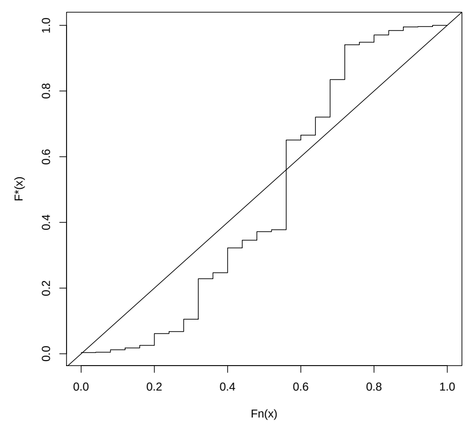
ACSC/STAT 4703, Actuarial Models II
Fall 2015
Toby Kenney
Sample Final Examination

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

For each question that asks you to simulate a small number of samples from a distribution, use the following simulated uniform values, starting from the first, and using as many numbers as needed for the question. Go back to the first value at the start of each part question.

```
0.58665797 0.12487271 0.87530540 0.49197147 0.55262301 0.14644543 0.89151074 0.46559276 0.42856173
0.63507522 0.78161985 0.69613284 0.37786683 0.51447243 0.48952100 0.28195163 0.62179048 0.66186936
0.42715830 0.70003263 0.59328856 0.97308150 0.14087141 0.08049598 0.98662077 0.91974635 0.56037580
0.07804151 0.48363702 0.33763780
```

1. 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?

(b) Which of the following statements best describes the fit of the Pareto distribution to the data:

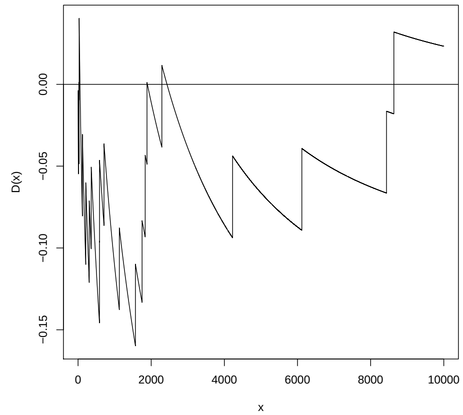
(i) The Pareto distribution assigns too much probability to high values and too little probability to low values.

(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.

2. 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:

(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:

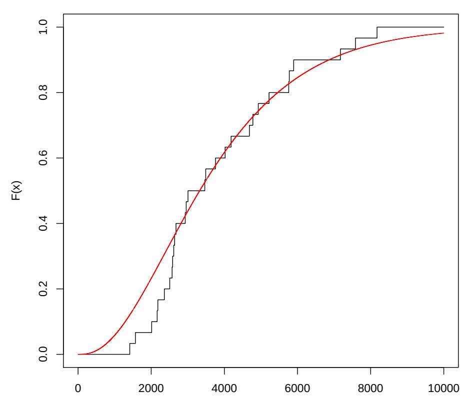
(i) The Weibull distribution assigns too much probability to high values and too little probability to low values.

(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.

3. 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

(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:

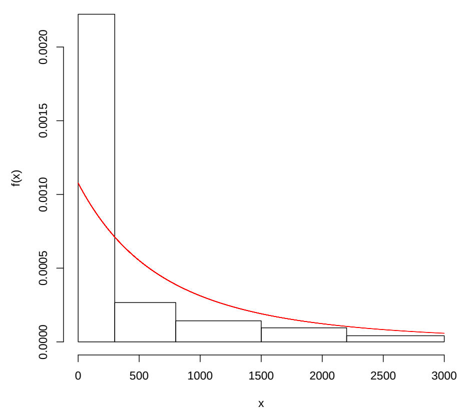
(i) The Gamma distribution assigns too much probability to high values and too little probability to low values.

(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.

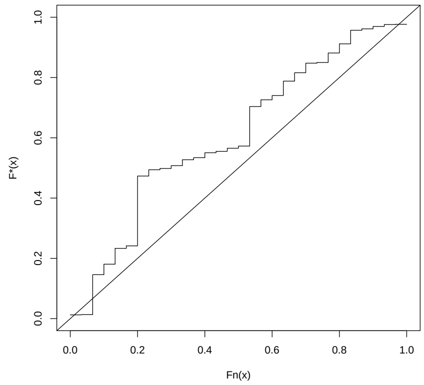
4. An insurance company collects a sample of 30 past claims, and attempts to fit a Pareto distribution to the claims. Based on experience with other claims, the company believes that a Pareto distribution with $\alpha = 2.8$ and $\theta = 2,600$ may be appropriate to model these claims. It compares the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



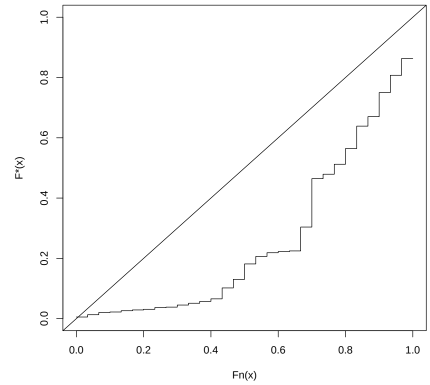
(a) How many data points in the sample were between 1500 and 3000?

