ECE 341 - Homework #14

Chi-Squared Tests.

Loaded Dice

1) The following Matlab code generates 240 random numbers from 1..6 (240 6-sided dice)

```
RESULT = zeros(1,6);
for i=1:240
   d6 = ceil(rand*6);
   RESULT(d6) = RESULT(d6) + 1;
end
```

Use a chi-squared test to determine if this is a fair die.

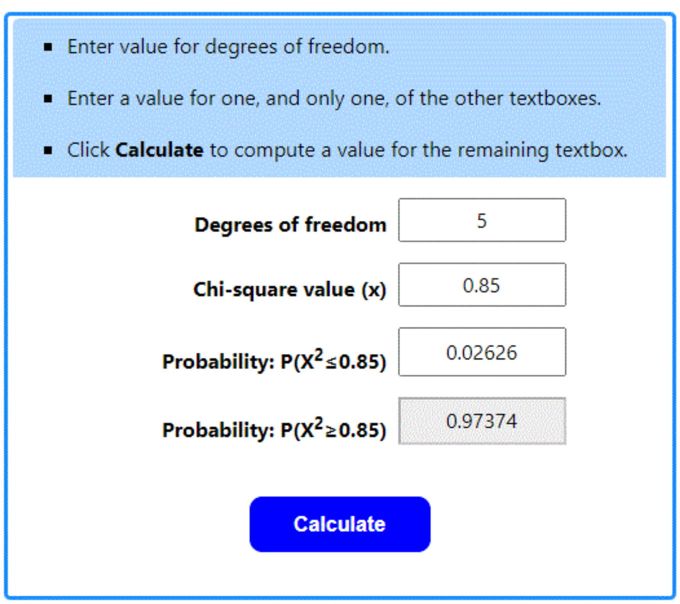
```
RESULT =

    40    44    42    38    39    37
```

Put this into a table and compute the chi-squared score:

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	40	40	0
2	1/6	40	44	0.4
3	1/6	40	42	0.1
4	1/6	40	38	0.1
5	1/6	40	39	0.025
6	1/6	40	37	0.225
			Total	0.85

Use a chi-squared table (or StatTrek) to convert this to a probability (p = 0.02626)

I can reject the null hypothesis (this is a fair die) with 52.626 certainty



2) The following Matlab code generates 240 random die rolls with 5% loading (5% of the time you roll a 6)

```
RESULT = zeros(1,6);
for i=1:240
   if(rand < 0.05) d6 = 6;
   else d6 = ceil(rand*6);
   end
   RESULT(d6) = RESULT(d6) + 1;
end
```

Use a chi-squared test to determine if this is a fair die.

```
RESULT =     37     37     33     43     45     45
```

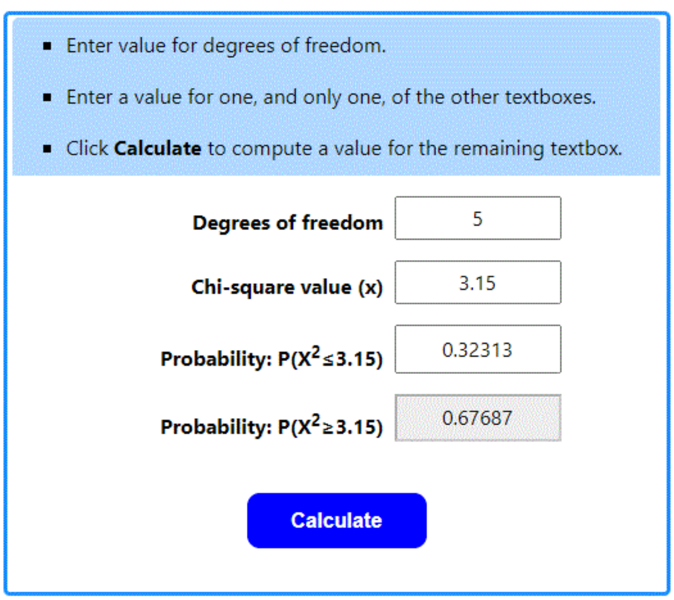
Put this into a table and compute the chi-squared score:

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	40	37	0.225
2	1/6	40	37	0.225
3	1/6	40	33	1.225
4	1/6	40	43	0.225
5	1/6	40	45	0.625
6	1/6	40	45	0.625
			Total	3.15

Convert this to a prbability using a chi-squared table (p = 0.32313)

