Chapter 16

Day 1

Discrete and Continuous Random Variables

Recap - Probability Rules

- $0 \le P(X) \le 1$ for any event X
- P(S) = 1 for the sample space S
- Addition Rule for Disjoint Events:
  - $P(A \cup B) = P(A) + P(B)$
- Complement Rule:
  - For any event A, $P(A^C) = 1 - P(A)$
- Multiplication Rule:
  - If A and B are independent, then $P(A \cap B) = P(A) \times P(B)$
- General Addition Rule (for nondisjoint) Events:
  - $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- General Multiplication rule:
  - $P(A \cap B) = P(A) \times P(B \mid A)$

Probability Terms

- Disjoint Events:
  - $P(A \cap B) = 0$
  - Events do not share any common outcomes
- Independent Events:
  - $P(A \cap B) = P(A) \times P(B)$ (Rule for Independent events)
  - $P(A \cap B) = P(A) \times P(B \mid A)$ (General rule)
  - $P(B) = P(B|A)$ (lines 1 and 2 implications)
  - Probability of B does not change knowing A
- At Least One:
  - P(at least one) = 1 – P(none)
  - From the complement rule [$P(A^C) = 1 - P(A)$]
- Impossibility: P(E) = 0
- Certainty: P(E) = 1

Math Phases in Probability

Math Symbol	Phrases		
$\ge$	At least	No less than	Greater than or equal to
$>$	More than	Greater than	
$<$	Fewer than	Less than	
$\le$	No more than	At most	Less than or equal to
$=$	Exactly	Equals	Is

Example 1

Write the following in probability format:

- **A. Exactly 6 bulbs are red** — P(red bulbs = 6)
- **B. Fewer than 4 bulbs were blue** — P(blue bulbs < 4)
- **C. At least 2 bulbs were white** — P(white bulbs ≥ 2)
- **D. No more than 5 bulbs were purple** — P(purple bulbs ≤ 5)
- **E. More than 3 bulbs were green** — P(green bulbs > 3)

Discrete Random Variable

Discrete Random Variable

A **discrete random variable** X has a countable number of possible values. The **probability distribution** of a discrete random variable X lists the values and their probabilities:

Value of X:	x_1	x_2	x_3	...	x_k
Probability:	p_1	p_2	p_3	...	p_k

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities is 1: $p_1 + p_2 + \ldots + p_k = 1$.

Find the probability of any event by adding the probabilities p_i of the particular values x_i that make up the event.

Discrete Random Variables

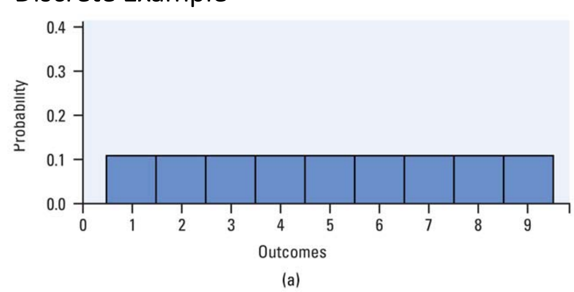
- **Variable's values follow a probabilistic phenomenon**
- **Values are countable**

- **Examples:**
  - **Rolling Die**
  - **Drawing Cards**
  - **Number of Children born into a family**
  - **Number of TVs in a house**

- **Distributions that we will study**

On AP Test	Not on AP
Uniform	Poisson
Binomial	Negative Binomial
Geometric	Hypergeometric

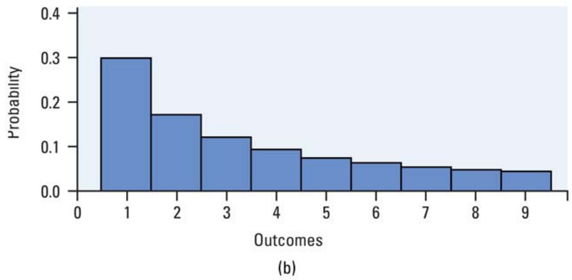
Discrete Example



(a)

- **Most people believe that each digit, 1-9, appears with equal frequency in the numbers we find**

Discrete Example cont



- **Benford's Law**
  - **In 1938 Frank Benford, a physicist, found our assumption to be false**
  - **Used to look at frauds**

Example 2

	1	2	3	4	5	6	7	8	9
P(x)	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

- **Verify Benford's Law as a probability model**

  Summation of P(x) = 1

- **Use Benford's Law to determine**
  - **the probability that a randomly selected first digit is 1 or 2**

    P(1 or 2) = P(1) + P(2) = 0.301 + 0.176 = 0.477
  - **the probability that a randomly selected first digit is at least 6**

