ECE 341 - Homework #14

Chi-Squared Tests. Due Thursday, June 11th

Please make the subject "ECE 341 HW#13" if submitting homework electronically to Jacob_Glower@yahoo.com (or on blackboard)

Refrigerator Door:

1) The time that a refrigerator door is held open is recorded over a 2-day periond. It is conjectured that this is an exponential distribution. Use a chi-squared test to determine whether this is or isn't.

```
day 1:
5.126, 5.720, 3.112, 12.250, 3.811, 32.847, 4.269,  5.085, 3.521, 3.552,
3.950, 8.417, 5.051,  5.868, 3.353,  3.959, 3.086, 50.000, 5.863, 3.531,
4.271, 6.421

day 2:
11.673,  2.425, 20.651,  4.796, 3.967,  2.836, 165.279, 1.156,  6.025, 5.884,
78.509, 56.183,  4.987,  5.047, 5.139, 28.309,   3.200, 2.620, 41.602, 7.147,
10.963,  6.559, 13.491, 18.940, 4.327,  6.277,   9.794, 7.398,  5.823, 7.126
```

In order for the assumed distribution to have a chance of matching the data, assume

- An exponential distribution
- With the same mean as the data
- With the data point at 165.279 seconds thrown out

Next comes the problem of how to divide the X-axis into N bins. Let's assume five bins, each with equal probability. The PDF and CDF are

```
TIME = [5.126, 5.720, 3.112, 12.250, 3.811, 32.847, 4.269,  5.085, 3.521, 3.552,
3.950, 8.417, 5.051,  5.868, 3.353,  3.959, 3.086, 50.000, 5.863, 3.531, 4.271,
6.421, 11.673,  2.425, 20.651,  4.796, 3.967,  2.836, 1.156,  6.025, 5.884,
78.509, 56.183,  4.987,  5.047, 5.139, 28.309,   3.200, 2.620, 41.602, 7.147,
10.963,  6.559, 13.491, 18.940, 4.327,  6.277,   9.794, 7.398,  5.823, 7.126
];

a = mean(TIME)

a =   11.0964
```

$$f(t) = \left(\frac{1}{11.09}\right) \exp\left(\frac{-t}{11.09}\right) u(t) \qquad \text{pdf}$$

$$F(t) = \left(1 - \exp\left(\frac{-t}{11.09}\right)\right) u(t) \qquad \text{cdf}$$

Now, determine where to place the bins.

- Assume 5 bins (51 data points gives 10 data points in each bin)
- Assume equal probabililty for each bin (could also do equal spacing)

That places the bins at

- 0% - 20%
- 20% - 40%
- 40% - 60%
- 60% - 80%
- 80% - 100%

To convert the probabilities to times, use the CDF

$$t = -11.09\ \ln(1 - F(t))$$

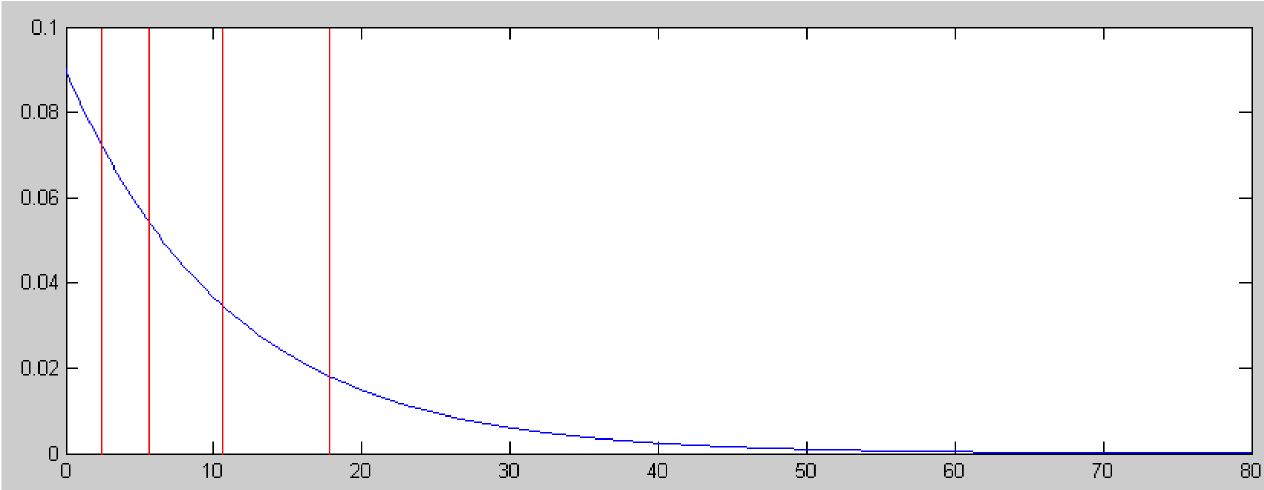
```
p =    0.2000    0.4000    0.6000    0.8000

T = -11.09*log(1-p)

T = 2.4747    5.6651   10.1617   17.8487
```

