Introduction to Probability and Statistics Fourteenth Edition

Chapter 5
Several Useful Discrete Distributions

5.1 Discrete RVs and their Probability distributions

Key ideas

- Random variable
- Discrete random variables
- Probability distribution
- The calculation of mean, variance, standard deviation.

Probability Distributions

- Chapters 1 and 2: Relative frequency
- Chapter 4: Probabilty
- Chatper 5 ad 6: Discrete and Continuous Random Variables

We define the probability distribution for a random variable X as the relative frequency distribution constructed for the entire population of measurements.

Random variables

- A **variable** X is a **random variable** if the value that it assumes, corresponding to the outcome of an experiment, is a chance or random event.
- It can be **discrete** and **continuous**.

Examples

- X: Number of defects on a randomly selected piece of furniture
- X: SAT score for a randomly selected college applicant
- X: Number of telephone calls received by a crisis interventi on hotline during a randomly selected time period

Probability Distributions for Discrete Random Variables

- The **probability distribution for a discrete random variable *x*** resembles the relative frequency distributions we constructed in Chapter 1.
- It is a graph, table or formula that gives the possible values of x and the probability $p(x)$ associated with each value.

We must have

$$0 \le p(x) \le 1 \text{ and } \sum p(x) = 1$$

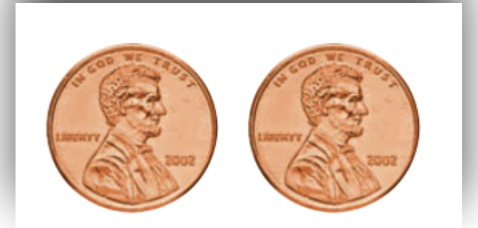
Here, p(x) is called the frequency function or the probability mass function (pmf).

EXAMPLE 5.1

- Toss two fair coins and let X equal the number of heads observed. Find the probability distribution for X.

Table 5.1 Simple Events and Probabilities in Tossing Two Coins

Simple Event	Coin 1	Coin 2	$P(E_i)$	x
E_1	H	H	1/4	2
E_2	H	T	1/4	1
E_3	T	H	1/4	1
E_4	T	T	1/4	0



-

Table 5.2 Probability Distribution for x (x = Number of Heads)

x	Simple Events in x	$p(x)$
0	E_4	1/4
1	E_2, E_3	1/2
2	E_1	1/4
		$\Sigma\, p(x) = 1$

