Rapid Learning Center

Chemistry :: Biology :: Physics :: Math

Rapid Learning Center Presents …

Teach Yourself College Statistics in 24 Hours

1/68

Discrete Probability Distributions

College Statistics Rapid Learning Series

Wayne Huang, PhD
Barry Monk, PhD
Linda Seeger, MA
Jessica Davis, MS
Steward Huang, PhD
Kelly Deters, PhD
Grace Antony, PhD

Rapid Learning Center

www.RapidLearningCenter.com/

2/68

Learning Objectives

By completing this tutorial, you will learn:

- Bernoulli distribution
- Binomial distribution
- Approximation of binomial distribution
- Geometric distribution
- Geometric experiment
- Geometric random variable
- Comparison and relationship with other distributions
- Memoryless property

3/68

Bernoulli Distribution

4/68

Definition: Bernoulli Trial

Bernoulli Trial - The experiment whose outcome is random and can be either of two possible outcomes.

e.g.:
Is a baby a boy?
Are a person's eyes black?
Did a citizen vote for a specific candidate?

Each of these question has two possible outcomes: yes ("**success**") or no ("**failure**")

5/68

Bernoulli Trial - More Examples

Some more examples of Bernoulli Trials:

Example	Success Outcome	Failure Outcome
Flipping a coin	Heads (P=0.5)	Tails (P=0.5)
Rolling a "4" on a die	4	Any other roll
Random polling a voter	"yes" vote	"no" vote

6/68

Bernoulli Trial Assumptions

Independent Trials—the outcome of one trial has no effect on other trials.

For the success state, P = constant for each trial

The success state is always denoted as X=1

Example	Success Outcome	Failure Outcome
Gender of survey participant	Female X = 1	Male
Product testing	Defective X = 1	No defect

7/68

Mean and Variance

Mean	$E(X) = px(1) + qx(0) = p$
Variance	$V(X) = E(X^2) - (E(X))^2$ $= [p(1)^2 + q(0)^2] - (p)^2$ $= p\text{-}p^2 = p(1\text{-}p)$ $= pq$

8/68

Binomial Distribution

9/68

Relationship of Distributions

1. n independent trials
2. Each trial result in only success and failure.
3. The probability of success of each trial p, is constant.

10/68

Relationship between Bernoulli and Binomial Distribution

Bernoulli 

Binomial    ……. 

- Binomial distribution = ∑(n independent Bernoullis)
- It is the number of "successes" in n trials.
- If $Y_1, Y_2, \ldots, Y_n$ are Bernoulli, then X is Binomial:

  $X = Y_1 + Y_2 + \ldots + Y_n$

11/68

Recognizing a Binomial Distribution

Example: According to Statistics, about 40% of men in the United Stated voted for Bush.
A random sample of 15 men is selected and the number who voted for Bush is recorded.
Is this an example of a binomial experiment?

n independent and identical trials:
- Trial: men
- n =15

Two outcomes, Success and Failure:
- Success=voted for Bush
- Failure=did not vote for Bush

Probability of success and failure:
1. P(S) = 0.4
2. P(F) = 0.6

x is the number of successes:

x = number of men who voted for Bush

12/68

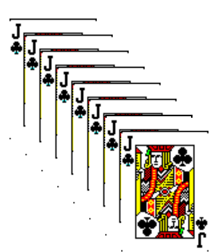
Binomial Distribution Examples

Examples of Binomial Distributions:

- Rolling a die to see if a 2 appears.
- Tossing a coin 10 times to see how many heads occur.
- Asking 100 people if they watch CNN news.

13/68

Non-Binomial Experiments



Examples which aren't binomial experiments:

- Rolling a die until a 1 appears (not a fixed number of trials).
- Asking 12 people how old they are (not two outcomes).
- Drawing 8 cards from a deck for a King (done without replacement, so not independent).

14/68

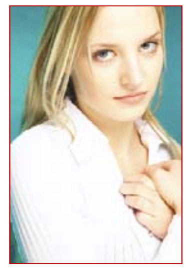
Typical Example

Example: Assume 10% of the population is blue-eyed. You pick 100 people randomly. How likely is it that you get 10 or more blue-eyed people?

The number of blue-eyed people you pick is a random variable X

X follows a binomial distribution with n = 100 and p = 0.1 (when picking the people with replacement).

