STAT 479
Spring 2013
Final
May 2, 2013

During 2012, Mao Medical provided dental insurance that covered 100% of all dental claims. In other words, there were no deductibles or upper limits.

For the dental claims incurred during 2012, Mao Medical randomly selected a sample of claims that were the following amounts:

100 100 200 300 450

This data will be used for questions 1-9.

Yi who is Mao's Chief Actuary hires Berry-Kaley Consulting to use this data to develop a continuous distribution that can be used for dental claim amounts. One of the partners of the consulting firm is Kristin who decides to use a Kernel Density model with a uniform kernel and a span of 100 to model this data.

1. (3 points) Determine the mean and variance of Kristen's Kernel Density model.

Solution:

$$Mean = E[X] \text{ for empirical distribution} = \frac{100+100+200+300+450}{5} = 230$$

$$Var = Var \text{ for empirical distribution + factor} =$$

$$\frac{(100-230)^2+(100-230)^2+(200-230)^2+(300-230)^2+(450-230)^2}{5}+\frac{100^2}{3}$$

$$=17,600+3333.33=20,933.33$$

2. (5 points) Determine $\hat{F}(220)$ under Kristen's Kernel Density model.

Solution:

$$F(220) = 0.4F_{100}(220) + 0.2F_{200}(220) + 0.2F_{300}(220) + 0.2F_{450}(220)$$

$$F_{100}(220) = 1 \text{ since } 220 > 200$$

$$F_{200}(220) = \frac{220 - 200 + 100}{200} = \frac{120}{200} \text{ since } 200 < x < 400$$

$$F_{300}(220) = \frac{220 - 300 + 100}{200} = \frac{20}{200} \text{ since } 300 < x < 500$$

$$F_{450}(220) = 0 \text{ since } 220 < 350$$

$$\widehat{F(220) =} \, 0.4(1) + 0.2\left(\frac{120}{200}\right) + 0.2F_{300}\left(\frac{20}{200}\right) + 0.2(0) = 0.54$$

For the dental claims incurred during 2012, Mao Medical randomly selected a sample of claims that were the following amounts:

100	100	200	300	450

This data is repeated from above for your convenience and will be used for questions 1-9.

The other partner in Berry-Kaley, Matthew, believes that using a triangular kernel would be more appropriate. He develops a continuous distribution based on the triangular kernel.

3. (2 points) Determine the mean and variance of Matthew's Kernel Density model with a span of 100.

Solution:

$$Mean = E[X] \text{ for empirical distribution} = \frac{100+100+200+300+450}{5} = 230$$

$$Var = Var \text{ for empirical distribution + factor =}$$

$$\frac{(100-230)^2+(100-230)^2+(200-230)^2+(300-230)^2+(450-230)^2}{5}+\frac{100^2}{6}$$

$$=17,600+1666.67=19,266.67$$

For the dental claims incurred during 2012, Mao Medical randomly selected a sample of claims that were the following amounts:

100	100	200	300	450

This data is repeated from above for your convenience and will be used for questions 1-9.

Yi reviews the work of Kristin and Matthew and decides that she really does not like using the Kernel Density model to model the claim amounts for dental insurance. She decides to consult with Cynthia who is well known as a dental insurance consultant. Cynthia states that dental claims should be modeled as an exponential distribution. Further, she suggests that best way to determine an estimate for θ is to use the Method of Percentile Matching with a percentile of 60%. She asks her partner, Kunyu, to complete the work.

4. (5 points) Determine $\hat{\theta}$ as determined by Kunyu.

   Solution:

   $$\lfloor (n+1)j \rfloor = \lfloor 6(.6) \rfloor = \lfloor 3.6 \rfloor = 3; h = 0.6$$

   $$x = (1-0.6)\,\mathrm{X}_3 + (0.6)X_4 = (0.4)(200) + (0.6)(300) = 160$$

   $$F(x) = 0.6 = 1 - e^{\frac{-x}{\theta}} \implies 0.6 = 1 - e^{\frac{-260}{\theta}}$$

   $$\theta = -260 / \ln(1-0.6) = 283.753$$

For the dental claims incurred during 2012, Mao Medical randomly selected a sample of claims that were the following amounts:

100 100 200 300 450

This data is repeated from above for your convenience and will be used for questions 1-9.