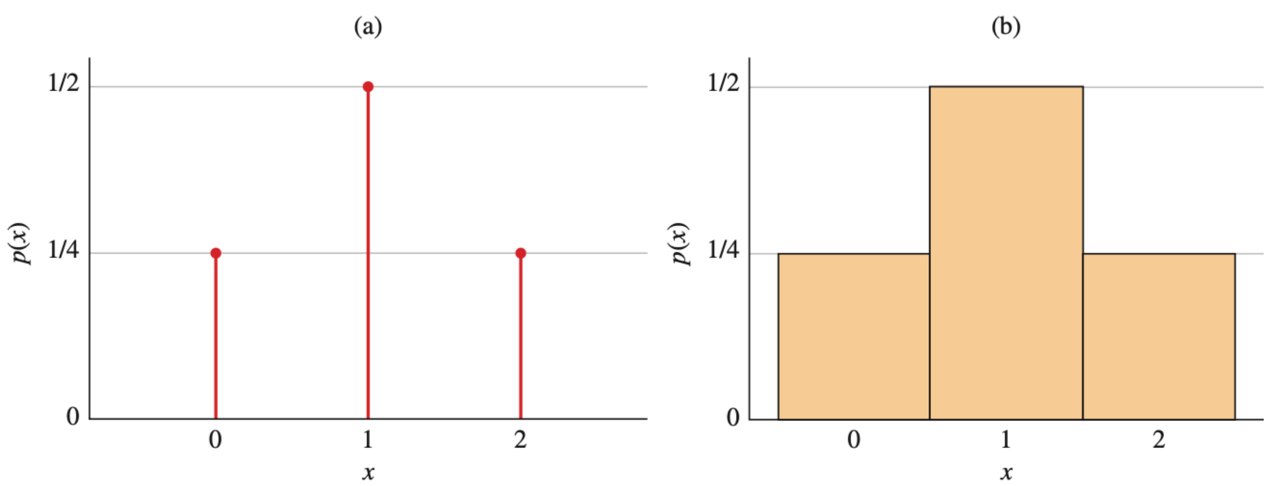


Figure 5.1
Probability histograms for Example 5.1

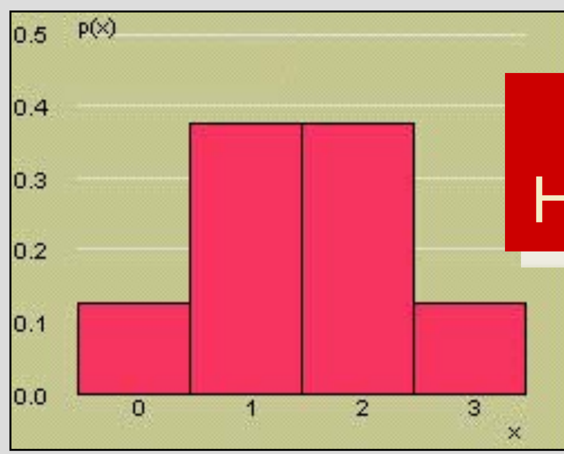
Example

- Toss a fair coin three times and define **x = number of heads.**

		x
HHH	1/8	3
HHT	1/8	2
HTH	1/8	2
THH	1/8	2
HTT	1/8	1
THT	1/8	1
TTH	1/8	1
TTT	1/8	0

$P(x = 0) = 1/8$
$P(x = 1) = 3/8$
$P(x = 2) = 3/8$
$P(x = 3) = 1/8$

x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8



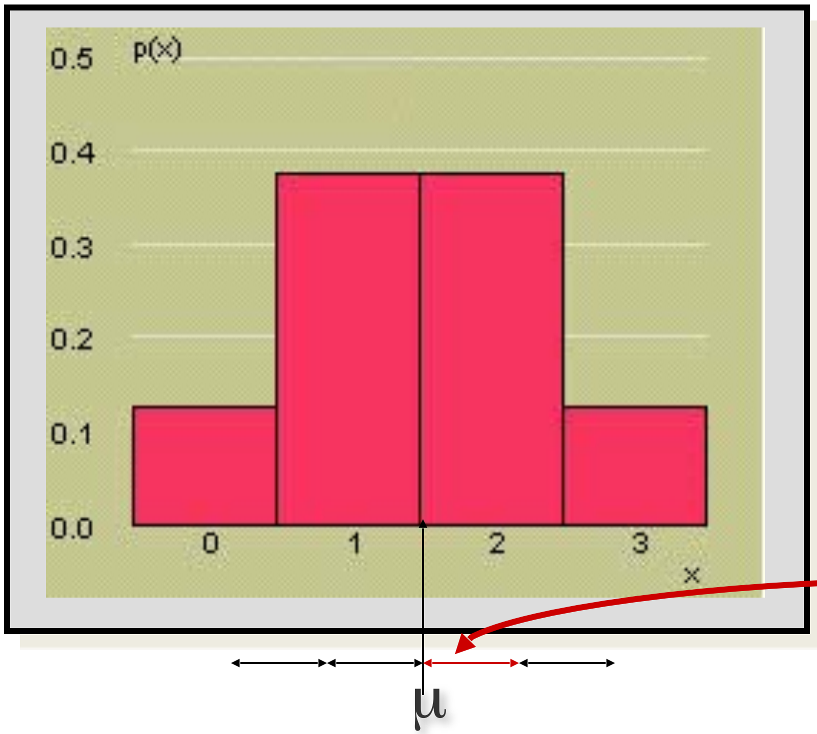
Probability Histogram for x

Probability Distributions

- Probability distributions can be used to describe the population, just as we described samples in Chapter 1.
  - **Shape:** Symmetric, skewed, mound-shaped…
  - **Outliers:** unusual or unlikely measurements
  - **Center and spread:** mean and standard deviation. A population mean is called μ and a population standard deviation is called σ.

Example

- The probability distribution for x the number of heads in tossing 3 fair coins.



- Shape? Symmetric; mound-shaped
- Outliers? None
- Center? $\mu = 1.5$
- Spread? $\sigma = .688$

THE MEAN AND STANDARD DEVIATION

- Let X be a discrete random variable with probability distribution $p(x)$.
- The mean or expected value of X is

$$\mu = \mathrm{E}[X] = \sum xp(x)$$

- The variance of X is

$$\sigma^2 = \mathrm{E}[(X - \mu)^2] = \sum (x - \mu)^2 p(x)$$

- The standard deviation of X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum (x - \mu)^2 p(x)}$$

Example

- Toss a fair coin 3 times and record x the number of heads.

