Math 3080 § 1.
Treibergs

Horse Kick Example: Chi-Squared Test for Goodness of Fit with Unknown Parameters

Name: Example
April 4, 2014

This **R**© program explores a goodness of fit test where the parameter is unknown. We ask whether the data rejects the null hypothesis that the underlying pmf is the Poisson Distribution. A previous discussion of this data was done in my program "`M3074HorseKickEg`, Horse Kick Example: Confidence Interval for Poisson Parameter." According to Bulmer, *Principles of Statistics,* Dover, 1979, Poisson developed his distribution to account for decisions of juries. But it was not noticed until von Bortkiewicz's book *Das Gesetz der kleinen Zahlen,* 1898, which applied it to the occurrences of rare events. We use one of his data sets: the number of deaths from horse kicks in the Prussian army, which is one of the classic examples. The counts give the number of deaths per army corps per year during 1875–1894.

Number of Deaths by Horse Kick per Corps per Year	0	1	2	3	4	Total
Frequency	144	91	32	11	2	280

The Poisson Random Variable describes the number of occurrances of rare events in a period of time or measure of area. For the rate constant $\lambda > 0$, the Poisson pmf is defined by the formula

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{for } x = 0, 1, 2, \ldots$$

Since $\mathrm{E}(X) = \mathrm{V}(X) = \lambda$, an estimator for λ and σ^2 is the sample mean

$$\bar{X} = S^2 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We show that the sample mean is the maximum likelihood estimator $\hat{\lambda}$. See Chapter 6 of Devore for further discussion. Assume that the distribution is given by a pmf that depends on vector of parameters θ. Suppose $X_1, \ldots, X_n$ is a random sample of observations from this distribution and

$$f(x_1, x_2, \ldots, x_n|\theta) \tag{1}$$

is the joint pmf or pdf where the parameters θ are unknown. If $x_1, \ldots, x_n$ are the observed values from this sample, (1) regarded as a function of θ is called the *likelihood function.* The *maximum likelihood estimators,* (MLE's) $\hat{\theta}$ are those values of θ that maximize the likelihood function, so that

$$f(x_1, x_2, \ldots, x_n|\hat{\theta}) \geq f(x_1, x_2, \ldots, x_n|\theta) \qquad \text{for all } \theta.$$

Let us compute the maximum likelihood estimator for λ, the rate constant of the Poisson distribution. The random variables $X_i \in \{0, 1, 2, \ldots\}$. Suppose we observe $x_1, x_2, \ldots, x_n$. Because of independence,

$$L(\lambda) = f(x_1, x_2, \ldots, x_n|\lambda) = \prod_{i=1}^{n} p(x_i, \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{(x_i)!} = ce^{-n\lambda}\lambda^{\sum x_i}$$

where we collect in the factor c all terms that don't involve λ. The maximum is found by taking logarithm and setting the derivative to zero.

$$\log L(\lambda) = \log c - n\lambda + \log(\lambda)\sum_{i=1}^{n} x_i$$

Setting the derivative to zero

$$0 = -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i$$

which yields the MLE

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

as claimed. This estimator is called the *full sample* estimator. For the horsekick data, we are told that the outcome j occured n_i times. Thus $x_i = 2$ occurred $n_2 = 32$ times. Thus the nutrials is $n = \sum n_j = 280$ and $\sum x_i = \sum j \cdot n_j = 196$ so

$$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{196}{280} = .7$$

The test whether a random sample comes from a Poisson distribution with unknown parameter using the χ^2 goodness of fit test, we use the data do obtain the MLE estimator for λ and then do the goodness of fit test with probabilities $p(i, \hat{\lambda})$. However, such a test will result in small expected cell counts and we must bin together small cells. If we expect $\lambda \approx 1$ then with a sample size of 280 trials, the expected cell counts will be $280 \cdot p(i, 1)$ or

Number of Deaths by Horse Kick per Corps per Year	0	1	2	3	4
Expected Frequency with $\lambda = 1$	103.00	103.01	51.50	17.18	4.29

Thus we decide to lump all observations $i \ge 3$ into one cell so that we have at least 5 per cell for a χ^2-test. Then the expected count of the last cell $nP(X \ge 3) = 280(1 - \sum_0^3 p(j, 1)) = 5.316684$. The lumped data is

Number of Deaths by Horse Kick per Corps per Year	$x_i = 0$	$x_i = 1$	$x_i = 2$	$x_i \ge 3$
Observed Frequency n_i	144	91	32	13