Yi likes the use of the exponential distribution, but would prefer to use $\hat{\theta} = 200$ since that is the median and she learned in STAT 479 that using the Percentile Matching approach has drawbacks. She hires a well known actuary from Kazakhstan to test her theory. Syrym decides to use the Kolmogorov-Smirnov test at a 90% significance level with the following hypothesis:

H_0: The data is from an exponential distribution with a mean of 200

H_1: The data is not from an exponential distribution with a mean of 200

5. (8 points) Find the Kolmogorov-Smirnov Test Statistic.

Solution:

x	$F_n(x^-)$	$F_n(x)$	$F_n^*(x)$	Greatest Absolute Value
100	0	0.4	$1-e^{-\frac{100}{200}} = 0.39346934$	0.39346934
200	0.4	0.6	$1-e^{-\frac{200}{200}} = 0.632120559$	0.232120559
300	0.6	0.8	$1-e^{-\frac{300}{200}} = 0.77686984$	0.17686984
450	0.8	1.0	$1-e^{-\frac{450}{200}} = 0.894600775$	0.105399225

$$D^* = Greatest\ Difference = 0.3935$$

6. (3 points) Determine the critical value for this Test.

Solution:

$$\frac{1.22}{\sqrt{5}} = 0.5456$$

7. (1 points) State Syrym's conclusion.

Solution:

We do not reject as 0.5456>0.3934

For the dental claims incurred during 2012, Mao Medical randomly selected a sample of claims that were the following amounts:

100	100	200	300	450

This data is repeated from above for your convenience and will be used for questions 1-9.

Yi would also like this hypothesis tested under the Anderson Darling Test. Ang, an expert in the Anderson Darling Test, determines that $A^2 = 0.4805$. Assume that this test is also completed at the 90% significance level.

8. (3 points) Determine the critical value for this Test.

   Solution:

   Critical value = 1.933

9. (1 points) State Ang's conclusion.

   Solution:

   Do not reject as 1.933>0.4805

During 2012, Mao Medical has 3600 dental policies in force. The number of claims per policy were distributed as follows:

Amount of Claims	Number of Claims
0	800
1	1000
2	1000
3	600
4	150
5	50

This data is used for questions 10 and 11.

Yi hires Binomial Distributions by Bryce to analyze the data. Not surprisingly, Bryce decides that this is best modeled with a Binomial distribution. Bryce uses the method of moment matching to estimate the parameters for the Binomial distribution:

10. (8 points) Determine the parameters estimated by Bryce.

Solution:

$$E[X]=\frac{(800)(0)+(1000)(1)+(1000)(2)+(600)(3)+(150)(4)+(50)(5)}{3600}=1.5694$$

$$E[X^2]=\frac{(800)(0^2)+(1000)(1^2)+(1000)(2^2)+(600)(3^2)+(150)(4^2)+(50)(5^2)}{3600}=3.903$$

$$Var(X)=3.903-(1.5694)^2=1.4396$$

$$mq=1.5694; mq(1-q)=1.4396$$

$$1-q=\frac{1.4396}{1.5694}=0.9173 ==> q=0.0827 ==> m=\frac{1.5694}{0.0827}=18.97$$

$$==> \widehat{m}=19 \text{ and } \widehat{q}=0.0826$$

During 2012, Mao Medical has 3600 dental policies in force. The number of claims per policy were distributed as follows:

Amount of Claims	Number of Claims
0	800
1	1000
2	1000
3	600
4	150
5	50

This data is repeated from question 10 for your convenience and is used for questions 10 and 11.

While Yi has heard that the Binomial distribution is a good distribution to use for modeling the number of dental claims, her favorite discrete distribution is the Poisson distribution. She asks her friend Yuxi to estimate the parameter λ for a Poisson distribution. Rather than providing Yi with a single value for $\hat{\lambda}$, Yuxi decides to provide a range of values. She does that by estimating $\hat{\lambda}$ using the Maximum Likelihood Estimator. She then estimates the range using a 99% linear confidence interval.

11. (8 points) Determine Yuxi's range of values for $\hat{\lambda}$.

B

Mao Medical Insurance also sells a medical insurance product. This product has an upper limit of 10,000 that is applied to each claim.

The following sample of **claim payments** paid under the medical insurance product are available:

500	10,000

Kaidan, another actuary at Mao Medical, believes that the claims should be modeled using a Pareto distribution with $\alpha = 2$. Kaidan uses the Maximum Likelihood Estimator to estimate θ.

12. (10 points) Determine the Maximum Likelihood Estimator of θ for the Pareto distribution.

