ACSC/STAT 4703, Actuarial Models II
FALL 2018
Toby Kenney
Sample Midterm Examination
Model Solutions

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 will 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

1. *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.

If the number of losses is n, then the aggregate loss follows a gamma distribution with $\alpha = n$ and $\theta = 60000$. The expected payment on the excess-of-loss insurance is therefore

$$\int_{150000}^{\infty}(x-150000)\frac{x^{n-1}e^{-\frac{x}{60000}}}{(n-1)!60000^n}\,dx$$
$$= \int_{150000}^{\infty}\frac{x^{n}e^{-\frac{x}{60000}}}{(n-1)!6000^n}\,dx - 150000\int_{150000}^{\infty}\frac{x^{n-1}e^{-\frac{x}{60000}}}{(n-1)!60000^n}\,dx$$
$$= \int_{2.5}^{\infty}\frac{60000nu^{n}e^{-u}}{n!}\,du - 150000\int_{2.5}^{\infty}\frac{u^{n-1}e^{-u}}{(n-1)!}\,du$$

This gives the following expected payments on the excess-of-loss reinsurance:

Number of Losses	Probability	Expected payment on excess-of-loss	product
0	0.04	0	0
1	0.54	$60000 \times 1 \times 0.2872975 - 150000 \times 0.0820850 = 4925.10$	2659.554
2	0.27	$60000 \times 2 \times 0.5438131 - 150000 \times 0.2872975 = 22162.95$	5983.996
3	0.15	$60000 \times 3 \times 0.7575761 - 150000 \times 0.5438131 = 54791.74$	8218.760

The total expected payment on the excess-of-loss reinsurance is therefore $2659.554 + 5983.996 + 8218.760 = \$16,862.31$.

2. *Aggregate payments have a compund distribution. The frequency distribution is negative binomial with $r = 4$ and $\beta = 12$. The severity distribution is a Gamma distribution with $\alpha = 8$ and $\theta = 3000$. Use a normal approximation to aggregate payments to estimate the probability that aggregate payments are more than \$2,000,000.*

   The frequency distribution has mean 48 and variance 624. The severity distribution has mean 24000 and variance 72000000.

   The mean of aggregate payments is therefore, $48 \times 24000 = 1152000$, and the variance is $624 \times 24000^2 + 48 \times 72000000 = 362880000000$, so the standard deviation is $\sqrt{362880000000} = 602395.2$. The probability of exceeding \$2,000,000 is therefore $1 - \Phi\left(\frac{2000000-1152000}{602395.2}\right) = 1 - Phi(1.407714) = 1 - 0.9203921 = 0.0796$.

3. *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.

   For the negative binomial distribution, we have $a = \frac{\beta}{1+\beta} = \frac{2.9}{3.9}$ and $b = \frac{(r-1)\beta}{1+\beta} = \frac{4\times 2.9}{3.9}$, so the recursive formula

$$f_S(x) = \frac{(p_1 - (a+b)p_0)f_X(x) + \sum_{i=1}^{x}\left(a + \frac{bi}{x}\right) f_X(i) f_S(x-i)}{1 - af_X(0)}$$

   becomes

$$f_S(x) = \sum_{i=1}^{x} \frac{2.9}{3.9}\left(1 + \frac{4i}{x}\right) f_X(i) f_S(x-i)$$

Since the severity distribution has no probability at zero, the only way for the aggregate loss to be zero is if the frequency is zero, the probability of which is $\left(\frac{1}{1+\beta}\right)^r = \frac{1}{3.9}^5 = 0.00110835$. We now use the recurrence:

$$f_S(1) = \frac{2.9}{3.9} \times 5 \times 0.600 \times 0.00110835 = 0.002472473$$
$$f_S(2) = \frac{2.9}{3.9} \times (3 \times 0.600 \times 0.002472473 + 5 \times 0.220 \times 0.00110835) = 0.004215883$$
$$f_S(3) = \frac{2.9}{3.9} \times \left(\frac{7}{3} \times 0.600 \times 0.004215883 + \frac{11}{3} \times 0.220 \times 0.002472473 + 5 \times 0.166 \times 0.00110835\right) = 0.00$$

The probability that the aggregate payments exceed 4 is therefore $1 - 0.00110835 - 0.002472473 - 0.004215883 - 0.006555954 = 0.9856473$.

4. *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.