But the full sample estimator is no longer the MLE for a binned sample. We must reevaluate the MLE. If we observe $x_i = j$ for n_j times, then given λ the probabilities that x_i falls into the ith cell become

$$\pi(x, \lambda) = p(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad \text{for } x \in \{0, 1, 2\}$$
$$\pi(x, \lambda) = 1 - \sum_{y=0}^{2} p(y, \lambda) = \left(1 - \sum_{y=0}^{2} \frac{e^{-\lambda}\lambda^y}{y!}\right), \qquad \text{if } x = 3$$

Thus the likelihood function for binned data is

$$L(\lambda) = e^{-\lambda n_0}\left(e^{-\lambda}\lambda\right)^{n_1}\left(\frac{e^{-\lambda}\lambda^2}{2}\right)^{n_2}\left(1 - e^{-\lambda} - e^{-\lambda}\lambda - \frac{e^{-\lambda}\lambda^2}{2}\right)^{n_3}$$

In the case our observed values,

$$L(\lambda) = ce^{-(144+91+32)\lambda}\lambda^{91+2\cdot 32}\left(1 - e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}\right]\right)^{13}$$
$$= ce^{-267\lambda}\lambda^{155}\left(1 - e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}\right]\right)^{13}$$

whose logarithm is

$$\log L(\lambda) = \log c - 267\lambda + 155\log\lambda + 13\log\left(1 - e^{-\lambda}\left[1+\lambda+\frac{\lambda^2}{2}\right]\right)$$

A way to maximize this function is to plot it and zoom in on region near its maximum. We tried the interval .6 to .8, then .65 to .75 and so on. The graphics tells us that the binned MLE is close to $\hat{\lambda} = .7019123$.

Now we run our chi-squared test of proportion. The observed and expected frequencies are

Number of Deaths by Horse Kick per Corps per Year	$x_i = 0$	$x_i = 1$	$x_i = 2$	$x_i \geq 3$
Observed Frequency n_i	144.00	91.00	32.00	13.00
Observed Proportion n_i/n	0.514	0.325	0.114	0.046
Theoretical Proportion $\pi(i, \hat{\lambda})$	0.496	0.348	0.122	0.034
Theoretical Frequency $n\pi(i, \hat{\lambda})$	138.78	97.41	34.19	9.62

Let p_i be the probability of the ith category. The null and alternative hypotheses are

$$\mathcal{H}_0 : \text{ for some } \lambda \text{ and for all } i,\ p_i = \pi(i,\lambda);$$
$$\mathcal{H}_a : \text{The hypothesis } \mathcal{H}_0 \text{ is not true}$$

The statistic is

$$\chi^2 = \sum_{j=0}^{3} \frac{\left(n_i - n\pi(i,\hat{\lambda})\right)^2}{n\pi(i,\hat{\lambda})}$$

The null hypothesis is rejected if $\chi^2 \geq \chi^2_{\alpha,k-m-1}$, where α is the level of significance and m is the number of parameters estimated. For this data and $\alpha = .05$ with $k = 4$ the number of bins, $m = 1$ the number of estimated parameters, we have $k - m - 1 = 2$ and $\chi^2_{.05,2} = 5.992$ from Table A7. Since our $\chi^2 = 1.942$ we fail to reject the null hypothesis. There is no significant indication that the horse kick data does not come from a Poisson distribution.

Devore talks about how to perform the test if the wrong estimator is used. The binned MLE was difficult to find in this Poisson case, and may be impossible to find for more complicated problems or other distributions. Thus a less powerful test can be obtained by estimating the parameter using the full sample estimator instead. If we compute the χ^2 statistic then the rejection region is reduced: The critical value c_α falls between

$$\chi^2_{\alpha,k-1-m} \leq c_\alpha \leq \chi^2_{\alpha,k-1}.$$

Thus using the full sample estimate, we reject $\mathcal{H}_0$ is $\chi^2 \geq \chi^2_{\alpha,k-1}$, do not reject $\mathcal{H}_0$ if $\chi^2 \leq \chi^2_{\alpha,k-1-m}$, and withhold judgement if $\chi^2_{\alpha,k-1-m} < \chi^2 < \chi^2_{\alpha,k-1}$.