We are interested in the probability $P(X \geq 10)$.



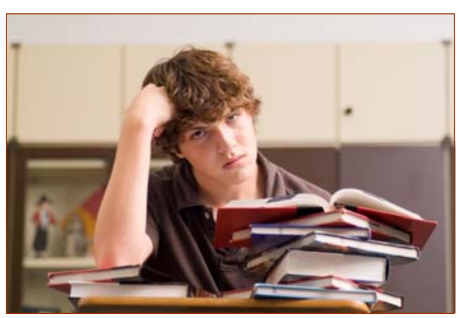
How is this probability found?

15/68

Probability Mass Function - 1

How to find the probability?

n	Probability of?	How can it happen?	Probability function
4	X=0	0000	$(1-p)(1-p)(1-p)(1-p)=(1-p)^4$ So $P(X=0) = (1-p)^4$



16/68

Probability Mass Function - 2

How do we find the probability?

n	Probability	How?	Probability function
4	X=0	0000	$(1-p)(1-p)(1-p)(1-p)=(1-p)^4$ So $P(X=0) = (1-p)^4$
4	X=1	1000, 0100, 0010, 0001	$p(1-p)(1-p)(1-p) + (1-p)p(1-p)(1-p) + (1-p)(1-p)p(1-p)+(1-p)(1-p)(1-p)p = 4p(1-p)^3$ So $P(X=1) = 4p(1-p)^3$

17/68

Probability Mass Function - 3

How do we find the probability?

n	Probability	How?	Probability function
4	X=0	0000	$(1-p)(1-p)(1-p)(1-p)=(1-p)^4$ So $P(X=0) = (1-p)^4$
4	X=1	1000, 0100, 0010, 0001	$p(1-p)(1-p)(1-p) + (1-p)p(1-p)(1-p) + (1-p)(1-p)p(1-p)+(1-p)(1-p)(1-p)p = 4p(1-p)^3$ So $P(X=1) = 4p(1-p)^3$
4	X=2	1100, 1010, 1001, 0110, 0101, 0011	$pp(1-p)(1-p) +p(1-p)p(1-p)+p(1-p)(1-p)p+(1-p)pp(1-p)+(1-p)p(1-p)p+(1-p)(1-p)pp = 6p^2(1-p)^2$ So $P(X=2) = 6p^2(1-p)^2$

18/68

Probability Mass Function - 4

How do we find the probability?

n	Probability	How?	Probability function
4	X=3	1110, 1101, 1011, 0111	ppp(1-p) +pp(1-p)p+p(1-p)pp+(1-p)ppp = $4p^3(1-p)$ So P(X=3) = $4p^3(1-p)$

19/68

Probability Mass Function - 5

How do we find the probability?

n	Probability	How?	Probability function
4	X=3	1110, 1101, 1011, 0111	ppp(1-p) +pp(1-p)p+p(1-p)pp+(1-p)ppp = $4p^3(1-p)$ So P(X=3) = $4p^3(1-p)$
4	X=4	1111	pppp = p^4 So P(X=4) = p^4

20/68

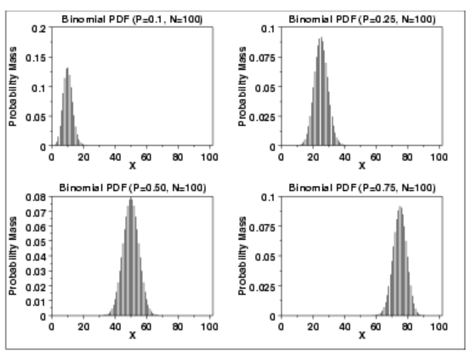
Formula for Probability Mass Function

A summary of the previous slides provides a picture of a general formula.

n = 4	
P(X = 0)	$(1-p)^4$
P(X = 1)	$4p(1-p)^3$
P(X = 2)	$6p^2(1-p)^2$
P(X = 3)	$4p^3(1-p)$
P(X = 4)	p^4
P(X = x)	(# of ways)$p^x(1-p)^{n-x}$

21/68

Formula for Probability



Binomial random variable's probability mass function:

$$P(X = x) = C_x^n p^x (1-p)^{n-x}$$
$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x},$$
$$\text{where } x = 0,1,2,\ldots n,\ 0 < p < 1$$

22/68

Another Example

Example: Suppose we have 100 marbles in a jar 30 of the marbles are green and 70 of them are blue. Define success S as drawing a green marble. If we sample with replacement, P(S)=0.3 for every trial. Find probability of X = 5 if n = 20.

