SOLUTION FOR HOMEWORK 10, STAT 4372

Welcome to your 10th homework. Here you have an opportunity to solve classical model selection problems based on hypotheses testing. These are absolutely classical statistical issues. Further, actuarial exams typically contain several questions on the topic.

1. Problem 16.5 Solution: In K-S test you use the cdf (not pdf), so you need to calculate it. Using the Table find that it is an inverse exponential distribution and then $F(x) = e^{-2/x}$. Then notice that the empirical cdf $\hat{F}_n(x)$ has jumps equal $1/n = 1/5 = .2$ at each observation. Draw a graphic (approximate) to see that the maximum difference is always at the points of observations where the empirical cdf has a jump. As a result, for each observation X_k you need to check the maximum of $|F(X_k) - \hat{F}_n(X_k - 0)|, |F(X_k) - \hat{F}_n(X_k)|$. Results are in the Table below.

x	$F(x)$	$\hat{F}_n(x-0)$	$\hat{F}_n(x)$	$MaxDifference$
1	.135	0	.2	.135
2	.368	.2	.4	.168
3	.513	.4	.6	.113
5	.670	.6	.8	.130
13	.857	.8	1	.143

The K-S statistic is .168 (the maximum of the right column). Note that you also say where the maximum occurs (here at point $x = .2 - 0$).

2. Problem 16.6 Solution: The problem is similar to the previous one so it is a good training for you. The cdf is

$$F(x) = \int_0^x 2(1+y)^{-3}dy = -(1+y)^{-2}|_{y=0}^{y=x} = 1 - (1+x)^{-2}.$$

Then the table contains the calculations.

x	$F(x)$	$\hat{F}_n(x-0)$	$\hat{F}_n(x)$	$MaxDifference$
.1	.174	0	.2	.174
.2	.306	.2	.4	.106
.5	.556	.4	.6	.156
1.0	.750	.6	.8	.150
1.3	.811	.8	1	.189

The K-S statistic is .189.

3. Problem 16.9 Solution: For the chi-square test we calculate 3 degrees of freedom: four groups minus zero estimated parameters (the underlying distribution is given explicitly) minus 1 (the latter is the rule because the total number of observations is fixed so frequencies in each cell are dependent). Then we calculate the chi-squared statistic using table. Note that $F(x) = 1 - S(X)$.

Interval	Observed	Expected	Chi-squared addend
[0,2]	21	$150F(1) = 150(2/20) = 15$	$6^2/15$
[1,2]	27	$150[F(2) - F(1)] = 150(4/20) = 30$	$3^2/30 = .3$
[2,3]	39	$150[F(3) - F(2)] = 150(6/20) = 45$	$6^2/45 = .8$
[3,4]	63	$150[F(4) - F(3)] = 150(8/20) = 30$	$3^2/60 = .15$
Total	150	150	3.65

From the chi-squared table we see that at .05 level of significance with 3 degrees of freedom the critical value is 7.81. The test-statistic is 3.65 and it is smaller, so the null hypothesis is accepted.

4. Problem 16.10 Solution: This problem is similar to the previous one only here you are estimating the parameter of the distribution under the null hypothesis (Poisson) so do not forget to subtract extra 1 when calculate the degrees of freedom for chi-square statistic.

Remember (or calculate) that for Poisson distribution (which belongs to exponential family of distributions), the MLE is the average sum of the number of claims (which is also the method of moments estimator) and thus

$$\hat{\lambda} = [(0)(50) + (1)(122) + (2)(101) + (3)(92)]/365 = 600/365 = 1.64$$

Now the table. Note that you combine cells with less than 5 observations.

Number of Claims	Observed	Expected	Chi-squared addend
0	50	$365e^{-1.64} = 70.53$	$(20.53)^2/70.53 = 5.98$
1	122	$365(1.64)e^{-1.64} = 115.94$	$(6.06)^2/115.94 = .32$
2	101	$365(1.64)^2e^{-1.64}/2 = 95.29$	$(5.71)^2/95.29 = .34$
≥ 3	92	$365 - 70.53 - 115.94 - 95.29 = 83.24$	$(8.76)^2/83.24 = .92$
Total		365	7.56

There are 2 degrees of freedom (4 cells minus 1 minus 1 for calculating the parameter). At .025 level of significance, the critical value is (from chi-squared table) is 7.38. Because $7.56 > 7.38$ the null hypothesis is rejected - the Poisson model is not a good fit for the data.

