Lecture Notes & Examples 6.1

Section 6.1 Discrete and Continuous Random Variables (pages 340 – 352)

1. Random Variables. Consider tossing a fair coin 3 times. The sample space would be:

S = { HHH, HHT, HTH, THH, HTT, THT, TTH, TTT } $n = 8$

Let **X** represent the number of heads obtained. We can depict this situation in a **probability distribution of X**:

Value	0	1	2	3
Probability	1/8	3/8	3/8	1/8

We can use the probability distribution to answer questions about the variable X such as what is P(X ≥ 1)?

$P(X \geq 1) = P(X=1) + P(X=2) + P(X=3)$
$= 1/8 + 3/8 + 3/8 = 7/8$

or $P(X \geq 1) = 1 - P(X=0)$
$= 1 - 1/8 = 7/8$

Definition: A **random variable** takes numerical values that describe the outcomes of some chance process. The **probability distribution** of a random variable gives its possible values and their probabilities.

2. Discrete Random Variables

Definition: A **discrete random variable** X takes on a fixed set of possible values with gaps between. The probability distribution of a discrete random variable X lists the values x_i and their probabilities p_i:

Value:	x_1	x_2	x_3	. . .
Probability:	p_1	p_2	p_3	. . .

Integer Values
Usually the result of counting

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.   ex: # of people; # of cars
2. The sum of the probabilities is 1.

To find the probability of any event, add the probabilities p_i of the particular values of x_i that make up that event.

Example - In 2010, there were 1319 games played in the National Hockey League's regular season. Imagine selecting one of these games at random and then randomly selecting one of the two teams that played in the game. Define the random variable X = number of goals scored by a randomly selected team in a randomly selected game. The table below gives the probability distribution of X:

Goals	0	1	2	3	4	5	6	7	8	9
Probability	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001

(a) Show that the probability distribution for X is legitimate.

✓ Legit

(1) all prob between 0 and 1 $0 \leq P_i \leq 1$. ✓ (2) $\Sigma P_i = 1$ ✓

(b) What is the probability that the number of goals scored by a randomly selected team in a randomly selected game is at least 6? $P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9)$

$.041 + .015 + .004 + .001 = .061$

Meaning: Team scored 6 or more goals about 6% of the time

Check Your Understanding

North Carolina State University posts the grade distributions for its courses online.[3] Students in Statistics 101 in a recent semester received 26% A's, 42% B's, 20% C's, 10% D's, and 2% F's. Choose a Statistics 101 student at random. The student's grade on a four-point scale (with A = 4) is a discrete random variable X with this probability distribution:

	E	D	C	B	A
Value of X:	0	1	2	3	4
Probability:	0.02	0.10	0.20	0.42	0.26

1. Say in words what the meaning of $P(X \geq 3)$ is. What is this probability?

$P(X \geq 3)$ is the probability that a student gets an A or a B.

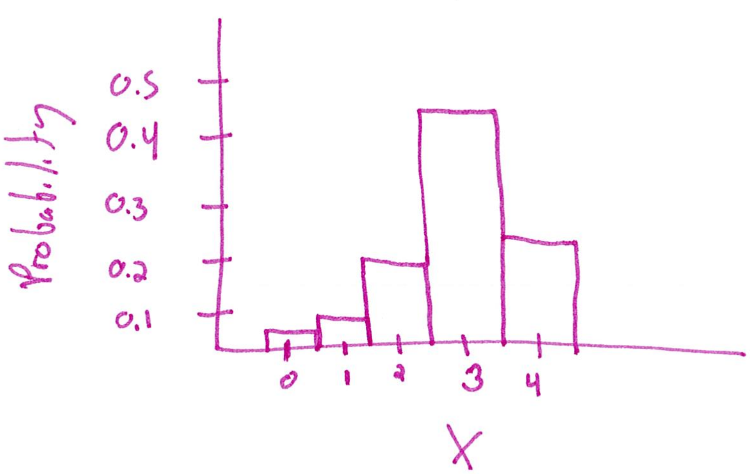
$$P(X \geq 3) = P(X=3) + P(X=4)$$
$$= 0.42 + .26 = 0.68$$

2. Write the event "the student got a grade worse than C" in terms of values of the random variable X. What is the probability of this event?

$$P(X < 2) = P(X=1) + P(X=0)$$
$$= 0.10 + 0.02 = 0.12$$

Histogram

3. Sketch a graph of the probability distribution. Describe what you see.



The histogram is skewed left. Higher grades are more likely, but there are a few low grades. Unimodal

3. The Mean (Expected Value) of a Discrete Random Variable

When analyzing shapes of distributions we used **SOCS**. If we want to know the center of a distribution of a discrete random variable we are going to have to compute the mean. The mean of a discrete random variable X is denoted by μ_x. It is an average of all possible values of the random variable X but we have to take into account how many times we *expect* the values to occur. For this reason the mean is also referred to as the **expected value** of the random variable.

Also $E(X)$

Example: Given the probability distribution of the discrete random variable X, find the expected value of X.

Value	1	2	3
Probability	0.5	0.2	0.3

$$E(x) = \mu_x = 1(0.5) + 2(0.2) + 3(0.3)$$
$$= 0.5 + 0.4 + 0.9$$
$$= 1.8$$

describes long run average outcome

Roulette Example

Value: Amt Win	$-\$1$	$\$1$ (Win: red)
Prob	$20/38$	$18/38$

$$E(x) = -\$1\left(\frac{20}{38}\right) + \$1\left(\frac{18}{38}\right)$$
$$= \frac{-20}{38} + \frac{18}{38} = \frac{-2}{38} \approx -\$.05$$

If play many times, on average I expect to lose $.05 each time I play

Definition: Suppose that X is a discrete random variable whose probability distribution is

Value:	x_1	x_2	x_3	. . .
Probability:	p_1	p_2	p_3	. . .

