the condition of "exactly 1 of them refuses to administer the shock":

Scenario 1: $\underset{\text{(A) refuse}}{0.35} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 2: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) refuse}}{0.35} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 3: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) refuse}}{0.35} \times \underset{\text{(D) shock}}{0.65} = 0.0961$

Scenario 4: $\underset{\text{(A) shock}}{0.65} \times \underset{\text{(B) shock}}{0.65} \times \underset{\text{(C) shock}}{0.65} \times \underset{\text{(D) refuse}}{0.35} = 0.0961$

The probability of exactly one 1 of 4 people refusing to administer the shock is the sum of all of these probabilities.

$$0.0961 + 0.0961 + 0.0961 + 0.0961 = 4 \times 0.0961 = 0.3844$$

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

*RR*SSSSSSS

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

*RR*SSSSSSS

S*RR*SSSSSS

Computing the # of scenarios

Earlier we wrote out all possible scenarios that fit the condition of exactly one person refusing to administer the shock. If *n* was larger and/or *k* was different than 1, for example, *n* = 9 and *k* = 2:

*RR*SSSSSSS

S*RR*SSSSSS

SS*RR*SSSSS

...

SS*R*SS*R*SSS

...

SSSSSSS*RR*

writing out all possible scenarios would be incredibly tedious and prone to errors.

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

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

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

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

$k = 1, n = 4$: $\binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4 \times 3 \times 2 \times 1}{1 \times (3 \times 2 \times 1)} = 4$

Computing the # of scenarios

Choose function

The *choose function* is useful for calculating the number of ways to choose k successes in n trials.

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

$k = 1, n = 4$: $\binom{4}{1} = \frac{4!}{1!(4-1)!} = \frac{4 \times 3 \times 2 \times 1}{1 \times (3 \times 2 \times 1)} = 4$

$k = 2, n = 9$: $\binom{9}{2} = \frac{9!}{2!(9-1)!} = \frac{9 \times 8 \times 7!}{2 \times 1 \times 7!} = \frac{72}{2} = 36$

Note: You can also use R for these calculations:

```
> choose(9,2)
[1] 36
```

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) There are $n - 1$ ways of getting $n - 1$ successes in n trials, $\binom{n}{n-1} = n - 1$.

Practice

Which of the following is false?

(a) There are n ways of getting 1 success in n trials, $\binom{n}{1} = n$.

(b) There is only 1 way of getting n successes in n trials, $\binom{n}{n} = 1$.

(c) There is only 1 way of getting n failures in n trials, $\binom{n}{0} = 1$.

(d) *There are $n - 1$ ways of getting $n - 1$ successes in n trials, $\binom{n}{n-1} = n - 1$.*

Binomial Distribution
Mean and Variance

- E(X)=np
- Var(X)=E(X-μ)^2 = np(1-p)
- Proof will be shown on the board.

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 100 Americans, how many would you expect to be obese?

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np$ = 100 x 0.262 = 26.2.

Expected value

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 100 Americans, how many would you expect to be obese?

- Easy enough, 100 x 0.262 = 26.2.
- Or more formally, $\mu = np$ = 100 x 0.262 = 26.2.
- But this doesn't mean in every random sample of 100 people exactly 26.2 will be obese. In fact, that's not even possible. In some samples this value will be less, and in others more. How much would we expect this value to vary?

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Going back to the obesity rate:

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

Expected value and its variability

Mean and standard deviation of binomial distribution

$$\mu = np \qquad \sigma = \sqrt{np(1-p)}$$

Going back to the obesity rate:

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.262 \times 0.738} \approx 4.4$$

We would expect 26.2 out of 100 randomly sampled Americans to be obese, with a standard deviation of 4.4.

Note: Mean and standard deviation of a binomial might not always be whole numbers, and that is alright, these values represent what we would expect to see on average.

Unusual observations

Using the notion that *observations that are more than 2 standard deviations away from the mean are considered unusual* and the mean and the standard deviation we just computed, we can calculate a range for the plausible number of obese Americans in random samples of 100.

26.2 ± (2 x 4.4) → (17.4, 35.0)

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

	Excellent	Good	Only fair	Poor	Total excellent/ good
	%	%	%	%	%
Independent private school	31	47	13	2	78
Parochial or church-related schools	21	48	18	5	69
Charter schools	17	43	23	5	60
Home schooling	13	33	30	14	46
Public schools	5	32	42	19	37

Gallup, Aug. 9-12, 2012

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1{,}000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1{,}000 \times 0.13 \times 0.87} \approx 10.6$$

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-

Practice

An August 2012 Gallup poll suggests that 13% of Americans think home schooling provides an excellent education for children. Would a random sample of 1,000 Americans where only 100 share this opinion be considered unusual?

A. Yes B. No

$$\mu = np = 1{,}000 \times 0.13 = 130$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1{,}000 \times 0.13 \times 0.87} \approx 10.6$$

Range of usual observations: $130 \pm 2 \times 10.6 = (108.8, 151.2)$

100 is outside this range, so would be considered unusual.

http://www.gallup.com/poll/156974/private-schools-top-marks-educating-

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. the number of desired successes, *k*, must be greater than the number of trials
E. the probability of success, *p*, must be the same for each trial

Practice

Which of the following is not a condition that needs to be met for the binomial distribution to be applicable?

