120

STAT 479
Test 3
Spring 2016

May 3, 2016

The Bergmann Statistical Institute collects and analyzes data from various insurance companies. Wang Warranty Company has retained Bergmann to collect and analyze data related to a warranty provided Amstutz Automobile Company. Wang provides a warranty to Amstutz with no deductibles or upper limits. During 2015, Wang paid the following five warranty claims to Amstutz:

6

3000 4000 5000 5000 7000

This data will be used for questions 1-3.

Cheng, who runs Wang Warranty, wants to develop a continuous distribution of the amount of claims using the Kernel Density Model using the uniform kernel. Mayfawny who is an expert in this area and the owner of Bergmann, models the data using the Kernel Density Model with the uniform kernel and a span of 1000.

1. (6 points) Calculate the 70th percentile under the resulting Kernel Density distribution.

For uniform kernel $k_y(x) = \begin{cases} 0 & \text{if } x < y-b \\ \frac{x-y+1000}{2\cdot 1000} & \text{if } y-b \le x \le y+b \\ 1 & \text{if } x > y+b \end{cases}$

X:	3000	4000	5000	7000
P.	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{1}{5}$

$$F(x) = 0.7 = \frac{1}{5}\cdot 1 + \frac{1}{5}\cdot 1 + \frac{2}{5}\cdot g \Rightarrow g = 0.75$$

$$g = 0.75 = \frac{x - 5000 + 1000}{2\cdot 1000}$$

$$\Rightarrow x = 5500$$

$\Rightarrow$ 70 th percentile $\rightarrow$ 5500

During 2015, Wang paid the following five warranty claims to Amstutz:

3000 4000 5000 5000 7000

This data will be used for questions 1-3 and is repeated here for your convenience.

5

Cheng is concerned that the use of the uniform distribution introduces too much variance into the Kernel Density Model.

2. (3 points) Calculate the variance under the model used above.

(uniform dist.

$\bar{x} = 4800 = \frac{3000+4000+5000+5000+7000}{5}$

$Var(x) = E[x^2] - \bar{x}^2$

$= \frac{3000^2+4000^2+5000^2+5000^2+7000^2}{5} - 4800^2$

$= 1760000$

variance $= E[x^2] - \bar{x}^2 + \frac{b^2}{3} = 1760000 + \frac{1000^2}{3}$

$= 2093333.333$

Cheng wants to know what the variance would be under the triangular kernel if the span was 1000.

3. (2 points) What did Mayfawny tell him?

variance $= E[x^2] - \bar{x}^2 + \frac{b^2}{6} = 1760000 + \frac{1000^2}{6}$

$= 1926666.667$

Wang Warranty also provides warranties to the Kexin Kar Kompany. The warranty provided to Kexin has an upper limit of 5000. Wang provides a sample of claims paid to Kexin as follows:

1000 2000 3000 4000 5000 5000 5000

The project manager for this project is Mengyun who decides to model claims for Kexin Kar using an exponential distribution.

4. (3 points) Calculate the maximum likelihood estimator for θ.

3

$$\hat{\theta}_{MLE} = \frac{1000 + 2000 + 3000 + 4000 + 5000 + 5000 + 5000}{4}$$

$$= \frac{25000}{4}$$

$$= 6250$$

Bergmann is also analyzing data associated with dental claims from Drew Dental Insurance Company. During 2015, Drew had 100 dental policies in force. The number of claims under each policy was distributed as follows:

Number of Claims	Number of Policies
0	10
1	25
2	40
3	14
4	7
5	0
6	4

100

This data is used for question 5-7.

Suyi, a Senior Vice President at Bergmann, believes that the claims should be modeled as a Poisson distribution. She wants to develop an 80% confidence interval for λ using the Maximum Likelihood Estimator for λ.

5. (6 points) Determine the confidence interval determined by Suyi.

6

$$\hat{\lambda} = \bar{x} = \frac{0\cdot10 + 1\cdot25 + 2\cdot40 + 3\cdot14 + 4\cdot7 + 5\cdot0 + 6\cdot4}{100}$$

$$= 1.99$$

$$var(\hat{\lambda}) = \frac{\hat{\lambda}}{n} = \frac{1.99}{100} = 0.0199$$

80% CI: $\hat{\lambda} \pm 1.282\sqrt{0.0199}$

$= (1.809151645, 2.170848355)$

The number of claims under each policy was distributed as follows:

Number of Claims	Number of Policies
0	10
1	25
2	40
3	14
4	7
5	0
6	4

9

This data is used for question 5-7 and is repeated here for your convenience.