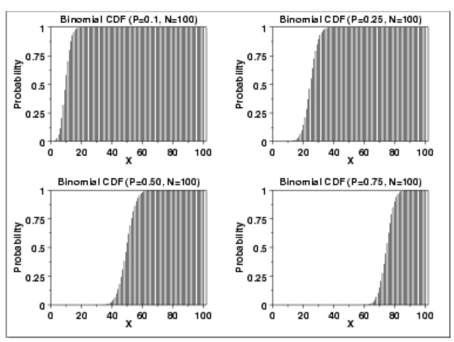
$$P(X=x)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x},$$

$$\text{where } x=5 \text{ and } n=20$$

$$P(x=5)=\frac{20!}{(5!15!)}0.3^5(1-0.3)^{15}=0.1789$$

23/68

Probability Cumulative Function



Binomial random variable's cumulative probability function is:

$$F(x)=\sum_{i=0}^{x} C_i^n p^i (1-p)^{n-i}$$

24/68

Example: Calculating the Probability

Example: A student claims that he gets grades better or equal to "A", 40% of the time. This quarter, he gets only one "A" out of 4 courses. How likely is it that he got one A, or worse, out of four courses given his claim?

$$P(x=0)=\frac{4!}{0!(4!)}(0.4)^0(0.6)^4=1(1)(0.006)=0.1296$$

$$P(x=1)=\frac{4!}{1!(3!)}(0.4)^1(0.6)^3=4(0.4)(0.216)=0.3456$$

$$P(x=0)+P(x=1)=0.5616$$

25/68

Using Binomial Formula - 1

Example: According to Times, about 40% of men in the United Stated voted for Bush. A random sample of 15 men is selected and the number who voted for Bush is recorded.

What is the probability that:

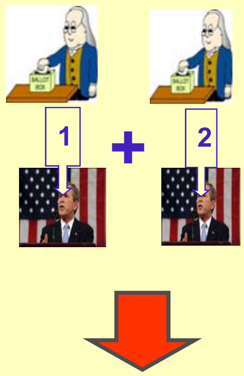
- ☐ exactly 1 of the 15 men voted for Bush?
- ☐ exactly 2 of the 15 men voted for Bush?

$$1.\ P(x=1)=\frac{15!}{1!(14!)}(0.4)^1(0.6)^{14}=0.0047$$

$$2.\ P(x=2)=\frac{15!}{2!(13!)}(0.4)^2(0.6)^{13}=0.0219$$

26/68

Using Binomial Formula - 2



- According to statistics, about **40%** of **men** in the United Stated **voted for Bush**.
- A random sample of **15** men is selected and the number who voted for Bush is recorded.
- What is the probability that either 1 or 2 men voted for Bush?

$$P(1 \le x \le 2) = 0.0047 + 0.0219 = 0.0266$$

27/68

Using Binomial Distribution Table - 1

$$B(x;n,p) = \sum_{0 \le y \le x} b(y;n,p)$$

The values of $B(x;n,p)$ for $0.5 < p < 1.0$ are obtained by using the formula

$$B(x;n,1-p) = 1 - B(n-1-x;n,p)$$

		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
5	0	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031
	1	0.977	0.919	0.835	0.737	0.633	0.528	0.428	0.337	0.256	0.188
	2	0.999	0.991	0.973	0.942	0.896	0.837	0.765	0.683	0.593	0.500
	3	1.000	1.000	0.998	0.993	0.984	0.969	0.946	0.913	0.869	0.813
	4	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.990	0.982	0.969
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
	1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011
	2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
	3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
	4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
	5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
	6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
	7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
	8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	0	0.463	0.206	0.087	0.035	0.013	0.005	0.002	0.000	0.000	0.000
	1	0.829	0.549	0.319	0.167	0.080	0.035	0.014	0.005	0.002	0.000
	2	0.964	0.816	0.604	0.398	0.236	0.127	0.062	0.027	0.011	0.004
	3	0.995	0.944	0.823	0.648	0.461	0.297	0.173	0.091	0.042	0.018
	4	0.999	0.987	0.938	0.836	0.686	0.515	0.352	0.217	0.120	0.059
	5	1.000	0.998	0.983	0.939	0.852	0.722	0.564	0.403	0.261	0.151
	6	1.000	1.000	0.996	0.982	0.943	0.869	0.755	0.610	0.452	0.304
	7	1.000	1.000	0.999	0.996	0.983	0.950	0.887	0.787	0.654	0.500
	8	1.000	1.000	1.000	0.999	0.996	0.985	0.958	0.905	0.818	0.696
	9	1.000	1.000	1.000	1.000	0.999	0.996	0.988	0.966	0.923	0.849
	10	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.991	0.975	0.941
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.982
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

