ECE 111 - Homework #12

Week #12: ECE 341 Random Processes. Due Tuesday, April 11th

Chi-Squared Tests

Problem 1: The following Matlab code generates 90 random die rolls for a six sided die

```
RESULT = zeros(1,6);
for i=1:90
   D6 = ceil( 6*rand );
   RESULT(D6) = RESULT(D6) + 1;
   end
RESULT
```

Determine whether this is a fair or loaded die using a Chi-Squared test.

```
             1     2     3     4     5     6
RESULT =    13    14    13    13    20    17
```

Set up a chi-squared table. The expected frequency (n*p) is

$$np = (90)\left(\frac{1}{6}\right) = 15$$

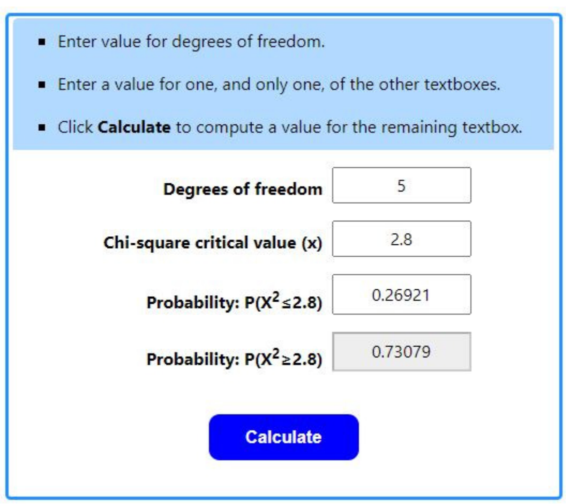
Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	15	13	0.27
2	1/6	15	14	0.07
3	1/6	15	13	0.27
4	1/6	15	13	0.27
5	1/6	15	20	1.67
6	1/6	15	17	0.27
			Total	**2.8**

From StatTrek. with 5 degrees of freedom, a chi-squared score of 2.80 corresponds to a probability of

p = 0.26921

There is a 26.921% chance that this is a loaded die

note: answers vary - this is a random process



Problem 2: The following Matlab code generates 90 rolls of a loaded six-sided die (12% of the time, you roll a 6):

```
RESULT = zeros(1,6);
for i=1:90
   if(rand < 0.12)
      D6 = 6;
   else
      D6 = ceil( 6*rand );
      end
   RESULT(D6) = RESULT(D6) + 1;
   end
RESULT
```

Determine whether this is a fair or loaded die using a Chi-Squared test.

```
RESULT =     13     14     16     13     12     22
```

Compute the chi-squared score:

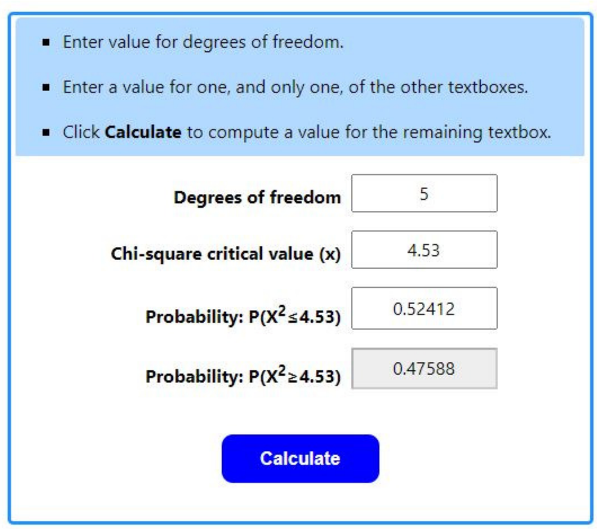
Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	15	13	0.27
2	1/6	15	14	0.07
3	1/6	15	16	0.07
4	1/6	15	13	0.27
5	1/6	15	12	0.6
6	1/6	15	22	3.27
			Total	**4.53**