x	$p(x)$	$xp(x)$	$(x-\mu)^2p(x)$
0	1/8	0	$(-1.5)^2(1/8)$
1	3/8	3/8	$(-0.5)^2(3/8)$
2	3/8	6/8	$(0.5)^2(3/8)$
3	1/8	3/8	$(1.5)^2(1/8)$

$$\mu = \sum xp(x) = \frac{12}{8} = 1.5$$

$$\sigma^2 = \sum (x-\mu)^2 p(x)$$

$$\sigma^2 = .28125 + .09375 + .09375 + .28125 = .75$$

$$\sigma = \sqrt{.75} = .688$$

FIND μ AND σ OF EXAMPLE 5.1

- $\mu = 0 \times 1/4 + 1 \times 1/2 + 2 \times 1/4 = 1$
- $\sigma^2 = (0-1)^2 \times 1/4 + (1-1)^2 \times 1/4 + (2-1)^2 \times 1/4 = 1/2$
- $\sigma = \sqrt{1/2} = 0.71$

x	0	1	2
$p(x)$	1/4	1/2	1/4

DISCRETE RVS

- Discrete random variables take on only a finite or countably infinite number of values.
- Three discrete probability distributions serve as models for a large number of practical applications:

✓Ch 5.2: The **binomial** random variable

✓Ch 5.3: The **Poisson** random variable

✓Ch 5.4: The **hypergeometric** random variable

5.2 Binomial Distribution

The Bernoulli Trail

1. Each trial results in **one of two outcomes**, success (S) or failure (F).
2. The probability of success on a single trial is p and **remains constant** from trial to trial. The probability of failure is $q = 1 - p$.
3. Bernoulli random variable defines X as 1 for a success and 0 for a failure in a Bernoulli trial, denoted as $X \sim Bernoulli(p)$. Its probability distribution is

$$p(1) = p, \quad p(0) = 1 - p.$$

Or,

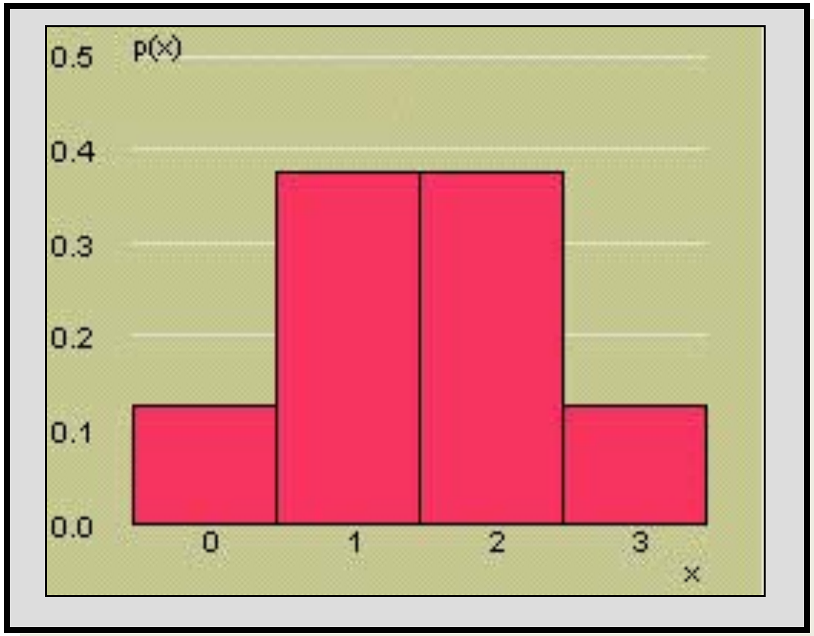
$$p(x) = p^x(1-p)^{1-x}, \quad for\ x = 0,1.$$

Jacob Bernoulli

1655-1705, Swiss Mathematician

The Binomial Random Variable

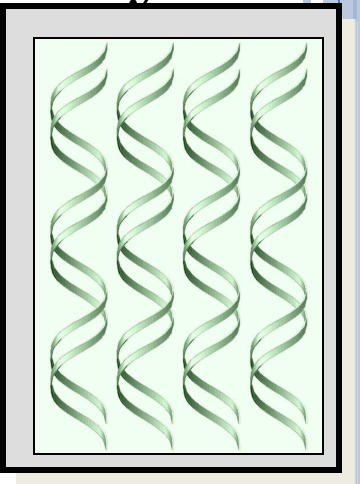
- The **coin-tossing experiment** is a simple example of a **binomial random variable.** Toss a fair coin $n = 3$ times and record x = number of heads.



x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

The Binomial Random Variable

- Many situations in real life resemble the coin toss, but the coin is not necessarily fair, so that $P(H) \neq 1/2$.



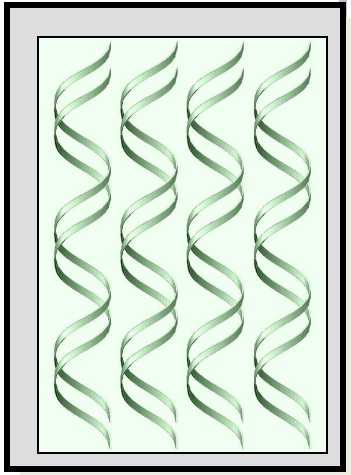
- **Example:** A geneticist samples 10 people and counts the number who have a gene linked to Alzheimer's disease.