- If n = 15, p = 0.4, use the above Table to calculate.
- P(x=1) =

  P(x≤1) -P(x=0) =

  0.005 -0.000 =

  0.005

28/68

Using Binomial Distribution Table - 2

$$B(x;n,p) = \sum_{0 \le y \le x} b(y;n,p)$$

The values of $B(x;n,p)$ for $0.5 < p < 1.0$ are obtained by using the formula

$$B(x;n,1-p) = 1 - B(n-1-x;n,p)$$

		p									
n	*x*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
5	0	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031
	1	0.977	0.919	0.835	0.737	0.633	0.528	0.428	0.337	0.256	0.188
	2	0.999	0.991	0.973	0.942	0.896	0.837	0.765	0.683	0.593	0.500
	3	1.000	1.000	0.998	0.993	0.984	0.969	0.946	0.913	0.869	0.813
	4	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.990	0.982	0.969
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
	1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011
	2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
	3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
	4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
	5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
	6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
	7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
	8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	0	0.463	0.206	0.087	0.035	0.013	0.005	0.002	0.000	0.000	0.000
	1	0.829	0.549	0.319	0.167	0.080	0.035	0.014	0.005	0.002	0.000
	2	0.964	0.816	0.604	0.398	0.236	0.127	0.062	0.027	0.011	0.004
	3	0.995	0.944	0.823	0.648	0.461	0.297	0.173	0.091	0.042	0.018
	4	0.999	0.987	0.938	0.836	0.686	0.515	0.352	0.217	0.120	0.059
	5	1.000	0.998	0.983	0.939	0.852	0.722	0.564	0.403	0.261	0.151
	6	1.000	1.000	0.996	0.982	0.943	0.869	0.755	0.610	0.452	0.304
	7	1.000	1.000	0.999	0.996	0.983	0.950	0.887	0.787	0.654	0.500
	8	1.000	1.000	1.000	0.999	0.996	0.985	0.958	0.905	0.818	0.696
	9	1.000	1.000	1.000	1.000	0.999	0.996	0.988	0.966	0.923	0.849
	10	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.991	0.975	0.941
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.982
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

- If n=15, p=0.4, use the above Table to calculate.
- P(x=2) =

  P(x≤2) -P(x≤1) =

  0.027 -0.005 =

  0.022

29/68

Using Binomial Distribution Table - 3

$$B(x;n,p) = \sum_{0 \le y \le x} b(y;n,p)$$

The values of $B(x;n,p)$ for $0.5 < p < 1.0$ are obtained by using the formula

$$B(x;n,1-p) = 1 - B(n-1-x;n,p)$$

		p									
n	*x*	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
5	0	0.774	0.590	0.444	0.328	0.237	0.168	0.116	0.078	0.050	0.031
	1	0.977	0.919	0.835	0.737	0.633	0.528	0.428	0.337	0.256	0.188
	2	0.999	0.991	0.973	0.942	0.896	0.837	0.765	0.683	0.593	0.500
	3	1.000	1.000	0.998	0.993	0.984	0.969	0.946	0.913	0.869	0.813
	4	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.990	0.982	0.969
	5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
10	0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
	1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011
	2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
	3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
	4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
	5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
	6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
	7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
	8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
	10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
15	0	0.463	0.206	0.087	0.035	0.013	0.005	0.002	0.000	0.000	0.000
	1	0.829	0.549	0.319	0.167	0.080	0.035	0.014	0.005	0.002	0.000
	2	0.964	0.816	0.604	0.398	0.236	0.127	0.062	0.027	0.011	0.004
	3	0.995	0.944	0.823	0.648	0.461	0.297	0.173	0.091	0.042	0.018
	4	0.999	0.987	0.938	0.836	0.686	0.515	0.352	0.217	0.120	0.059
	5	1.000	0.998	0.983	0.939	0.852	0.722	0.564	0.403	0.261	0.151
	6	1.000	1.000	0.996	0.982	0.943	0.869	0.755	0.610	0.452	0.304
	7	1.000	1.000	0.999	0.996	0.983	0.950	0.887	0.787	0.654	0.500
	8	1.000	1.000	1.000	0.999	0.996	0.985	0.958	0.905	0.818	0.696
	9	1.000	1.000	1.000	1.000	0.999	0.996	0.988	0.966	0.923	0.849
	10	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.991	0.975	0.941
	11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.994	0.982
	12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.996
	13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