Use StatTrek to conver this to a probability: p = 0.52412

There is a 52.412% chance this is a loaded die

It's not easy to spot 12% loading

note: answers vary - this is a random process



Sidelight: Roll the dice 900 times (not required - just playing with Matlab)

```
RESULT = zeros(1,6);
for i=1:900
   if(rand < 0.12)
      D6 = 6;
   else
      D6 = ceil( 6*rand );
      end
   RESULT(D6) = RESULT(D6) + 1;
   end
RESULT

   136    131    125    137    155    216
```

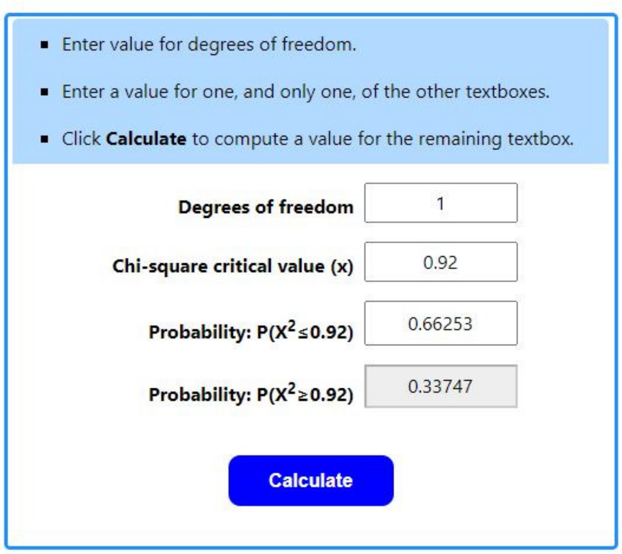
Calculated the chi-squared score:

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	150	136	1.31
2	1/6	150	131	2.41
3	1/6	150	125	4.17
4	1/6	150	137	1.13
5	1/6	150	166	1.71
6	1/6	150	216	29.04
			Total	**39.75**

Now you can be almost 100% certain that the die is loaded

- Given enough data, you can spot even 1% loading
- It might take a **lot** of rolls to detect small amounts of loading, but you do it.

Chi-squared tests also let you calculate how many rolls you need in order to detect a given amount of loading.



Am I Psychic?

Problem #3: Shuffle a deck of 52 playing cards and place it face down on a table.

- Predict the suit of the top card then reveal it. If correct, place the card in one pile (correct). If incorrect, place it in another pile.
- Repeat for all 52 cards.

Use a chi-squared test to test the hypothesis that you're just guessing (probability of being correct is 25%)

I was correct 10 times out of 52

Calculating the chi-squared score:

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
Correct	1/4	13	10	0.69
Incorrect	3/4	39	42	0.23
			Total	0.92

Use StatTrek to convert this to a probabiliy od 0.66253

There is a 66.253% chance that I'm not just guessing

- 99% tells me I'm not guessing (psychic)
- 1% tells me my guesses are not random (such as I always guess spades)
- Inbetween tells me I'm just guessing. Sigh.

