Math 3080 § 1. **Kerrich Example: Chi-Squared** Name: Example
Treibergs **Test for Goodness of Fit** April 1, 2014

This **R**© program is about a test of goodness of fit for a finite discrete distribution. In this simplest case, we assume that the underlying distribution is known and we test whether the experiment indicates that the distribution is otherwise.

This story was taken from Bulmer, *Principles of Statistics,* Dover Publications, New York, 1979; originally published by Oliver and Boyd, Edinburgh, 1965. While interned in Denmark in World War II, J. E. Kerrich performed statistical experiments. He tossed five coins $n = 2000$ times and and recorded the number of heads in five. He published a book on statistical experiments after the war. Here is the data Assuming that the coin is fair, the number of heads should

Number of Heads	0	1	2	3	4	5	Total
Frequency	59	316	596	633	320	76	2000
Expected Count	62.5	312.5	625	625	312.5	62.5	2000
Relative Frequency	.030	.158	.298	.316	.160	.038	1.000
Theoretical Probability	.031	.156	.312	.312	.156	.031	1.000

theoretically follow the binomial distribution $p(x) = \binom{5}{x}/32$.

Suppose that the probability of a head actually p_i for $i = 0, \ldots, 5$. If we let $\pi(i) = p(i)$ for $i = 0, \ldots, 5$. Then we test

$$\mathcal{H}_0 : \qquad p_i = \pi(i) \text{ for all } i = 0, \ldots, 5;$$
$$\mathcal{H}_a : \qquad p_i \neq \pi(i) \text{ for some } i = 0, \ldots, 5;$$

The test statistic devised by Karl Pearson in 1900 is

$$\chi^2 = \sum_{j=0}^{5} \frac{\big(X_i - n\pi(i)\big)^2}{n\pi(i)}$$

where X_i is the observed count and $k = 6$ is the number of cells. It is asymptotically distributed as a χ^2 random variable with $k-1$ degrees of freedom. By the rule of thumb, if all of the expected cell counts exceed five, then the chi-squared approximation is appropriate. The expected cell counts here are all at least 65.5. We reject $\mathcal{H}_0$ if $\chi^2 \geq \chi^2_{\alpha,k-1}$. The statistic works out to $\chi^2 = 4.772$. The $\alpha = .05$ critical value is $\chi^2_{\alpha,k-1} = 11.070$ thus we cannot reject $\mathcal{H}_0$. The p-value is .4434. Thus the experiment did not provide significant evidence that five tosses gives any other distribution than the standard binomial.

R Session:

```
R version 2.13.1 (2011-07-08)
Copyright (C) 2011 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.41 (5874) i386-apple-darwin9.8.0]

[History restored from /Users/andrejstreibergs/.Rapp.history]

> ############### INPUT THE DATA  ############################
> x=c(59,316,596,633,320,76)
> k=length(x); k
[1] 6
> n=sum(x); n
[1] 2000
> #########  MAKE A TABLE OF OBSERVATIONS  ###################
> pbinom(0:5,5,.5)
[1] 0.03125 0.18750 0.50000 0.81250 0.96875 1.00000
> ex=n*dbinom(0:5,5,.5);ex
[1]  62.5 312.5 625.0 625.0 312.5  62.5
> M=rbind(x, ex, x/n, dbinom(0:5,5,.5))
> rownames(M)=c("Observed","Expected","Relative Freq","Theoretical Freq")
> colnames(M)=0:5; M
                        0         1        2        3         4        5
Observed         59.00000 316.00000 596.0000 633.0000 320.00000 76.00000
Expected         62.50000 312.50000 625.0000 625.0000 312.50000 62.50000
Relative Freq     0.02950   0.15800   0.2980   0.3165   0.16000  0.03800
Theoretical Freq  0.03125   0.15625   0.3125   0.3125   0.15625  0.03125
```

```
> ############  COMPUTE THE CHI-SQ "BY HAND"  #########################
> chisq=sum((x-ex)^2/ex);chisq
[1] 4.7792
> ############ ALPHA = .05  CRITICAL CHI-SQ FOR  K-1  D.F.  ############
> critchi2 = qchisq(.05,k-1,lower.tail=F); critchi2
[1] 11.0705
> pvalue = pchisq(chisq,5,lower.tail=F); pvalue
[1] 0.4434169
> ######  WE FAIL TO REJECT H0: OBSERVED DATA ARE PLAUSIBLY BINOMIAL  ##
>
> ######  RUN THE CANNED CHI-SQ TEST  #################################


>  c1 = chisq.test(x,p=dbinom(0:5,5,.5)); c1

Chi-squared test for given probabilities

data:  x
X-squared = 4.7792, df = 5, p-value = 0.4434
```