Before starting new chapter: brief <u>Review from Algebra – Combinations</u>

In how many ways can we select x objects out of n objects?

In how many ways you can select 5 numbers out of 45 numbers ballot to win a lottery?

Combinations Rule

A sample of *n* elements is to be drawn from a set of *N* elements. The, the number of different samples possible is denoted by $\binom{N}{n}$

and is equal to $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ where n! = n(n-1)(n-1)…·5·4·3·2·1

For example 5! = 5·4·3·2·1 Note: 0! = 1! = 1

Example: select 5 cases out of 40 possible choices (the order of selections does NOT matter!)

$$_{40}C_5 = \frac{40!}{5!(40-5)!} = \ldots$$

*Or click on TI-83/84 click **40 MATH PRB 3: 5***

Combination Rule assumes that sample is taken *without replacement* and *without regarding the order* (that means, the order in which elements are selected does not matter).

Example. A "poker hand" consists of 5 cards selected from a standard deck of 52 cards , order does not matter. The number of different "poker hands" is

$$\binom{52}{5} = \frac{52!}{5!47!} = {}_{52}C_5 = 2,598,960$$

TI-83: 52 – MATH – PRB –3:nCr – ENTER – 5 – ENTER

<u>**Chapter 4** Random Variable and Probability Distribution:</u>

The Topics:

- Two Types of Random Variables: discrete and continuous
- Probability Distributions for Discrete Random Variables
- The Binomial Distribution
- Poisson and Hypergeometric Distributions
- Probability Distributions for Continuous Random Variables
- The Normal Distribution
- Uniform and Exponential Distributions

Chapter 4.1 Random Variables – Discrete and Continuous

Chapter 4.2 Probability Distributions for Discrete Random Variable

A **random variable** is a function that to each outcome (sample point) of an experiment assigns a numerical value associated with this outcome.
In other words: a **random variable** is numerical description of outcomes of a random experiment. Random variables are classified as

- **Discrete:** a set of possible values is countable
- **Continuous** a set of possible values is an interval

Example: Is a random variable below continuous or discrete?

1. X = sum of dots on two dice
2. U = world population
3. Y = capacity of a compact disc
4. W = lifetime of a light bulb

Exercises:

1. A coin is tossed and we assign x = 1 if it turns "heads" and x = 0 if it turns "tails" (discrete, possible values are 0,1)
2. A die is rolled and we assign x = number of dots shown on the face of the die (discrete, possible values 1, 2, 3, 4, 5, 6)
3. Two dice are rolled and we define x = sum of dots on the two dice (discrete, possible values 2, 3, 4, …, 12)
4. A light bulb is randomly selected and let t = lifetime of a randomly chosen light bulb (continuous, set of possible values ion an interval (0, ∞))
5. A student is selected and x = his/her height is recorded (continuous, set of possible values in an interval [?, ?] from the shortest to the tallest human)

Example: Consider the different possible orderings of boy (B) and girl (G) in four sequential births. There are **$2*2*2*2=2^4$ = 16** possibilities, so the sample space is:

BBBB	BGBB	GBBB	GGBB
BBBG	BGBG	GBBG	GGBG
BBGB	BGGB	GBGB	GGGB
BBGG	BGGG	GBGG	GGGG

If girl and boy genders are equally likely [P(G) = P(B) = 1/2], and the gender of each child is independent of that of the previous child, then **the probability of each of these 16 possibilities is:**

(1/2)(1/2)(1/2)(1/2) = 1/16.

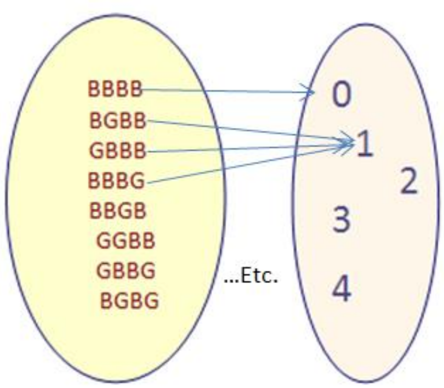
Assign the number of girls to each event:

BBBB (0)	BGBB (1)	GBBB (1)	GGBB (2)
BBBG (1)	BGBG (2)	GBBG (2)	GGBG (3)
BBGB (1)	BGGB (2)	GBGB (2)	GGGB (3)
BBGG (2)	BGGG (3)	GBGG (3)	GGGG (4)

Note that:

- for each possible outcome a single numeric value is assigned
- all outcomes are assigned a numeric value.