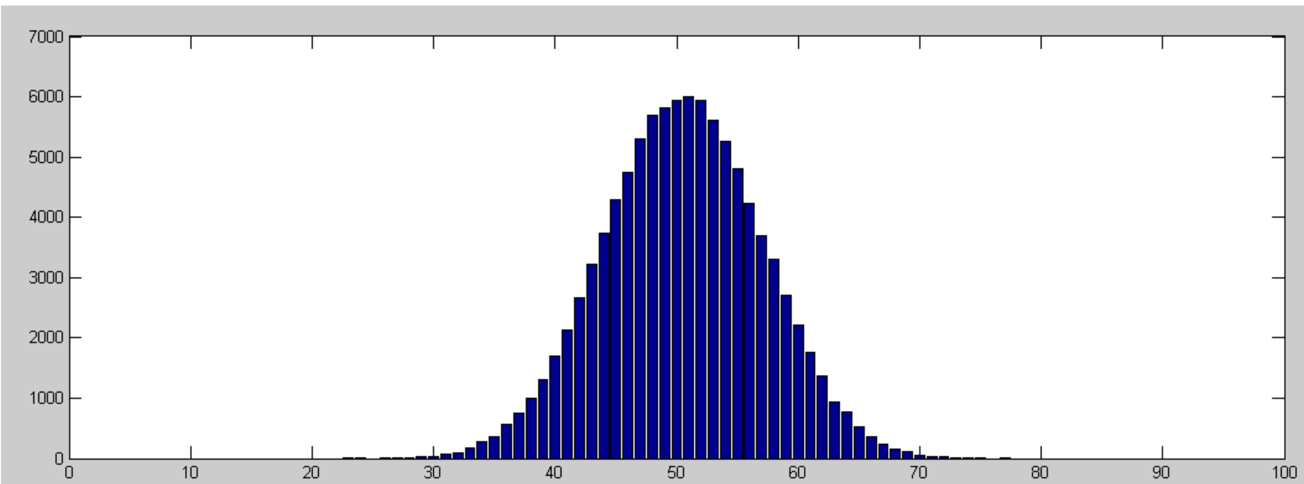
Monte-Carlo Simulation

Problem #4: Let y be the sum of six 4-sided dice plus five 6-sided dice plus four 8-sided dice

$$y = 6d4 + 5d6 + 4d8$$

a) Generate 100,000 values for y using Matlab and plot the frequency of each number on a bar chart

```
RESULT = zeros(90,1);
for i=1:1e5
   d4 = ceil( 4*rand(6,1) );
   d6 = ceil( 6*rand(5,1) );
   d8 = ceil( 8*rand(4,1) );
   Y = sum(d4) + sum(d6) + sum(d8);
   RESULT(Y) = RESULT(Y) + 1;
   end;
bar(RESULT)
```



Frequnecy of Each Result

Note: The resuult is a bell-shaped curve (central-limit theorem in action)

b) From your results, determine the probability that y > 59.5 (the number of times the sum is more than 59.5)

```
>> sum(RESULT(60:90)) / 1e5

ans =    0.0857
```

The sum is 60 or more 8.57% of the time

c) From your results, determine 'a' such that y < a 5% of the time

```
>> sum(RESULT(1:40)) / 1e5

ans =    0.0634

>> sum(RESULT(1:39)) / 1e5

ans =    0.0465
```

5% happens somewhere between 39 and 40. Call it 39.5

d) From your results, determine 'b' such that y > b 5% of the time

```
>> sum(RESULT(61:90)) / 1e5

ans =    0.0636

>> sum(RESULT(62:90)) / 1e5

ans =    0.0459
```

5% happens somewhere between 61 and 62 (call it 61.5)

Note: the 90% confidence interval for y is a < y < b.

The 90% confidnece interval is 39.5 < y < 61.5

Normal Approximation

The mean and standard deviation for a fair 6-sided die and 4-sided die are:

$$\mu_{d4} = 2.5 \qquad \mu_{d6} = 3.5 \qquad \mu_{d8} = 4.5$$

$$\sigma_{d4} = 1.118 \qquad \sigma_{d6} = 1.7078 \qquad \sigma_{d10} = 2.2913$$

Problem 5: Let Y be the sum of rolling six 4-sided dice (6d4) plus five 6-sided dice (5d6) plus four 8-sided dice.

Y = 6d4 + 5d6 + 4d8

a) What is the mean and standard deviation of Y?