   The method of rounding preserves this probability, since it assigns all values between 3.5 and 4.5 to 4, etc. Therefore this probability is $e^{-\left(\frac{3.5}{12}\right)^3} - e^{-\left(\frac{8.5}{12}\right)^3} = 0.2745978$.

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

   Using local moment matching, the probabilities of the intervals $[3.5, 5.5]$ and $[5.5, 7.5]$ are preserved, so the probability of these intervals is $e^{-\left(\frac{3.5}{12}\right)^3} - e^{-\left(\frac{7.5}{12}\right)^3} = 0.1921159$.

   For the interval $[7.5, 9.5]$, the probability of this interval is $e^{-\left(\frac{7.5}{12}\right)^3} - e^{-\left(\frac{9.5}{12}\right)^3} = 0.174517$, while the conditional mean times this probability is

$$\begin{aligned}
\int_{7.5}^{9.5} x\left(\frac{3x^2}{12^3}e^{-\left(\frac{x}{12}\right)^3}\right)dx &= \int_{\left(\frac{7.5}{12}\right)^3}^{\left(\frac{9.5}{12}\right)^3} 12\sqrt[3]{u}e^{-u}\,du \\
&= 12\int_{\left(\frac{7.5}{12}\right)^3}^{\left(\frac{9.5}{12}\right)^3} u^{\frac{1}{3}}e^{-u}\,du \\
&= 12\Gamma\left(\frac{4}{3}\right)(0.2507382 - 0.1117140) \\
&= 12 \times 0.8929795 \times (0.2507382 - 0.1117140) \\
&= 1.489749
\end{aligned}$$

We are now trying to solve for p_8 and p_9 such that

$$\begin{aligned}
p_8 + p_9 &= 0.174517 \\
8p_8 + 9p_9 &= 1.489749 \\
p_8 &= 9 \times 0.174517 - 1.489749 = 0.080904
\end{aligned}$$

So the probability of the interval $[3.5, 8.5]$ is therefore $0.1921159 + 0.080904 = 0.273020$.

5. *An insurance company has the following portfolio of auto insurance policies:*

Type of driver	*Number*	*Probability claim*	*mean of claim*	*standard deviation*
Good driver	*60*	*0.02*	*\$2,500*	*\$14,000*
Average driver	*140*	*0.06*	*\$3,800*	*\$19,200*
Bad driver	*50*	*0.13*	*\$7,000*	*\$22,600*

Calculate the cost of reinsuring losses above \$500,000, if the loading on the reinsurance premium is one standard deviation above the expected claim payment on the reinsurance policy, using a Pareto approximation for the aggregate losses on this portfolio.

The expected aggregate loss is $60 \times 0.02 \times 2500 + 1400 \times 0.06 \times 380 + 50 \times 0.13 \times 7000 = 80420$, and the variance is

$60 \times 0.02 \times 0.98 \times 2500^2 + 60 \times 0.02 \times 14000^2 + 140 \times 0.06 \times 0.94 \times 3800^2 + 140 \times 0.06 \times 19200^2 + 50 \times 0.13 \times 0.87 \times 7000^2 + 50 \times 0.13 \times 22600^2 = 7050179240$

We solve for the parameters of a Pareto distribution with these moments:

$$\begin{aligned}
\frac{\theta}{\alpha - 1} &= 80420 \\
\frac{\theta^2 \alpha}{(\alpha - 1)^2(\alpha - 2)} &= 7050179240 \\
\frac{\alpha - 2}{\alpha} &= \frac{80420^2}{7050179240} = 0.91733503218 \\
\alpha &= \frac{2}{1 - 0.91733503218} = 24.1940455884 \\
\theta &= 23.1940455884 \times 80420 = 1865265.14622
\end{aligned}$$

The expected payment on the excess-of-loss reinsurance for losses above $a = 5000000$ is

$$\int_a^\infty \left(\frac{\theta}{\theta+x}\right)^\alpha dx = \int_{a+\theta}^\infty \theta^\alpha u^{-\alpha}\,du = \left[-\frac{\theta^\alpha u^{1-\alpha}}{\alpha-1}\right]_{a+\theta}^\infty = \frac{\theta^\alpha(a+\theta)^{1-\alpha}}{\alpha-1} = 325.947338998$$

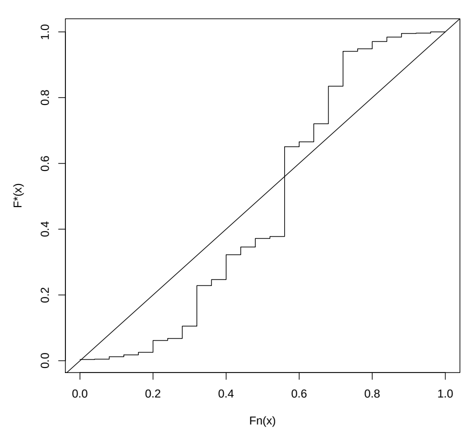
where we have used the substitution $u = x+\theta$.