Yi decides that this sample of claim payments is too small and expands that sample to the following five **claim payments** paid under the medical insurance product:

500	1250	2750	3250	10,000

The upper limit of 10,000 still applies.

This data is used for questions 13-14.

Yi is not sure which distribution to use to model claims incurred. She thinks that it is potentially a uniform distribution on a range of 0 to U.

13. (4 points) Determine the Maximum Likelihood Estimator of U for the uniform distribution.

Alternatively, Yi thinks it might be an exponential distribution with a parameter of θ.

14. (4 points) Determine the Maximum Likelihood Estimator of θ for the exponential distribution.

Yi completes additional investigative work and discovers the amount of the 5th claim was actually 15,000. Therefore, the **amount incurred** for the five claims was actually:

500	1250	2750	3250	15,000

This data is used for questions 15 and 16.

15. (4 points) Based on this additional data, what would be your Maximum Likelihood Estimators of U for the uniform distribution and for θ if it is an exponential distribution.

Therefore, the **amount incurred** for the five claims was actually:

500	1250	2750	3250	15,000

This data is repeated from question 15 for your convenience and is used for questions 15 and 16.

Yi hires Zhang Consultants to model this data. Zhang decides to model the data as a Gamma distribution. Zhang determines the parameters for the Gamma distribution using the Method of Moments.

16. (8 points) Determine the parameters determined by Zhang.

Yi thinks that the distribution of claims looks more like an exponential and asks Salisbury & Associates to test the data to determine if the exponential distribution is reasonable. One of the associates at Salisbury, Alex, is tasked with completing the project. He decides to complete a Chi Square Test at a 95% significance level on the following hypothesis:

H_0: The data is from an Exponential distribution with $\theta = 5000$.

H_1: The data is not from an Exponential distribution with $\theta = 5000$.

Alex informs Yi that in order to complete such a test, he will need more data. Yi is able to produce the following data for Alex:

Amount of Payment	Number of Payments
0 – 1000	950
1000 - 4000	1750
4000 - 8000	1300
8000+	1000

This data is used for questions 17-19.

17. (10 points) Determine the χ^2 .

18. (3 points) Determine the critical value for this test.

19. (2 points) State Alex's conclusion.

Beginning in 2013, Mao Medical will begin offering a new coverage which is Hospital Indemnity Insurance. The coverage only pays for costs associated with hospital visits. The coverage that Mao will offer will have a deductible of 500 for each claim. Additionally, it will have a maximum out of pocket of 1200. In other words, the most that an insured could pay is 1200. If an insured pays 1200, then Mao will pay all additional costs.

Yi expects the number of claims to be distributed as a Poisson distribution with a mean of 1. Yi also expects the amount of each claim to be distributed as a Pareto distribution with $\theta = 6000$ and $\alpha = 4$.

Mao Medical hires Emily from Scully Simulators to simulate **claim payments** for 2013.

Using simulation, Emily 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. Julia and Ben are the first two insureds. First, Emily determines the number of claims for Julia and then the amount of each claim for Julia. Next, Emily determines the number of claims for Ben. Finally, Emily simulates the amount of each of Ben's claims.

The random numbers used in the simulation are:

0.93 0.22 0.75 0.43 0.52 0.92 0.14 0.67 0.55 0.30 0.66 0.95 0.71 0.04

20. (12 points) Calculate the simulated aggregate claim payments paid by Mao for Julia and the simulated aggregate claim payments paid by Mao for Ben.

After completing these two simulations, Scully asks Mao Medical how many simulations the Company would like completed. Yifei who is the president of Mao Medical wants the standard deviation of the estimate of $E[X]$ to be less than 100. He asks Scully to determine the number of simulations based on that criteria. In order to do this, Scully completed two more simulations. Those simulations result in aggregate claims payments of:

2000 6000

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

21. (6 points) Determine n.

Yi is worried that the average claim for the Hospital Indemnity may be significantly different from that assumed above. She has decided to take the first three claims that are received and to use those to estimate the expected claim amount per claim. The estimator for μ that she has decided to use is the average of the sample mean and the sample median. Expressed in symbols, the estimator is:

$$\hat{\mu} = \frac{\bar{X} + \pi_{0.5}}{2}$$

The first three claims received were:

900	900	3900

Yi asks Cici, another actuary at Mao Medical, to use the Bootstrap method to determine the mean square error in this estimator.

22. (10 points) What was Cici's answer?