5. Problem 16.11 Solution: Note that the distribution of the number of accidents is per day, but the counts are per year with 365 days. Keeping this in mind, the expected count $E(N)$ for k accidents is (I use the Poisson pmf)

$$365\Pr(N = k) = \frac{365e^{-.6}(.6)^k}{k!}.$$

Now the table.

Number of Accidents	Observed	Expected	Chi-squared addend
0	209	200.32	.38
1	111	120.19	.70
2	33	36.06	.26
≥ 3	12	8.43	1.51
Total	365	365	2.85

There are 3 degrees of freedom (4 groups minus 1 minus zero number of estimated parameters), and this yields the critical value 7.81. Thus the null hypothesis as accepted.

6. Problem 16.12 Solution: We first calculate the test-statistics, and note that the expected number of observations in each cell is $1000(1/20) = 50$. Also, the number of degrees of freedom is $20 - 1 = 19$. Write

$$\hat{\chi}^2_{19} = \sum_{j=1}^{20} \frac{O_j - 50)^2}{50}$$

$$= .02[\sum_{j=1}^{20} O_j^2 - 100 \sum_{j=1}^{20} O_j + (20)(50)^2]$$

$$= .02[51,850 - (100)(1,000) + 50,000] = 37.$$

The probability $\Pr(\chi^2_{19} \geq 37) = .0079$. This is the observed level of significance, also called the p-value.

7. Problem 16.13 Solution: Using the Table I find that

$$f(x) = \alpha\theta^\alpha/(x+\theta)^{\alpha+1},$$

and the likelihood function is

$$L(\alpha, \theta) = \frac{\alpha^{20}\theta^{20\alpha}}{\prod_{j=1}^{20}(x_j + \theta)^{\alpha+1}}.$$

Now remember our trick — calculate and maximize the log-likelihood,

$$l(\alpha, \theta) = 20\ln(\alpha) + 20\alpha\ln(\theta) - (\alpha+1)\sum_{j=1}^{20}\ln(x_j + \theta).$$

Now we can use the given statistics and calculate the likelihoods under the two hypotheses. Under the null hypothesis $L_0 = L(2, 3.1) = -58.78$. Under the alternative hypothesis you need to use the maximum likelihood estimate for θ and it is the given value 7. This yields that $L_1 = L(2, 7) = -55.33$. Then the test statistic is

$$\hat{\chi}^2_1 = 2(L_1 - L_0) = 6.90.$$

Note that we have only one degree of freedom because the null hypothesis is fully specified and the alternative has only one free parameter θ which was estimated via MLE. Then using table we get the answer

$$\mathrm{p-value} = \Pr(\chi^2_1 \geq 6.90) = .0086.$$

8 Problem 16.22. Solution: Note that the number of accidents is a discrete random variable (number of accidents) . As a result, only 3 candidates for the model are binomial, Poisson and negative binomial. Then you should remember the discussion on page 109. If you look at the sequence kn_k/n_{k-1} then the numbers are 2.67, 2.33, 2.01, 1.67, 1.32 and 1.04. The sequence is decreasing indicating a binomial distribution.

An alternative approach is to calculate the sample mean equal to 2 and the variance 1.49. The variance is significantly smaller than the mean — using the Table you can see that only Binomial has this property.

9. Problem 16.24 Solution: The loglikelihood values are -385.9 for the Poisson and -382.4 for the negative binomial. The test statistic is

$$\hat{\xi}_1^2 = 2(L_1 - L_0) = 2(-382.4 + 385.9) = 7.$$

Note that there is just 1 degree of freedom for the chi-square test because L_1 has two free parameters and L_0 only one, so the difference is 1. Then from chi-square table

$$\Pr(\chi_1^2 > 3.84) = .05.$$

Because $7 > 3.84$ the null hypothesis (Poisson distribution) is rejected on .05 level of significance in favor of the negative binomial.

10. Problem 16.25 Solution: Sample size is $n = 100$, and then the SBC subtracts $(r/2)\ln(n) = r(2.3)$ from the likelihood. Then for the models in Table 16.24 the penalized SBC criteria are:

Generalized Pareto: $-219.1 - 6.9 = -226$

Burr: $-219.2 - 6.9 = -226.1$

Pareto: $-221.2 - 4.6 = -225.8$

Lognormal: $-221.4 - 4.6 = -226$

Inverse Exponential: $-224.3 - 2.3 = -226.6$.

The largest value points on the Pareto distribution as the better model for the data.