The expected square of the payment is

$$\begin{aligned}\int_a^\infty 2(x-a)\left(\frac{\theta}{\theta+x}\right)^\alpha dx &= \int_{a+\theta}^\infty 2(u-\theta-a)\theta^\alpha u^{-\alpha}\,du = 2\left[-\frac{\theta^\alpha u^{2-\alpha}}{\alpha-2}\right]_{a+\theta}^\infty - 2\times 325.947338998(\theta+a\\ &= 2\frac{\theta^\alpha(a+\theta)^{2-\alpha}}{\alpha-2} - 651.894677996(\theta+a)\\ &= 69473758.38\end{aligned}$$

The variance is $69473758.38 - 325.947338998^2 = 69367516.7122$, so the mean plus one standard deviation is $325.947338998 + \sqrt{69367516.7122} = \$8,654.66$.

6. *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:*

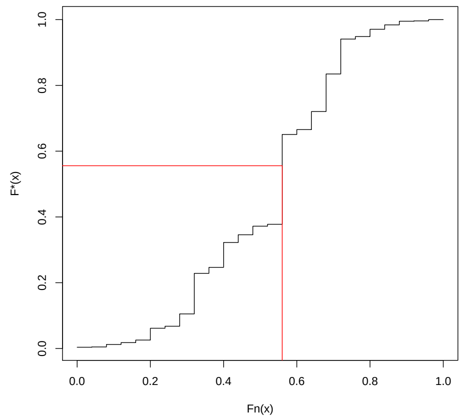


(a) How many of the points in their sample were less than 1,200? [5 mins.]

We have

$$F^*(1200) = 1 - \left(\frac{46}{58}\right)^{3.5} = 0.5557224$$

so we look for the point on the graph with $F^*(x) = 0.5557224$.



We see that the corresponding value of $F_n(x)$ is 0.56. (The values of $F_n(x)$ are in increments of 0.04, since there are 25 data points. The value corresponding to $F^*(x)$ is one increment before 0.6, so is 0.56). Thus the number of samples less than 1,200 is 14.

(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.

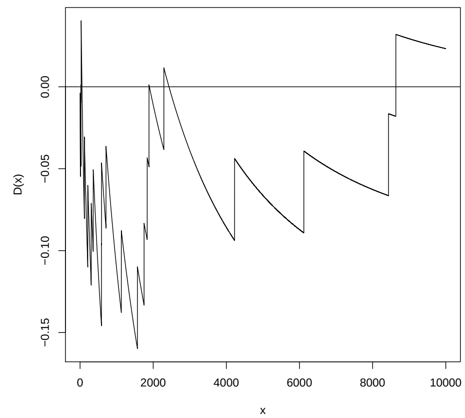
(ii) The Pareto distribution assigns too much probability to low values and too little probability to high values.

(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.

We see that there are 8 data points with $F^*(x) < 0.1$ approximately. The expected number is 2.5. There are 7 data points with $F^(x) > 0.9$. Again, the expected number is 2.5. The Pareto distribution has therefore underestimated the probabilities of these tail regions, and overestimated the probability of the region in between. Therefore, statement (iv) best describes the fit.

7. *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

The data points in the sample correspond to vertical line segments on the plot. We see for example, that there are 3 data points above 6000, so $F_{20}(6000) = \frac{17}{20} = 0.85$. Reading from the graph, we get that $D(6000) \approx -0.09$. This means $F^*(6000) = 0.85 - (-0.09) = 0.94$. This gives:

$$1 - e^{-\left(\frac{6000}{\theta}\right)^{0.6}} = 0.94$$

$$\left(\frac{6000}{\theta}\right)^{0.6} = -\log(0.06)$$

$$\frac{6000}{\theta} = (-\log(0.06))^{\frac{1}{0.6}}$$

$$\theta = \frac{6000}{(-\log(0.06))^{\frac{1}{0.6}}} = 1070.112$$

This is clearly closest to (ii), so (ii) is the value of θ used. (The difference between this answer and the 1,100 is because we only have limited accuracy reading the graph.)