    P(≥6) = P(6) + P(7) + P(8) + P(9)

    = 0.067 + 0.058 + 0.051 + 0.046 = 0.222

Example 3

Write the following in probability format with discrete RV (25 colored bulbs):

A. **Exactly 6 bulbs are red**

   P(red bulbs = 6) = P(6)

B. **Fewer than 4 bulbs were blue**

   P(blue bulbs < 4) = P(0) + P(1) + P(2) + P(3)

C. **At least 2 bulbs were white**

   P(white bulbs ≥ 2) = P(≥ 2) = 1 – [P(0) + P(1)]

D. **No more than 5 bulbs were purple**

   P(purple bulbs ≤ 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5)

E. **More than 3 bulbs were green**

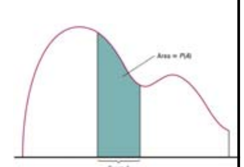
   P(green bulbs > 3) = P(> 3) = 1 – [P(0) + P(1) + P(2) + P(3)]

Continuous Random Variables

- **Variable's values follow a probabilistic phenomenon**
- **Values are uncountable (infinite)**
- **P(X = any value) = 0 (area under curve at a point)**

- **<u>Examples</u>:**
  - **Plane's arrival time -- minutes late (uniform)**
  - **Calculator's random number generator (uniform)**
  - **Heights of children (apx normal)**
  - **Birth Weights of children (apx normal)**

- **Distributions that we will study**
  - **Uniform**
  - **Normal**



Continuous Random Variables

- **We will use a normally distributed random variable in the majority of statistical tests that we will study this year**
- **We need to be able to**
  - **Use z-values in Table A**
  - **Use the normalcdf from our calculators**
  - **Graph normal distribution curves**

Example 4

Determine the probability of the following random number generator:

A. Generating a number equal to 0.5

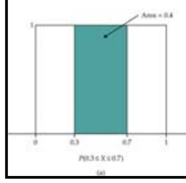
P(x = 0.5) = 0.0

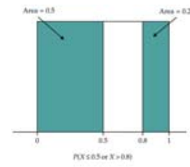
B. Generating a number less than 0.5 or greater than 0.8

P(x ≤ 0.5 or x ≥ 0.8) = 0.5 + 0.2 = 0.7

C. Generating a number bigger than 0.3 but less than 0.7

P(0.3 ≤ x ≤ 0.7) = 0.4





Example 5

In a survey the mean percentage of students who said that they would turn in a classmate they saw cheating on a test is distributed N(0.12, 0.016). If the survey has a margin of error of 2%, find the probability that the survey misses the percentage by more than 2% [P(x<0.1 or x>0.14)]

Change into z-scores to use table A

$$z = \frac{0.10 - 0.12}{0.016} = -1.25$$

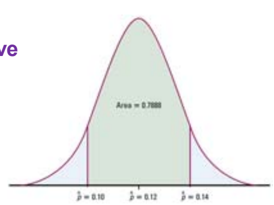
(for x = 0.14, since it is 2 σ's above the mean, z = 1.25)

0.8944 – 0.1056 = 0.7888

1 – 0.7888 = 0.2112

ncdf(0.1, 0.14, 0.12, 0.016) = 0.7887

1 – 0.7887 = 0.2112



Means

Mean of a Discrete Random Variable

Suppose that X is a discrete random variable whose distribution is

Value of X:	x_1	x_2	x_3	. . .	x_k
Probability:	p_1	p_2	p_3	. . .	p_k

To find the **mean** of X, multiply each possible value by its probability, then add all the products:

$$\mu_X = x_1p_1 + x_2p_2 + \ldots + x_kp_k$$
$$= \Sigma x_ip_i$$

Mean of a random variable is also known as its expected value

Discrete Random Variable - Mean

The mean, or expected value [E(x)], of a discrete random variable is given by the formula

$$\mu_x = \sum [x \cdot P(x)]$$

where x is the value of the random variable and P(x) is the probability of observing x

Mean of a Discrete Random Variable Interpretation:

If we run an experiment over and over again, the law of large numbers helps us conclude that the difference between $\bar{x}$ and μ_x gets closer to 0 as n (number of repetitions) increases

Summary

- **Random variables (RV) values are a probabilistic**
- **RV follow probability rules**
- **Discrete RV have countable outcomes**
- **Continuous RV has an interval of outcomes (∞)**

Chapter 16

Day 2

Variance

Variance of a Discrete Random Variable

Suppose that X is a discrete random variable whose distribution is

Value of X:	x_1	x_2	x_3	. . .	x_k
Probability:	p_1	p_2	p_3	. . .	p_k

and that μ is the mean of X. The **variance** of X is

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \ldots + (x_k - \mu_X)^2 p_k$$
$$= \Sigma (x_i - \mu_X)^2 p_i$$

The **standard deviation** σ_X of X is the square root of the variance.