- **Coin:** Person
- **Head:** Has gene
- **Tail:** Doesn't have gene
- **Number of tosses:** $n = 10$
- **P(H):** P(has gene) = proportion in the population who have the gene.

The Binomial Experiment

1. The experiment consists of ***n* identical Bernoulli trials**.
2. Each trial results in **one of two outcomes**, success (S) or failure (F).
3. The probability of success on a single trial is p and **remains constant** from trial to trial. The probability of failure is $q = 1 - p$.
4. The trials are **independent**.
5. We are interested in X, **the number of successes in *n* trials.**

Binomial or Not?



- Very few real life applications satisfy these requirements exactly.

- Select two people from the U.S. population, and suppose that 15% of the population has the Alzheimer's gene.
  - For the first person, p = P(gene) = .15
  - For the second person, $p \approx$ P(gene) = .15, even though one person has been removed from the population.

The Binomial Probability Distribution

- For a binomial experiment with n trials and probability p of success on a given trial, the probability of k successes in n trials is

$$P(x=k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \text{ for } k = 0,1,2,\ldots n.$$

$$\text{Recall} \quad C_k^n = \frac{n!}{k!(n-k)!}$$

$$\text{with } n! = n(n-1)(n-2)\ldots(2)1 \text{ and } 0! \equiv 1.$$

The Mean and Standard Deviation

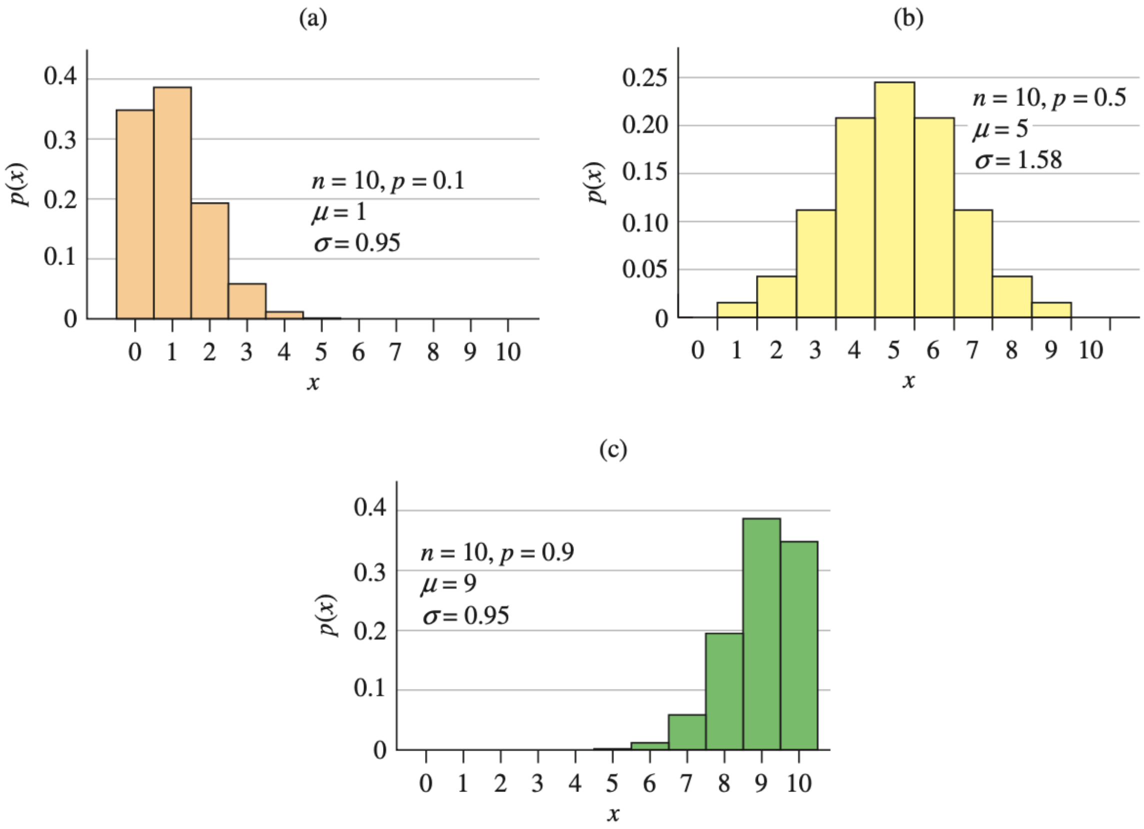
- For a binomial experiment with n trials and probability p of success on a given trial, the measures of center and spread are:

$$\text{Mean}: \mu = np$$

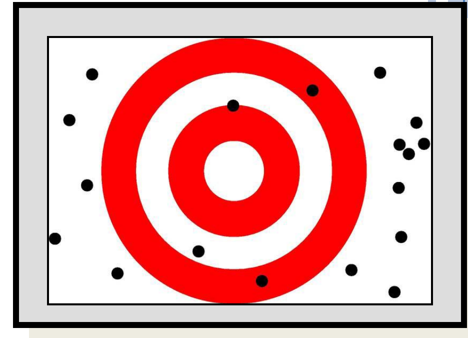
$$\text{Variance}: \sigma^2 = npq$$

$$\text{Standard deviation}: \sigma = \sqrt{npq}$$

Fig 5.3 Binomial Probability Distributions



EXAMPLE



A marksman hits a target 80% of the time. He fires five shots at the target. What is the probability that exactly 3 shots hit the target?

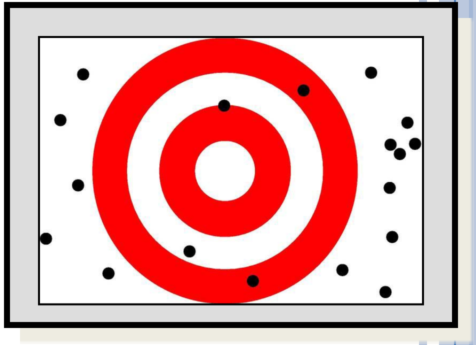
X~ Binomial(n=5, p =0.8)

n **= 5** ***success*** **= hit** ***p*** **= .8** ***x*** **= # of hits**

$$P(x=3) = C_3^5 p^3 q^{5-3} = \frac{5!}{3!2!}(.8)^3(.2)^{5-3}$$

$$= 10(.8)^3(.2)^2 = .2048$$

Example

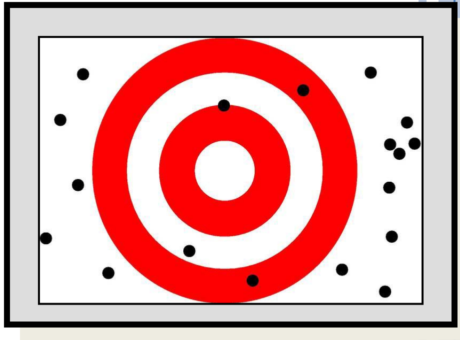


What is the probability that more than 3 shots hit the target?

$$P(x>3)=C_4^5 p^4 q^{5-4}+C_5^5 p^5 q^{5-5}$$

$$=\frac{5!}{4!1!}(.8)^4(.2)^1+\frac{5!}{5!0!}(.8)^5(.2)^0$$

$$=5(.8)^4(.2)+(.8)^5=.7373$$



Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected binomial distributions.

✓Find the table for the correct value of ***n*.**

✓Find the column for the correct value of ***p*.**

✓The row marked "*k*" gives the cumulative probability, $P(x \leq k) = P(x = 0) + \ldots + P(x = k)$

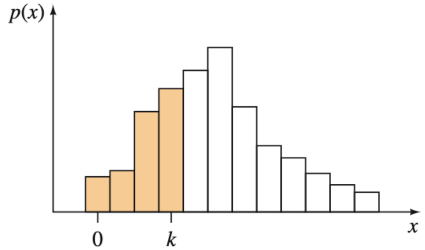


Table 1 Cumulative Binomial Probabilities

Tabulated values are $P(x \le k) = p(0) + p(1) + \cdots + p(k)$.
(Computations are rounded to the third decimal place.)

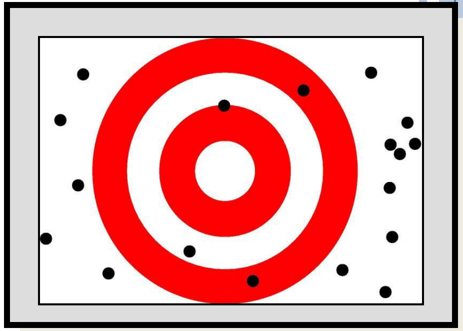
$n = 2$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	.000	0
1	1.000	.998	.990	.960	.910	.840	.750	.640	.510	.360	.190	.098	.020	1
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2

$n = 3$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	.000	.000	0
1	1.000	.993	.972	.896	.784	.648	.500	.352	.216	.104	.028	.007	.000	1
2	1.000	1.000	.999	.992	.973	.936	.875	.784	.657	.488	.271	.143	.030	2
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3

EXAMPLE



k	p = .80
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000

What is the probability that exactly 3 shots hit the target?