Sort the data and count how many data points fall into each bin

0 - 1.4747	1.4747 - 5.6651	5.6651 - 10.1617	10.1617 - 17.8487	17.8487 - infinity
1.1560	2.4250	5.7200	10.9630	18.9400
	2.6200	5.8230	11.6730	20.6510
	2.8360	5.8630	12.2500	28.3090
	3.0860	5.8680	13.4910	32.8470
	3.1120	5.8840		41.6020
	3.2000	6.0250		50.0000
	3.3530	6.2770		56.1830
	3.5210	6.4210		78.5090
	3.5310	6.5590		
	3.5520	7.1260		
	3.8110	7.1470		
	3.9500	7.3980		
	3.9590	8.4170		
	3.9670	9.7940		
	4.2690			
	4.2710			
	4.3270			
	4.7960			
	4.9870			
	5.0470			
	5.0510			
	5.0850			
	5.1260			
	5.1390			



Now form a Chi-squared table

Die Roll (bin)	p theoretical probability	np expected frequency	N actual frequency	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
0 - 1.4747	1/5	10.2	1	8.298
1.4747 - 5.6651	1/5	10.2	24	18.6706
5.6651 - 10.1617	1/5	10.2	14	1.4157
10.1617 - 17.8487	1/5	10.2	4	3.7686
17.8487 - infinity	1/5	10.2	8	0.4745
			Total:	32.6275

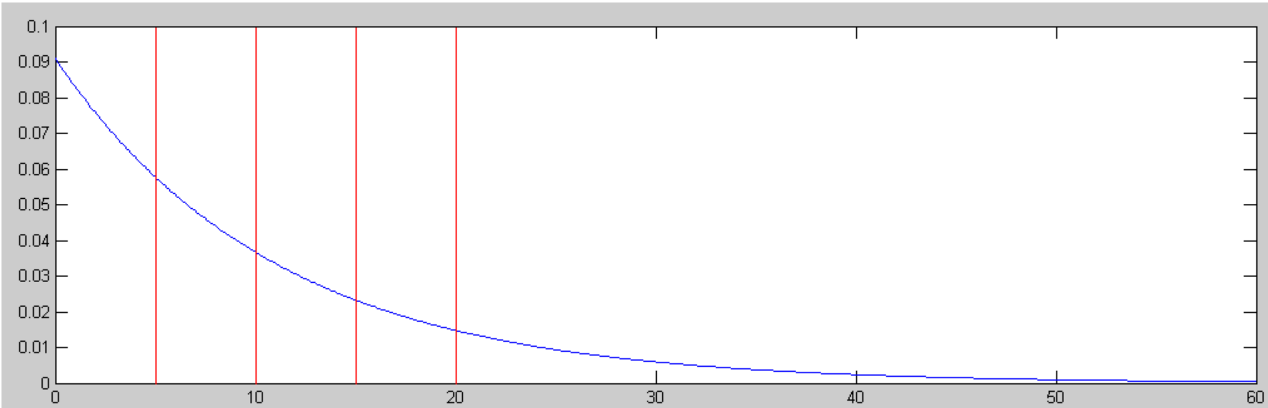
Use a chi-squared table to convert this to a probability

p = 1.0000 (meaning p > 0.99995)