When you sum normal distributions, the means add

$$\mu_y = 6 \cdot 2.5 + 5 \cdot 3.5 + 4 \cdot 4.5$$

$$\mu_y = 50.5$$

and the variance adds

$$\sigma_y^2 = 6 \cdot (1.118)^2 + 5 \cdot (1.7078)^2 + 4 \cdot (2.2913)^2$$

$$\sigma_y^2 = 43.0833$$

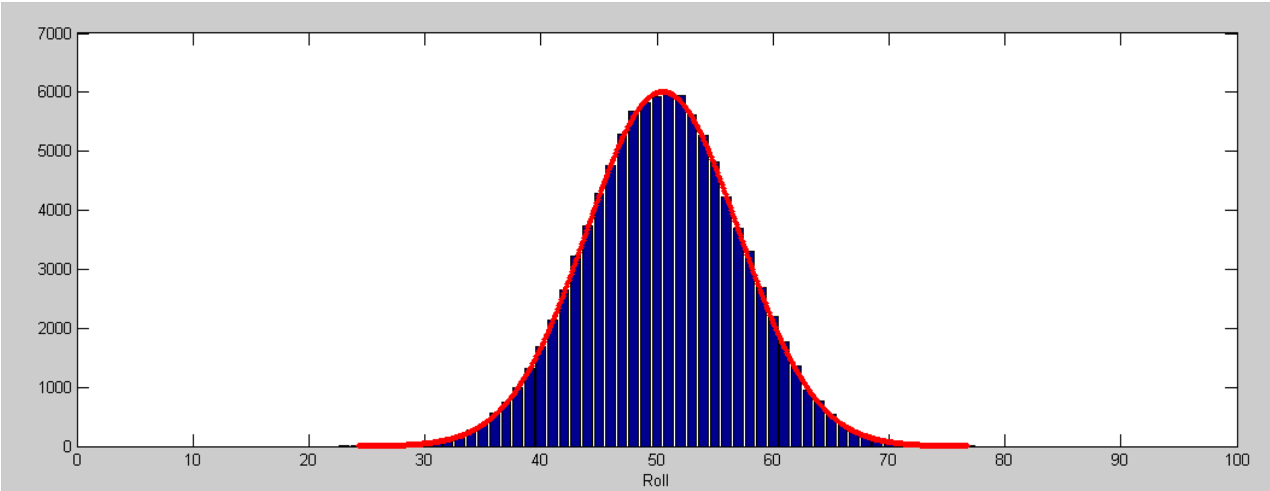
resulting in the standard deviation being

$$\sigma_y = \sqrt{43.0833}$$

$$\sigma_y = 6.5638$$

Just for fun, plot the normal distribution on top of the Monte-Carlo simulation (scaled so that the peak is the same). The normal approximation is almost dead on (with requiring zero die rolls)

```
>> hold on
>> x = 50.5;
>> s = 6.5638;
>> s1 = [-4:0.01:4]';
>> p = exp(-s1.^2 / 2);
>> p = p * max(RESULT);
>> plot(s1*s+x, p, 'r.')
>> xlabel('Roll');
```



b) Using a normal approximation, what is the 90% confidence interval for Y?

From StatTrek, the z-score for 5% tails is 1.64485

$$\mu - 1.64485\sigma < roll < \mu + 1.64485$$

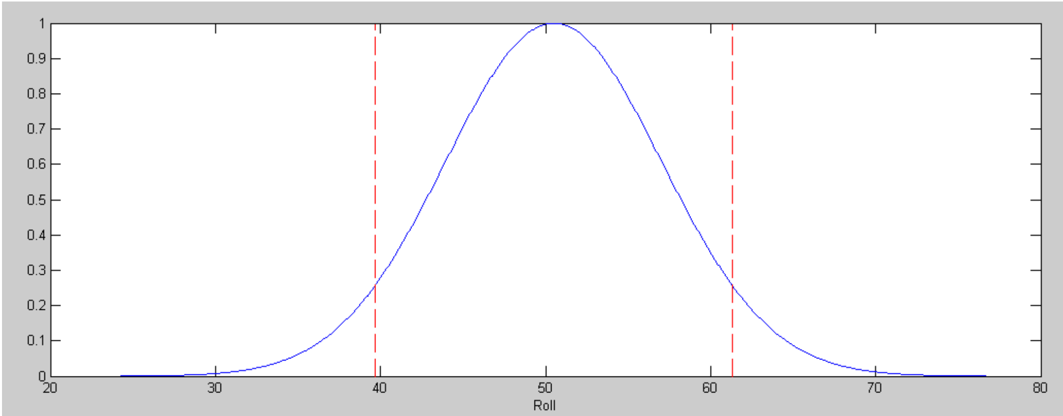
$$39.704 < roll < 61.296$$ *zero rolls (normal approximation)*

note: From a Monte-Carlo simulation, the result was

$$39.5 < roll < 61.5$$ *100,000 rolls*

Note that with a normal approximation, I got the same result without needing *any* die rolls.

```
>> clf
>> p = p / max(p);
>> plot(s1*s+x, p, 'b', 39.704*[1,1],[0,1],'r--', 61.296*[1,1],[0,1],'r--')
>> xlabel('Roll');
```



90% confidence interval calculated using a normal distribution.

c) Using a normal approximation, what is the probability that the sum the dice will be more than 59.5?

