ACSC/STAT 4703, Actuarial Models II

FALL 2023

Toby Kenney

Practice Midterm Examination

This Sample examination has more questions than the actual midterm, in order to cover a wider range of questions. Estimated times are provided after each question to help your preparation.

Here are some values of the Gamma distribution function with $\theta = 1$ that may be needed for this examination:

x	α	$F(x)$
245	255	0.2697208
$\left(\frac{7.5}{12}\right)^3$	$\frac{4}{3}$	0.1117140
$\left(\frac{9.5}{12}\right)^3$	$\frac{4}{3}$	0.2507382
2.5	1	0.917915
2.5	2	0.7127025
2.5	3	0.4561869
2.5	4	0.2424239
0.3542	3	0.005692012
5.6458	3	0.9202284

1. An insurer is assessing a model. Under the model, a certain statistic X should follow a gamma distribution with parameters θ_1 and $\alpha = 3$, and another statistic Y should follow a Weibull distribution with $\tau = 2$ and scale parameter θ_2. They compute the statistic $\frac{X}{\theta_1} + \left(\frac{Y}{\theta_2}\right)^2$. What is the probability that this statistic exceeds 7? [10 mins]

2. An insurer models claims as following a Pareto distribution with $\theta = 2000$ and α varying between individuals. They model $\alpha = 2 + 2A$ where A follows a gamma distribution with $\alpha = 3$ and $\theta = 1$. What is the VaR at the 0.95 level of the loss distribution for a random individual?

   (i) 1243
   (ii) 8445
   (iii) 9290
   (iv) 15919

3. An insurance company models aggregate losses following a Pareto distribution with $\alpha = 8$ and $\theta = 9600$ for $x < \$50,000$ and a Pareto distribution with $\alpha = 3$ for $x > \$50,000$. The probability that a loss exceeds \$50,000 is 0.00001. The scale parameters of the Pareto distributions are chosen so that the overall density function is continuous. What is the expected aggregate loss? [15 mins]

4. An insurance company has the following data on its policies:

Policy limit	Losses Limited to			
	20,000	50,000	100,000	500,000
20,000	1,400,000			
50,000	7,540,000	8,010,000		
100,000	22,600,000	24,100,000	28,700,000	
500,000	5,900,000	6,220,000	6,650,000	6,920,000

Use this data to calculate the ILF from \$20,000 to \$500,000 using

(a) The direct ILF estimate. [5 mins]

(b) The incremental method. [5 mins]

5. An insurance company charges a risk charge equal to the square of the average loss amount, divided by 100,000. It has the following data on a set of 1,200 claims from policies with limit \$1,000,000.

Losses Limited to	50,000	100,000	500,000	1,000,000
Total claimed	16,700,000	20,880,000	27,030,000	32,410,000

Calculate the ILF from \$100,000 to \$1,000,000. [10 mins]

6. An insurer calculates the ILF on the pure premium from \$1,000,000 to \$2,000,000 on a particular policy is 1.092. A reinsurer offers excess-of-loss reinsurance of \$1,000,000 over \$1,000,000 for a loading of 30%. The original insurer uses a loading of 20% on policies with limit \$1,000,000. If the insurer buys the excess-of-loss reinsurance, what is the loading on its premium for policies with a limit of \$2,000,000? [10 mins]

7. An insurer models a loss as following a Weibull distribution with $\tau = 4$ and $\theta = 100$. What are the parameters c_n and d_n that make the distribution of $\frac{M_n - d_n}{c_n}$ converge, where M_n are block maxima of a block of n samples, and what is the limiting distribution? [15 mins]

8. An insurer models aggregate daily losses with a distribution in the MDA of a Fréchet distribution with $\xi = 0.8$. In the past 100 years, there have been 21 years including daily losses exceeding \$500,000, and 9 years including daily losses exceeding \$1,000,000. What is the probability of a daily loss exceeding \$2,000,000 during the next year? [10 mins]

9. A reinsurer offers an excess-of-loss reinsurance contract on a portfolio with attachment point \$10,000,000 and no policy limit. The aggregate loss distribution is estimated to lie in the MDA of a Gumbel distribution. The reinsurer estimates that the probability of paying a claim is 0.08 and the expected payment on the contract is \$4,800. What is the expected square of the payment on the contract. [10 mins]

10. An insurer estimates that the time to completion of a claim comes from a distribution in the MDA of a GEV distribution with $\xi = -1.8$. The maximum time to completion is 20 years. They find that 2% of claims are incomplete after 5 years. Assuming the GPD approximation applies above 5 years, what percentage of claims are incomplete after 10 years?

