Homework: Chapter 8: #1,14, 27
Due *Wed*, October 27

Announcements:
- Quiz starts after class today, ends Wed.
- Order switched from original plan – start Chapter 8 today, finish Chapter 7 on Wed.

Random Variables

Chapter 8

What we will cover this week and early next week:

Today: 8.1 to 8.3
Wed: 7.7 + extra material
Friday: 8.4 + extra material, start 8.5 if time
Monday: 8.5 to 8.7
Skip Section 8.8

8.1 What is a Random Variable?

Random Variable: assigns a number to each outcome of a random circumstance, or, equivalently, to each unit in a population.

Two different broad classes of random variables:
1. A **continuous random variable** can take any value in an interval or collection of intervals.
2. A **discrete random variable** can take one of a countable list of distinct values.

Notation for either type: X, Y, Z, W, etc.

Examples of Discrete Random Variables

Assigns a number to each outcome in the sample space for a random circumstance, or to each unit in a population.

1. Couple plans to have 3 children.
The *random circumstance* includes the 3 births, specifically the sexes of the 3 children.
Possible outcomes (sample space): BBB, BBG, etc.
X = number of girls
X is *discrete* and can be 0, 1, 2, 3
For example, the number assigned to BBB is X=0

2. Population consists of UCI students (*unit* = student)
Y = number of siblings a student has
Y is *discrete* and can be 0, 1, 2, …??

Examples of Continuous Random Variables

Assigns a number to each outcome of a random circumstance, or to each unit in a population.

1. Population consists of UC female students
Unit = female student
W = height
W is *continuous* – can be anything in an interval, even if we report it to nearest inch or half inch

2. You are waiting at a bus stop for the next bus
Random circumstance = when the bus arrives
Y = time you have to wait
Y is *continuous* – anything in an interval

Today: Discrete Random Variables

X = the *random variable* (r.v.), such as number of girls.
k = a *number* the discrete r. v. could equal (0, 1, etc).
$P(X = k)$ is the *probability* that X equals k.

***Probability* distribution function (pdf)** for a discrete r.v. X is a table or rule that assigns probabilities to possible values of X.

***Cumulative* distribution function (cdf)** is a rule or table that provides $P(X \leq k)$ for every real number k. (More useful for continuous random variables than for discrete, as we will see.)

NOTE: Sometimes the probabilities are given or observed, and sometimes you have to compute them using rules from Ch. 7.

Conditions for Probabilities for Discrete Random Variables

Condition 1

The *sum of the probabilities* over all possible values of a discrete random variable must equal 1.

Condition 2

The *probability of any specific outcome* for a discrete random variable, *P(X = k),* must be between 0 and 1.

Note: The possible values *k* are *mutually exclusive*

Example on Board: 2 Clicker questions with 4 choices each, X = points earned if you are just guessing. Find *pdf* and *cdf.*

Another example of computing the PDF and CDF from Chapter 7 Rules

Example: You buy 2 tickets for the Daily 3 lottery (different days)
Probability that you win each time is 1/1000 = .001, independent.
X = number of winning tickets you have, could be 0, 1, 2.

$P(X = 0) = (.999)^2 = .998001$ (Rule 3b) (998,001 in a million)

$P(X = 2) = (.001)^2 = .000001$ (Rule 3b) (1 in a million)

$P(X = 1) = 1 - P(X = 0 \text{ or } X = 2) = 1 - (.998001 + .000001)$
$= .001998$ (Rule 1) (1998 in a million)

k	pdf *P(X=k)*	cdf *P(X ≤ k)*
0	.998001	.998001
1	.001998	.999999
2	.000001	1.0

Example of using *observed proportions* to create a pdf

Survey of 173 students in introductory statistics:

k	Number with *k* siblings	pdf *P(X=k)*	cdf *P(X ≤ k)*
0	14	14/173 = .08	.08
1	68	68/173 = .39	.39 + .08 = .47
2	53	.31	.47 + .31 = .78
3	21	.12	.90
4	8	.05	.95
5	6	.03	.98
6	3	.02	1.00

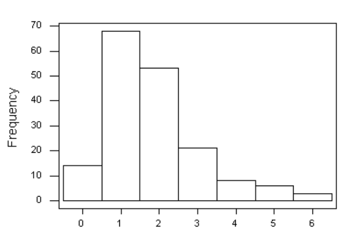
Clicker data collection (non credit)

How many siblings (brothers and sisters) do you have? Count half-siblings (share one parent), but not step siblings.