- If n=15, p=0.4, use the above Table to calculate.
- P(1≤x≤2) =

  P(x≤2) - P(x=0) =

  0.027-0.000 =

  0.027

30/68

Mean and Variance

Mean and variance are as follows:

$$\mu = E(X) = E\left(\sum_{i=1}^{n} Z_i\right) = \sum_{i=1}^{n} E(Z_i)$$
$$= \sum_{i=1}^{n} p = np$$

$$\sigma^2 = Var(X) = Var\left(\sum_{i=1}^{n} Y_i\right) = \sum_{i=1}^{n} Var(Y_i)$$
$$= \sum_{i=1}^{n} p(1-p) = np(1-p)$$

31/68

Example: Mean and Variance

Example: If n = 20, p = 0.3, find the mean and variance.

E(x) = np
E(x) = 20(0.3) = 6

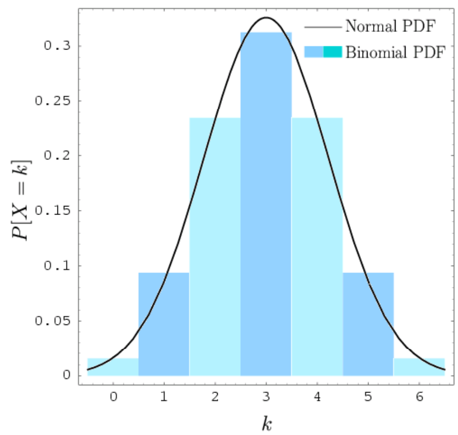
Var(x) = np(1-p)
Var(x) = 20(0.3)(0.7) = 4.2

32/68

Normal Approximation

The normal distribution can give an excellent approximation to B(*n*, *p*) when the following conditions are met:

1. $np \geq 5$
2. $n(1 - p) \geq 5$



33/68

Normal Approximation of a Binomial Distribution - 1

Records show that 30% of all payments to a certain company are late. Let r be a random variable that represents the number of late payments. Suppose 50 payments are submitted this week. Estimate the probability that between 20 and 25 payments are late this week.

Find $P(20 \leq r \leq 25)$

First check to see that both np and nq are greater than 5. Then calculate the mean and the standard deviation:

$$\mu = np = 50 \times .30 = 15$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{50(.3)(.7)} = 3.24$$

34/68

Normal Approximation of a Binomial Distribution - 2

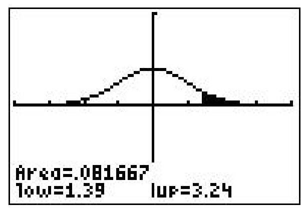
Continuity Correction:
The continuity correction is needed because we are approximating a discrete probability distribution with a continuous distribution.

If r is the left endpoint we subtract .5 to get the corresponding normal variable x, if r is a right endpoint of an interval we add .5 to get the corresponding variable x.

P(20≤ r ≤25) is approximated with P(19.5 ≤ x ≤ 25.5)
Where x is the corresponding normal variable.

$$z = \frac{19.5 - 15}{3.24} = 1.39 \qquad z = \frac{25.5 - 15}{3.24} = 3.24$$

$$p(1.39 < z < 3.24) = .082$$



35/68

Poisson Approximation

- The Binomial distribution converges towards the Poisson distribution as the number of trials goes to infinity while the product *np* remains fixed.

Therefore the Poisson distribution with parameter λ = *np* can be used as an approximation to B(*n*, *p*) of the binomial distribution

- if n is sufficiently large and p is sufficiently small.