I am at least 99.995% certain that this is not an exponential distribution.

it looks more like a Gamma distribution with small probability near t = 0

Option #2: equal spacing in time



0 - 5	5 - 10	10 - 15	15 - 20	20+
1.1560	5.0470	10.9630	18.9400	28.3090
2.4250	5.0510	11.6730	20.6510	32.8470
2.6200	5.0850	12.2500		41.6020
2.8360	5.1260	13.4910		50.0000
3.0860	5.1390			56.1830
3.1120	5.7200			78.5090
3.2000	5.8230			
3.3530	5.8630			
3.5210	5.8680			
3.5310	5.8840			
3.5520	6.0250			
3.8110	6.2770			
3.9500	6.4210			
3.9590	6.5590			
3.9670	7.1260			
4.2690	7.1470			
4.2710	7.3980			
4.3270	8.4170			
4.7960	9.7940			
4.9870				

Die Roll (bin)	p theoretical probability	np expected frequency	N actual frequency	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
0 - 5	0.3629	18.51	20	0.1199
5 - 10	0.2312	11.79	19	4.4092
10 - 15	0.1473	7.51	4	1.6405
15 - 20	0.0938	4.78	2	1.6168
20+	0.1647	8.4	6	0.6857
			Total:	8.4721

From StatTrek, a Chi-squared score of 8.47 corresponds to a probability of 0.92

I am 92% certain that this is not an exponential distribution with a mean of 11.09

Am I psychic?

Person A rolled a 6-sided die 100 times and predicted the result before each roll. What is the probability that person A is psychic? (i.e. reject the null hypothesis that guessing was random)

Prediction (in order)

```
2 6 1 6 1 6 2 2 5 2 2 5 2 5 6 4 3 4 6 5 1 1 5 2 5 5 1 6 1 5 5 3 3 1 4 1 6 3 1
3 3 1 4 3 4 4 1 1 5 3 5 3 3 4 2 3 3 1 4 1 1 1 6 4 6 5 1 1 6 5 4 3 6 1 4 5 3 5
4 4 6 4 4 4 4 2 4 1 5 4 3 4 1 4 3 4 1 6 2 5
```

Die Roll (in order)

```
2 4 6 2 3 2 4 5 5 1 4 1 2 6 4 4 4 2 5 2 4 3 6 6 4 4 4 2 5 1 4 4 6 2 1 4 2 1 1
6 6 2 4 6 1 3 6 6 3 3 4 6 2 5 5 3 4 2 4 1 5 1 4 5 3 6 1 6 5 6 2 6 2 3 6 5 4 1
5 5 1 4 3 3 4 4 3 6 2 2 4 4 4 6 2 4 4 5 4 2
```

2) Is this a fair die? (use a Chi-squared to test the null hypothesis: all numbers have a 1/6 chance of coming up)

Die Roll (bin)	p theoretical probability	np expected frequency	N actual frequency	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	16.67	22	1.7042
2	1/6	16.67	10	2.6688
3	1/6	16.67	16	0.0269
4	1/6	16.67	22	1.7042
5	1/6	16.67	17	0.0065
6	1/6	16.67	13	0.808
			Total:	6.9186

A chi-squared score of 6.91 corresponds to a probability of 77%

I am 77% certain that this is not a fair die.

I cannot say with 90% certainty that this is not fair die - no conclusion.

3) Is the prediction random? (each number has equal probability). Check with a Chi-squared test.

Die Roll (bin)	p theoretical probability	np expected frequency	N actual frequency	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
1	1/6	16.67	12	1.3083
2	1/6	16.67	18	0.1061
3	1/6	16.67	11	1.9285
4	1/6	16.67	28	7.7006
5	1/6	16.67	13	0.808
6	1/6	16.67	18	0.1061
			Total:	11.9576

This corresponds to a probability of 97%

I am 97% certain that the guesses are not uniformly distributed

4) Is the person psychic? (does the predicted number match the actual die roll more than it should?). Check with a Chi-squared test.