The count of the number of girls is an example of a **discrete random variable**: a function assigning a number (in this case a number of girls) to each outcome (the gender order in each family) from sample space



Domain: the list of all possible orders of genders

Range: the list of all possible numbers of girls

X = 3 means "there are three girls in the family with four kids" which happens when any of the four outcomes BGGG, GBGG, GGBG, or GGGB occurs. The events are mutually exclusive, so, we use the addition rule to find:

P(X = 3) = P(BGGG) + P(GBGG) + P(GGBG) + P(GGGB) = 4/16

The probability distribution of a **discrete random variable** can be given as **a table that lists the possible values of the random variables and their associated probabilities.** Or can be given as **a formula, or as a bar or line graph** (like below).

Notation: For a discrete random variable *x*, *p(x)* will stand for the probability that the observed value of the random variable is a number *x,* for example *p(2) = P(x = 2) = probability that x =2*

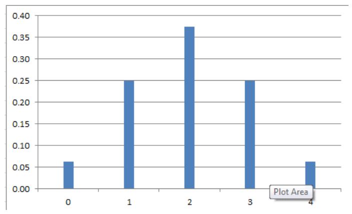
Requirements for a Discrete Probability Distribution

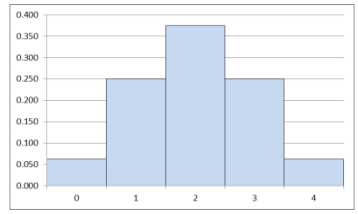
The probability distribution of a **discrete** random variable X must satisfy the following two conditions.

$\sum P(x) = 1$ for all x, and 0≤P(x)≤1 for every x

Probability distribution and the graphs for the random variable “number of girls in a family with 4 kids”:

x	P(x)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16





Is it a valid distribution? Check.

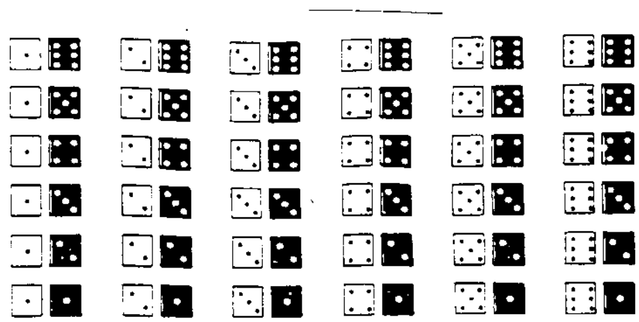
Example:

X=x	P(X=x)
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

Consider the experiment of **tossing two six-sided dice**.

Let random variable **X represent the sum of the numbers on the two dice**

Find and graph the probability distribution of the sums of dots on the dice.

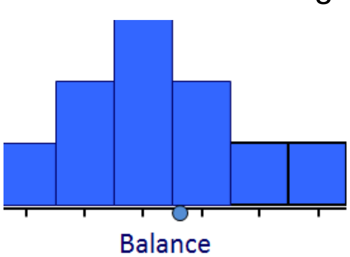


Exercise

4.8 Hotel management. Give an example of a discrete random variable that would be of interest to the manager of a hotel.

Expected Value of Discrete Variable - μ (The mean of the probability distribution):

The mean of a probability distribution is a **weighted average**, with the values of the random variable weighted by their probabilities.



The mean is known as the **expected value** of a random variable, because it is the value that is expected to occur, on average.

The **expected value** of a discrete random variable X is equal to the sum of the products of values multiplied by their probabilities.

$$\mu = E(X) = \sum_{all\ x} xP(x)$$

Interpretation:

- µ is a measure of a center of the distribution of x
- µ is the average value of x observed in a very large (more precisely, as n → ∞, $\bar{x} \to \mu$) number of repetitions of the experiment

  Example: Find expected number of girls in a family with 4 children (use the probability distribution

x	P(x)
0	1/16
1	4/16
2	6/16
3	4/16
4	1/16

Answer: ………………………………….