11. An actuary is reviewing a sample of 483,230 observations that she believes comes from the MDA of a Fréchet distribution. She uses the Hill estimator to estimate ξ. She uses the $j = 481000$th order statistic as the threshold for the Hill estimator. Using this threshold, she gets the estimate $\xi = 1.45$. The order statistics near to this one are given in the following table:

j	$x_{(j)}$
480999	594303
481000	599045
481001	615667
481002	630520
481003	649402
481004	682034
481005	684215
481006	690144

What value of ξ would she have calculated if she had used $j = 481005$? [15 mins]

12. Loss amounts follow an exponential distribution with $\theta = 60,000$. The distribution of the number of losses is given in the following table:

Number of Losses	Probability
0	0.04
1	0.54
2	0.27
3	0.15

Assume all losses are independent and independent of the number of losses. The insurance company buys excess-of-loss reinsurance on the part of the loss above $150,000. Calculate the expected payment for this excess-of-loss reinsurance. [15 mins]

13. Claim frequency follows a negative binomial distribution with $r = 5$ and $\beta = 2.9$. Claim severity (in thousands) has the following distribution:

Severity	Probability
0	0
1	0.600
2	0.220
3	0.166

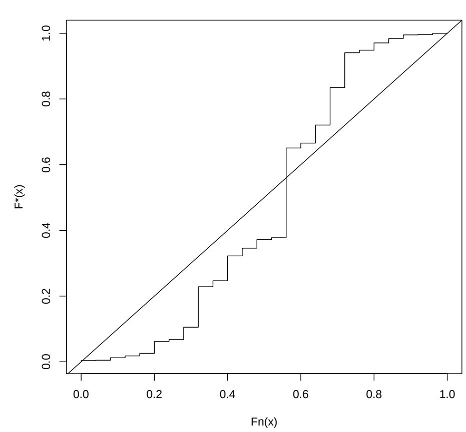
Use the recursive method to calculate the exact probability that aggregate claims are at least 4. [15 mins]

14. Using an arithmetic distribution ($h = 1$) to approximate a Weibull distribution with $\tau = 3$ and $\theta = 12$, calculate the probability that the value is between 3.5 and 8.5, for the approximation using:

(a) The method of rounding. [10 mins]

(b) The method of local moment matching, matching 1 moment on each interval. [$\Gamma\left(\frac{4}{3}\right) = 0.8929795$.] [15 mins]

15. An insurance company collects a sample of 25 past claims, and attempts to fit a Pareto distribution to the claims. Based on experience with other claims, the company believes that a Pareto distribution with $\alpha = 3.5$ and $\theta = 4,600$ may be appropriate to model these claims. It constructs the following p-p plot to compare the sample to this distribution:



(a) How many of the points in their sample were less than 1,200? [5 mins.]

(b) Which of the following statements best describes the fit of the Pareto distribution to the data: [5 mins.]

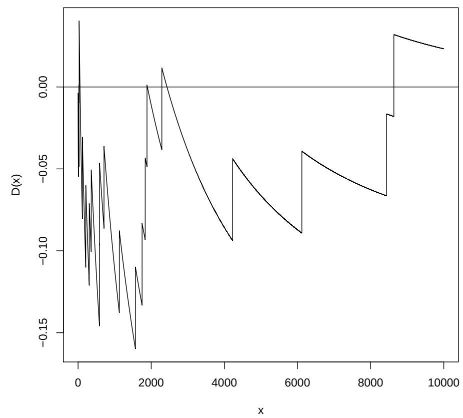
(i) The Pareto distribution assigns too much probability to high values and too little probability to low values.

(ii) The Pareto distribution assigns too much probability to low values and too little probability to high values.

(iii) The Pareto distribution assigns too much probability to tail values and too little probability to central values.

(iv) The Pareto distribution assigns too much probability to central values and too little probability to tail values.

16. An insurance company collects a sample of 20 claims. Based on previous experience, it believes these claims might follow a Weibull distribution with $\tau = 0.6$ and a known value of θ. To test this, it obtains a plot of $D(x)$.



(a) Which of the following is the value of θ used in the plot: [5 mins.]

(i) 800

(ii) 1,100

(iii) 2,200

(iv) 3,500

(b) Which of the following statements best describes the fit of the Weibull distribution to the data: [5 mins.]

