STAT 479
Final
December 13, 2010

1. A health insurance company has the following sample of five claims:

   500, 500, 700, 800, 1500

   The company wants to model their claims using a Gamma distribution.

   Calculate the parameters for the Gamma distribution using the Method of Moments.

Solution

$$E(X) = \alpha\theta = \bar{X} = \frac{500+500+700+800+1500}{5} = 800$$

$$Var(X) = \alpha\theta^2 = E(X^2) - (E(X))^2 = \frac{500^2+500^2+700^2+800^2+1500^2}{5} - 800^2 = 136,000$$

$$\frac{Var(X)}{E(X)} = \frac{\alpha\theta^2}{\alpha\theta} = \theta = \frac{136,000}{800} = 170$$

$$\alpha = \frac{E(X)}{\theta} = \frac{800}{170} = 4.70588$$

2. The variance of a distribution will be estimated using:

$$\hat{\sigma}^2 = \frac{\sum (X - \bar{X})^2}{n}$$

The following sample is selected from the distribution:

5, 5, 8

Using the bootstrap method, estimate the mean square error in the above estimator.

Solution

$$Var\ of\ \ Empirical\ Distribution = E(X^2) - (E(X))^2 = \frac{5^2 + 5^2 + 8^2}{3} - \left(\frac{5+5+8}{3}\right)^2 = 2$$

Resample

Possibilities	*Probabilities*	*Estimator*
5,5,5	$(2/3)^3 = 8/27$	$\frac{(5-5)^2 + (5-5)^2 + (5-5)^2}{3} = 0$
5,5,8	$(3)(2/3)^2(1/3) = 12/27$	$\frac{(5-6)^2 + (5-6)^2 + (8-6)^2}{3} = 2$
5,8,8	$(3)(2/3)(1/3)^2 = 6/27$	$\frac{(5-7)^2 + (5-7)^2 + (8-7)^2}{3} = 2$
8,8,8	$(3)(1/3)^3 = 1/27$	$\frac{(8-8)^2 + (8-8)^2 + (8-8)^2}{3} = 0$

$$MSE = E[(\sigma^2 - \hat{\sigma}^2)^2] =$$

$$\frac{8}{27}(2-0)^2 + \frac{12}{27}(2-2)^2 + \frac{6}{27}(2-2)^2 + \frac{1}{27}(2-0)^2 = \frac{36}{27} = \frac{4}{3}$$

3. During the last calendar year, Purdue incurred 100 worker's compensation claims which are summarized in the following table:

Amount of Claim	**Number of Claims**
0 – 5000	14
5000-15,000	16
15,000-45,000	40
More than 45,000	30

Sally wants to test the following hypothesis using the Chi Square Test at a 10% significance level:

H_0: Workers Compensation claims are uniformly distributed with θ = 60,000.
H_1: Workers Compensation claims are not uniformly distributed with θ = 60,000.

Calculate the Chi Square test statistic.

Solution

Range	O_j	$\hat{p}_j$	E_j	$\frac{(O_j - E_j)^2}{E_j}$
0-5000	14	5/60	8.3333	3.85334
5000-15000	16	10/60	16.6667	0.02667
15000-45000	40	30/60	50	2
45000+	30	15/60	25	1
			TOTAL= χ^2 ➔	6.88

Determine the critical value for this test.

Solution

Degrees of Freedom =k-1-Estimated Parameters = 4-1-0=3

The critical value is 6.251

State your conclusion with regard to the hypothesis.

Solution

Reject the Hypothesis

4. Using simulation, Farhana estimates the claim **payments** that will be made under a dental insurance policy. The policy has a $50 deductible for each claim. The maximum that will be paid under the dental policy for all claims in a year is $500.

   You are given that the number of claims is distributed as a Poisson distribution with $\lambda = 2$.

   The amount of each claim is distributed as an exponential distribution with $\theta = 200$.

   Using simulation, Farhana wants to estimate the total claims that will need to be paid under the dental policy. She does so by estimating the claims for each insured. Erin and Ure are the first two insureds. First, Farhana determines the number of claims for Erin and then the amount of each claim for Erin. Next, Farhana determines the number of claims for Ure. Finally, Farhana simulates the amount of each of Ure's claims.

   The random numbers used in the simulation are:

   0.20 0.50 0.85 0.50 0.30 0.90 0.45 0.60 0.95 0.75 0.05

   Calculate the amount paid for Erin and the amount paid for Ure.

