Math 140
Introductory Statistics

Next test on Oct 19th

At the Hockey games

Team	Arena Seating (thousands)
New Jersey Devils	19
New York Islanders	16
New York Rangers	18
Philadelphia Flyers	18
Pittsburgh Penguins	17

Construct the probability distribution for X, the probability for the total number of people that can attend two distinct games

P(X), attendance at 2 hockey games

Teams	Maximum Attendance (thousands)
Devils/Islanders	19 + 16 = 35

So here we found X = 35,

Find the rest and then calculate
P(X) for all X

How many possible X values will we have?

P(X), attendance at 2 hockey games

Teams	Maximum Attendance (thousands)
Devils/Islanders	19 + 16 = 35

How many possible X values will we have?
We are selecting 2 arenas out of 5

The number of X values is

$$\binom{5}{2} = \frac{5!}{3!\ 2!} = \frac{5*4*3*2*1}{3*2*1*2*1} = 10$$

P(X), attendance at 2 hockey games

Teams	Maximum Attendance (thousands)
Devils/Islanders	19 + 16 = 35
Devils/Rangers	19 + 18 = 37
Devils/Flyers	19 + 18 = 37
Devils/Penguins	19 + 17 = 36
Islanders/Rangers	16 + 18 = 34
Islanders/Flyers	16 + 18 = 34
Islanders/Penguins	16 + 17 = 33
Rangers/Flyers	18 + 18 = 36
Rangers/Penguins	18 + 17 = 35
Flyers/Penguins	18 + 17 = 35

P(X), attendance at 2 hockey games

Total Possible Attendance (thousands), *x*	Probability, *p*
33	0.1
34	0.2
35	0.3
36	0.2
37	0.2

What is the probability that attendance is at least 36,000?

P(X), attendance at 2 hockey games

Total Possible Attendance (thousands), *x*	Probability, *p*
33	0.1
34	0.2
35	0.3
36	0.2
37	0.2

P(X > or = 36) = 0.2+0.2 = 0.4   40% probability

What do you think is the average number of cars per household?

Vehicles per Household, *x*	Proportion of Households, *p*
0	0.087
1	0.331
2	0.381
3	0.201

What do you think is the average number of cars per household?

Vehicles per Household, x	Proportion of Households, p
0	0.087
1	0.331
2	0.381
3	0.201

Average = 0 * 0.087 + 1 * 0.331 + 2 * 0.381 + 3 * 0.201 = 1.696

In a more abstract way

Average = Sum over all (X* P(X))

The mean of a probability distribution for the random variable X is called its **expected value** and is denoted by μ_X or $E(X)$.

The mean gets called expected value of the distribution
Sometimes indicated as E(X) or μ_x

Getting insured

The probability that you get burglarized is 26.9 per 1000 people

Outcome	Payout, x	Probability, p	$x \cdot p$
No burglary	0	0.9731	0
Burglary	5000	0.0269	134.50

What the company will pay is $5,000

Getting insured

The expected payout the company will pay is

0 if there are no burglaries
$5000 if there is a burglary

For each policy we will pay on average
0*0 + 5000 *0.0269

Outcome	Payout, x	Probability, p	$x \cdot p$
No burglary	0	0.9731	0
Burglary	5000	0.0269	134.50

= Sum (X * P(X)) = 134.50

Getting insured

The company will break even if they charge 134.50

97% of the time people will pay,
not be burglarized and not get anything,

The rest of the time they will be burglarized and get 5K

The company breaks even, since

$$E(X) = X\,P(X)$$

What everyone pays = What the company pays

The Wisconsin lottery

Winnings, x	Probability, p
\$1	1/10
\$2	1/14
\$3	1/24
\$18	1/200
\$50	1/389
\$150	1/20,000
\$900	1/120,000

A ticket costs \$1.
Why don’t the probabilities add to one?
What is missing?

The Wisconsin lottery

Winnings, x	Probability, p
$1	1/10
$2	1/14
$3	1/24
$18	1/200
$50	1/389
$150	1/20,000
$900	1/120,000

What is the expected value for the probability distribution?

The Wisconsin lottery

Winnings, x	Probability, p
\$1	1/10
\$2	1/14
\$3	1/24
\$18	1/200
\$50	1/389
\$150	1/20,000
\$900	1/120,000

P(0) = 1 - Sum of the above = 0.7793

$\mu_x = 0.6014$

If we spend \$1 we will get back, on average, 60 cents.

The Wisconsin lottery

We are given 10 tickets.
How much do we expect to win?

0.6014 for each ticket

10* 0.6014 = $6.01 for 10 tickets

Calculate the expected value

Value, x	Frequency, f
5	12
6	23
8	15

Value, x	Proportion, f/n
5	0.24
6	?
8	?