A. 0
B. 1
C. 2
D. 3
E. 4 or more

Graph of pdf for number of siblings (with frequency instead of relative frequency)

Compare class results.



More Complicated Examples for Discrete R.V.s

Probability distribution function (pdf) *X* is a table or rule that assigns probabilities to possible values of *X.*

Using the sample space to find probabilities:

Step 1: List all simple events in sample space.

Step 2: Find probability for each simple event.

Step 3: List possible values for random variable *X* and identify the value for each simple event.

Step 4: Find all simple events for which *X* = *k,* for each possible value *k.*

Step 5: *P*(*X* = *k*) is the sum of the probabilities for all simple events for which *X* = *k.*

Example: Sibling blood types

Blood types and possible alleles:

Type O: Must be OO
Type A: Could be AA or OA
Type B: Could be BB or OB
Type AB: Must be AB

Suppose father has OO (type O) and mother has OA (type A).

They have 3 children. Let X = number with Blood type A.

Each child equally likely to inherit:

Father	Mother	Child blood type
O	O	Blood type O
O	A	Blood type A

So, child has Type O or Type A, each with probability 1/2

Example: Sibling blood types

Family has 3 children. Probability of type A is ½ for each child. What are the probabilities of 0, 1, 2, or 3 with type A?

Sample Space: For each child, write either O or A. There are eight possible arrangements of O and A for three births. These are the *simple events.*

S = {OOO, OOA, OAO, AOO, OAA, AOA, AAO, AAA}

Sample Space and Probabilities: The eight simple events are equally likely. Each has probability (1/2)(1/2)(1/2) = 1/8

Random Variable X: number of Type A in three children. For each simple event, the value of *X* is the number of A's listed.

How Many Children with Type A?

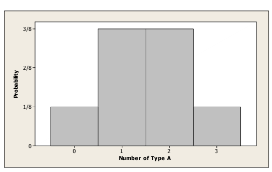
Value of *X* for each simple event:

Simple Event	OOO	OOA	OAO	AOO	OAA	AOA	AAO	AAA
Probability	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
X = # Type A	0	1	1	1	2	2	2	3

Probability distribution function for X = # of Type A:

k	0	1	2	3
P(*X* = *k*)	1/8	3/8	3/8	1/8

Graph of the pdf of *X*:



Cumulative Distribution Function for number of Type A:

Cumulative distribution function (cdf) provides the probabilities $P(X \leq k)$ for any real number *k*.

Cumulative probability = probability that *X* is less than or equal to a particular value.

Example: ***Cumulative Distribution Function for the Number Kids with Type A***

k	0	1	2	3
$P(X \leq k)$	1/8	4/8	7/8	1

For example, the probability is 7/8 that ≤ 2 kids have Type A.

8.3 Expected Value (Mean) for Random Variables

The **expected value** of a random variable is the **mean** value of the variable *X* in the sample space, or population, of possible outcomes.

If *X* is a random variable with possible values $x_1, x_2, x_3, \ldots,$ occurring with probabilities $p_1, p_2, p_3, \ldots,$ then the **expected value** of *X* is calculated as

$$\mu = E(X) = \sum x_i p_i$$

Example of expected value

Number of siblings for intro stat students:

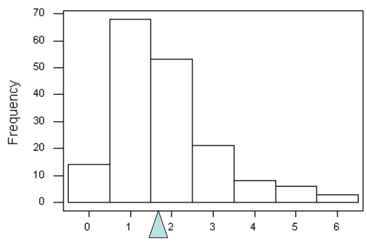
x_i	p_i	$x_i p_i$
0	14/173 = .08	.00
1	68/173 = .39	.39
2	.31	.62
3	.12	.36
4	.05	.20
5	.03	.15
6	.02	.12

$\mu = E(X) = \sum x_i p_i$

= 1.84

= mean number of siblings

Expected value = mean value is where the picture of the pdf "balances"



1.84

Other examples of expected value

Ex 1: Just guessing for 2 questions on clicker, quiz

X = clicker, Y = quiz; E(X)=2.5, E(Y)=1, (on board)

Ex 2: Raffle ticket costs $2.00. You win

$5.00 with probability 1/10, so net gain = $3

$100 with probability 1/100, so net gain = $98

Nothing with probability 89/100, so net gain = –$2

X = net gain. What is E(X)?