[We can find the value of θ by reading off the value of $D(x)$ for any X on the graph. If it is difficult to count the number of vertical line segments, we could compare $D(x_1)$ and $D(x_2)$ for values of x_1 and x_2

with no vertical line segments in between. For example, we can read the value $D(4200) \approx -0.04$, which leads us to solve

$$F^*(6000) - F^*(4200) = 0.05$$

We can try the values given to see which is closer to the solution.]

(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.

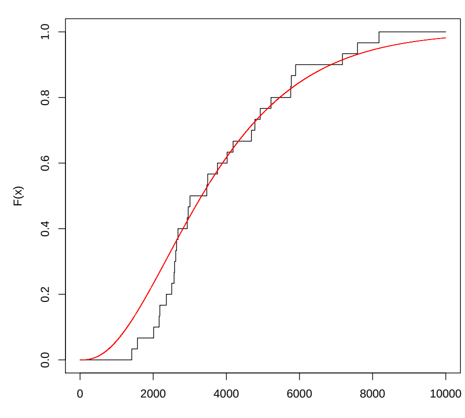
(ii) The Weibull distribution assigns too much probability to low values and too little probability to high values.

(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.

Recall that $D(x) = F_n(x) - F^*(x)$, so if $D(x) < 0$, we have $F^*(x) > F_n(x)$, while if $D(x) > 0$, we have $F^*(x) < F_n(x)$. On the graph shown, we have that $D(x)$ is nearly always negative for the range of the data. [Technically, it is positive for all values larger than the data sample, but this always happens, because for the largest value of the data sample, we have $F_n(x) = 1 > F^*(x)$.] This means that $F^*(x) > F_n(x)$ for most x in the range. This means that the Weibull distribution assigns more probability to smaller values of x, and less probability to larger values of x, which is statement (ii).

8. *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.]

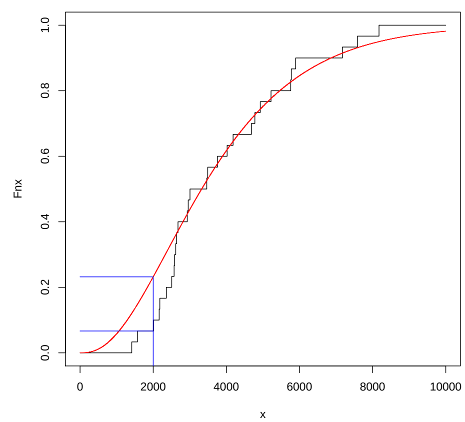
(i) 0.0102432

(ii) 0.0450353

(iii) 0.0924252

(iv) 0.1678255

The Kolmogorov-Smirnov test statistic is the maximum value of the absolute difference between the empirical and model distribution functions, that is $|F_n(x) - F^*(x)|$. On the graph, we see this happens at around 2000, and read the values from the graph:



We read $F_n(x) = 0.066667$ (we know the possible values of $F_n(x)$, since we know there are 30 data points), and $F^*(x) = 0.23$ (the actual value is 0.2318889.) The difference is therefore about 1.6, so (iv) is the correct answer.

(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.

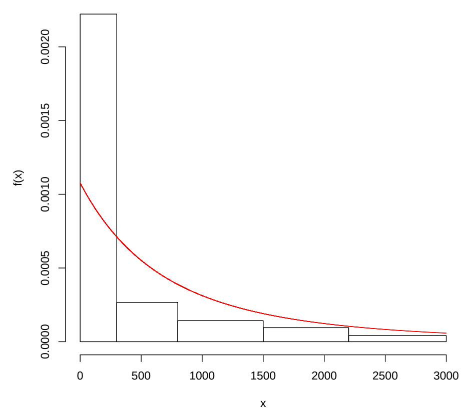
(ii) The Gamma distribution assigns too much probability to low values and too little probability to high values.

(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.

From the graph, we see that $F*(x)$ is too large for small values less than about 2500, and about correct for larger values. This means that the gamma model assigns too little probability in the range 0–2,000 and too much in the range 2,000–2,500. We also see that $F^*(x)$ is slightly too low at values above 6,000. This means that the gamma distribution assigns too little probability to values larger than 6,000. This means that (iv) is probably the best description of the fit. However, a case could be made for (ii) being a good description, since the difference between $F^*(x)$ and $F_n(x)$ for $x > 6000$ is very small.