I can reject the null hypothesis (this is a fair die) with 32.3% certainty

Wih only 200 die rolls, I can't detect 5% loading



3) Repeat problem #2 with 1000 die rolls

```
RESULT = zeros(1,6);
for i=1:1000
   if(rand < 0.05) d6 = 6;
   else d6 = ceil(rand*6);
   end
   RESULT(d6) = RESULT(d6) + 1;
end
```

Use a chi-squared test to determine if this is a fair die.

```
RESULT =     163    162    144    156    173    202
```

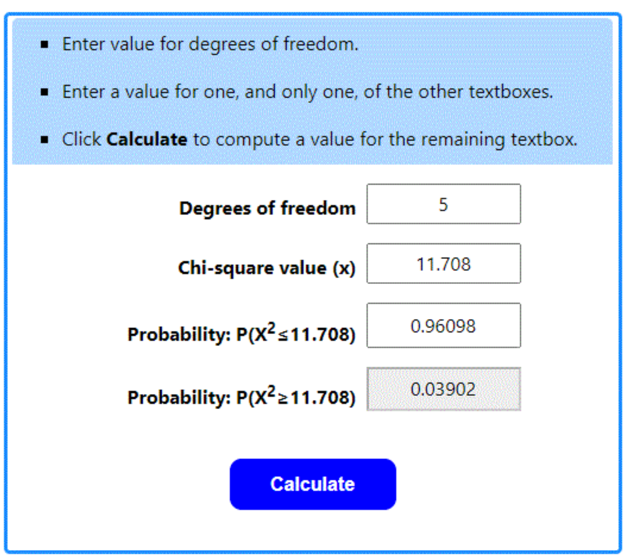
Compute the chi-squred score

Roll	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	166.67	163	0.0808
2	1/6	166.67	162	0.1309
3	1/6	166.67	144	3.0835
4	1/6	166.67	156	0.6831
5	1/6	166.67	173	0.2404
6	1/6	166.67	202	7.4891
			Total	11.7078

Convert this to a probability using StatTrek. (p = 0.96098)

I can reject the null hypothesis (this is a fair die) with 96.1% certainty

Note: With a large enough sample size, you can detect 5% loading. It might take a *lot* of die rolls though.



Am I psychic?

4) Shuffle a deck of 52 playing cards. Without looking at the top card, predict the suit (clubs, diamonds, hearts, and spades). Repeat for all 52 cards, keeping track of how many you got right and how many you got wrong.

- From the results, use a chi-squared test to determine if you are just guessing (25% chance of getting the suit correct.)

I was correct 10 times out of 52

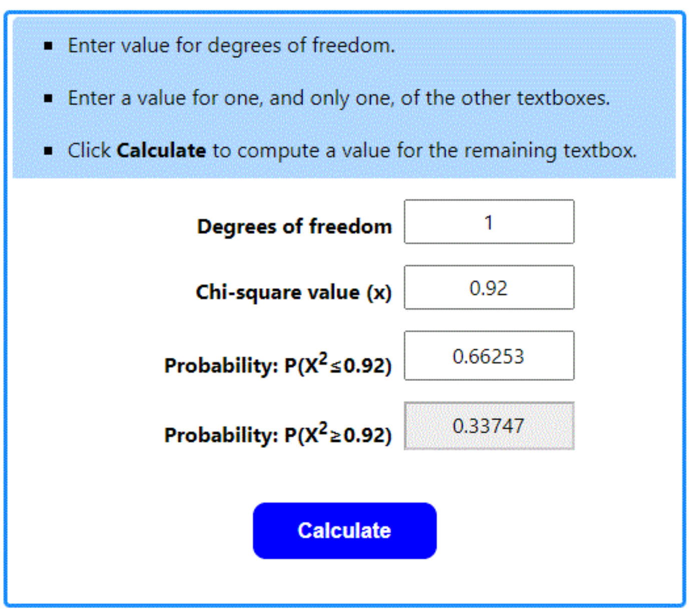
Calculating the chi-squared score:

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

I can reject the null hypothesis (I'm just guessing) with 66.2% certainty

Promising, but not enough to bet the house



Central Limit Theorem:

5) The following code sums four uniform distributions

Y = sum(rand(3,1));

The Central Limit Theorem states that this will converge to a normal distribution with

- mean = 1.5
- variance = 3/12
- standard deviation = 1/2

Use a chi-squared test to determine if Y does / does not have a normal distribution.