The other Senior Vice President at Bergmann is Yang. Yang believes that the data should be modeled using a binomial distribution.

6. (3 points) Determine the Maximum Likelihood Estimate of q given that $m=9$.

$$\hat{q}_{MLE} = \frac{\bar{x}}{m} = \frac{1.99}{9} = 0.22111111$$

Chengtao, as the peer reviewer of the work for Yang, decides that he likes the binomial distribution as a model for this data. However, he decides to develop parameters using the method of moments.

7. (6 points) Determine m and q using the method of moments.

$$\bar{x} = mq = 1.99 \qquad var(x) = E[x^2] - \bar{x}^2 = \frac{567}{100} - 1.99^2 = 1.7099$$

$$= mq(1-q)$$

$$1-q = \frac{var(x)}{\bar{x}} = 0.859246231 \Rightarrow q = 0.140753769$$

$$m = \frac{\bar{x}}{q} = 14.13816494 \approx 14$$

as m must be integer

$$\Rightarrow \hat{m} = 14 \quad \hat{q} = \frac{\bar{x}}{\hat{m}} = 0.142142857$$

In addition to analyzing the number of claims for Drew Dental, Bergmann is analyzing the amount of each claim. Due to data issues, Drew is not able to provide the amount of all the claims. However, Drew was able to provide the following sample of claim amounts:

100 120 180 300 500

This data is used for questions 8-15.

Connor has been assigned the task of determining an acceptable model for the amount of each claim for Drew Dental. Connor wants to model the claim amount as an exponential distribution. She asked one of her team members, Tong, to develop the parameter for the exponential distribution using the smoothed empirical distribution and the 55th percentile.

6

8. (6 points) Determine the $\hat{\theta}$ determined by Tong.

$g = 0.55 \Rightarrow j = \lfloor 0.55 \cdot (5+1) \rfloor = \lfloor 3.3 \rfloor = 3$

$h = 3.3 - 3 = 0.3$

$\rightarrow \pi_g = 0.7 \cdot 180 + 0.3 \cdot 300 = 216$

$F^*(\pi_g) = F^*(216) = 0.55 = 1 - e^{-216/\theta}$

$e^{-\frac{216}{\theta}} = 0.45$

$\hat{\theta} = \frac{-216}{\ln 0.45} = 270.5045938$

Drew was able to provide the following sample of claim amounts:

100 120 180 300 500

This data is used for questions 8-15 and is repeated here for your convenience.

Connor decided that she would also like to model the amount of each claim as a Pareto distribution. She asked another team member, Jieyu, to estimate the parameters for the Pareto distribution using the sample data and the Method of Moments Matching.

8

9. (8 points) Determine the estimated parameters for the Pareto as determined by Jieyu.

$$\bar{x} = \frac{100+120+180+300+500}{5} = 240 = \frac{\theta}{\alpha-1}$$

$$E[x^2] = \frac{100^2+120^2+180^2+300^2+500^2}{5} = 79360 = \frac{\theta^2 \cdot 2}{(\alpha-1)(\alpha-2)}$$

$$\rightarrow \frac{E[x^2]}{\bar{x}^2} = \frac{\frac{2\theta^2}{(\alpha-1)(\alpha-2)}}{\frac{\theta^2}{(\alpha-1)^2}} = \frac{2\alpha-2}{\alpha-2} = \frac{79360}{57600}$$

$$79360\alpha - 158720 = 115200\alpha - 115200$$

$$-35840\alpha = 43520$$

$$\begin{cases} \hat{\alpha} = -1.214285714 \\ \hat{\theta} = -531.4285714 \end{cases}$$

Drew was able to provide the following sample of claim amounts:

100 120 180 300 500

This data is used for questions 8-15 and is repeated here for your convenience.

Connor decides to test whether an exponential model is appropriate using hypothesis testing. She develops the following hypothesis:

H_0: The data is distributed as an exponential distribution with a mean of 250.
H_1: The data is not distributed as an exponential distribution with a mean of 250.

12

Connor asks Jackson to test this hypothesis. Jackson decides to use the Kolmogorov-Smirnov Test at a 5% significance level.

10. (8 points) Determine the Kolmogorov-Smirnov Test Statistic. $\theta = 250$.

x	$F_n(x^-)$	$F_n(x)$	$F_n^*(x) = 1-e^{-\frac{x}{\theta}}$	Absolute value of Difference
100	0	0.2	0.329679954	0.329679954
120	0.2	0.4	0.381216608	0.181216608
180	0.4	0.6	0.513247744	0.113247744
300	0.6	0.8	0.698805188	0.101194212
500	0.8	1	0.864664717	0.1353352833

$\Rightarrow$ Test statistic $D = \max(|\text{Difference}|) = 0.329679954$

11. (2 points) Determine the critical value.