The **variance** of a discrete random variable x is defined as

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

The **standard deviation** of a discrete random variable x is equal to the square root of the variance, i.e. $\sigma = \sqrt{\sigma^2}$

Example: Finding an Expected Value and standard deviation of probability distribution

Two fair coins are tossed and the number of heads x is observed. Complete the table, and compute the expected value $\mu = E(x)$, variance σ^2, and the standard deviation σ of x

x	*0*	*1*	*2*
p(x)			

TI-83: Enter the values of *x* in L_1 and corresponding values of *p(x)* in L_2,
then use CALC →1-Var Stats L_1 , L_2 → ENTER

Fair Game

Suppose you are playing a coin toss game in which you are paid $1 if the coin turns up heads and you lose $1 when the coin turns up tails. The expected value of this game is *E(X) = 0.* A game of chance with an expected payoff of 0 is called a ***fair game***.

x	*P(x)*	*xP(x)*
-1	0.5	-0.50
1	0.5	0.50
	1.0	0.00

E(X)=μ=mean=0 (a fair game)

More Examples:

Find the mean and standard deviation of the following probability distribution

Bob, an owner of College Painters, studied his records for the past 20 weeks and reports the following number of houses painted per week (able below).

Solution: First find the mean: $\mu = \Sigma[xP(x)]$

Then find the variance: $\sigma^2 = \Sigma[(x-\mu)^2 P(x)] =$

Standard deviation: take the square root of the variance: σ=__________

# houses painted (x)	*# of weeks*	*Probability P(x)*	*x*P(x)*
10	5	.25	2.5
11	6	.30	3.3
12	7	.35	4.2
13	2	.10	1.3
		$\mu =$	11.3

Exercise: An insurance policy costs $100 and will pay policyholder $10,000 if they suffer a major injury or $3,000 if they suffer a minor injury. The company estimates that each year 1 in every 2,000 policyholders may have a major injury, and 1 in 500 may have a minor injury

a. Create a model for x = profit of the company on one policy

b. What is the company expected profit on one policy? What is its standard deviation?

c. If the company sells 900 these policies, what is the company expected profit?

Solution:

1. Complete the probability distribution table

x			
P(x)			

2. Use the formulas: $\mu = \Sigma\, x\, p(x)$ and $\sigma^2 = \Sigma\, (x-\mu)^2\, p(x)$

x	$p(x)$	$x\,p(x)$	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^2 p(x)$
Total:	1	89			67,879

μ = E(X) = ….., σ^2 =……….., σ = ……………… [do it again using TI-83]

3. Find the expected profit for 900 policies: ……………………….

Recall: Chebyshev and Empirical Rules for Discrete Random Variables

Let x be a discrete random variable with probability distribution $p(x)$, mean μ, and standard deviation σ. Then, depending on the shape of $p(x)$, the following probability statements can be made:

	Chebyshev's Rule	**Empirical Rule**
$P(x-\sigma < x < \mu+\sigma)$	≥ 0	$\approx .68$
$P(x-2\sigma < x < \mu+2\sigma)$	$\geq 3/4$	$\approx .95$
$P(x-3\sigma < x < \mu+3\sigma)$	$\geq 8/9$	≈ 1.00

Chebyshev's Theorem applies to probability distributions just as it applies to any frequency distributions. For a random variable X with mean m, standard deviation s, and for any number k > 1: $P(|X-\mu|<k\sigma) \geq 1-\frac{1}{k^2}$

More exercises:

Graph the distribution; answer the questions.

Exercise 4.8, p. 193 Consider the probability distribution for the random variable *x* shown below:

x	0	1	2	3	4	5
$p(x)$	.002	.029	.132	.309	.360	.168

a. What is the probability that *x* is less than 2
b. What is the probability that $1 < x \leq 4$
c. Find μ, σ^2, and σ by hand, and then use TI-83
d. Graph *p(x)*.
e. Locate μ and the interval $\mu \pm 2\sigma$ on your graph. What is the probability that x will fall in this interval. Compare this result with estimates obtained using Chebyshev's and Empirical Rules.

4.18 The random variable x has the following discrete probability distribution:

x	1	3	5	7	9
$p(x)$	.1	.2	.4	.2	.1

a. List the values x may assume.
b. What value of x is most probable?
c. Display the probability distribution as a graph.
d. Find $P(x = 7)$.
e. Find $P(x \geq 5)$.
f. Find $P(x > 2)$.
g. Find E(x)

Exercise [based on 4.28, p. 197] The USDA reports that one in every 100 slaughtered chickens has fecal contamination. Consider a random sample of three slaughtered chickens. Let *x* equal the number of chickens in the sample that have fecal contamination.

a. Find the probability distribution *p(x)* of *x*

b. Find the probability $P(x \leq 1)$
c. What is the probability that at least one of the three have fecal contamination
d. Find the mean and the standard deviation of x

4.36 **Expected lotto winnings.** Most states offer weekly lotteries to generate revenue for the state. Despite the long odds of winning, residents continue to gamble on the lottery each week. In SIA, Chapter 3 (p. 116), you learned that the chance of winning Florida's Pick-6 Lotto game is 1 in approximately 23 million. Suppose you buy a $1 Lotto ticket in anticipation of winning the $7 million grand prize. Calculate your expected net winnings. Interpret the result.

