BACKGROUND FOR THE TEST

This test is set as a case study of Santiago Substandard Auto Company. This means that you will be using the same set up for all the problems. It also means that you are using the same data for several problems. Here is the case study scenario.

Santiago sells automobile insurance to drivers who have a bad driving record. Each policy only covers one automobile. During the exam, you will be studying the results of three years – 2018, 2019, and 2020.

During 2018, Santiago had 10,000 policies. Each policy paid for 100% of all claims. In other words, there were no deductibles, no upper limit, and no coinsurance. A policy can have more than one claim on an automobile. You will then be given the results for 2018 and will be fitting models or the parameters for models to this data.

During 2019, Santiago modifies its auto insurance policy by implementing an upper limit per claim of 10,000. There remains no deductible and no coinsurance. Once again, you are given the data for 2019 and will be fitting models or the parameters for the models to this data.

Finally, in 2020, Santiago has again revised the insurance policy sold by instituting a deductible **per claim** of 500. The Company also instituted an upper limit **per insurance policy** of 25,000. You will be expected to project the claims for 2020 using simulation.

There are 11 questions and 120 points on the test. While the stem to the problems all revolve around the case study scenario, each problem stands on its own. In other words, your answer to one problem does not get used in another problem. However, since you are dealing with the same data, there may be some time savings. For example, if you are using the same sample data within multiple problems, you should only need to calculate the mean and variance of the sample once.

In order to complete all the tasks necessary, Alisa, who is the chief actuary for Santiago, hires various consultants to complete each of the 11 problems. All in all, as you take the test, you will see Alisa hires ten consultants from this class to complete various tasks. In the end, we can conclude that Santiago's biggest problem is TOO MANY CONSULTANTS!

STAT 479
Spring 2020
Test 2

Santiago Substandard Auto Company sells automobile insurance to drivers who have a bad driving record. Each policy only covers one automobile.

During 2018, Santiago had 10,000 policies. Each policy paid for 100% of all claims. In other words, there were no deductibles, no upper limit, and no coinsurance.

During 2018, the number of claims per insured automobile for Santiago Substandard Auto Company was:

Number of Accidents in 2018	**Number of Policies**
0	2400
1	3400
2	2400
3	1500
4	300

1. (10 points) Alisa, the company's actuary, hires Jaden from Cornell Consultants. Jaden models the data as a Poisson distribution. Calculate the 90% linear confidence interval for λ based on the above data.
   Solution:

$$\hat{\lambda} = \bar{X} = \frac{3400 + (2)(2400) + (3)(1500) + (4)(300)}{10,000} = 1.39$$

$$Var[\hat{\lambda}] = \frac{\hat{\lambda}}{n} = \frac{1.39}{10,000} = 0.000139$$

$$90\% \; Confidence \; Interval \; => 1.39 \pm 1.645\sqrt{0.000139} = (1.3706, 1.40904)$$

2. Alisa also retains another consultant, Tom, to test the following hypothesis using the Chi Square Test with a 99% significance level:

   H_0: The data is from a Poisson distribution.
   H_1: The data is not from a Poisson distribution.

   Tom wants to group the data so that there are at least 1000 policies in each cell. Therefore, he used the data in the following table to complete the Chi Square Test:

Number of Accidents in 2018	Number of Policies
0	2400
1	3400
2	2400
3+	1800

   (10 points) Calculate the Chi Square test statistic.

   Solution:
   Since we are not given λ in the hypothesis, we have to estimate using the MLE.
   $\lambda = \overline{X} = 1.39$ from question number 1.

Number of Accidents in 2011	**Number of Policies** (O_j)	E_j	$\frac{(E_j - O_j)^2}{E_j}$
0	2400	$10{,}000e^{-1.39} = 2490.753$	3.307
1	3400	$10{,}000e^{-1.39}(1.39) = 3462.147$	1.116
2	2400	$10{,}000e^{-1.39}(1.39)^2 / 2 = 2406.192$	0.016
3+	1800	$10{,}000 - 2490.753 - 3462.147 - 2406.192 = 1640.908$	15.425
		χ^2 ➔	19.864

   (4 points) Calculate the critical value for this test.