$\alpha = 5\% \Rightarrow$ critical value $= \frac{1.36}{\sqrt{5}} = 0.60821049$

12. (2 points) State Jackson's conclusion regarding Connor's hypothesis.

As $0.60821049 > 0.329679954$

$\Rightarrow$ Fail to reject H_0.

$\Rightarrow$ The dist could be exponential distribution with mean of 250 at 5% significant level

Drew was able to provide the following sample of claim amounts:

100 120 180 300 500

This data is used for questions 8-15 and is repeated here for your convenience.

Connor also wants to test the following hypothesis using the Likelihood Ratio Test:

H_0: The data is distributed as an exponential distribution with a mean of 250.
H_1: The data is distributed as a gamma distribution with $\alpha = 2$.

14

Connor asks Tianyu to complete this test at a 5% significance level.

In completing his work, Tianyu calculated L_1 which is the value of the maximum likelihood estimator under the alternative hypothesis.

13. (2 points) Determine the Maximum Likelihood Estimate of θ for the alternative hypothesis.

For Gamma $\hat{\theta}_{MLE} = \frac{\bar{X}}{\alpha} = \frac{240}{2} = 120$

14. (10 points) Tianyu determines that $L_1 = \frac{X}{(120)^5} e^Y$. Determine the numeric values of X and Y. Remember that $\Gamma(\alpha) = (\alpha - 1)!$ provided that α is a positive integer.

under Alternative (gamma)

$$L_1 = L(\theta_1) = f(100) \cdot f(120) \cdot f(180) \cdot f(300) \cdot f(500)$$

$$= \frac{(100/\theta)^2 \cdot e^{-100/\theta}}{100 \cdot (2-1)!} \cdot \frac{(120/\theta)^2 \cdot e^{-120/\theta}}{120 \cdot (2-1)!} \cdot \ldots \cdot \frac{(500/\theta)^2 e^{-500/\theta}}{500 (2-1)!}$$

$$= \frac{3.24 \times 10^{11} e^{-1200/\theta}}{\theta^{10}} = \frac{13.02083333 \cdot e^{-10}}{(120)^5}$$

$\to$ $X = 13.02083333$ $Y = -10$

15. (2 points) Determine the critical value for this test.

$df = 1 - 0 = 1$ $\alpha = 5\%$ $\chi^2_{critical} = 3.841$

Dr. Ge, who has been named the Chief Statistician at Bergmann after earning her PhD at Harvard, sees the project that Connor has been working on and concludes that more data is necessary to develop an appropriate model. Dr. Ge asks Michael to see if he can extract more data from Drew Dental. Michael is unable to develop complete data, but is able to derive the following grouped data:

Amount of Claim	Number of Claims
0 – 200	8
200 – 400	9
400 +	3

10

This data is used for questions 16-21.

Using Michael's data, Dr. Ge models the claims as an exponential distribution with $\hat{\theta}$ derived using the Maximum Likelihood Estimator.

16. (10 points) Determine the $\hat{\theta}$ used by Dr. Ge. $F(x) = 1 - e^{-\frac{x}{\theta}}$

$$L(\theta) = (F(200) - F(0))^8 [F(400) - F(200)]^9 [1 - F(400)]^3$$

$$= (1 - e^{-\frac{200}{\theta}})^8 \cdot [e^{-\frac{200}{\theta}} - e^{-\frac{400}{\theta}}]^9 [e^{-\frac{400}{\theta}}]^3$$

$$= (1 - e^{-\frac{200}{\theta}})^8 (e^{-\frac{200}{\theta}})^9 [1 - e^{-\frac{200}{\theta}}]^9 [e^{-\frac{400}{\theta}}]^3$$

$$= (1 - e^{-\frac{200}{\theta}})^{17} \cdot e^{-\frac{3000}{\theta}}$$

$$\ell(\theta) = 17 \ln(1 - e^{-\frac{200}{\theta}}) + (-\frac{3000}{\theta})$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{17 \cdot (-e^{-\frac{200}{\theta}}) \cdot 200}{(1 - e^{-\frac{200}{\theta}}) \cdot \theta^2} + \frac{3000}{\theta^2} = 0$$

$$(1 - e^{-\frac{200}{\theta}}) 3000 = 3400 e^{-\frac{200}{\theta}} \Rightarrow e^{-\frac{200}{\theta}} = 0.46875 \qquad \hat{\theta} = \frac{-200}{\ln 0.46875} = 263.9616922$$

$$\hat{\theta} = 263.9616922$$

You are given the following grouped data:

Amount of Claim	Number of Claims
0 – 200	8
200 – 400	9
400 +	3

5

This data is used for questions 16-21 and repeated here for your convenience.