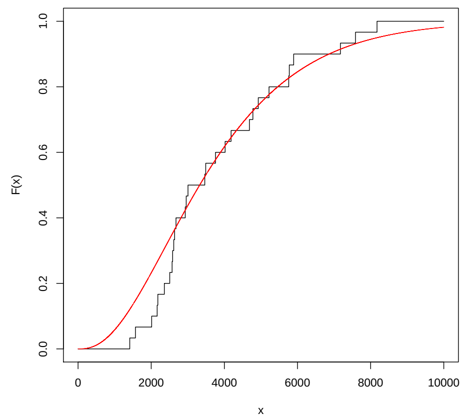
(i) The Weibull distribution assigns too much probability to high values and too little probability to low values.

(ii) The Weibull distribution assigns too much probability to low values and too little probability to high values.

(iii) The Weibull distribution assigns too much probability to tail values and too little probability to central values.

(iv) The Weibull distribution assigns too much probability to central values and too little probability to tail values.

17. An insurance company collects a sample of 30 claims. Based on previous experience, it believes these claims might follow a gamma distribution with $\alpha = 2.7$ and $\theta = 1400$. To test this, it compares plots of $F_n(x)$ and $F_*(x)$.



(a) Which of the following is the value of the Kolmogorov-Smirnov statistic for this model and this data [5 mins.]

(i) 0.0102432

(ii) 0.0450353

(iii) 0.0924252

(iv) 0.1678255

(b) Which of the following statements best describes the fit of the Gamma distribution to the data: [5 mins.]

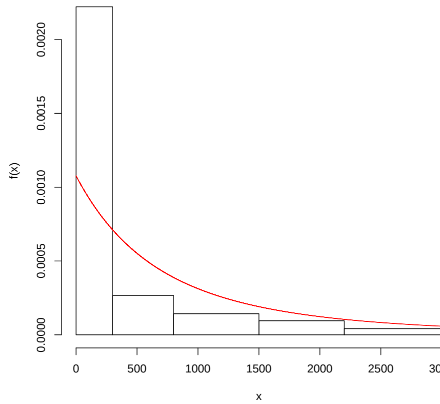
(i) The Gamma distribution assigns too much probability to high values and too little probability to low values.

(ii) The Gamma distribution assigns too much probability to low values and too little probability to high values.

(iii) The Gamma distribution assigns too much probability to tail values and too little probability to central values.

(iv) The Gamma distribution assigns too much probability to central values and too little probability to tail values.

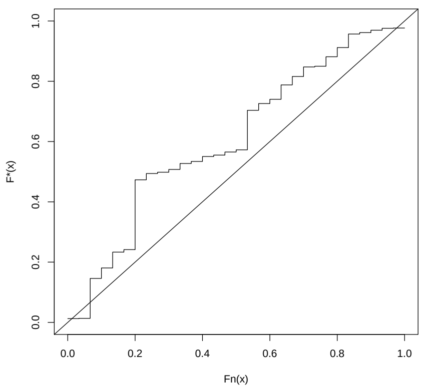
18. An insurance company collects a sample of 30 past claims, and attempts to fit a Pareto distribution to the claims. Based on experience with other claims, the company believes that a Pareto distribution with $\alpha = 2.8$ and $\theta = 2,600$ may be appropriate to model these claims. It compares the density functions in the following plot:



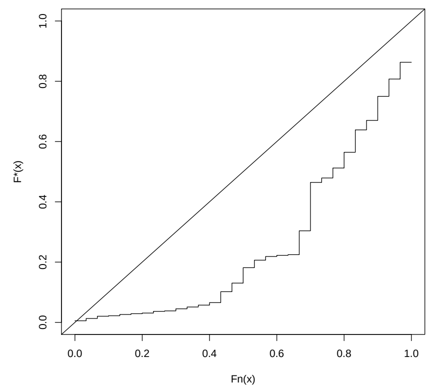
(a) How many data points in the sample were between 1500 and 3000? [5 mins.]

(b) Which of the following plots is the p-p plot for this data and model? [10 mins.]

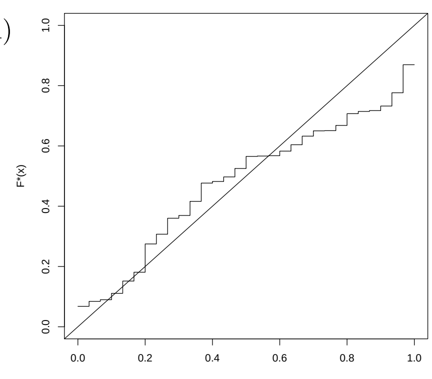
(i)



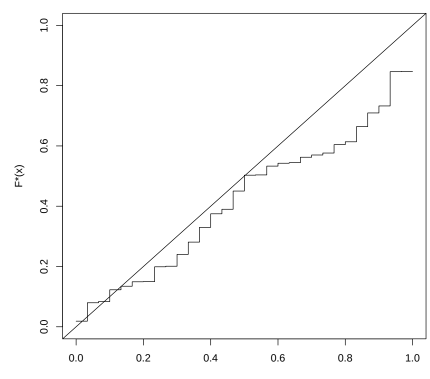
(ii)



(iii)



(iv)



[15 mins]

19.