9. *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:*

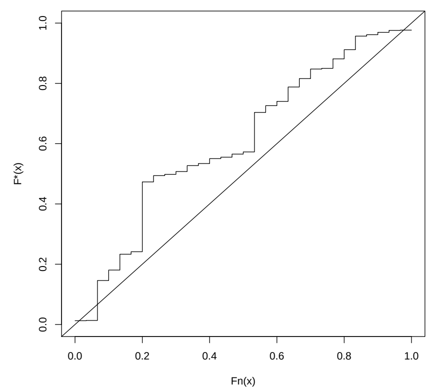


(a) How many data points in the sample were between 1500 and 3000? [5 mins.]

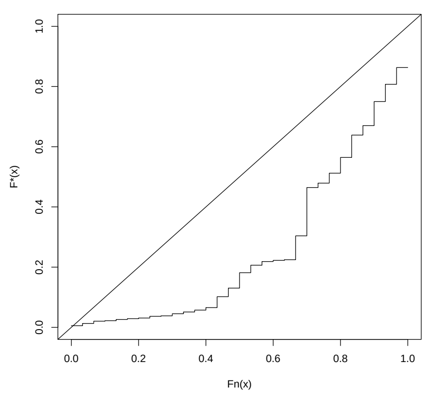
We are asking how many data points are in the last two bars. The height of the fourth bar (from 1,500–2,200) is about 0.0001, and the height of the fifth bar (from 2,200–3,000) is about 0.00005, so the areas of these two bars are $700 \times 0.0001 = 0.07$ and $800 \times 0.00005 = 0.04$ respectively. Since there are 30 claims in the sample, these correspond to 2 data points and 1 data point respectively, (which would give accurate heights of 0.00009524 and 0.00004167 respectively). Therefore, the number of data points between 1,500 and 3,000 is 3.

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

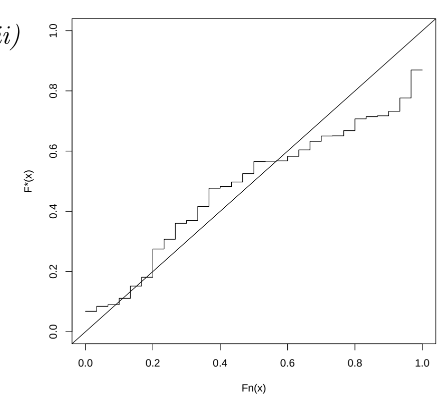
(i)


F*(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0

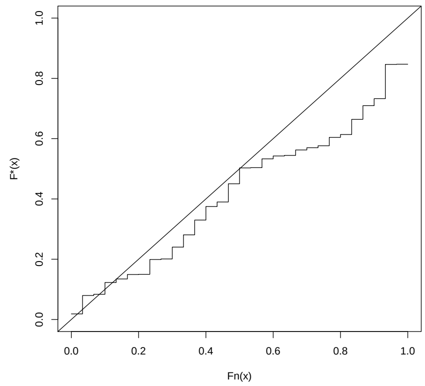
(ii)


F*(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0

(iii)


F*(x)
Fn(x)
0.0
0.2
0.4
0.6
0.8
1.0

(iv)



From the histogram, we see that the model assigns too little probability to small values less than 300, and too much probability to values more than 500. The p-p plot should therefore have slope less than 1 for the first part, then slope more than 1. We would expect $F_n(x) > F^*(x)$ for all x, so the p-p plot should be entirely below the line $y = x$ (it is in theory possible there could be some small values with $F*(x) > F_n(x)$, since the histogram only shows grouped data, so it is possible for example that all samples in the range 0–300 actually fell in the range 200–300). It seems that the largest difference between $F_n(x)$ and $F^*(x)$ should happen at around $x = 500$, and it looks like the area of the bar 0–300 on the histogram is approximately equal to the combined area of the other 3 bars. More accurately, it looks like the height of this bar is about 0.0022, and the width is about 300, so the area is about 0.66, so the largest difference between $F_n(x)$ and $F^*(x)$ should occur at about $F_n(x) = 0.66$. Also, after the first bar, the model is overestimating the probability density, which means that after this point, the slope of the p-p plot should be more than 1.