   Solution:
   Degrees of freedom = 4 – 1 – 1 = 2 ➔ Critical Value = 9.21

   (1 point) State your conclusion.

   Solution:
   Since χ^2 > 9.21, we reject H_0.

3. Alisa also wants to develop a model for the amount of each claim that Santiago Substandard Auto Company will incur. She hires Jiafu from Niu Number Crunchers. The only data that Alisa can provide to Jiafu is the following sample of five claims received during 2018:

200 1000 5000 10,000 100,000

Jiafu wants to test the following hypothesis using the Kolmogov-Smirnov Test with a 95% significance level:

H_0: The data is from a Pareto distribution with a $\theta = 75{,}000$ and $\alpha = 4$.
H_1: The data is not from a Pareto distribution with a $\theta = 75{,}000$ and $\alpha = 4$.

(10 points) Calculate the Kolmogorov-Smirnov Test statistic.

Solution:

x	$F_n(x^-)$	$F_n(x)$	$F^*(x) = 1 - \left(\frac{75{,}000}{75{,}000 + x}\right)^4$	Absolute Value of Maximum Difference
200	0	0.2	$1 - \left(\frac{75{,}000}{75{,}200}\right)^4 = 0.01060$	0.18940
1000	0.2	0.4	$1 - \left(\frac{75{,}000}{76{,}000}\right)^4 = 0.05160$	0.34840
5000	0.4	0.6	$1 - \left(\frac{75{,}000}{80{,}000}\right)^4 = 0.22752$	0.37248
10,000	0.6	0.8	$1 - \left(\frac{75{,}000}{85{,}000}\right)^4 = 0.39387$	0.40613
100,000	0.8	1.0	$1 - \left(\frac{75{,}000}{175{,}000}\right)^4 = 0.96626$	0.16626

D =Maximum absolute value = 0.40613

(4 points) Calculate the critical value for the Kolmogov-Smirnov test.

Solution:

Critical Value = $\frac{1.36}{\sqrt{5}} = 0.6082$

(1 points) State Jiafu's conclusion.

Solution:
Since D<0.6082, we do not reject H_0

4. Santiago also wants to test the following hypothesis using the likelihood ratio method:

   H_0: The data is from an exponential distribution with a $\theta = 25,000$.
   H_1: The data is from a gamma distribution.

   Alisa wants to test this hypothesis at a 90% significance level.

   Use the same sample that was provided to Jiafu which was the following sample of five claims received during 2018:

   200 1000 5000 10,000 100,000

   (10 points) When Alisa calculated L_0 which is $L(\boldsymbol{\theta_0})$, she got an equation in the form of $\dfrac{e^A}{(25,000)^B}$. Determine A and B.

Solution:

$$L(\boldsymbol{\theta_0}) = f(200) \cdot f(1000) \cdot f(5000) \cdot f(10,000) \cdot f(100,000) =$$

$$\frac{e^{-\frac{200}{\theta}}}{\theta} \cdot \frac{e^{-\frac{1000}{\theta}}}{\theta} \cdot \frac{e^{-\frac{5000}{\theta}}}{\theta} \cdot \frac{e^{-\frac{10,000}{\theta}}}{\theta} \cdot \frac{e^{-\frac{100,000}{\theta}}}{\theta} = \frac{e^{-\frac{116,200}{\theta}}}{\theta^5} = \frac{e^{-\frac{116,200}{25,000}}}{(25,000)^5} = \frac{e^{-4.648}}{(25,000)^5}$$

$$A = -4.648 \quad and \quad B = 5$$

(4 points) Calculate the critical value for this test.

Solution:

Degrees of freedom = 2 – 0 = 2 ➔ Critical Value = 4.605

5. Santiago is also concerned with the variance around the Maximum Likelihood Estimators being used by the various consultants. Alisa hires Keyi to calculate the maximum likelihood estimate for the exponential distribution with a parameter of θ.

Use the same sample that was provided to Jiafu which was the following sample of five claims received during 2018:

200 1000 5000 10,000 100,000

(4 points) Determine Keyi's maximum likelihood estimate for θ.