E(X) = $3 ×(10/100) + $98 ×(1/100) – $2.00 ×(89/100)

= $(30 + 98 – 178)/100 = –$50/100

This is a *loss* of 50 cents on average for each $2.00 ticket.

This means the people running the raffle *gain* 50 cents per ticket.

Should you buy extended warranties?

You buy a new appliance, computer, etc.

- Extended warranty for a year costs $10.
- Unknown to you, the probability you will need a repair is 1/50, and it will cost $200 if you do.

Is the warranty a good deal?

X = your cost to repair the item.

k	*P(X = k)*	*k P(X=k)*
$200	*1/50*	*$200/50*
$0	*49/50*	*0/50*

E(X) = $200/50 = $4.00

If you buy the warranty your cost is fixed at $10. If you don't, your cost is either $200 or $0, but the long run *average* is $4.00

Notes about expected value

- It's the *average* or *mean* value of the random variable *over the long run*.
- It may not be an actual possible value for the random variable (usually it isn't; e.g. 1.84 sibs).
- In gambling, lotteries, insurance, extended warranty, etc., you can be pretty sure that your "expected" cost per event if you play or buy is more than if you don't – the house wins!
- However, for insurance, for example, you might prefer the peace of mind of knowing your fixed cost. For lottery, you might want the thrill of the possibility of winning, even though you lose on average.

Standard Deviation for a Discrete Random Variable

The **standard deviation** of a random variable is essentially the average distance the random variable falls from its mean over the long run.

If X is a random variable with possible values $x_1, x_2, x_3, \ldots$, occurring with probabilities $p_1, p_2, p_3, \ldots$, and **expected value** $E(X) = \mu$, then

$$\text{Variance of } X = V(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i$$

$$\text{Standard Deviation of } X = \sigma = \sqrt{\sum (x_i - \mu)^2 p_i}$$

Example 8.13 *Stability or Excitement*

Two plans for investing $100 – which would you choose?

Plan 1		Plan 2	
X = Net Gain	*Probability*	*Y = Net Gain*	*Probability*
$5,000	.001	$20	.3
$1,000	.005	$10	.2
$0	.994	$4	.5

Expected Value for each plan:

Plan 1:

$E(X) = \$5{,}000 \times (.001) + \$1{,}000 \times (.005) + \$0 \times (.994) = \$10.00$

Plan 2:

$E(Y) = \$20 \times (.3) + \$10 \times (.2) + \$4 \times (.5) = \$10.00$

Example 8.13 *Stability or Excitement (cont)*

Variability for each plan:

Plan 1			Plan 2		
$(X-\mu)^2$	p	$(X-\mu)^2p$	$(Y-\mu)^2$	p	$(Y-\mu)^2p$
$(\$5,000 - \$10)^2 = \$24,900,100$	.001	\$24,900.1	$(\$20 - \$10)^2 = \$100$	.3	\$30
$(\$1,000 - \$10)^2 = \$980,100$	.005	\$4,900.5	$(\$10 - \$10)^2 = 0$	.2	0
$(\$0 - \$10)^2 = 100$	.994	\$99.4	$(\$4 - \$10)^2 = \$36$	.5	\$18

Plan 1: *Variance* of X = \$29,900.00 and σ = \$172.92

Plan 2: *Variance* of X = \$48.00 and σ = \$6.93

The possible outcomes for Plan 1 are much more variable. If you wanted to ***invest cautiously***, you would choose **Plan 2**, but if you wanted to have the ***chance to gain a large amount of money***, you would choose **Plan 1**.

Notes about standard deviation

- Similar to when we used standard deviation for data in Chapter 2, it is most useful for *normal* random variables, which we will cover on Friday.
- In general, useful for comparing two random variables to see which is more spread out. *Examples*:
  - Two cities both have average yearly temperature of 65 degrees, but one has s.d of 5 degrees and the other has s.d. of 20 degrees. Which would you prefer?
  - One investment fund has average rate of return over many years of 8%, and s.d. of 2%. The other has average of 10%, but s.d. of 20%. The second one is higher on average, but is much more volatile.