Looking at the options, plots (i) and (iii) are above the $y = x$ line for some values of x. Plot (iv) is close to the line for values less than $F_n(x) = 0.5$, and does not deviate so much from the line, and its furthest point from the line is around $F_n(x) = 0.9$, so it is not correct. Therefore, plot (ii) is the correct plot.

10. *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.]

x	$F*(x)$	$\|D(x^-)\|$	$\|D(x^+)\|$	
2.31	0.002306004	0.002306004	0.09769400	0.09769400
8.65	0.008594205	0.091405795	0.19140580	0.19140580
35.29	0.034377462	0.165622538	0.26562254	0.26562254
42.27	0.040966725	0.259033275	0.35903327	0.35903327
151.51	0.135881599	0.264118401	0.36411840	0.36411840
194.99	0.169776735	0.330223265	0.43022327	0.43022327
523.50	0.371864450	0.228135550	0.32813555	0.32813555
1262.01	0.624085208	0.075914792	0.17591479	0.17591479
1402.72	0.654532208	0.145467792	0.24546779	0.24546779
6063.74	0.938484160	0.038484160	0.06151584	0.06151584

So the Kolmogorov-Smirnov statistic is 0.4302.

11. *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

The Anderson-Darling statistic for complete data with no truncation or censorship can be calculated as

$$A^2 = -n + n\sum_{j=0}^{k-1}(1 - F_n(y_j))^2\left(\log(1 - F^*(y_j)) - \log(1 - F^*(y_{j+1}))\right) + n\sum_{j=1}^{k}(F_n(y_j))^2\left(\log(F^*(y_{j+1})) - \mathrm{l}\right.$$

We compute the terms in the following table:

j	y_j	$n(1-F_n(y_j))^2\left(\log(1-F^*(y_j))-\log(1-F^*(y_{j+1}))\right)$	$n(F_n(y_j))^2\left(\log(F^*(y_{j+\right.$
0	0.00	0.311762095	
1	0.27	0.345036701333	0.0537909266667
2	2.03	0.491444564001	0.1688056266667
3	9.89	0.221886528	0.1292382000000
4	16.96	0.225038542667	0.2190801066680
5	28.38	1.48709673333	1.3969901666700
6	236.46	0.12106449	0.1178877600000
7	268.36	0.474605439999	0.6564291533340
8	453.19	0.295214416001	0.5353877333330
9	633.26	0.093793512	0.2520768600000
10	718.68	0.449547516666	1.5963013333400
11	1414.59	0.0629061973335	0.3116266266660
12	1588.19	0.174861072	1.4057990400000
13	2535.69	0.166850634666	1.9338534800000
14	4937.93	0.00785521533335	0.2608198133330
15	5431.13		2.20090500000001.9172
total	4.92896366333	11.2389918267	

This gives $A^2 = 4.928964 + 11.2389918267 - 15 = 1.1679558267$.

12. *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*

For the inverse gamma distribution, the log-likelihood of the data point x is

$$\log\left(\frac{705.7664^{1.695545}e^{-\frac{705.7664}{x}}}{x^{2.695545}\Gamma(1.695545)}\right) = 1.695545\log(705.7664) - \log(\Gamma(1.695545)) - 2.695545\log(x) - \frac{705.76}{x}$$

$$= 11.21829 - 2.695545\log(x) - \frac{705.7664}{x}$$

The total log-likelihood of the data is therefore

$$11.21829 \times 15 - 2.695545\,(\log(105.13) + \log(304.10) + \log(323.11) + \log(359.09) + \log(360.43) + \log(368.$$
$$\log(448.81) + \log(606.88) + \log(612.58) + \log(930.35) + \log(1002.37) + \log(1161.78) + \log(1205.$$
$$-705.7664\left(\frac{1}{105.13} + \frac{1}{304.10} + \frac{1}{323.11} + \frac{1}{359.09} + \frac{1}{360.43} + \frac{1}{368.63} + \frac{1}{413.47} + \frac{1}{448.81} + \frac{1}{606.88} + \right.$$
$$\frac{1}{1002.37} + \frac{1}{1161.78} + \frac{}{1}$$

For the inverse exponential, the log-likelihood of the data point x is

$$\log\left(\frac{416.2476}{x^2}e^{-\frac{416.2476}{x}}\right) = 6.03128 - 2\log(x) - \frac{416.2476}{x}$$