If we follow the scientific method, we should set the bins and sample size BEFORE we collect data. Our choice of n in the experiment will depend on our preliminary guess, say $\lambda = 1$, and the consideration that the χ^2-test requires expected cell sizes to exceed five. Once the bins are set, the $\pi(i, \lambda)$'s may be obtained. Once the data is collected, the MLE may be computed. Or the test may be run using the approximating Full MLE instead with the corresponding decrease in the rejection region for the test.

Number of Deaths by Horse Kick per Corps per Yearr	$x_i = 0$	$x_i = 1$	$x_i = 2$	$x_i \geq 3$
Full Theoretical Proportion	0.50	0.35	0.12	0.03
Full Theoretical Frequency	139.04	97.33	34.07	9.56

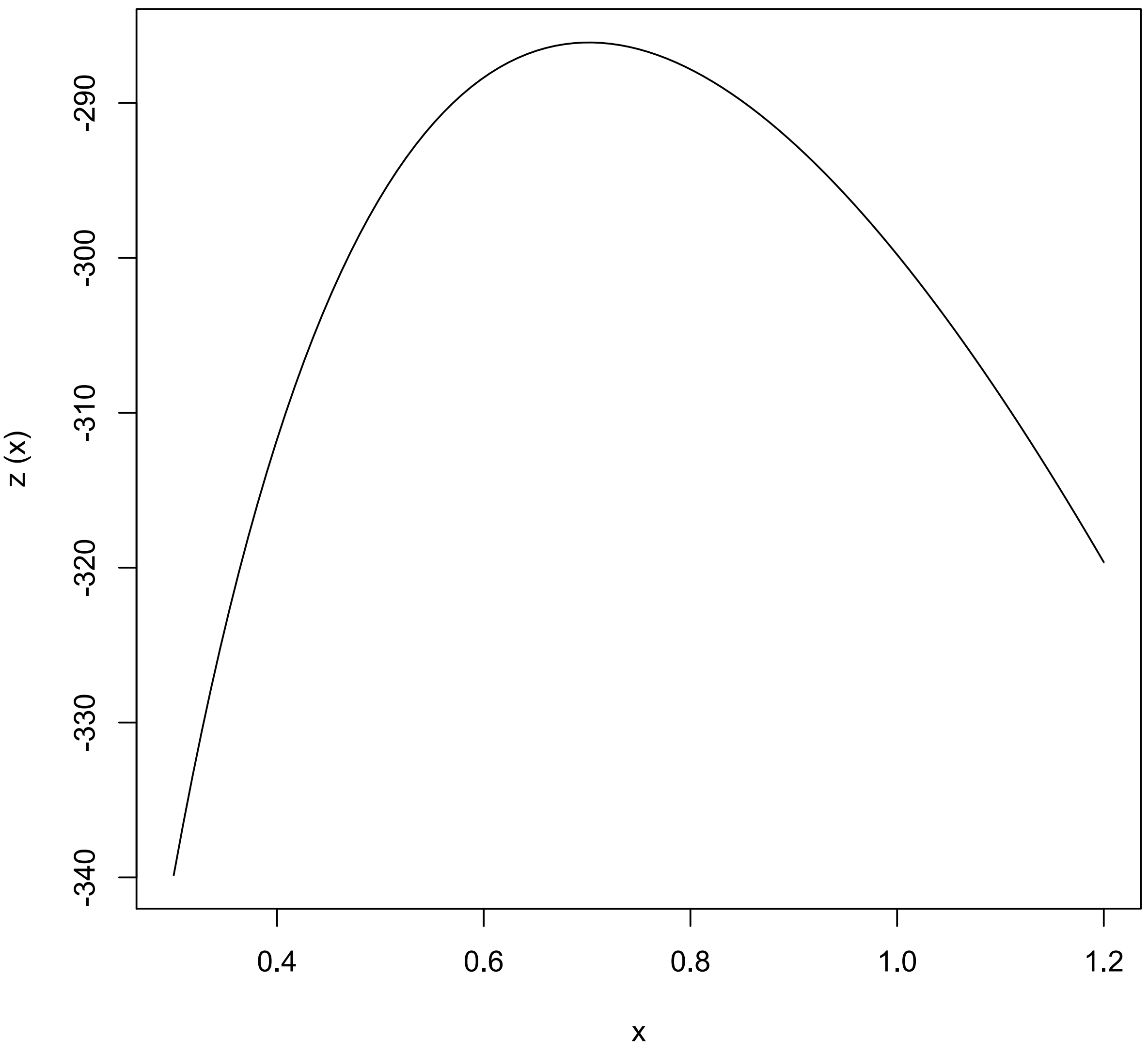
Because $\chi^2_{.05,2} = 7.815$ and the full statistic works out to be $\chi^2 = 1.951823$, which is less that $\chi^2_{\alpha,k-1-m}$, we do not reject $\mathcal{H}_0$.