Chapter 4.3 Binomial Distribution

Key concepts: *Binomial experiment, success, failure, number of successes, binomial distribution*

Characteristics of a Binomial Experiment

1. The experiment consists of ***n* identical trials** (repetitions).

2. There are **only two possible outcomes** on each trial. We will denote one outcome by S (for success) and the other by F (for failure).

3. The **probability** of S remains **the same from trial to trial**. This probability is denoted by p, and the probability of F is denoted by q. Note that $q = 1 - p$.

4. The **trials are independent**.

The binomial random variable x is the number of successes S in n trials

Example: Roll a die *10 times* (ten trials). *Success*: rolling a "six". *Failure*: not rolling a "six*". Probability stays the same* per trial (there is 1/6 chance to roll a six in each trial). Trials are *independen*t.

Problems to answer: What is the probability that:
a. "we're out of luck" and did not roll six even once (zero successes in 10 trials)
b. We rolled 6 only once (one success)
c. We rolled it twice (two successes in 10 trials)…three times… four times…, … etc. up to rolling a six every time in all 10 trials.

Example: Select randomly a sample of 10 products from a very large batch. If you know that 7% of the products are defective, then what is the probability of finding: no bad product? One? Two?…all ten?
Success: selecting defective product; *failure*: selecting not defective product
Probability of success 7%; the same in each trial
Number of trials: 10
Independent trials (selecting less than 5% of the population of all products)

Example: Taking multiple choice (choice A-E) exam completely unprepared, x = number of correct answers, *success* = correct answer, p = 1/5, n = number of questions

Example (based on Example 4.10, p. 201) A computer retailer sells both desktop and laptop personal computers (PCs) online. Assume that 80% of the PCs that the retailer sells online are desktops and 20% are laptops, and that sells are independent. Let x represent the number of the next four online PC purchases that are laptops.

1. Explain why x is a binomial random variable; state what is the *success*, what is the *failure*, and what are the values of p, q, and n. What are possible values of x
2. Derive formula for the probability distribution $p(x)$ of x.
3. Find the expected value and the standard deviation of x.

Extend (guess) the formula for $p(x)$ to the case of an arbitrary Binomial experiments with arbitrary n and p.

Table 4.2 Sample Points for PC Experiment of Example 4.9

DDDD	*LDDD*	*LLDD*	*DLLL*	*LLLL*
	DLDD	*LDLD*	*LDLL*	
	DDLD	*LDDL*	*LLDL*	
	DDDL	*DLLD*	*LLLD*	
		DLDL		
		DDLL		

N

P

Number of “successes” X=

P(X=0)=

P(X=1)=

P(X=2)=

P(X=3)=

P(X=4)=

Note that the coefficients represent the number of ways that we may have x successes in n trials = $\binom{n}{x}$.

Indeed, if n = 4, then $\binom{4}{0} = 1, \binom{4}{1} = 4, \binom{4}{2} = 6, \binom{4}{3} = 4. \binom{4}{4} = 1.$

Generalizing:

$p(x) = \binom{4}{x} (0.2)^x (0.8)^{4-x}$ x = 0, 1, 2, 3, 4

From TI-83: mean = μ = 0.8, standard deviation = σ = 0.8

<u>Consider a procedure meeting all the following requirements</u>

Repeat a procedure or experiment a fixed number of times

The probability of x successes occurring is

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

$p(x)$ = Probability of x ‘Successes’

p = Probability of a ‘Success’ on a single trial

q = $1 - p$

n = Number of trials

x = Number of ‘Successes’ in n trials
(x = 0, 1, 2, ..., n)

$n - x$ = Number of failures in n trials

Exercise: Tossing 1 coin 5 times in a row (or five coins once) and noting number of tails. What’s the probability of 3 tails?

<u>The Mean and Standard Deviation of Binomial Distribution</u>

Mean (expected value) E(X)= $\mu = n \bullet p$

Variance Var(X) = $\sigma^2 = n \bullet p \bullet (1-p) = n \bullet p \bullet q$

Std. Dev. $\sigma = \sqrt{\text{npq}}$

Example: Find expected number of successes in “rolling a die 10 times and the success is rolling an even number”: Find standard deviation of the distribution of numbers of successes in this experiment.