According to one rule of thumb, this approximation is good.
if $n \geq 20$ and $p \leq 0.05$, and also
if $n \geq 100$ and $np \leq 10$

36/68

Limits of Binomial Distribution - 1

- As *n* approaches ∞ and
- *p* approaches 0
- while *np* remains fixed at λ > 0 or
- at least *np* approaches λ > 0,

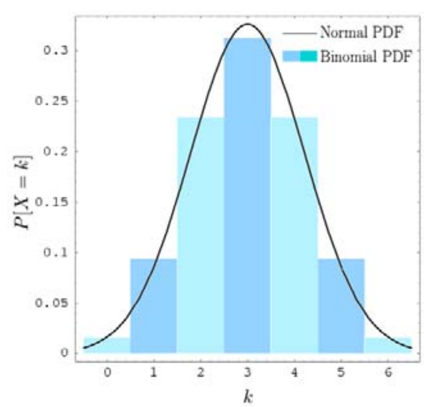
Then the Binomial(*n*, *p*) distribution approaches the Poisson distribution with expected value λ.

37/68

Limits of Binomial Distribution - 2

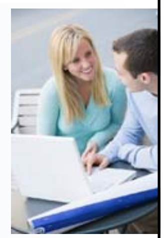
As *n* approaches ∞ while *p* remains fixed, the following distribution approaches the normal distribution with mean 0 and variance 1.

$$\frac{X - np}{\sqrt{np(1-p)}}$$



38/68

Geometric Distribution

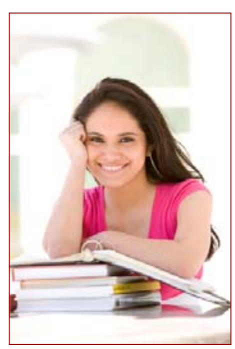


39/68

Definition: Geometric Distribution

Geometric Distribution – The property of the number of times needed to do something until getting a desired result.

Examples:
How many times will I throw a coin until it lands on a *head*?
How many children will I have until I get a boy?
How many cards will I draw from a pack until I get a King?



40/68

Geometric Distributions

A geometric distribution meets the following requirements:

1. Each observation falls into one of two categories, either "**success**" or "**failure**."
2. The probability of a success, call it **p**, is the same for each observation.
3. The observations are all **independent** (this allows us to multiply probabilities).
4. The variable of interest is the number of trials required to obtain **the first success.**

41/68

Recognizing a Geometric Distribution - 1

Example: An experiment consists of rolling a single die. The event of interest is rolling a 2; this event is called a success. Is this a geometric experiment?

- Rolling a 2 will represent a success, and rolling any other number will represent a failure.
- The probability of rolling a 2 on each roll is the same, p = 1/6.
- The observations are independent.
- A trial consists of rolling the die once. We roll the die until the first 2 appears.
- Since all of the requirements are satisfied, this experiment describes a geometric distribution.

42/68

Recognizing a Geometric Distribution - 2

Example: An experiment consists of tossing a single coin. The event of interest is tossing a head; this event is called a success. Is this a geometric experiment?

- Tossing a Head will represent a success, and Tails will represent a failure.
- The probability of tossing a Head on each toss is the same, p = 1/2.
- The observations are independent.
- A trial consists of tossing a Head once. We toss the coin until the first Head appears.
- Since all of the requirements are satisfied, this experiment describes a geometric distribution.

43/68

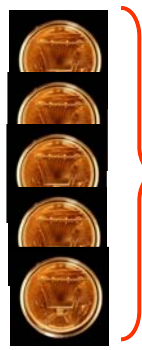
Recognizing a Geometric Distribution - 3

Example: An experiment consists of drawing cards from a single deck without replacement. The event of interest is drawing a King; this event is called a success. Is this a geometric experiment?

- Draw a King from a pack will represent a success, and drawing any others will represent a failure.
- The probability of drawing a King in each draw will not be the same as there is not replacement.
- This does not describe a geometric distribution.

44/68

Geometric Random Variable



Failures

Success

Geometric Random Variables

- This is an example of a *waiting time problem.*
- wait until a certain event occurs…
- or the number of Bernoulli trials which must be conducted before a trial results in a success.