20. An insurance company collects the following sample:

    `2.31 8.65 35.29 42.27 151.51 194.99 523.50 1262.01 1402.72 6063.74`

    They model this as following a Pareto distribution with $\alpha = 2$ and $\theta = 2000$. Calculate the Kolmogorov-Smirnov statistic for this model and this data. [10 mins.]

21. An insurance company collects the following sample:

    `0.27 2.03 9.89 16.96 28.38 236.46 268.36 453.19 633.26 718.68 1414.59 1588.19 2535.69 4937.93 5431.13`

    They model this as following a gamma distribution with $\alpha = 0.4$ and $\theta = 6000$. Calculate the Anderson-Darling statistic for this model and this data. [10 mins.]

    You are given the following values of the Gamma distribution used in the model:

x	$F(x)$	$\log(F(x))$	$\log(1-F(x))$
0.27	0.02056964	−3.8839392	−0.02078414
2.03	0.04609387	−3.0770753	−0.04719001
9.89	0.08680820	−2.4440542	−0.09080935
16.96	0.10767291	−2.2286572	−0.11392253
28.38	0.13222244	−2.0232696	−0.14181987
236.46	0.30572308	−1.1850755	−0.36488438
268.36	0.32111513	−1.1359556	−0.38730373
453.19	0.39258278	−0.9350079	−0.49853938
633.26	0.44506880	−0.8095264	−0.58891114
718.68	0.46633756	−0.7628455	−0.62799177
1414.59	0.59250242	−0.5234003	−0.89772028
1588.19	0.61583950	−0.4847689	−0.95669484
2535.69	0.71295893	−0.3383315	−1.24812996
4937.93	0.84646394	−0.1666877	−1.87381984
5431.13	0.86352967	−0.1467270	−1.99164807

22. An insurance company collects the following sample:

    `105.13 304.10 323.11 359.09 360.43 368.63 413.47 448.81 606.88 612.58 930.35 1002.37 1161.78 1205.25 5585.37`

    They want to decide whether this data is better modeled as following an inverse gamma distribution, or an inverse exponential distribution. They calculate that the MLEs for the inverse gamma distribution as $\alpha = 1.695545$ and $\theta = 705.7664$, and the MLE for the inverse exponential distribution as $\theta = 416.2476$. They also calculate, for this data that $\sum_{i=1}^{15} \log(x_i) = 95.31415$ and $\sum_{i=1}^{15} \frac{1}{x_i} = 0.03603625$, and that $\Gamma(1.695545) = 0.9078021$. You are given the following table of critical values for the chi-squared distribution at the 5% significance level. Indicate in your answer which critical value you are using. [15 mins.]

Degrees of Freedom	95% critical value
1	3.841459
2	5.991465
3	7.814728
4	9.487729
5	11.070498

23. An insurance company collects the following sample:

    ```
    0.1 0.2 0.3 2.1 16.8 28.4 45.7 53.5 74.2 99.5 159.3 183.5 206.3 273.9 461.9 482.9 1118.5
    1444.7 2084.3 3984.8
    ```

    They want to decide whether this data is better modeled as following an inverse exponential distribution or a Weibull distribution. They calculate that the MLE for the inverse exponential distribution is $\theta = 1.052901$, and the corresponding likelihood is -183.51. They also calculate that for the Weibull distribution, the MLE is $\tau = 0.48$, $\theta = 255.2235$. The log-likelihood is therefore -141.8325. Use AIC and BIC to determine which distribution is a better fit for the data. [5 mins.]

24. An insurance company collects the following data sample on claims data

Claim Amount	Number of Claims
Less than $5,000	1,026
$5,000–$10,000	850
$10,000–$20,000	1,182
$20,000–$50,000	942
More than $50,000	573

    Its previous experience suggests that the distribution should be modelled as following a Pareto distribution with $\alpha = 3$ and $\theta = 28,000$. Perform a chi-squared test to determine whether this distribution is a good fit for the data at the 95% level. [10 mins.]

    You may use the following critical values for the chi-squared distribution:

Degrees of Freedom	95% critical value
1	3.841459
2	5.991465
3	7.814728
4	9.487729
5	11.070498