Discrete Random Variable - Variance

Variance and Standard Deviation of a Discrete Random Variable:

The variance of a discrete random variable is given by:

$$\sigma_x^2 = \sum[(x-\mu_x)^2 \cdot P(x)] = \sum[x^2 \cdot P(x)] - \mu_x^2$$

and standard deviation is $\sqrt{\sigma^2}$

Note: round the mean, variance and standard deviation to one more decimal place than the values of the random variable

Uniform PDF

An experiment is said to be a Uniform experiment provided:

1.The probability of each value of the random variable is equal (like in a six-sided die)

2.The trials are independent of each other (what happened last does not affect what happens next)

Uniform PDF

If X is a value of the uniform random variable, then probability formula for X is

$$P(x) = \frac{1}{n}$$ for x = 0, 1, 2, 3, … , n

where n is the total number of discrete values of the random variable x

Mean: $\mu_x = \sum[x \cdot P(x)] = \frac{1}{n}\sum x$

Standard Deviation:

$$\sigma_x^2 = \sum[(x-\mu_x)^2 \cdot P(x)] = \frac{1}{n}\sum[(x-\mu_x)^2] =$$

$$\sum[x^2 \cdot P(x)] - \mu_x^2 = \frac{1}{n}\sum[x^2] - \mu_x^2$$

Using The Graphing Calculators

We can use 1-Var-Stats to calculate the mean and standard deviation of a discrete random variable given it's outcomes and probability

- **Type in outcomes in L1 (name a list on nSpire)**
- **Type in corresponding probabilities in L2**
- **Use 1-Var-Stats L1, L2 to get statistics (nSpire set "frequency list" as list of probabilities)**

Example 1

You have a fair 10-sided die with the number 1 to 10 on each of the faces.

Determine the mean and standard deviation.

Mean: $\sum [x \cdot P(x)] = (1/10)(\sum x) = (1/10)(55) = 5.5$

Var: $\sum[x^2 \cdot P(x)] - \mu^2_x = (1/n) \sum [x^2] - \mu_x^2$
$= (1/10)(385) - 30.25)$
$= (38.5 - 30.25)$
$= 8.25$

St Dev $= 2.8723$

Example 2

Below is a distribution for number of visits to a dentist in one year.

X = # of visits to a dentist

x	0	1	2	3	4
P(x)	.1	.3	.4	.15	.05

Determine the expected value, variance and standard deviation.

Mean: $\sum [x \cdot P(x)] = (.1)(0) + (.3)(1) + (.4)(2) + (.15)(3) + (.05)(4)$
$= 0 + .3 + .8 + .45 + .2 = 1.75$

Var: $\sum[x^2 \cdot P(x)] - \mu^2_x = \sum [x^2 \cdot P(x)] - \mu_x^2$
$= (0 + .3 + .4(4) + .15(9) + .05(16)) - 3.0625)$
$= 4.05 - 3.8626$
$= 0.9875$

St Dev $= 0.9937$

Example 3

What is the average size of an American family? Here is the distribution of family size according to the 1990 Census:

# in family	2	3	4	5	6	7
p(x)	.413	.236	.211	.090	.032	.018

Mean: $\sum [x \cdot P(x)] = (.413)(2) + (.236)(3) + (.211)(4) + (.09)(5) + (.032)(6) + (.018)(7)$

$= .826 + .708 + .844 + .45 + .192 + .126$

$= 3.146$

Example 4

You are trying to decide whether to take out a $250 deductible which will cost you $90 per year. Records show that for this community the average cost of repair is $900. Records also show that 10% of the drivers have an accident during the year. If you have sufficient assets so that you will not be financially handicapped if you had to pay out the $900 or more for repairs, should you buy the policy?

P(accident) = 0.1 so P(no accident) = 0.9
ave repair cost = $900 Yearly cost = $90

Expected Yearly with Policy: $\sum [x \cdot P(x)] = (.1)(250) + (.9)(0) + 90$
$= 25 + 90$ = $115

Expected Yearly without: $\sum [x \cdot P(x)] = (.1)(900) + (.9)(0)$ = $90

Summary

- **Summary**
  - **Expected value is the mean** $\sum[x \cdot P(x)]$
  - **Variance is** $\sum\left[(x-\mu_x)^2 \cdot P(x)\right]$
  - **Standard Deviation is** $\sqrt{variance}$