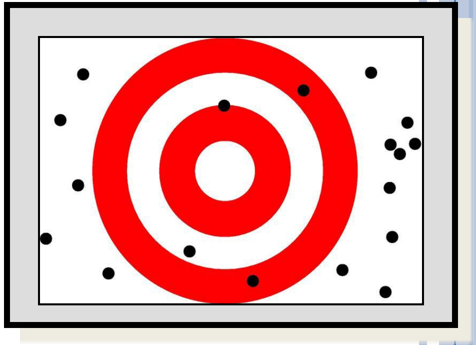
P(x = 3) = P($x \leq 3$) – P($x \leq 2$)
= .263 - .058
= .205

Check from formula:
P(x = 3) = .2048

$n = 5$

	p												
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5

EXAMPLE



k	$p = .80$
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000

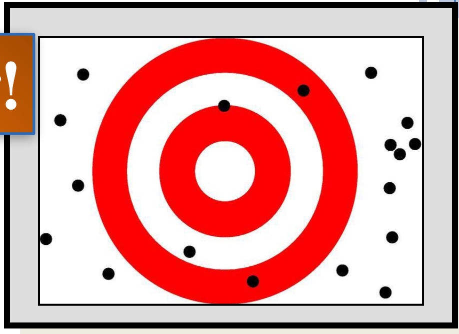
What is the probability that more than 3 shots hit the target?

$P(x > 3)$ $= 1 - P(x \leq 3)$

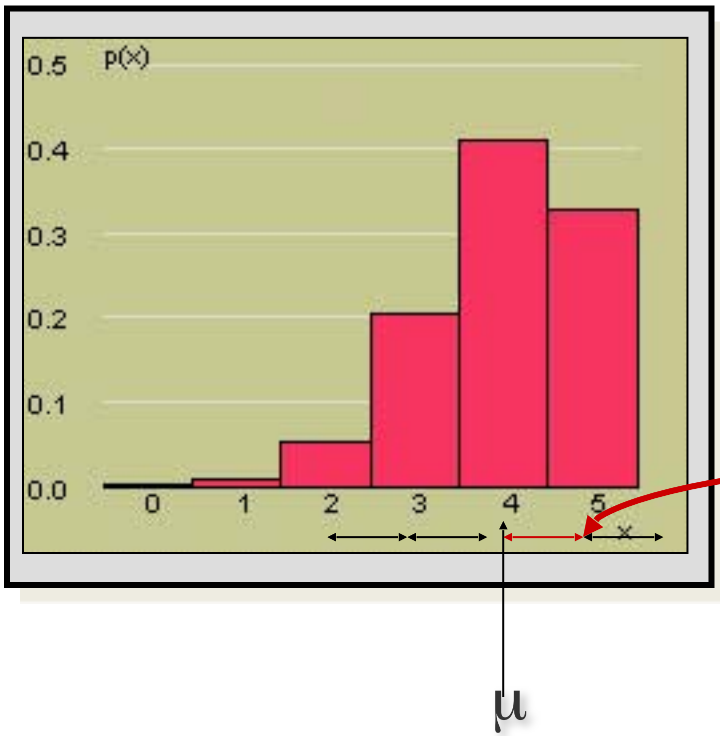
$= 1 - .263 = .737$

Check from formula:
$P(x > 3) = .7373$

Example



- Here is the probability distribution for x = **number of hits.** What are the mean and standard deviation for x?