45/68

Binomial and Geometric Distributions

Binomial Distribution

- Fixed (pre-determined) number of trials
- Variable number of success

Geometric Distributions

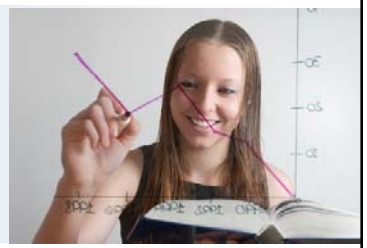
- Variable number of trials (until a single success, no matter how long)
- Fixed number of successes (1)

Both distributions

- Independent trials
- 2 possible random outcomes
- Probability of success in any one trial is *p* and is constant.

46/68

Probability and Geometric Distributions



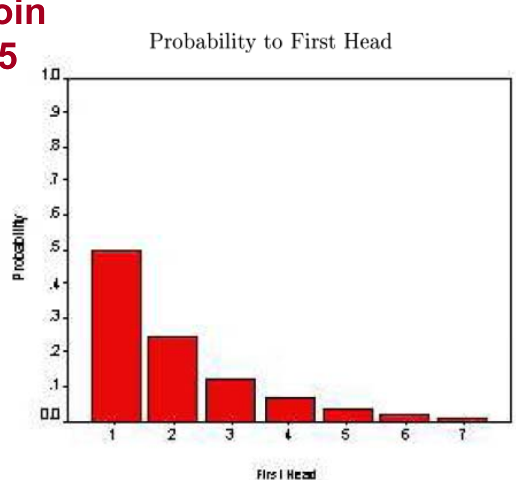
47/68

Probability to First Head

Fair Coin
p=0.5

Failures

Success



First Head	1	2	3	4	5	6	7	8	9	10
Probability	.5	.25	.125	.0625	.0313	.0156	.0078	.0039	.0020	.0010

48/68

Memoryless Property

The geometric distribution is the only discrete memoryless random distribution.

- It is a discrete analog of the exponential distribution.
- This means that the chance of getting a heads up on the 7th trial after failing the first 6 times is the same probability as getting a heads on any of the first 6 trials.
- The random process does not "remember" the number of failures.

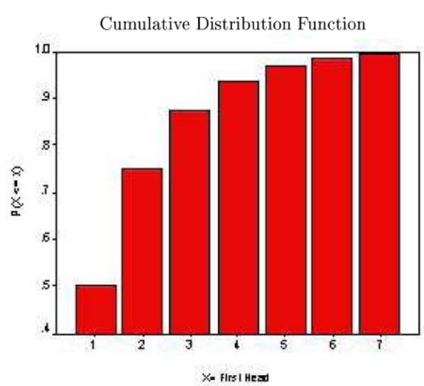
Notice the graph on the previous slide showed the probability of a toss being the FIRST head decreases with increasing number of tosses…but the probability of getting a heads on each toss is the same.

49/68

Cumulative Probability

Fair Coin
p=0.5

Failures



Success

First Defective	1	2	3	4	5	6	7	8
Probability	.5	.75	.875	.09375	.9688	.9844	.9922	.9961

50/68

Probability Distributions

Failures

Success

Geometric distribution can be either one of the following:

- the probability distribution of the number *X* of Bernoulli trials needed to get one success, supported on the set { 1, 2, 3, ...},
- the probability distribution of the number *Y* = *X* − 1 of failures before the first success, supported on the set { 0, 1, 2, 3, ... }.

51/68

The Probability Mass Function

Suppose that John is at a party and he starts asking girls to dance.

Let X be the number of girls that John needs to ask in order to find a partner.

If the ___ girl accepts	X is	And the probability of that happening is ___
first	1	P = p
second	2	P = (1-p)p

Probability of a single trial being a success is p

Probability that the first trial failed.

Probability that the second trial is a success.

52/68

The Probability Mass Function

Suppose that John is at a party and he starts asking girls to dance.

Let X be the number of girls that John needs to ask in order to find a partner.

If the ___ girl accepts	X is	And the probability of that happening is ___
first	1	P = p
second	2	P = (1-p)p
third	3	P = (1-p)(1-p)p P = $(1-p)^2p$
n	n	P = $(1-p)^{n-1}p$

All girls through "n-1" were a failure

Probability that "n" girl is a success

53/68

Probability Mass Function

For the probability of success on each trial *p*,

- The probability that *k* trials are needed to get one success is:
  - $P(X=k)=(1-p)^{k-1}p$ for *k* = 1, 2, 3, ...