Solution:

$$MLE = \bar{X} = \frac{200+1000+5000+10,000+100,000}{5} = 23,240$$

(4 points) Determine the variance of this maximum likelihood estimate.

Solution:

$$Var(\hat{\theta}) = \frac{\left(\hat{\theta}\right)^2}{n} = \frac{(23,240)^2}{5} = 108,019,520$$

(12 points) Using the delta method, calculate the approximate variance of the maximum likelihood estimator for $S(10,000)$.

Solution:

$$S(10,000) = 1 - F(10,000) = e^{-\frac{10,000}{\theta}}$$

$$g(\theta) = e^{-\frac{10,000}{\theta}} \implies g'(\theta) = \left(\frac{10,000}{\theta^2}\right) e^{-\frac{10,000}{\theta}}$$

$$Var[S(10,000)] = [g'(\theta)]^2 \left[\frac{\sigma^2}{n}\right] = \left[\left(\frac{10,000}{(23,240)^2}\right) e^{-\frac{10,000}{23,240}}\right]^2 [108,019,520] = 0.01566$$

6. (4 points) While Jiafu is completing his work, Alisa decides to take the same sample and model the amount of claims as uniformly distributed on the range of (0,U). To do this, she hires Tague from Sode Soothsayers Inc.

   The sample was the following five claims received during 2018:

   200 1000 5000 10,000 100,000

   Calculate the maximum likelihood estimate of U.

   Solution:

   MLE of U is MAX(200 1000 5000 10,000 100,000) = 100,000

7. (5 points) Following this work, Alisa hires Dotterer Data Scrubbers to develop more information regarding the amount of each claim. Brian from Dotterer Data Scrubbers develops the following distribution of claim amounts for 2018:

Amount of Claim	**Number of Claims**
0-5000	3000
5000-20,000	5000
20,000-100,000	5000
100,000+	900

If Alisa continued to model the data as a uniform distribution, what would Alisa's maximum likelihood estimate of U be based on the data developed by Brian.

Solution:

$$MLE\ for\ U = \frac{Total\ Number\ of\ Claims}{Total\ Claims\ Below\ Censoring\ Point}(Censoring\ Point) =$$

$$\frac{13{,}900}{13{,}000}(100{,}000) = 106{,}923.08$$

During 2019, Santiago modifies its auto insurance policy by implementing an upper limit per claim of 10,000. There remains no deductible and no coinsurance.

The first four claims received during 2019 resulted in **payments** of:

800 4000 10,000 and 10,000

This sample will be used for questions 8-9.

8. (6 points) Santiago hired Elizabeth's Exponential Consulting Firm. Liz assumes that the **total claim amount** is distributed as an exponential distribution with a mean of θ.

   Calculate the maximum likelihood estimate of θ.

   Solution:

$$\hat{\theta} = \frac{Total\ \ Amount\ Paid}{Number\ of\ Uncensored\ Observations} = \frac{800 + 4000 + 10,000 + 10,000}{2} = 12,400$$

During 2019, Santiago modifies its auto insurance policy by implementing an upper limit per claim of 10,000. There remains no deductible and no coinsurance.

The first four claims received during 2019 resulted in **payments** of:

800 4000 10,000 and 10,000

This sample will be used for questions 8-9.

9. (10 points) Santiago also hires The Miller Consulting Group to analyze the claims. One of the iconic consultants for The Miller Consulting Group is T'Fuego. He is assigned to complete this project and he assumes that the **total claim amount** is distributed as a Weibull distribution with parameters $\tau = 1$ and θ.

   T'Fuego uses the maximum likelihood estimate to estimate θ. What was T'Fuego's estimate of θ.