Step 1: Collect data

- collected 100 values for Y

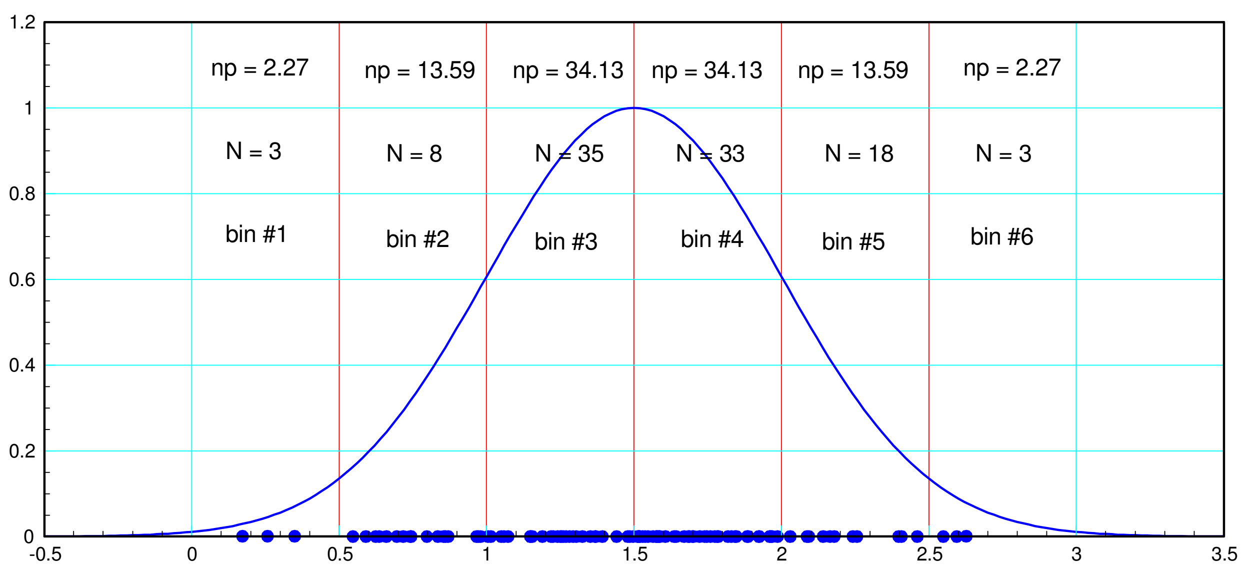
Step 2: Split the range space into N bins

- This is somewhat arbitrary
- I'll make each bin one standard deviation (1/2)

Step 3: Count how many times the data falls into each bin

Matlab Code

```
RESULT = zeros(1,6);
for n=1:100
    Y = sum(rand(3,1));
    if(Y < 0.50) bin = 1;
    elseif(Y < 1.00) bin = 2;
    elseif(Y < 1.50) bin = 3;
    elseif(Y < 2.00) bin = 4;
    elseif(Y < 2.50) bin = 5;
    else bin = 6;
    end
    RESULT(bin) = RESULT(bin) + 1;
end
RESULT =       3      8     35     33     18      3
```



Calculate the chi-squared score

bin (Y)	bin (z-score)	p	n*p	N	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
< 0.50	< -2	0.02275	2.275	3	0.231
(0.50, 1.00)	(-2, -1)	0.13590	13.59	8	2.2993
(1.00, 1.50)	(-1, 0)	0.34135	34.135	35	0.0219
(1.50, 2.00)	(0, +1)	0.34135	34.135	33	0.0377
(2.00, 2.50)	(+1, +2)	0.13590	13.59	18	1.4311
> 2.50	(> +2)	0.02275	2.275	3	0.231
				Total	4.2522