Find the z-score

$$z = \left(\frac{59.5-\mu_y}{\sigma_y}\right) = \left(\frac{59.5-50.5}{6.5638}\right) = 1.3712$$

From StatTrek, this corresponds to a probability of 0.08516

There is am 8.561% chance of rolling more than 59.5

note: From the Monte-Carlo simulation, the odds are 8.57%. With a normal approximation, I got the same result with zero die rolls.

Student-t Test

Problem 6: Using Matlab, determine four values for Y

Y = 6d4 + 5d6 + 4d8

6a) From this, determine the mean and standard deviation of your data set.

```
DATA = [];
for i=1:4
   d4 = ceil( 4*rand(6,1) );
   d6 = ceil( 6*rand(5,1) );
   d8 = ceil( 8*rand(4,1) );
   Y = sum(d4) + sum(d6) + sum(d8);
   DATA = [DATA, Y];
   end
DATA
x = mean(DATA)
s = std(DATA)

x =   52.5000
s =    8.5829
DATA =    49    59    42    60
```

6b) Use a t-test to determine

The 90% confidence interval

From StatTrek, 5% tails with 3 degrees of freedom corresponds to a t-score of 2.3534

$$\bar{x} - 2.3534s < roll < \bar{x} + 2.3524s$$

$$32.309 < roll < 72.690$$ *sample size = 4*

The actual 90% confidence interval is

$$39.704 < roll < 61.296$$ *sample size = infinity*

With a sample size of four, the results are close but a little off

The probabillity of scoring more than 59.5 points.

Find the t-score

$$t = \left(\frac{59.5 - \bar{x}}{s}\right) = 0.81563$$

From StatTrek, this corresponds to a tail with an area of 23.72

There is a 23.72% chance of rolling 59.5 or higher

The actual odds are 8.561%. The odds are a little off, but then the sample size is only four.

Problem 7: Using Matlab, determine ten values for Y

Y = 2d4 + 3d6 + 4d8

```
DATA = [];
for i=1:10
   d4 = ceil( 4*rand(6,1) );
   d6 = ceil( 6*rand(5,1) );
   d8 = ceil( 8*rand(4,1) );
   Y = sum(d4) + sum(d6) + sum(d8);
   DATA = [DATA, Y];
   end
DATA

DATA =    60    49    42    55    48    53    62    46    44    53
```

7a) From this, determine the mean and standard deviation of your data set.

```
x = mean(DATA)
s = std(DATA)

x =   51.2000
s =    6.6131
```

7b) Use a t-test to determine

The 90% confidence interval

5% tails and nine degrees of freedom corresponds to a t-score of 1.8331

$$\bar{x} - 1.8331s < roll < \bar{x} + 1.8331s$$

$39.0775 < roll < 63.3225$ ten rolls

Note: the actual 90% conficence interval is

$39.704 < roll < 61.296$ infinite rolls

With only ten rolls, you're pretty close. More rolls gets closer.

The probabillity of scoring more than 59.5 points.

Find the t-score

$$t = \left(\frac{59.5 - \bar{x}}{s}\right) = 1.2551$$

From StatTrek, with 9 degrees of freedom, this t-score corresponds to a tail with an area of 12.053%

p(y > 59.5) = 12.053% ten rolls

p(y > 59.5) = 8,561% infinte rolls

Summary:

Method	# Die Rolls	90% Conf Interval	p(roll > 59.5)
Monte-Carlo	100,000	39.5 < roll < 61.5	8.57%
Normal Dist	0	39.7 < roll < 61.3	8.56%
t-Test	4	32.3 < roll < 72.7	23.7%
t-Test	10	39.1 < roll < 63.3	12.1%

Comment: With statistics, you can get similar results using only a few measurements.

- More measurements give closer results
- You don't need a huge sample size: four to ten is actually pretty good.