Number of Claims

n	$\Pr(N=n)$	$F_N(n)$
0	$e^{-2}=0.1353$	*0.1353*
1	$2e^{-2}=0.2707$	0.4060
2	$2e^{-2}=0.2707$	0.6767
3	$(4/3)e^{-2}=0.1804$	*0.8571*
4	$(4/3)e^{-2}=0.1804$	*0.8571*

Amount of Claim

$$F(x)=1-e^{\frac{-x}{\theta}} \Rightarrow u^{**}=1-e^{\frac{-x^{**}}{\theta}} \Rightarrow e^{\frac{-x^{**}}{\theta}}=1-u^{**} \Rightarrow x^{**}=-200\ln(1-u^{**})$$

Erin

Random Number 0.20 => *N=1*

Random Number 0.50 => $X_1=-200\ln(0.5)=138.63$ *Amount Paid* $=138.63-50.00=88.63$

Ure

Random Number 0.85 => *N=3*

Random Number 0.50 => $X_1=-200\ln(0.5)=138.63$ *Amount Paid* $=138.63-50.00=88.63$

Random Number 0.30 => $X_1=-200\ln(0.7)=71.33$ *Amount Paid* $=71.33-50.00=21.33$

Random Number 0.90 => $X_1=-200\ln(0.1)=460.52$ *Amount Paid* $=460.52-50.00=410.52$

$Total=88.63+21.33+410.52=520.48$

$Max=500 \Rightarrow Amount\ Paid=500$

5. Yu Fire Company has experienced the following four claims this year:

5000 25,000 50,000 60,000

The company's actuary, Kicho, wants to test the following hypothesis using the Kolmogorov-Smirnov Test with a 10% significance level:

H_0: Claims are distributed as an exponential distribution with θ = 40,000.
H_1: Claims are not distributed as an exponential distribution with θ = 40,000.

Calculate the Kolmogorov-Smirnov test statistic.

Solution

x	$F_4(x^-)$	$F_4(x)$	$F^*(x) = 1 - e^{\frac{-x}{40,000}}$	Absolute Value of Difference
5,000	0	0.25	0.1175	0.1325
25,000	0.25	0.50	0.4647	0.2147
50,000	0.50	0.75	0.7135	0.2135
60,000	0.75	1.00	0.7769	0.2231

Maximum Absolute Value of the Difference = 0.2231

Calculate the critical value.

Solution

$$\frac{1.22}{\sqrt{n}} = \frac{1.22}{\sqrt{4}} = 0.61$$

State your conclusion with regard to Kicho's hypothesis.

Solution

Do Not Reject

6. The following ten dental claim payments were made by Hennessey Dental HMO:

   50, 50, 60, 100, 100, 120, 150, 150, 150, 150

   Claims under Hennessey HMO are subject to an upper limit of 150.

   Hennessey models claims using an Exponential distribution.

   Determine the Maximum Likelihood Estimator for θ.

Solution

$$MLE = \frac{Total\ \ of\ All\ Payments}{Number\ of\ Uncensored\ Payments} = \frac{1080}{6} = 180$$

7. Dunham Dental Insurance Company has the following data from 1000 insureds:

Number of Claims in One Year	Number of Insureds
1	200
2	700
3	80
4	15
5	5

Dunham believes that the number of claims in one year is distributed as a binomial distribution.

Use the method of moments to estimate the parameters for a binomial distribution.

Solution

$$E(X) = \frac{200 + 1400 + 240 + 60 + 25}{1000} = 1.925$$

$$E(X^2) = \frac{200(1) + 700(4) + 80(9) + 15(16) + 5(25)}{1000} = 4.085$$

$$Var(X) = E(X^2) - (E(X))^2 = 4.085 - (1.925)^2 = 0.379375$$

$$E(X) = mq \qquad Var(X) = mq(1-q)$$

$$\frac{Var(X)}{E(X)} = \frac{mq(1-q)}{mq} = 1 - q = \frac{0.379375}{1.925} = 0.1970779 ==> q = 0.802922$$

$$m = \frac{E(X)}{q} = \frac{1.925}{0.802922} = 2.397. \text{ But } m \text{ must be an integer so } m=2$$

$$q = \frac{E(X)}{m} = \frac{1.925}{2} = 0.9625$$

8.

a. State the Anderson-Darling critical value at a 5% significance level.

   2.492

b. State the principle of parsimony.

   Unless there is overwhelming evidence to the contrary, a simple model is better.

c. Which score based approach to determining a model is consistent with the principle of parsimony?

   Using the highest p-value of the Chi Square Goodness of Fit test.

9. An automobile insurance company has the following information regarding three auto claims:

   a. One claim is equal to 5000.
   b. Two claims exceed 5000.

   The claims are assumed to follow a Pareto distribution which α = 3.

   Calculate the Maximum Likelihood Estimator for θ.

Solution

$$Pareto: f(x) = \frac{\alpha\theta^{\alpha}}{(x+\theta)^{\alpha+1}} = \frac{3\theta^3}{(x+\theta)^4} \quad and \quad F(x) = 1-\left(\frac{\theta}{x+\theta}\right)^{\alpha} = 1-\left(\frac{\theta}{x+\theta}\right)^{3}$$

$$L(\theta) = f(5000)[1-F(5000)]^2 = \left(\frac{3\theta^3}{(5000+\theta)^4}\right)\left[\left(\frac{\theta}{5000+\theta}\right)^3\right]^2 = \frac{3\theta^3}{(5000+\theta)^4}\cdot\frac{\theta^6}{(5000+\theta)^6} = \frac{3\theta^9}{(5000+\theta)^{10}}$$

$$l(\theta) = \ln(3) + 9\ln(\theta) - 10\ln(5000+\theta)$$

$$l^{/}(x) = \frac{9}{\theta} - \frac{10}{\theta+5000} = \frac{9(5000+\theta)-10(\theta)}{\theta(5000+\theta)} = \frac{45,000-\theta}{\theta(5000+\theta)} = 0$$

$$=> 45,000-\theta = 0 ==> \hat{\theta} = 45,000$$

10. Note: Part a. is worth one question and Part b. and c. are each worth ½ question.

You are given the following sample of five claim amounts:

10 20 30 40 50

a. If the claims are assumed to be distributed as normal, determine the maximum likelihood estimator of μ and σ.

$$\hat{\mu} = \bar{X} = \frac{10+20+30+40+50}{5} = 30$$

$$\hat{\sigma}^2 = \frac{\sum (X_i - \hat{\mu})^2}{n} = \frac{(10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2}{5} = 200$$

$$\hat{\sigma} = \sqrt{200} = 14.142$$

b. If the claims are assumed to be uniform, calculate the maximum likelihood estimator of θ.

$$\hat{\theta} = Max[X_1, X_2, X_3, X_4, X_5] = 50$$

c. If claims are assumed to be distributed gamma with α = 2, calculate the maximum likelihood estimator of θ.

$$\hat{\theta} = \frac{\bar{X}}{\alpha} = \frac{30}{2} = 15$$

11. You are given the following four disability claims which are believed to come from a population which is distributed following a Weibull distribution with $\tau = 2$:

500, 1200, 2400, 5000

Estimate θ using the Method of Percentile Matching and the 25th percentile.

$$j = \lfloor (n+1)(0.25) \rfloor = \lfloor 5(0.25) \rfloor = \lfloor 1.25 \rfloor = 1$$

$$n = 1.25 - 1 = 0.25$$

$$\hat{\pi}_{0.25} = 0.75X_1 + 0.25X_2 = (0.75)(500) + (0.25)(1200) = 675$$

$$F(X) = 1 - e^{-(x/\theta)^{\tau}}$$

$$F(675) = 0.25 = 1 - e^{-(675/\theta)^2}$$

$$e^{-(675/\theta)^2} = 0.75 \Rightarrow -(675/\theta)^2 = \ln(0.75) \Rightarrow 675/\theta = \sqrt{-\ln(0.75)}$$

$$\hat{\theta} = \frac{675}{\sqrt{-\ln(0.75)}} = 1258.48$$

12. Using the inversion method of simulation with pseudo random numbers selected on a uniform distribution from 0 to 1, you are to simulate claims for the following three distributions:

    i. **Single Parameter** Pareto distribution with α = 3 and θ = 1000

    ii. $$F(x) = \begin{cases} 0.002x & 0 \le x < 300 \\ 0.002x + 0.1 & 300 \le x \le 450 \end{cases}$$

    iii. $$F(x) = \begin{cases} 0.001x & 0 \le x < 650 \\ 0.650 & 650 \le x \le 850 \\ 0.002x - 1.05 & 850 < x \le 1025 \end{cases}$$

Complete the following table (Show your work!)

u **	**x** for Distribution i**	**x** for Distribution ii**	**x** for Distribution iii**
0.30	1126.25	$0.3=0.002x^{**}$ $x^{**}=150$	$0.3=0.001x^{**}$ $x^{**}=300$
0.65	1418.98	F(x) jumps at 300 from 0.60 to 0.70 so 0.65=> $x^{**}=300$	0.65=0.65 Use Max=850
0.90	2154.43	$0.9=0.002x^{**}+0.1$ $x^{**}=400$	$0.9=0.002x^{**}-1.05$ $x^{**}=975$

$$F(x) = 1 - \left(\frac{\theta}{x}\right)^{\alpha} \Rightarrow u^{**} = 1 - \left(\frac{1000}{x^{**}}\right)^{3} \Rightarrow \left(\frac{1000}{x^{**}}\right) = (1-u^{**})^{1/3} \Rightarrow x^{**} = \frac{1000}{(1-u^{**})^{1/3}}$$

$$0.30 \Rightarrow x^{**} = \frac{1000}{(1-0.3)^{1/3}} = 1126.25$$

$$0.65 \Rightarrow x^{**} = \frac{1000}{(1-0.65)^{1/3}} = 1418.98$$

$$0.90 \Rightarrow x^{**} = \frac{1000}{(1-0.9)^{1/3}} = 2154.53$$

13. During the last 30 day month, the following data on the number of automobile accidents on the Purdue campus has been collected:

Number of Accidents	Number of Days
0	8
1	12
2	5
3	4
4	0
5	1

The Police Department believes that the number of accidents in a day is distributed as a Poisson distribution.

What is the 95% linear confidence interval for $\hat{\lambda}$?

$$\hat{\lambda} = \overline{X} = \frac{0+12+10+12+0+5}{30} = 1.3$$

$$Var(\hat{\lambda}) = \frac{\hat{\lambda}}{n} = \frac{1.3}{30} = 0.04333$$

$$95\%\ \mathit{Confidence\ Interval} = 1.3 \pm 1.96\sqrt{0.04333} \Rightarrow (0.89199, 1.70801)$$