**Homework#3b(3b) CSC148, Prof Mitchell 10/26/19**

**dueDates classTime: for extra credit, Oct 31 OR normal dueDate, Nov 5**

**#1. Introduction**

This assignment is a follow on to h3. Data sample d’s arrival process was represented by h3’s first required histogram (refer to it here as H). H is based on selecting a small (preferably between 2.5 to 5 minutes) fixed time interval length “sis” (as in h3).

*Generally, a busier system should use a smaller sis size*, and a smaller sis results in a *hypothesized Poisson distribution with a smaller rather than larger . A larger  typically has a flatter shape, and does not model “rare” events as well as smaller .*

Total data collection time “T” is divided into T/sis intervals “I”, each interval of equa length sis.

The x-axis for H is the small non-negative integers Xi = 0, 1, 2, … , m. m is the largest arrivals count occurring in any sis interval. Recall that (H’s bar height for xi) =

(the number of sis intervals having i arrivals).

The fitting process to be used requires evaluation of the Poisson pmf for each of the (m+1) points xi. Therefore, be sure that your histogram has a corresponding frequency bar for xi=0 (and xi=1, etc. if your histogram happens to be skewed to the right of the y-axis, such that x0=x1=x2= …=xp=0, for some small integer p). Omitting x0=x1=x2= …=xp=0 might skew the  value, reducing the likelihood of fit success.

**#2. The distribution fitting calculations**

For distribution fitting, H can be represented as a discrete function fH:

domain values are the xi of section #1, and fH(xi) is the number of sis intervals in which there were xi arrivals.

Then, in the Chi-square notation used in course module P4 (as used in most textbooks),

Oi is fH(i) and Ei=n\*pi, where n is the number of subIntervals (NOT the number of vehicles observed!).

Specifying d’s mean x^ as  in the Poisson pmf determines the one pmf from the Poisson family for which  is calculated.

In h3b, you can keep the sis subinterval size value or change it to another value for this assignment (Note: changing the size of sis will require re-calculating the fH(xi) values and x^. This will determine a different Poisson pmf).

After calculating the test statistic , formulate the null hypothesis H0:

“d fits a Poisson distribution whose parameter is the mean x^ of d”,

where .05 is the confidence level *(explained next class)* of the H0 decision.

*It is possible that d does NOT fit a Poisson distribution. Class discussion has identified ways that d might slightly violate Poisson process assumptions. Assignment focus is conducting a proper fitting process/steps and correct H0 interpretation. Your grade is based on properly applying the fitting calculation, and not on whether your data d fits.*

**#3. Hand Ins – ALL content must be word processed**

1. The following items, numbered, and in the order shown below:

1) sis size value and its units (for example, 4 minutes)

2) n\_sis, that is, the number of subIntervals in [0,T] (this is n factor in n\*pi = Ei)

3) avg(number of arrivals over all sis subIntervals), denoted x^ in h3

4) Your corrected histogram, ONLY if you were email-notified that your original h3 histogram was incorrectly built AND have not yet re-submitted a correctly-built one

If you did NOT get an email notification about redoing your histogram, you can skip/omit step 4).

1. Letting k = final number of combined classes, a table “T” with the same format as Table 3, module P4, Slide#17:

Include 4 column header row: < xi, Oi, Ei,(Oi-Ei)2/Ei > for this table

The header row is followed by rows with the 4 values for each xi,

and a last line has totals for 2nd (Oi), 3rd (Ei) and 4th (Oi-Ei)2/Ei columns.

*The Appendix describes Python modules that you can use for finalizing*  *classes. You can use any tool, or do it manually, to calc. combined classes. Using the Appendix modules is optional, but helpful if you have Python experience.*

1. a) 1. Test statistic value  = ∑ (Oi-Ei)2/Ei

2.  the degrees of freedom: display calculation of  = k-s-1

3. Xcritical the Chi-square critical value for your d, and its cell

(row, column) of Table 4, module P4, Slide#21

b) Short 1-2 sentence write up that states your conclusion for accepting or

rejecting null hypothesis H0.

**Appendix**

**2 utilities for Chi-square fitting for a hypothesized Poisson distribution**

**# Calculating p(i) and Ei values**

If you have Python3 experience, and access to a Python3 system, you can copy and use source file posted on athena at /gaia/home/faculty/mitchell/simpy3demos/ chisquareCalcs\_f19.py

# A**utomating Combining Classes for Chi-square fitting**

The source code posted on athena at /gaia/home/faculty/mitchell/simpy3demos/**poissonPMFfit.py**

consists of two small loops, each with hard-coded initializations that

1. Iterates through values xi and calculate the corresponding pmf values pi (not Ei)
2. Transforms a given list of Chi-square Ei values by combining classes according to a

minimal class size.

Item b) above can easily be adapted to other situations, as sketched below.[[1]](#endnote-1)

**Review the class combining criterion**

Given the initial classes E1, … , Ek for the Chi-square fitting algorithm, the module’s last loop computes the values Ej’ resulting from combining adjacent classes so that each Ej’ is larger than a recommended threshold (most textbooks cite a value such as “4.0 or 5.0”) for Chi-square fitting.

1. Given any list L1 of numbers n1, n2, …, nm, a More General Problem is to compute list L2 of numbers c1, c2, …, ck

   such that: for a given, fixed boundary value v, c1 = the smallest partial sum (n1+n2 + …+ nj1) >= v, and

   c2 = smallest partial sum (nj1+1+ … +nj2) >=v, etc. The kth (i.e., last) partial sum might be a small residual < v. In such a case, simply combine ck, that is, add ck to ck-1.

   Example: Given L1 = [2,3,6,4,2,7,4,5,2] and v=6, then L2 = [11,6,7,11]. The right-most value 11 has absorbed (added in) residual value 2, thus “combining” the last two classes (and avoiding the result [11,6,7,9,2] having a small class 2). [↑](#endnote-ref-1)