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

$$\mu_x = E(x) = x_1 p_1 + x_2 p_2 + \dots x_n p_n = \sum x_i p_i$$

Formula Sheet

Example: Find the expected value of the random variable X in the NHL example and interpret the value in context.

Goals	0	1	2	3	4	5	6	7	8	9
Probability	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001

$$E(x) = \mu_x = (0)(.061) + (1)(.154) + 2(.228) + 3(.229) + 4(.173) + 5(.094) + 6(.041) + 7(.015) + 8(.004) + 9(.001)$$
$$= 2.851$$

Mean # of goals = 2.851 goals/game. If we repeat random selection process over and over again, the mean # of goals scored would be about 2.851 in the long run.

Note: *A common error on the AP Exam is that students incorrectly believe that the expected value of a random variable must be equal to one of the possible values of the variable. This is not the case.*

4. The Standard Deviation (and Variance) of a Discrete Random Variable

In order to describe the *spread* of the distribution of a discrete random variable, we are going to use the standard deviation. In order to find the standard deviation, we first compute the variance and then find its square root. The variance is the *average of the squared deviation of the possible X values from the mean.* Again, however, we must take into account how often we expect the different values of X to occur.

Definition: Suppose that X is a discrete random variable whose probability distribution is

Value:	x_1	x_2	x_3	...
Probability:	p_1	p_2	p_3	...

ON Formula Sheet

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

$$Var(X) = \sigma_x^2 = (x_1-\mu_x)^2 p_1 + (x_2-\mu_x)^2 p_2 + (x_3-\mu_x)^2 p_3 \dots$$

The **standard deviation** of X, σ_x is the square root of the variance. $= \Sigma (x_i - \mu_x)^2$

Example. Compute and interpret the standard deviation of the random variable X in the NHL example and interpret its meaning in context. From previous example $E(X) = \mu_x = 2.851$

Goals	0	1	2	3	4	5	6	7	8	9
Probability	0.061	0.154	0.228	0.229	0.173	0.094	0.041	0.015	0.004	0.001

$$\sigma_x^2 : (0-2.851)^2(.061) + (1-2.851)^2(.154) + (2-2.851)^2(.228) + \dots (9-2.851)^2(.001)$$

$$\sigma_x^2 = 2.66$$

so $\sigma_x : \sqrt{2.66} \approx 1.63$

standard deviation

On average, a randomly selected team's # of goals in a randomly selected game will differ from the mean by about 1.63 goals.

Check Your Understanding

A large auto dealership keeps track of sales made during each hour of the day. Let X = the number of cars sold during the first hour of business on a randomly selected Friday. Based on previous records, the probability distribution of X is as follows:

Cars sold:	0	1	2	3
Probability:	0.3	0.4	0.2	0.1

1. Compute and interpret the mean of X.

$\mu_X = 0(0.3) + 1(0.4) + 2(0.2) + 3(0.1) = 1.1$

The long run average, over many Friday mornings, will be about 1.1 cars sold.

2. Compute and interpret the standard deviation of X.

$\sigma_X^2 = (0-1.1)^2(0.3) + (1-1.1)^2(.4) + (2-1.1)^2(.2) + (3-1.1)^2(0.1)$

$\sigma_X^2 = 0.89$

$\sigma_X = \sqrt{.89} = .943$

(stand dev)

On average, the # of cars sold on a randomly selected Friday will differ from the mean (of 1.1) by .943 cars sold.

5. Continuous Random Variables

Definition: A **continuous random variable** X takes all values in an interval of numbers. The probability distribution of X is described by a *density curve*. The probability of any event is the area under the density curve and above the values of X that make up that event.

* Use Real #'s
* Usually From Measuring

example) weight, age, time, etc

The most familiar continuous probability distribution is the (vaunted) **Normal distribution**.

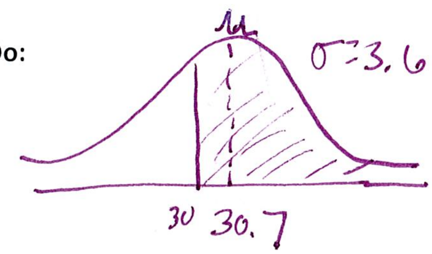
$N(\mu, \sigma) = N(30.7, 3.6)$

Example: The weights of three-year-old females closely follow a *Normal distribution* with a mean of μ = 30.7 pounds and a standard deviation of 3.6 pounds. Randomly choose one three-year-old female and call her weight X. Find the probability that the randomly selected female weighs at least 30 pounds.

State: What is the probability that a randomly selected 3 year old female weights at least 30 pounds?

Plan: Let X represent the weight of 3 year old females.
X is N(30.7, 3.6) We want to find $P(x \geq 30)$

Do:



Calculate: normalcdf(30, 10,000, 30.7, 3.6) ≈ .577

$\uparrow$ lower bound, $\uparrow$ upper bound, $\uparrow$ μ, $\uparrow$ σ

You may also convert to z score and use Table A

z-score = $\frac{30-30.7}{3.6} = \frac{-.7}{3.6} = -.19$ From Table A $P(z < -.19) = .4247$

so $P(z \geq -.19) = 1 - .4247 = .5753$

Conclude:

There is about a 58% chance that the randomly selected 3 year old female will weigh at least 30 pounds.