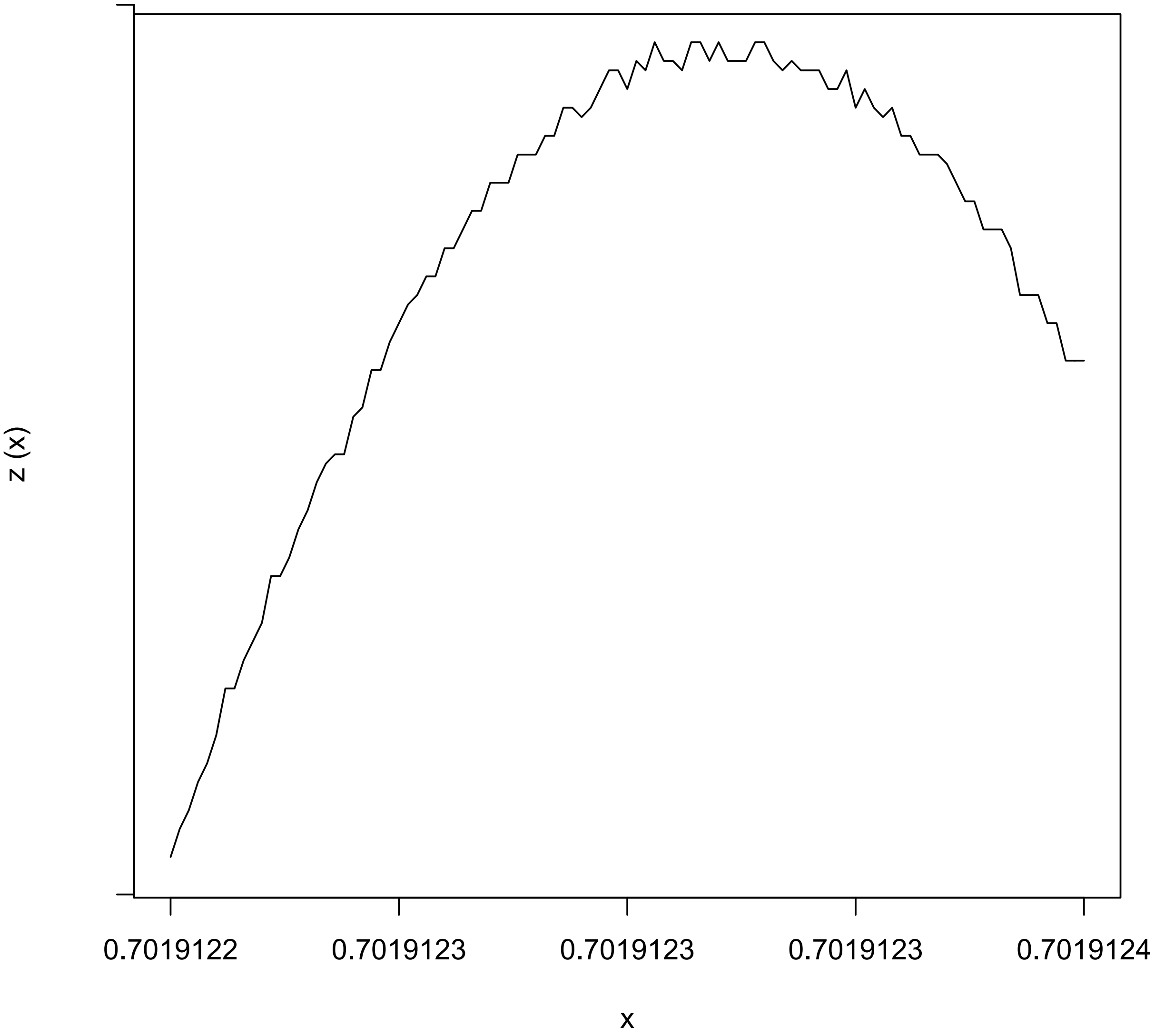
R Session:

```
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]

> #################  READ IN THE HORSEKICK FREQUENCIES  ######
> kick=scan()
1: 144 91 32 11 2
6:
Read 5 items
> n=sum(kick); n
[1] 280
> sxi=sum(0:4*kick);sxi
[1] 196
```

```
> ######################  FULL MLE FOR LAMBDA  ###############
> sxi / n
[1] 0.7

> ###########  EXPECTED CELL SIZES WITH LAMBDA = 1  ##########
> 280*dpois(0:4,1)
[1] 103.006244 103.006244  51.503122  17.167707   4.291927
> ###  EXPECTED LUMPED FOURTH CELL SIZE WITH LAMBDA = 1    ###
> 280* ppois(3,1,lower.tail=F)
[1] 5.316684

> ##############  LOG OF LIKELIHOOD FUNCTION  ################
> z=function(t){-267*t +155*log(t)+13*log(1-exp(-t)*(1+t+t^2/2))}

> ##############  DO GRAPHIC SEARCH FOR MAXIMUM  #############
> plot(z,.3,1.2)
> plot(z,.6,.8)
> plot(z,.65,.75)
> plot(z,.68,.72)
> plot(z,.7,.71)
> plot(z,.701,.703)
> plot(z,.7015,.7017)
> plot(z,.7017,.7019)
> plot(z,.7019,.7020)
> plot(z,.7019,.70192)
> plot(z,.70191,.701915)
> plot(z,.701911,.701912)
> plot(z,.701912,.701913)
> plot(z,.7019122,.7019124)
Warning message:
In plot.window(...) :
  relative range of values =  47 * EPS, is small (axis 2)
> lh=.7019123


> ##  THEORETICAL PROBABILITIES WITH MLE EST. FOR LAMBDA  ##
> tf=c(dpois(0:2,lh),ppois(2,lh,lower.tail=F))
> tf
[1] 0.49563659 0.34789342 0.12209534 0.03437465
> sum(tf)
[1] 1
> ##############  LUMPED FREQUENCIES  ######################
> fr=c(kick[1:3],kick[4]+kick[5]);
> fr
[1] 144  91  32  13
> n=sum(fr)
> n
[1] 280
```

```
> ############ CANNED TEST HAS WRONG DF  ###################