(b) Which of the following plots is the p-p plot for this data and model?

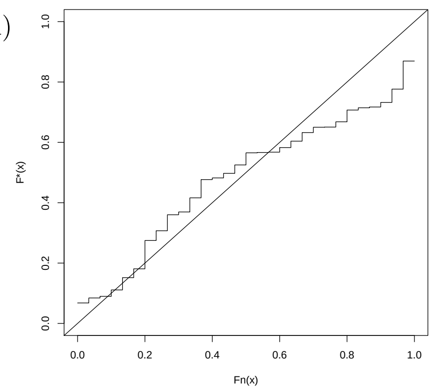
(i)



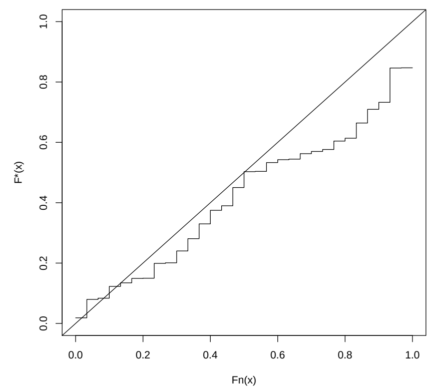
(ii)



(iii)



(iv)



5. 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.

6. 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.

   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

7. 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.

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

8. 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.

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

9. An insurance company sells home insurance. It estimates that the standard deviation of the aggregate annual claim is $5,326 and the mean is $1,804.

   (a) How many years history are needed for an individual or group to be assigned full credibility? (Use $r = 0.05$, $p = 0.95$.)

   (b) What is the Credibility premium, using limited fluctuation credibility, for an individual who has claimed a total of $42,381 in the past 19 years?

10. For a car insurance policy, the book premium for claim severity is $2,300. An individual has made 7 claims in the past 12 years, with average claim severity $1,074. Calculate the credibility estimate for claim severity for this individual using limited fluctuation credibility, if the standard for full credibility is:

    (a) 157 claims.

    (b) 284 years.

11. A worker's compensation insurance company classifies workplaces as "safe" or "hazardous". Claims from hazardous workplaces follow a Gamma distribution with $\alpha = 0.1021749$, $\theta = 1066798$ (mean $109,000 and standard deviation $341,000). Claims from safe workplaces follow a Gamma distribution with $\alpha = 0.01209244$, $\theta = 2646281$ (mean $32,000 and standard deviation $261,000). 94% of workplaces are classified as safe.

    (a) Calculate the expectation and variance of claim size for a claim from a randomly chosen workplace.

    (b) The last 2 claims from a particular workplace are $488,200 and $17,400. Calculate the expectation and variance for the next claim size from this workplace.

12. An insurance company sets the book pure premium for its home insurance at $791. The expected process variance is 6,362,000 and the variance of hypothetical means is 341,200. If an individual has no claims over the last 8 years, calculate the credibility premium for this individual's next year's insurance using the Bühlmann model.

13. An insurance company is reviewing the premium for an individual with the following past claim history:

Year	1	2	3	4	5
Exposure	0.2	1	1	0.4	0.8
Aggregate claims	0	$2,592	0	$147	$1,320

    The usual premium per unit of exposure is $2,700. The expected process variance is 123045 and the variance of hypothetical means is 36403 (both per unit of exposure). Calculate the credibility premium for this individual if she has 0.6 units of exposure in year 6.

14. An insurance company has 3 years of past history on a homeowner, denoted X_1, X_2, X_3. Because the individual moved house at the end of the second year, the third year has a different mean and variance, and is not as correlated with the other two years. It has the following

$$\mathbb{E}(X_1) = 1,322 \qquad \mathrm{Var}(X_1) = 226,000$$
$$\mathbb{E}(X_2) = 1,322 \qquad \mathrm{Var}(X_2) = 226,000$$
$$\mathbb{E}(X_3) = 4,081 \qquad \mathrm{Var}(X_3) = 1,108,000$$
$$\mathbb{E}(X_4) = 4,081 \qquad \mathrm{Var}(X_4) = 1,108,000$$
$$\mathrm{Cov}(X_1, X_2) = 214 \qquad \mathrm{Cov}(X_1, X_3) = 181$$
$$\mathrm{Cov}(X_2, X_3) = 181 \qquad \mathrm{Cov}(X_1, X_4) = 181$$
$$\mathrm{Cov}(X_2, X_4) = 181 \qquad \mathrm{Cov}(X_3, X_4) = 861$$

It uses a formula $\hat{X}_4 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3$ to calculate the credibility premium in the fourth year. Calculate the values of α_0, α_1, α_2 and α_3.