A. the trials must be independent
B. the number of trials, *n*, must be fixed
C. each trial outcome must be classified as a *success* or a *failure*
D. the number of desired successes, k, must be greater than the number of trials
E. the probability of success, *p*, must be the same for each trial

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

A. pretty high
B. pretty low

Gallup: http://www.gallup.com/poll/160061/obesity-rate-stable-2012.aspx, January 23, 2013.

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

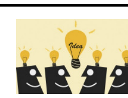
(a) $0.262^8 \times 0.738^2$
(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$
(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2$
(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

Practice

A 2012 Gallup survey suggests that 26.2% of Americans are obese. Among a random sample of 10 Americans, what is the probability that exactly 8 are obese?

(a) $0.262^8 \times 0.738^2$
(b) $\binom{8}{10} \times 0.262^8 \times 0.738^2$
(c) $\binom{10}{8} \times 0.262^8 \times 0.738^2 = 45 \times 0.262^8 \times 0.738^2 = 0.0005$
(d) $\binom{10}{8} \times 0.262^2 \times 0.738^8$

Practice



- 40% of students admitted to university A will actually enroll.

a. What is the probability that at most 1 student will enroll if the college offers admission to 5 students?

$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$

a. What is the probability that at most 1 student will enroll if the college offers admission to 5 students?

$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$

- n=5
- p=0.4
- X: at most 1 (X≤1)

- P(X≤1|n=5,p=0.4) = P(X=0)+P(X=1)
- =[5!/(0!5!)]*(0.4^0)*(0.6^5)
- +[5!/(1!4!)]*(0.4^1)*(0.6^4)
- =0.337

b. What is the probability that between 2 and 4 students (including 2 and 4) will enroll if the college offers admissions to 5 students?

$$P(X) = \frac{n!}{x!\,(n-x)!} P^x (1-P)^{n-x}$$

- n=5
- p=0.4
- X: between 2 and 4 (including 2 and 4)

- P(2≤X≤4|n=5, p=0.4)
- =P(X=2)+P(X=3)+P(X=4)
- =5!/(2!3!)*(0.4)^2*(0.6)^3
- +5!/(3!2!)*(0.4)^3*(0.6)^2
- +5!/(4!1!)*(0.4)^4*(0.6)^1
- =0.653

c. What is the probability that at most 6 students will enroll if the college offers admission to 10 students?

- P(X≤6|n=10,p=0.4))
- =P(0)+P(1)+P(2)+P(3)+P(4)+P(5)+P(6)
- …

⇒Use "Cumulative Binomial Probabilities (Table 3)" table for this calculation.

- By using the table, look for n = 10, x = 6, P=0.4 and read the corresponding P(X≤6|n=10,p=0.4)= 0.945.

d. What is the probability that more than 9 will enroll if 13 students was offered the admissions?

- P(X>9|n=13, p=0.4)

- How can we solve this probability???

- = 1- P(X≤9|n=13,p=0.4)
- =1-0.992 = 0.008

- Since more than 9 will enroll = 10,11,12 or 13 students will enroll = 1 – P(0~9 students will enroll)

X: Number of Success (Enrollment)
0
1
2
3
4
5
6
7
8
9
10
11
12
13

= 1- P(At most 9 OR Less than 10)

More than 9
(or At least 10)

If 70% of these students admitted actually enroll, what is the probability that at least 8 out of 13 students will actually enroll?

- P(X≥8|n=13,p=0.7)

- How can we solve this problem?
- We do not have the table with P>0.5.

⇒Redefine the definition of p as "probability of non-enrollment".

- P(X=8,9,10,11,12 or 13 enrolling| n=13, p=0.7)
- =P(X=5,4,3,2,1 or 0 not enrolling| n=13,p=0.3)
- =P(X≤5| n = 13, p=0.3) = by using the table

X: Number of Success (Enrollment) P(X)=0.7	X' : Number of Failure (Non-Enrollment) P(X')=0.3
0	13
1	12
2	11
3	10
4	9
5	8
6	7
7	6
8	5
9	4
10	3
11	2
12	1
13	0

In summary…

- At least a = $P(X \geq a)$
- At most a = $P(X \leq a)$
- Fewer than a = $P(X<a)$
- More than a = $P(X>a)$

- In order to use the table,
- $P \leq 0.5$ and $P(X \leq a)$

- Otherwise…

- If p > 0.5, then switch "success" and "failure" definitions and use "failure" probability.
- $P(X \geq a | n, p = p_0)$
- $=P(X \leq n-a | n, p= (1-p_0))$

- If $P(X>a|n, p)$, then use
- $1-P(X \leq a|n, p)$

- If $P(X \geq a|n, p)$, then use
- $1-P(X<a|n, p)$
- $=1-P(X \leq a-1|n, p)$

- If $P(X<a|n, p)$ then use
- $P(X \leq a-1|n, p)$

Shapes of binomial distributions

For this activity you will use a web applet. Go to http://socr.stat.ucla.edu/htmls/SOCR_Experiments.html and choose Binomial coin experiment in the drop down menu on the left.

- Set the number of trials to 20 and the probability of success to 0.15. Describe the shape of the distribution of number of successes.

- What happens to the shape of the distribution as *n* stays constant and *p* changes?

- What happens to the shape of the distribution as *p* stays constant and *n* changes?

Distributions of number of successes [start here]

Hollow histograms of samples from the binomial model where *p* = 0.10 and *n* = 10, 30, 100, and 300. What happens as *n* increases?