Several ways to do this. Group the data in to two bins:

- correct guesses
- # incorrect guesses

Guess (bin)	p theoretical probability	np expected frequency	N actual frequency	$\chi^2 = \left(\frac{(np-N)^2}{np}\right)$
Correct	1/6	16.67	17	0.0065
Incorrect	5/6	83.33	83	0.0013
			Total:	0.0078

A chi-squared score of 0.0078 corresponds to a probability of 7%

There is a 7% chance the person is psychic

There is a 93% chance that 17 correct guesses was just chance

Poisson approximation for a binomial distribution.

5) Let X be the number of 1's you get when you roll 60 dice. The Poisson approximation for the pdf is

$$\begin{pmatrix} 60 \\ x \end{pmatrix} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{60-x} \approx \left(\frac{1}{x!}\right) 10^x e^{-10}$$

- Use Matlab to count the number of 1's you get when you roll 60 dice
- Repeat 100 times
- Check whether the result is consistent with a Poisson distribution with $\lambda = Np = 10$ using a Chi-squred test

First, roll 60 dice 100 times to collect data

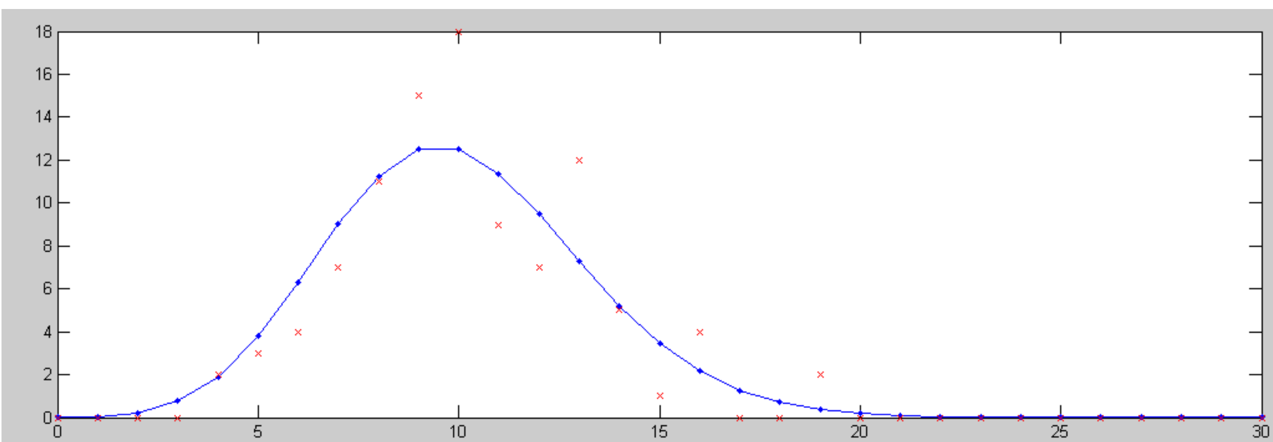
```
RESULT = zeros(61,1);
N = 100;
for i=1:N
   X = sum( ceil( 6*rand(60,1) ) == 1)
   RESULT(X + 1) = RESULT(X + 1) + 1;
end

bar(RESULT)

x = [0:60]';
p = 0*x;
for i=1:61
    p(i) = 1 / factorial(x(i)) * 10^x(i) * exp(-10);
end

plot(x,N*p,'b.-',x,RESULT,'rx')

xlim([0,30])
```



Next, decide how many bins you want and define what those bins are.

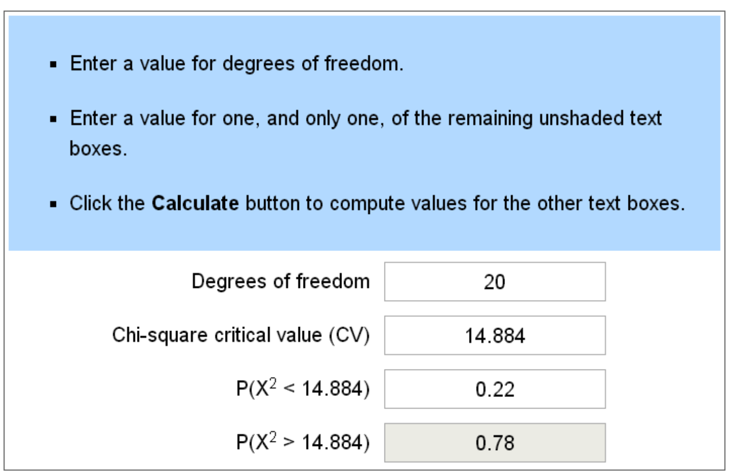
- One bin for each number for numbers between 0 and 20 isn't too bad