Chapter 16

Day 3

Probability Laws

- **Law of Large Numbers – True**
  - **Sample mean, x, approaches population mean, μ, as sample size increases**
- **Law of Small Numbers – False**
  - **Random behavior in short term does not mimic long-term behavior**
- **Law of Averages – Bad Statistics**
  - **eventually everything evens out**

Rules for Means

Rules for Means

Rule 1. If X is a random variable and *a* and *b* are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$

Rule 2. If X and Y are random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

- **Means follow the rules for linear combinations (from Algebra)**
- **When you linearly combine two or more (rules give only the 2 case example) random variables, you combine their means in the same manner**

Rules for Variances

Rules for Variances

Rule 1. If X is a random variable and *a* and *b* are fixed numbers, then

$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

Rule 2. If X and Y are *independent* random variables, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

This is the **addition rule for variances of independent random variables.**

- **Adding a number to a random variable does not change its variance**
- **Multiply a random variable by a number changes the variance by the square of that number**
- **When you combine random variables, you always add the variances**

Rules for Standard Deviations

- **Follow the rules for variances and then take the square root to find the standard deviation**
- **In general standard deviations do not add**
- **Note: independence is required for the calculations of combined variances, but not for means**
  - **Methods for combining non-independent variables' variances involve covariance terms and are not part of this course**

Example 1

Scores on a Math test have a distribution with μ = 519 and σ = 115. Scores on an English test have a distribution with μ = 507 and σ = 111. If we combine the scores
a) what is the combined mean

b) what is the combined standard deviation?

$\mu_M + \mu_E = 519 + 507 = 1016$

Scores are not independent so the following is not correct!

$\sigma^2_{M+E} = \sigma^2_M + \sigma^2_E = 115^2 + 111^2 = 25546$

$\sigma_{M+E} = \sqrt{25546} = 159.83$

Example 2

Suppose you earn \$12/hour tutoring but spend \$8/hour on dance lessons. You save the difference between what you earn and the cost of your lessons. The number of hours you spend on each activity is independent. Find your expected weekly savings and the standard deviation of your weekly savings.

Hrs Dancing / week	Probability
0	0.4
1	0.3
2	0.3

Hrs Tutoring / week	Probability
1	0.3
2	0.3
3	0.2
4	0.2

Example 2 cont

Hrs Dancing / week	Probability
0	0.4
1	0.3
2	0.3

Expect value for Dancing, μ_X, is
0(0.4) + 1(0.3) + 2(0.3) = 0.9

Variance: $\sum [P(x) \cdot x^2] - \mu_x^2$
= $(.4(0) + .3(1) + .3(4)) - 0.9^2)$
= 1.5 – 0.81
= 0.69
St Dev = 0.8307

Example 2 cont

Hrs Tutoring / week	Probability
1	0.3
2	0.3
3	0.2
4	0.2

Expect value for Tutoring, μ_Y, is
1(0.3) + 2(0.3) + 3(0.2) + 4(0.2) = 2.3

Variance: $\sum [x^2 \cdot P(x)] - \mu_x^2$
= $(.3(1) + .3(4) + .2(9) + .2(16)) - 2.3^2)$
= 6.5 – 5.29
= 1.21
St Dev = 1.1

Example 2 cont

Expect value for Weekly Savings, μ_{12Y-8X}, is

$12\ \mu_Y - 8\ \mu_X = 12\ (2.3) - 8\ (0.9)$
= 27.6 – 7.2 = \$20.4

Variance of Weekly Savings, σ^2_{12Y-8X}, is

$\sigma^2_{12Y} + \sigma^2_{8X} = 12^2(1.21) + 8^2(0.69)$
= 174.24 + 44.16 = 218.4

so standard deviation = \$14.79

Combining Normal Random Variables

- **Any linear combination of independent Normal random variables is also Normally distributed**
- **For example: If X and Y are independent Normally distributed random variables and *a* and *b* are any fixed numbers, then *a*X + *b*Y is also Normally distributed**
- **Mean and standard deviations can be found by using the rules from previous slides**

Example 3

Tom's score for a round of golf has a N(110,10) distribution and George's score for a round of golf has a N(100,8) distribution. If they play independently, what is the probability that Tom will have a better (lower) score than George?

Let X be Tom's score and Y be George's score

$\mu_{X-Y} = \mu_X - \mu_Y = 110 - 100 = 10$

$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y = 10^2 + 8^2 = 164 \approx (12.8)^2$

so X – Y is N(10,12.8)

P(X-Y<0) = P(z < Z) with Z = (0 – 10) / 12.8 = -0.78

Example 3 cont

We could have used our calculator, ncdf(-E99,0,10,12.8), or Table A to get the probabilities illustrated in the graph below