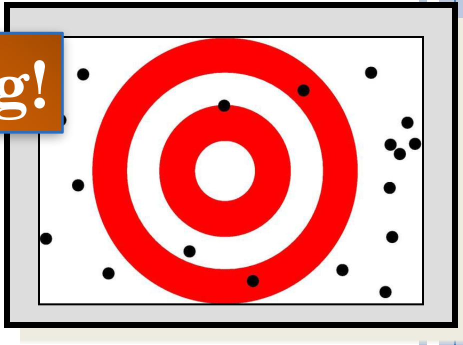
$$\text{Mean}: \mu = np = 5(.8) = 4$$

$$\text{Standard deviation}: \sigma = \sqrt{npq}$$

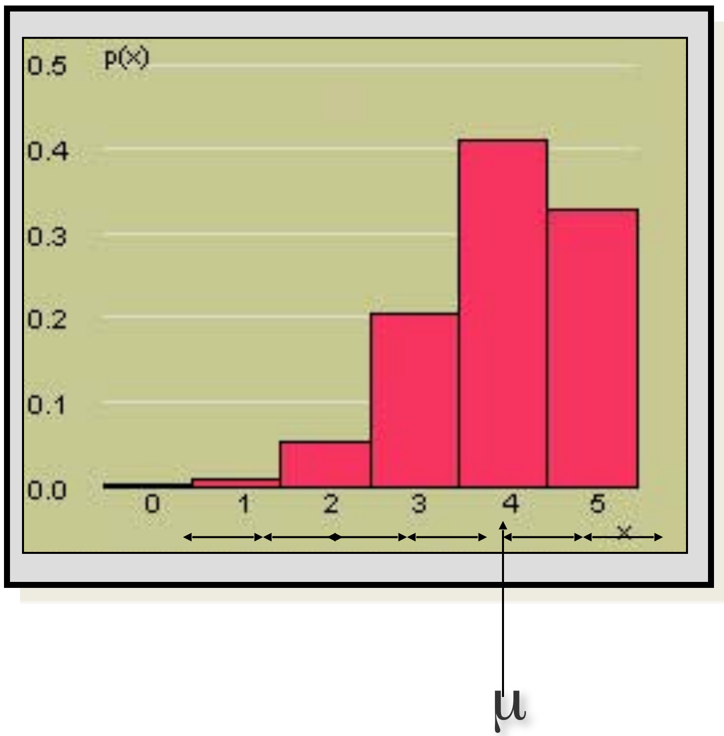
$$= \sqrt{5(.8)(.2)} = .89$$

Example

Link to Hypothesis Testing!



- Would it be unusual to find that none of the shots hit the target?



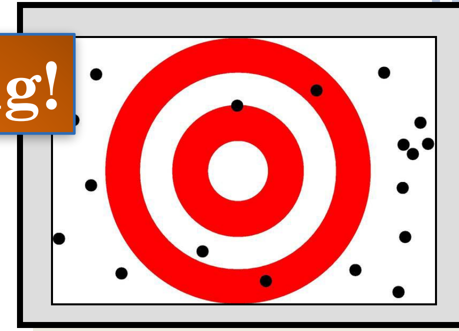
The value:

$P(x = 0) = 0.2\text{^}5 = 0.0$,

This is very unusual.

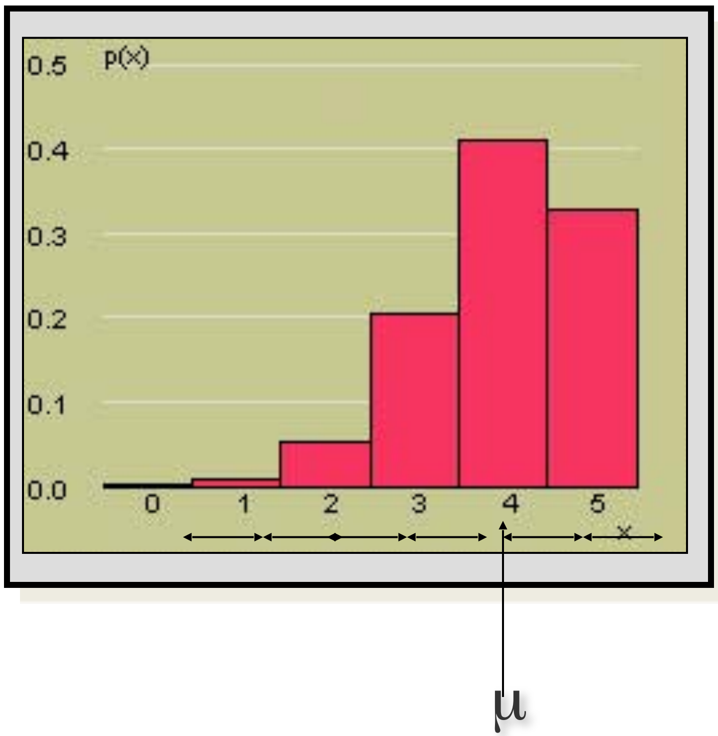
EXAMPLE

Link to Hypothesis Testing!



- Would it be unusual to find that none of the shots hit the target?

$$\mu = 4; \sigma = .89$$



- The value $x = 0$ lies

$$z = \frac{x - \mu}{\sigma} = \frac{0 - 4}{.89} = -4.49$$

- more than 4 standard deviations below the mean. Very unusual.

5.3 Poisson distribution (skipped)

The Poisson Random Variable

- The Poisson random variable x is a model for data that represent the number of occurrences of a specified event in a given unit of time or space.

- **Examples:**
  - The number of calls received by a switchboard during a given period of time.
  - The number of machine breakdowns in a day
  - The number of traffic accidents at a given intersection during a given time period.

The Poisson Probability Distribution

- x is the number of events that occur in a period of time or space during which an average of μ such events can be expected to occur. The probability of k occurrences of this event is

  X~Poisson(μ)

$$P(x = k) = e^{-\mu}\frac{\mu^k}{k!}$$

For values of k = 0, 1, 2, … The mean and standard deviation of the Poisson random variable are

Mean: μ

Standard deviation: $\sigma = \sqrt{\mu}$

EXAMPLE

The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident during a one-week period.