X	p	np	N	Chi-Squared
0	0.0000	0.0045	0	0.0045
1	0.0005	0.0454	0	0.0454
2	0.0023	0.227	0	0.227
3	0.0076	0.7567	0	0.7567
4	0.0189	1.8917	2	0.0062
5	0.0378	3.7833	3	0.1622
6	0.0631	6.3055	4	0.843
7	0.0901	9.0079	7	0.4476
8	0.1126	11.2599	11	0.006
9	0.1251	12.511	15	0.4952
10	0.1251	12.511	18	2.4082
11	0.1137	11.3736	9	0.4954
12	0.0948	9.478	7	0.6479
13	0.0729	7.2908	12	3.0417
14	0.0521	5.2077	5	0.0083
15	0.0347	3.4718	1	1.7598
16	0.0217	2.1699	4	1.5435
17	0.0128	1.2764	0	1.2764
18	0.0071	0.7091	0	0.7091
19	0.0037	0.3732	2	7.0913
20+	0.0035	0.3454	0	0.3454
			Total	**14.884**

From StatTrek, a Chi-squared score of 14.88 corresponds to a probability of 22%

I am only 22% certain that this is not a Poisson distribution

(I can't tell it's different with only 100 rolls)



Next, decide how to split up the data. If you plot a Poisson distribution with Np = 10

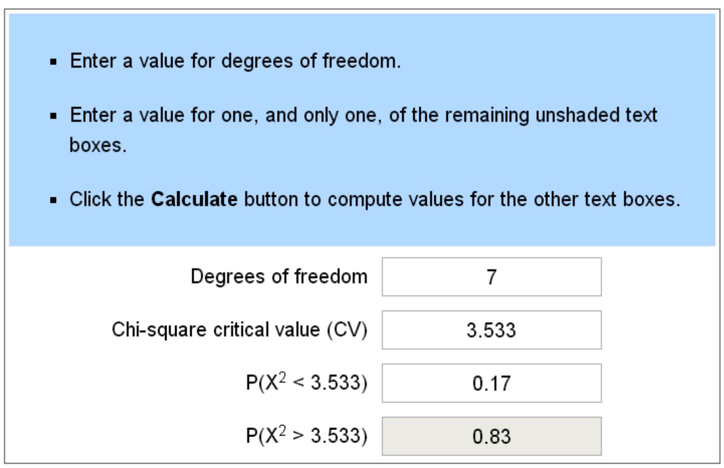
Option #2: Group the numbers in groups of 3 (reduces the number of bins)

Bin range of X	p	np	N	Chi-Squared
0 - 2	0.0028	0.28	0	0.28
3 - 5	0.0643	6.43	5	0.318
6 - 8	0.2657	26.57	22	0.786
9 - 11	0.364	36.4	42	0.8615
12 - 14	0.2198	21.98	24	0.1856
15 - 17	0.0692	6.92	5	0.5327
18 - 20	0.0127	1.27	2	0.4196
20+	0.0015	0.15	0	0.15
			Total	**3.5336**

From StatTrek, a chi-squared value of 3.5336 corresponds to a probability of 17%

I am only 17% certain that this is not a Poisson distribution

(I can't tell it's different with only 100 rolls)




```
Chi2 =    3.5790
Chi2 =   11.0776
Chi2 =    3.3048
Chi2 =    6.9674
Chi2 =   10.3505
Chi2 =    6.8304
Chi2 =    5.6311
Chi2 =    1.7137
Chi2 =    4.1792
Chi2 =    3.4463
Chi2 =    2.5340
Chi2 =   11.7599
```