   Solution:

$$L(\theta) = f(800)\cdot f(4000)\cdot[1 - F(10,000)] = \left[\frac{\frac{800}{\theta}e^{-\frac{800}{\theta}}}{800}\right]\cdot\left[\frac{\frac{4000}{\theta}e^{-\frac{4000}{\theta}}}{4000}\right]\cdot\left[e^{-\frac{10,000}{\theta}}\right]^2 = \frac{e^{-\frac{24,800}{\theta}}}{\theta^2}$$

$$\ell(\theta) = \frac{-24,800}{\theta} - 2\ln(\theta)$$

$$\frac{d}{d\theta}\ell(\theta) = \frac{24,800}{\theta^2} - \frac{2}{\theta} = 0 ==> 24,800 = 2\theta ==> \hat{\theta} = 12,400$$

In 2020, Santiago has again revised the insurance policy sold by instituting a deductible **per claim** of 500. The Company also instituted an upper limit **per insurance policy** of 25,000.

10. (12 points) Santiago Substandard Auto Company hires Selbo Simulators to simulate **claim payments** for 2020.

    Jason, the consultant from Selbo Simulators, assumes that the number of claims is distributed as a binomial distribution with $m = 5$ and $q = 0.4$.

    Jason also assumes that the amount of each claim is distributed as an exponential distribution with θ = 20,000.

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

    The random numbers used in the simulation are:

    0.40 0.02 0.20 0.80 0.25 0.50 0.30 0.60 0.95 0.75 0.05

    Calculate the simulated aggregate claim payments for Jordyn and the simulated aggregate claim payments for Katie.

Solution:

$N ==>$	0	1	2	3	4	5
$\Pr(N=n)$	0.0778	0.2592	0.3456	0.2304	0.0768	0.0102
$F(N)$	0.0778	0.3370	0.6826	0.9130	0.9898	1.0000

For claim amount:

$$u^{**} = 1 - e^{-\frac{x^{**}}{20,000}} ==> 1 - u^{**} = e^{-\frac{x^{**}}{20,000}} ==> x^{**} = -20,000[\ln(1-u^{**})]$$

For Jordyn, a random number of 0.40 means that N = 2.

For Claim 1,

$X_1 = 20,000[\ln(1-0.02)] = 404.05$. *Total payment is 0 since the amount is less than deductible.*

For Claim 2,

$X_2 = 20,000[\ln(1-0.2)] = 4462.87$. *Total payment is* $4462.87 - 500 = 3962.87$

Total Payments for Jordyn = 0 + 3962.87 = 3962.87 which does not exceed the per policy limit of 25,000.

For Katie, a random number of 0.80 means that N=3

For Claim 1,

$X_1 = 20,000[\ln(1-0.25)] = 5753.64$. *Total payment is* $5753.64 - 500 = 5253.64$

For Claim 2,

$X_2 = 20,000[\ln(1-0.5)] = 13,862.94$. *Total payment is* $13,862.94 - 500 = 13,362.94$

For Claim 3,

$X_3 = 20,000[\ln(1-0.3)] = 7133.50$. *Total payment is* $7133.50 - 500 = 6633.50$

Total amount prior to applying the per policy limit is 5253.64 + 13,362.94 + 6633.50=25250.08

Therefore, the amount would be limited by the per policy limit of 25,000 so the amount that would have been paid for Katie is 25,000.

In 2020, Santiago has again revised the insurance policy sold by instituting a deductible **per claim** of 500. The Company also instituted an upper limit **per insurance policy** of 25,000.

11. (9 points) Evan is retained to estimate the number of claim payments during 2020 if Santiago sells 10,000 policies with each policy covering one auto. Alisa asks Evan to ignore the upper limit as this complicates that calculation and to just take into account the deductible per claim.

    Evans assumes the assumptions in the simulation, namely that the number of claims is distributed as a binomial distribution with $m=5$ and $q=0.4$ and the amount of claims is distributed as an exponential distribution with θ = 20,000.

    Calculate the expected number of claim payments.

    Solution:

$$N^P = vN^C$$

$$N^C = (n)(m)(q) = (10,000)(5)(0.4) = 20,000$$

$$v = \Pr(X > d) = \Pr(X > 500) = e^{\frac{-500}{20,000}} = 0.975309912$$

$$N^P = (20,000)(0.975309912) = 19,506.2$$

12. BONUS: What is Santiago Substandard Auto Company's biggest problem?

Solution:

TOO MANY CONSULTANTS.