Dr. Ge then asks Mengying and Ningzhu to create a model for this data assuming a uniform distribution on the range of $(0, U)$. Mengying uses the information given and determines $\hat{U}$ using the Maximum Likelihood Estimator.

17. (3 points) Determine the $\hat{U}$ used by Mengying.

$$\hat{u}_{MLE} = \frac{20}{17} \cdot 400 = 470.5882353$$

Ningzhu decides to dig deeper into data and finds that the three claims that exceeded 400 were actually 450, 500, and 520. Using this additional data, Ningzhu calculates $\hat{U}$ using the Maximum Likelihood Estimator.

18. (2 points) Determine the $\hat{U}$ used by Ningzhu.

$$\hat{u}_{MLE} = \max(X_1 \cdots X_n) = 520$$

You are given the following grouped data:

Amount of Claim	Number of Claims
0 – 200	8
200 – 400	9
400 +	3

20

12

This data is used for questions 16-21 and repeated here for your convenience.

Dr. Ge also decides that they should test whether the uniform distribution is an appropriate fit to this data using hypothesis testing. Dr. Ge wants to test the following hypothesis:

H_0: The data is distributed as a uniform distribution over $(0,U)$.

H_1: The data is not distributed as a uniform distribution over $(0,U)$.

Dr. Ge is not comfortable with either of the estimates for U so she estimates $U = 525$.

Dr. Ge asks Shunan to complete a hypothesis test using the Chi-Square test at a significance level of 5%. She further instructs Shunan to use the grouped data and not the additional data developed by Ningzhu.

19. (8 points) Determine the χ^2 test statistic.

Range.	Obs	E_j	$\frac{(E_j - O_j)^2}{E_j}$
0 – 200	8	$= \frac{200}{525} \cdot 20 = 7.619047619$	0.019047619
200 – 400	9	$= \frac{200}{525} \cdot 20 = 7.619047619$	0.250297619
400 +	3	$= \frac{125}{525} \cdot 20 = 4.761904762$	0.651904762

$$\chi^2 = \sum \frac{(E_j - O_j)^2}{E_j} = 0.92125$$

20. (2 points) Determine the critical value for this test.

$df = 3 - 1 - 1 = 1 \quad \alpha = 5\% \quad \Rightarrow \chi^2$ critical value $= \underline{3.841}$

21. (2 points) State Shunan's conclusion with regard to the hypothesis.

As $0.92125 < 3.841 \Rightarrow$ Fail to reject H_0.

$\Rightarrow$ The data could be distributed over $[0, 525)$ at 5% significant level.

Michael Henry is the owner of The Henry Health Insurance Company. Henry has been selling a hospital indemnity plan that has no deductible and no upper limits. Henry decides that this product is too risky. Henry will now begin selling a hospital indemnity plan that has a ordinary deductible of 300 for each claim. Additionally, it will have a maximum out of pocket of 750 for a calendar year per policy. In other words, the most that an insured could pay is 750 for any calendar year. If an insured pays 750, then Henry will pay all additional costs.

Based on past experience. Michael knows that the number of claim is distributed as a binomial distribution with $m=4$ and $q=0.2$. Michael also expects the amount of each claim to be distributed as an exponential distribution with $\theta=2000$.

Michael contacts Huining who works for Bergmann Statistical Bureau to simulate claims payments for the new hospital indemnity policy. Huining has published numerous papers on simulation and is considered the leading expert in the field.

10

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

The random numbers used in the simulation are:

0.98 0.22 0.75 0.10 / 0.52 0.92 / 0.14 0.67 0.55 0.30 0.66 0.95 0.71 0.04

22. (10 points) Calculate the simulated aggregate claim payments paid by Henry for Lindsay and the simulated aggregate claim payments paid by Henry for Rehan.

N.	0.	1	2	3	4
$P(N)$	$0.8^4 = 0.4096$	$4\cdot 0.2\cdot 0.8^3 = 0.4096$	$6\cdot 0.2^2\cdot 0.8^2 = 0.1536$	$4\cdot 0.2^3\cdot 0.8 = 0.0256$	$0.2^4 = 0.0016$
$F(N)$	0.4096	0.8192	0.9728	0.9984	1.