15. An insurance company has the following previous data on aggregate claims:

Policyholder	Year 1	Year 2	Year 3	Year 4	Mean	Variance
1	1,210	246	459	1,461	944.00	340158.00
2	0	0	0	0	0.00	0.00
3	0	2,185	0	0	548.25	1202312.25
4	809	0	0	1,725	633.50	674939.00
5	0	0	0	0	0.00	0.00

Calculate the Bühlmann credibility premium for policyholder 3 in Year 5.

16. An insurance company collects the following claim frequency data for 7,000 customers insured for the past 3 years:

No. of claims	Frequency
0	1,491
1	2,461
2	1,810
3	831
4	302
5	72
6	30
7	2
8	1
> 8	0

It assumes that the number of claims an individual makes in a year follows a Poisson distribution with parameter Λ, which may vary between individuals.

Find the credibility estimate for the expected number of claims per year for an individual who has made 4 claims in the past 3 years.

17. Use the method of inversion to simulate two random samples from a Pareto distribution with $\alpha = 4$, $\theta = 6,200$.

18. An insurance company classifies individuals into three classes, each with a different claim severity distribution, as shown in the following table:

Class	Probability	Severity Distribution	Parameters
1	0.20	Pareto	$\alpha = 4$, $\theta = 7,000$
2	0.35	Weibull	$\tau = 1.7$, $\theta = 800$
3	0.45	Inverse Weibull	$\tau = 2.8$, $\theta = 590$

Simulate 2 claim severities from 2 random individuals.

19. A pension plan has three types of exit with probabilities in the table below:

Exit Type	Probability
Retirement	0.65
Withdrawl	0.25
Death	0.10

Simulate the number of each type from a sample of 634 plan members. [You may use a normal approximation to the binomial distribution.]

20. Use a stochastic process method to simulate 2 samples from each of the following distributions:

(a) A Poisson distribution with $\lambda = 3$.

(b) A negative binomial distribution with $r = 7$ and $\beta = 0.52$.

21. Simulate 2 samples from a normal distribution with $\mu = 3$ and $\sigma = 7$ using

(a) A Box-Muller transformation.

(b) The polar method.

22. An insurance company is simulating its aggregate losses. It is attempting to estimate the probability that its aggregate losses exceed $1,000,000.

(a) How many aggregate losses does it need to simulate to ensure that there is a 99% probability that the estimated probability of exceeding $1,000,000 is within 0.001 of the true probability, regardless of the true probability?

(b) Suppose the true probability of aggregate losses exceeding $1,000,000 is 0.05. How many simulations does the company need to perform in order for the relative error in this estimated probability to be less than 1% with probability 0.95?

23. A reinsurance company is using a simulation to calculate the premium for a stop-loss insurance contract. It simulates 100,000 outcomes, and finds that the mean payment is $492,384, and the standard deviation of the payments is $2,643,000. It wants to calculate the net premium with a 99% chance that the relative error in its net premium is less than 1%. Assuming the mean and standard deviation are similar to the results it already has, how many more simulations does it need to perform to achieve this accuracy?

24. An insurance company is estimating its aggregate losses. It simulates 1000 claim frequencies, and finds a total of 749 claims. It therefore simulates 749 claim severities, and simulates the aggregate losses by adding the claim severities in groups corresponding to the simulated claim frequencies. The insurance company has a second line of insurance which also has the same severity distribution, but a different frequency distribution. It simulates 1000 new frequencies and gets a total of 749 claims again. It uses the same simulated claim severities to model aggregate losses for the second line of insurance. Based on these simulated values, it calculates a 95% confidence interval for the aggregate losses. Which of the following statements best describes this procedure? Explain your answer.

(i) The procedure is sound and should produce an accurate confidence interval.

(ii) The procedure is unsound and will produce a narrower confidence interval than it should (so the confidence interval will contain the true value less than 95% of the time).

(iii) The procedure is unsound and will produce a wider confidence interval than it should (so the confidence interval will contain the true value more than 95% of the time).

(iv) The procedure is unsound, and the confidence interval will be wider than it should in some cases and narrower than it should in others.