Example: A die is thrown repeatedly until the first time a "5" appears. What is the probability that the first time a "5" appears is the third roll?

It's a geometric distribution with *p* = 1/6

k = 3

$P(X=3) = (1-1/6)^{3-1}(1/6)$

P(X=3) = 0.12

54/68

Geometric Sequence

What is the probability that it will take more than n tries to succeed?

- If John asks an infinite number of girls to dance, eventually one of them will accept.
- So, the probability that it will take more than n tries is the same as the probability that John fail n times. That is …
- $P(X > n) = (1-p)^n$

55/68

Example #1

Example: We toss a certain biased coin and get heads 75 % of the time.
The random variable X (the number of the first toss that results in heads), is geometrically distributed with p = 0.75.
What is the probability mass function?

A Biased Coin p=0.75

$p = 0.75$

$n = n$

$P = (1-p)^{n-1}p$

$P = (1-0.75)^{n-1}(0.75)$

$P = (0.25)^{n-1}(0.75)$

56/68

Example #2

Example: Find the probability that the coin in the previous example is heads-up on the third trial.

A Biased Coin p=0.75

$p = 0.75$

$n = 3$

$P = (1-p)^{n-1}p$

$P = (1-0.75)^{3-1}(0.75)$

$P = (0.25)^{2}(0.75) = 0.047$

57/68

Example #3

Example: The probability of rolling a 2 on a die is 1/6 for each trial.
What is the probability of rolling a 2 for the first time on the 6th roll?

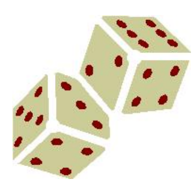
$p = 0.17$

$n = 6$

$P = (1-p)^{n-1}p$

$P = (1-0.17)^{6-1}(0.17)$

$P = (0.83)^{5}(0.17) = 0.067$



58/68

Example #4

Example: A boy is trying to find the key to his father's car on a dark night, out of a keychain with 10 different keys.
What is the probability of the boy succeeding in the 4th trial?

$p = 1/10 = 0.10$

$n = 4$

$P = (1-p)^{n-1}p$

$P = (1-0.10)^{4-1}(0.10)$

$P = (0.90)^{3}(0.10) = 0.073$

59/68

Mean and Variance

60/68

Mean and Variance

Mean and variance are as follows:

$$\mu = E(X) = \frac{1}{p}$$

$$\sigma^2 = Var(X) = \frac{(1-p)}{p^2}$$

61/68

Mean and Variance Example #1

Example: If the probability of John's request for a dance being accepted is 0.5, how many girls, on average, will he have to ask?

p = 0.5

E(X) = 1/p

E(X) = 1/0.5

On average, he'll need to ask 2 girls before he is successful

62/68

Mean and Variance Example #2

Example: What if the probability of any one girl agreeing to dance with John is only 0.25? How many girls will he have to ask, on average?

$p = 0.25$

$E(X) = 1/p$

$E(X) = 1/0.25$

On average, he'll need to ask 4 girls before he is successful

63/68

Negative Binomial Distributions

64/68

Negative Binomial Distribution

Negative binomial distribution is the sum of geometric distributions.

Geometric distribution:
How many trials will be needed before you have one success?

→

Sum of multiple geometric distributions:
How many trials will you need before you have "n" number of successes?

→

Negative binomial distribution:
X_n = number of trials needed until the n^{th} success

65/68

Learning Summary

- Binomial and geometric distributions.
- Geometric relates to Negative Binomial distribution.
- The difference between geometric and binomial experiments.
- How to calculate binomial and geometric probabilities.
- Geometric distributions are memoryless.
- The Normal distribution and Poisson distribution are used to approximate Binomial distributions.

66/68

Congratulations

You have successfully completed the tutorial

Discrete Probability Distributions

Rapid Learning Center

67/68

Rapid Learning Center

Chemistry :: Biology :: Physics :: Math

What's Next …

Step 1: Concepts – Core Tutorial (Just Completed)

→ **Step 2: Practice – Interactive Problem Drill**

Step 3: Recap – Super Review Cheat Sheet

Go for it!

http://www.RapidLearningCenter.com

68/68