Lindsay. $u_N^{**} = 0.98 \Rightarrow N_{Lindsay} = 3.$ $\quad u_x^{**} = F(x) = 1-e^{-\frac{x}{\theta}} \Rightarrow x^{**} = -\theta \cdot \ln(1-u^{**})$

$\Rightarrow u_{x_1}^{**} = 0.22 \Rightarrow x_{x_1}^{**} = -2000 \ln(1-0.22) = 496.9227186 > 300.$

$\rightarrow$ Lindsay pay 300 and Henry pay 196.9227186 For Lindsay.

$u_{x_2}^{**} = 0.75 \Rightarrow x_{x_2}^{**} = -2000 \ln(1-0.75) = 2772.588722 > 300$

Lindsay pay 300 and Henry pay 2472.588722 For Lindsay

$u_{x_3}^{**} = 0.1 \Rightarrow x_{x_3}^{**} = 210.7210313 < 300$

while $210.7210313 > 750 - 600 = 150$

$\Rightarrow$ Lindsay pay 150 and Henry pay 60.72103132. For Lindsay

$\Rightarrow$ Henry in total pay 2730.232472 For Lindsay

For Rehan $u_N^{**} = 0.52 \Rightarrow N^{**} = 1$.

$u_X^{**} = 0.92 \Rightarrow X^{**} = -2000 \ln(1-0.92) = 5051,457289 > 300$

Rehan pay 300 and Henry pay 4751,457289

$\Rightarrow$ Henry pay 4751,457289 For Rehan.

conclusion

$\Rightarrow$ Henry pay 2730,232472 For Lindsay and 4751,457289 For Rehan.

4

After completing these two simulations, Huining asks Henry Health Insurance how many ~~simulations the~~ Company would like completed. Jacob, who is the Chief Actuary of Henry Health Insurance, wants the standard deviation of the estimate of $E[X]$ to be less than 2% of the estimate of $E[X]$. In other words, he wants $\sqrt{Var(\overline{X})} < 0.02\overline{X}$. He asks Huining to determine the number of simulations based on that criteria. In order to do this, Huining completed two more simulations. Those simulations result in aggregate claims payments of:

2000 5000

Using these last two simulations only, Huining determined that n simulations were needed.

23. (4 points) Determine n.

$$\bar{x} = 3500 \qquad var(\bar{x}) = \frac{Var(x)}{n} = \frac{E[x^2] - \bar{x}^2}{n} = \frac{2250000}{n}$$

$$\Rightarrow \sqrt{\frac{2250000}{n}} < 2\% \cdot 3500 = 70$$

$$n > 459.1836735 \quad \text{or} \quad n \geq 460$$

$\Rightarrow$ n need to be at least $\underline{\underline{460}}$

Michael is concerned about the variance of the claims under the revised hospital indemnity plan. The main reason to change the hospital indemnity plan was to reduce the volatility of the claims. Michael asks Joanna, who is Vice President of claims, to use the first three claims that are received to estimate the variance. Joanna estimates the variance using the following estimator:

$$\hat{Var}(X) = \frac{\sum_{i=1}^{3}\left(X_i - \overline{X}\right)^2}{2}$$

10

The first three claims received were:

1600 2500 2500

Joanna asks Christian, another actuary at Henry, to use the Bootstrap method to determine the mean square error in this estimator.

24. (10 points) Determine Christian's answer divided by 100,000. $Var(X) = E[X^2] - (E[X])^2 = 180000.$

sample	probability	$\hat{\mu}$	$\hat{Var}(\hat{X})$	$[Var(\hat{X}) - Var(X)]^2$
1600 1600 1600	$(\frac{1}{3})^3 = \frac{1}{27}$	1600	0	$180000^2 = 3.24\times10^{10}$
1600, 2500, 1600	$(\frac{1}{3})^2 \cdot \frac{2}{3} \cdot 3 = \frac{6}{27}$	1900	$\frac{\Sigma(X_i-\bar{X})}{2} = 270000$	$90000^2 = 8.1\times10^9$
1600, 2500, 2500	$(\frac{1}{3})\cdot(\frac{2}{3})^2 \cdot 3 = \frac{12}{27}$	2200	270000	8.1×10^9
2500 2500 2500	$(\frac{2}{3})^3 = \frac{8}{27}$	2500	0	3.24×10^{10}

$$\Rightarrow MSE = \frac{1}{27}\cdot 3.24\times10^{10} + \frac{8}{27}\cdot 3.24\times10^{10} + \frac{18}{27}\cdot 8.1\times10^9$$
$$= 1.08\times10^{10} + 5.4\times10^9$$
$$= 1.62\times10^{10}$$

$$\Rightarrow \text{Answer} = \frac{MSE}{100{,}000} = \underline{162000}$$