Convert the chi--squared score to a probability using a ch-squared table

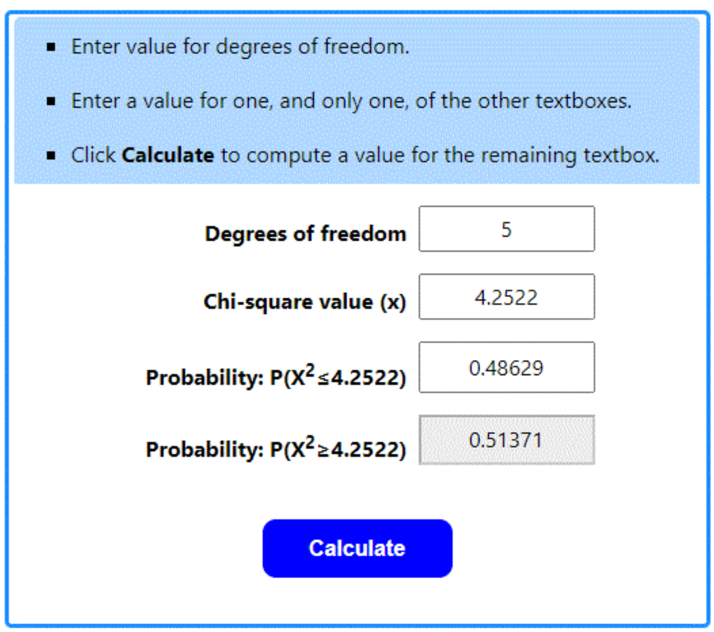
p = 0.48629

I can reject the null hypothesis (Y has a normal distribution) with 48.6% certainty

With only 100 data points, I don't have enough evidence to claim that Y is not a Normal distribution

central limit theorem in action...

With more data, I might be able to tell the difference...



Poisson approximation for a binomial distribution.

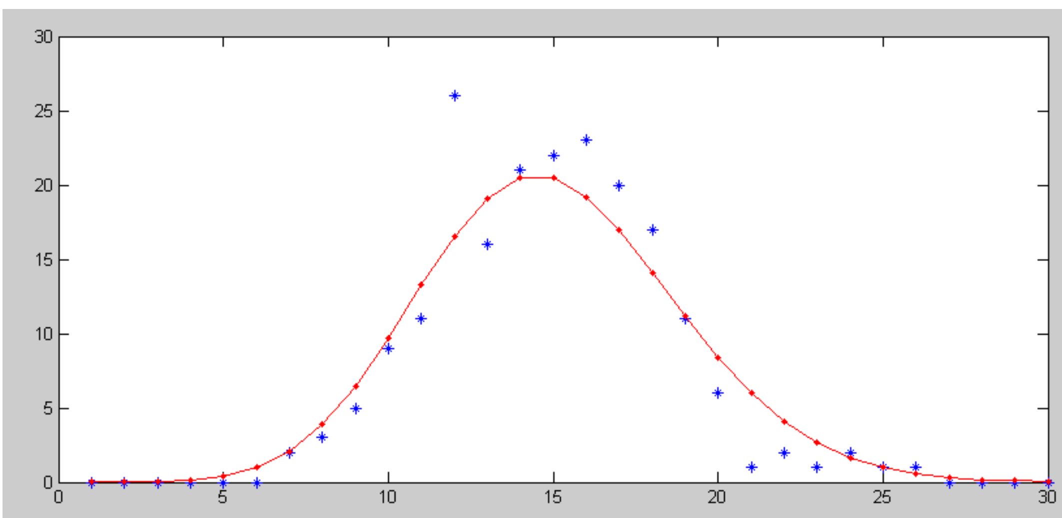
6) Let X be the number of 1's you get when you roll 90 dice. The Poisson approximation for the pdf is

$$\binom{90}{x}\left(\frac{1}{6}\right)^{x}\left(\frac{5}{6}\right)^{90-x} \approx \left(\frac{1}{x!}\right) 15^{x}\, e^{-15}$$

- Use Matlab to count the number of 1's you get when you roll 90 dice
- Repeat 200 times
- Check whether the result is consistent with a Poisson distribution with $\lambda = Np = 15$ using a Chi-squred test

Code:

```
Result = zeros(90,1);
for i=1:200
    Dice = ceil(6*rand(90,1));
    N = sum(Dice == 1);
    Result(N) = Result(N) + 1;
end

k = [1:30]';
Result = Result(k);

Poisson = 1./ factorial(k) .* (15.^k) * exp(-15) * 200;

plot(k,Result,'b*',k,Poisson,'r.-')
```



Step 2: Split the data into 10 bins (groups of three)

bin	k values	n*p Poisson	N Result	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	{0,1,2}	0.0422	0	0.0422
2	{3,4,5}	1.4841	0	1.4841
3	{6,7,8}	12.4444	10	0.4801
4	{9,10,11}	39.5515	46	1.0514
5	{12,13,14}	60.0957	59	0.02
6	{15,16,17}	50.2764	60	1.8806
7	{18,19,20}	25.4844	18	2.1981
8	{21,22,23}	8.3883	5	1.3686
9	{24,25,26}	1.8898	2	0.0064
10	{27,28,29,30}	0.3037	0	0.3037
			Total	**8.8352**

Convert the chi-squared score to a probability (p = 0.54738)

I can reject the null hypothesis (data has a Poisson distribution) with 54.738% certainty

With only 200 data points, I can't tell the difference between a binomial and a Poisson distribution