> chisq.test(fr,p=tf)

Chi-squared test for given probabilities

data:  fr
X-squared = 1.9417, df = 3, p-value = 0.5846

> #############  CHI SQ TEST "BY HAND"  ####################
>
> chsq = sum((fr-n*tf)^2/(n*tf));chsq
[1] 1.941692
> qchisq(.05,2,lower.tail=F)
[1] 5.991465
> pchisq(chsq,2,lower.tail=F)
[1] 0.3787624
>
>
> ############# CANNED FULL MLE LAMBDA CHI SQ TEST  ########
> fula = .7
> fultf=c(dpois(0:2,fula),ppois(2,fula,lower.tail=F)); fultf
[1] 0.49658530 0.34760971 0.12166340 0.03414158
> chisq.test(fr,p=fultf)

Chi-squared test for given probabilities

data:  fr
X-squared = 1.9518, df = 3, p-value = 0.5825


> ###########  FULL CHI SQ TEST "BY HAND"  #################
> chsq = sum((fr-n*fultf)^2/(n*fultf));chsq
[1] 1.951823
> qchisq(.05,3,lower.tail=F)
[1] 7.814728

> ###########  GENERATE MATRIX OF FREQ AND PROB  ###########
> M=matrix(c(fr,fr/n,tf,n*tf,fultf,n*fultf),ncol=4,byrow=T)
> colnames(M)=c("=0","=1","=2","3")
> rownames(M)=c("Obs. Freq.","Obs. Prop.","Th. Prop.",
  "Th. Freq.","Full Th. Prop.","Full Th. Freq.")
> M
                        =0         =1         =2           3
Obs. Freq.     144.0000000 91.0000000 32.0000000 13.00000000
Obs. Prop.       0.5142857  0.3250000  0.1142857  0.04642857
Th. Prop.        0.4956366  0.3478934  0.1220953  0.03437465
Th. Freq.      138.7782455 97.4101575 34.1866938  9.62490314
Full Th. Prop.   0.4965853  0.3476097  0.1216634  0.03414158
Full Th. Freq. 139.0438851 97.3307195 34.0657518  9.55964356
>
```


-290
-300
-310
-320
-330
-340
z (x)
0.4
0.6
0.8
1.0
1.2
x


z (x)
0.7019122
0.7019123
0.7019123
0.7019123
0.7019124
x