E(x)=10·0.5=5 In the average we expect to succeed 5 times when we roll a die 10 times.

Standard deviation: $\sqrt{(10 \cdot 0.5 \cdot 0.5)}=\sqrt{2.5}=1.58$

Exercise: (like Ex. 4.13, p. 206) Suppose a poll of 20 employees is taken in a large company. The purpose is to determine *x*, the number who favor unionization. Suppose that 60% of all the company's employees favor unionization (p = .6, q = .4, n = 20)

a. Find the probability that x=0, =1, …, =20
b. Find the probability that $x \leq 10$.
c. Find the probability that $x \geq 8$.
d. Find the probability that $x=8$ (use the formula).
e. Find the mean and standard deviation of x.
f. Graph the probability distribution of x and locate the interval $\mu \pm 2\sigma$ on the graph

Calculator: if n=20, p=0.6, X=0, 1, …, 20

Then to find the probability of an exact number of successes use the function Binompdf(n, p, x) (2nd – DISTR – A: binomialpdf

to find the probability of no more than given number of successes use the function Binomcdf(n, p, x) (2nd – DISTR – B: binomialcdf

to find the probability of more than given number of successes use the function

1-Binomcdf(n, p, x)

P(X=0)= Binomial(20, 0.6, 0)=…

P(X=1)=Binompdf(n, p, x)=Binomial(20, 0.6, 0)=…

…

b) Binomcdf(20, 0.6, 10)=

c) 1- Binomcdf(20, 0.6, 8)=

The Binomial Distribution Tables are harder to use, and are limited to some cases only. We will not use the tables. We will compute the "BINOMIAL" problems using calculators, and occasionally the formula.

<u>Below: More of "At least…" and "At most…" problems:</u>

(Find the probability P(x is at least .. successes) or P(x is at most …successes) or P(x is more than… successes) or P(x is less than… successes))

Class Exercises: p.209

4.42 Suppose *x* is a binomial random variable with $n = 3$ and $p = .3$.

a. Calculate the value of $p(x)$, $x = 0, 1, 2, 3$, using the formula for a binomial probability distribution.

b. Using your answers to part **a**, give the probability distribution for *x* in tabular form, and graph the distribution.

4.44 If *x* is a binomial random variable, use ~~Table I in Appendix D~~ your calculator to find the following probabilities:

a. $P(x = 2)$ for $n = 10$, $p = .4$
b. $P(x \leq 5)$ for $n = 15$, $p = .6$
c. $P(x > 1)$ for $n = 5$, $p = .1$

4.45 If *x* is a binomial random variable, calculate μ, σ^2, and σ for:

b. $n = 80$, $p = .2$

4.49 Paying for music downloads. A *Pew Internet & American Life Project Survey* (October 2010) revealed that half of all U.S. adults use the Internet and have paid to download music. In a random sample of 250 U.S. adults, let *x* be the number of adults in the sample who used the Internet and paid to download music.

a. Explain why *x* is a binomial random variable (to a reasonable degree of approximation).
b. What is the value of *p*? Interpret this value.
c. What is the expected value of *x*? Interpret this value.

Exercise 4.51, p. 210. According to the Canadian Journal of Information and Library Science (Vol. 33, 2009), nearly 90% of workers in law libraries are satisfied with their job. Assume the true proportion of law librarians in Canada who are satisfied with their job is .9. In a random sample of 20 law librarians in Canada, what is the probability that at most 2 are unsatisfied with their job?

Exercise 4.55 **Bridge inspection ratings**. According to the National Bridge Inspection Standard (NBIS), public bridges over 20 feet in length must be inspected and rated every 2 years. The NBIS rating scale ranges from 0 (poorest rating) to 9 (highest rating). University of Colorado engineers used a probabilistic model to forecast the inspection ratings of all major bridges in Denver (*Journal of Performance of Constructed Facilities*, Feb. 2005). For the year 2020, the engineers forecast that 9% of all major Denver bridges will have ratings of 4 or below.

a. Use the forecast to find the probability that in a random sample of 10 major Denver bridges, at least 3 will have an inspection rating of 4 or below in 2020.

b. Suppose that you actually observe 3 or more of the sample of 10 bridges with inspection ratings of 4 or below in 2020. What inference can you make? Why?