There are $n=50$ outcomes

Calculate the expected value

Value, x	Frequency, f
5	12
6	23
8	15

There are n=50 outcomes

Value, x	Proportion, f/n
5	0.24
6	?
8	?

$$\mu_x = \sum \left(x \frac{f}{n} \right)$$

~ 6.3

6.2 Variances of Probability distributions

$$\sigma^2{}_n = \sum \frac{\left(x - \bar{x}\right)^2}{n}$$

$$SD = \sqrt{\sigma^2{}_n} = \sqrt{\sum \frac{\left(x - \bar{x}\right)^2}{n}}$$

For a set of data - note the n instead of n-1

6.2 Variances of Probability distributions

For a probability distribution P(X)

$$\sigma^2{}_x = \sum \frac{\left(x - \bar{x}\right)^2}{n}$$

$$SD = \sqrt{\sigma^2{}_n} = \sqrt{\sum \frac{\left(x - \bar{x}\right)^2}{n}}$$

Try calculating the variance and SD

Winnings, *x*	Probability, *p*
\$1	1/10
\$2	1/14
\$3	1/24
\$18	1/200
\$50	1/389
\$150	1/20,000
\$900	1/120,000

Recall, the average is 60.14 cents or 0.6014 dollars

Try calculating the variance and SD

Winnings, *x*	Probability, *p*
$1	1/10
$2	1/14
$3	1/24
$18	1/200
$50	1/389
$150	1/20,000
$900	1/120,000

The way to think about this is: for each deviation from the mean, what is the probability? For example The contribution for the $1 winning is

$$(1 - 0.6014)^2 * 1/10$$

Try calculating the variance and SD

Winnings, x	Probability, p
\$1	1/10
\$2	1/14
\$3	1/24
\$18	1/200
\$50	1/389
\$150	1/20,000
\$900	1/120,000

So all together, the variance is

$$(1 - 0.6014)^2 * 1/10 + (2 - 0.6014)^2 * 1/14 + \text{etc etc} = 16.32$$

The standard deviation is its square root = \$4.04

Variances

Formula for the Variance of a Probability Distribution

The variance of a discrete probability distribution is given by

$$Var(X) = \sigma_X^2 = \sum (x - \mu_X)^2 \cdot p$$

Here, p is the probability that the random variable X takes on the specific value x and μ_X is the expected value. To get the standard deviation, take the square root of the variance.

Significance

The average payout is \$0.60
With a SD of \$4.04

This means on average you win little
but there is a chance of winning a lot
(that is why the SD is large)

Let's triple the lottery

Original Winnings, x	Winnings in Special Promotion, $3x$	Probability, p
\$0	\$0	0.7793
\$1	\$3	1/10
\$2	\$6	1/14
\$3	\$9	1/24
\$18	\$54	1/200
\$50	\$150	1/389
\$150	\$450	1/20,000
\$900	\$2700	1/120,000

Calculate mean and standard deviation
Compare to original lottery

Let's triple the lottery

Original lottery winnings \$0.6014
New winnings \$1.804

Is this smart on Wisconsin's part?

Original SD = \$4.04
New SD = \$12.12

How are they related?

They are both multiplied by 3!
SD = σ_X

Let's triple the lottery

μ_x turns into 3 μ_x

σ_X turns into 3 σ_X

This is true in general.
Multiplying data will rescale average and SD
of the distribution

What if I had decided to add
50 cents to each winning?

Let's add 50 cents to the payout

μ_x turns into $0.50 + \mu_x$

σ_X stays σ_X

Just like rescaling and recentering
So, adding C and multiplying by D gives

μ_x turns into $\mu_{C+DX} = C + D^*\mu_x$

σ_X turns into $\sigma_{C+DX} = D^*\sigma_X$

In general

Linear Transformation Rule: The Effect of a Linear Transformation of X on μ_X and σ_X

Suppose you have a probability distribution for random variable X, with mean μ_X and standard deviation σ_X. If you transform each value by multiplying it by d and then adding c, where c and d are constants, then the mean and the standard deviation of the transformed values are given by

$$\mu_{c+dX} = c + d\mu_X$$
$$\sigma_{c+dX} = |d|\sigma_X$$

Now, this was for TRIPLING the lottery

What if we kept the same lottery and bought 3 tickets?

What do you think?

If every time I play my average payout is \$0.6014
What do I get after buying 3 tickets?

What do you think?

If every time I play my average payout is $0.6014
What do I get after buying 3 tickets?

Duh - 3 * 0.6014 = 1.804!

Just like before!

It does not matter if I triple the lottery or
if I buy three tickets, the result is the same.

My take-home on average is tripled.

What do you think?

We can conclude that when we select three items from the same distribution we find

$$\mu_{3,X} = 3\,\mu_x = \mu_x + \mu_x + \mu_x$$

In general, for different distributions we get

$$\mu_{X,Y} = \mu_x + \mu_y$$

Two tickets from two lotteries

Let's buy a ticket from the lottery of California and of Texas

California $\mu_x = \$0.50$
Texas $\mu_Y = \$0.75$

What are the expected total winnings?

$$\mu_{CA,TX} = \mu_{CA} + \mu_{TX} = \$0.50 + \$0.75 = \$1.25$$

What about the standard deviation?

We will look at that next time.

Practice and hk

Page 286
E11, E12, E13, E14, E15, E16

Page 295
P10, P13, E19, E20,