Let X denote the number of accident on a certain section of highway per week: X~ Poisson($\mu = 2$).

$$P(x=1) = e^{-\mu}\frac{\mu^{k}}{k!} = e^{-2}\frac{2^{1}}{1!} = e^{-2}2 = .2707$$

Cumulative Probability Tables

You can use the **cumulative probability tables** to find probabilities for selected Poisson distributions.

✓Find the column for the correct value of μ.

✓The row marked "k" gives the cumulative probability, $P(x \le k) = P(x = 0) + \ldots + P(x = k)$

EXAMPLE

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that there is exactly 1 accident?

k	2.0	2.5	3.0	3.5	4.0	4.5
0	.135					
1	.406					
2	.677					
3	.857					
4	.947					
5	.983					
6	.995					
7	.999					
8	1.000					
9						
10			1.000	.999	.997	.993
11						
12						
13						

P(x = 1) = P($x \le 1$) – P($x \le 0$)

= .406 - .135

= .271

Check from formula: P(x = 1) = .2707

Example

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that 8 or more accidents happen?

$P(x \geq 8)$ $= 1 - P(x < 8)$

$= 1 - P(x \leq 7)$

$= 1 - .999 = .001$

This would be very unusual (small probability) since $x = 8$ lies

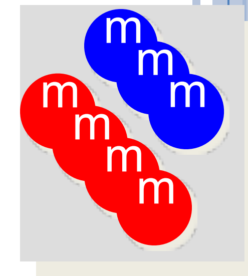
$$z = \frac{x - \mu}{\sqrt{\mu}} = \frac{8 - 2}{1.414} = 4.24$$

standard deviations above the mean.

5.4 Hypergeometric distribution

(skipped)

The Hypergeometric Probability Distribution



- The “M&M® problems” from Chapter 4 are modeled by the **hypergeometric distribution**.
- A bowl contains M red candies and N-M blue candies. Select *n* candies from the bowl and record *x* the number of red candies selected. Define a “red M&M®” to be a “success”.

The probability of exactly *k* successes in *n* trials is

$$P(x = k) = \frac{C_k^M C_{n-k}^{M-N}}{C_n^N}$$

The Mean and Variance

The mean and variance of the hypergeometric random variable *x* resemble the mean and variance of the

$$\text{Mean}: \mu = n\left(\frac{M}{N}\right)$$

$$\text{Variance}: \sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

EXAMPLE

A package of 8 AA batteries contains 2 batteries that are defective. A student randomly selects four batteries and replaces the batteries in his calculator. What is the probability that all four batteries work?

Success = working battery

N = 8

M = 6

***n* = 4**

$$P(x=4)=\frac{C_4^6 C_0^2}{C_4^8}$$

$$=\frac{6(5)/2(1)}{8(7)(6)(5)/4(3)(2)(1)}=\frac{15}{70}$$

EXAMPLE

What are the mean and variance for the number of batteries that work?

$$\mu = n\left(\frac{M}{N}\right) = 4\left(\frac{6}{8}\right) = 3$$

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$

$$= 4\left(\frac{6}{8}\right)\left(\frac{2}{8}\right)\left(\frac{4}{7}\right) = .4286$$

SUMMARY

Key Concepts

I. The Binomial Random Variable

1. Five characteristics: n identical independent trials, each resulting in either success S or failure F; probability of success is p and remains constant from trial to trial; and x is the number of successes in n trials.

2. Calculating binomial probabilities

    a. Formula: $\boxed{P(x = k) = C_k^n p^k q^{n-k}}$

    b. Cumulative binomial tables

    c. Individual and cumulative probabilities using Minitab

3. Mean of the binomial random variable: $\mu = np$

4. Variance and standard deviation: $\sigma^2 = npq$ and $\boxed{\sigma = \sqrt{npq}}$

Key Concepts

II. The Poisson Random Variable

1. The number of events that occur in a period of time or space, during which an average of μ such events are expected to occur

2. Calculating Poisson probabilities

   a. Formula: $$P(x = k) = \frac{\mu^k e^{-\mu}}{k!}$$

   b. Cumulative Poisson tables

   c. Individual and cumulative probabilities using Minitab

3. Mean of the Poisson random variable: $E(x) = \mu$

4. Variance and standard deviation: $\sigma^2 = \mu$ and $\sigma = \sqrt{\mu}$

5. Binomial probabilities can be approximated with Poisson probabilities when $np < 7$, using $\mu = np$.

Key Concepts

III. The Hypergeometric Random Variable

1. The number of successes in a sample of size n from a finite population containing M successes and N – M failures
2. Formula for the probability of k successes in n trials:

$$P(x = k) = \frac{C_k^M C_{n-k}^{M-N}}{C_n^N}$$

3. Mean of the hypergeometric random variable:

$$\mu = n\left(\frac{M}{N}\right)$$

4. Variance and standard deviation:

$$\sigma^2 = n\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-n}{N-1}\right)$$