The log-likelihood of the data is therefore

$$6.03128 \times 15 - 2\,(\log(105.13) + \log(304.10) + \log(323.11) + \log(359.09) + \log(360.43) + \log(368.63) +$$
$$\log(448.81) + \log(606.88) + \log(612.58) + \log(930.35) + \log(1002.37) + \log(1161.78) + \log(1205.25) +$$
$$-416.2476\left(\frac{1}{105.13} + \frac{1}{304.10} + \frac{1}{323.11} + \frac{1}{359.09} + \frac{1}{360.43} + \frac{1}{368.63} + \frac{1}{413.47} + \frac{1}{448.81} + \frac{1}{606.88} + \frac{1}{612.}\right.$$
$$\frac{1}{1002.37} + \frac{1}{1161.78} + \frac{1}{1205.2}$$

The likelihood ratio statistic is therefore $2(-114.0824 - (-115.1591)) = 2.1534$. This should be compared to the chi-square distribution with one degree of freedom (since the inverse gamma has 2 degrees of freedom, and the inverse exponential has 1). The critical value for this is 3.841459, so the statistic is not significant. This means there is not sufficient evidence that the inverse gamma distribution fits the data better.

13. *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.]*

The AIC is $l(x) - 2p$, while the BIC is $l(x) - \frac{p}{2}\log(n)$. For the inverse exponential distribution, we have $p = 1$, while for the Weibull distribution, we have $p = 2$. For this data set, we have $n = 20$, so the AIC and BIC are:

Model	AIC	BIC
Inverse Exponential	$-183.51 - 2 \times 1 = -185.51$	$-183.51 - \frac{1}{2}\log(20) = -185.007866137$
Weibull	$-141.8325 - 2 \times 2 = -145.8325$	$-141.8325 - \frac{2}{2}\ln(20) = -144.828232274$

14. *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*

The expected frequencies of each interval are:

$$4573\left(1-\left(\frac{28}{33}\right)^3\right) = 1779.598$$
$$4573\left(\left(\frac{28}{33}\right)^3-\left(\frac{28}{38}\right)^3\right) = 963.9355$$
$$4573\left(\left(\frac{28}{38}\right)^3-\left(\frac{28}{48}\right)^3\right) = 921.7474$$
$$4573\left(\left(\frac{28}{48}\right)^3-\left(\frac{28}{78}\right)^3\right) = 696.1798$$
$$4573\left(\frac{28}{78}\right)^3 = 211.5395$$

Therefore, the chi-squared statistic is

$$\frac{(1026-1779.598)^2}{1779.598}+\frac{(850-963.9355)^2}{963.9355}+\frac{(1182-921.7474)^2}{921.7474}+\frac{(942-696.1798)^2}{696.1798}+\frac{(573-211.5395)^2}{211.5395} = 1$$

Since the parameters are not estimated the number of degrees of freedom is $5-1=4$, so the critical value is 9.487729. The null hypothesis is rejected. The data do not fit the model well.

15. *A homeowner's house is valued at \$560,000. However, the home is insured only to a value of \$360,000. The insurer requires 80% coverage for full insurance. The home sustains \$6,000 of fire damage. The deductible is \$5,000, decreasing linearly to zero for losses of \$8,000. How much does the insurer reimburse?*

    The insurer pays $\frac{360000}{560000\times 0.8} = 80.3571428571\%$ of the costs. For a loss of \$6,000, the deductible is $5000\times\frac{2}{3}$, so the insurer pays $\left(6000-\frac{10000}{3}\right)\times 0.803571428571 = \$2,142.86$.

16. *An auto insurance company uses an expected loss ratio of 0.81. In accident year 2014, the earned premiums were \$1,420,000. In 2014, the insurance company made a total of \$189,300 in loss payments for accident year 2014, a total of \$152,500 in 2015, and a total of \$239,600 in 2016. What loss reserves should the company hold for this accident year at the end of 2016.*

    The total loss payments made are \$581,400. The expected total payments are $1420000\times 0.81 = \$1,150,200$. Therefore the reserve should be $1150200-581400 = \$568,800$.

17. *The following table shows the cumulative losses (in thousands) on claims from one line of business of an insurance company over the past 6 years.*

	Development year					
Accident year	*0*	*1*	*2*	*3*	*4*	*5*
2011	*751*	*1,022*	*1,448*	*1,133*	*1,473*	*1,493*
2012	*1,337*	*1,297*	*1,460*	*1,537*	*1,679*	
2013	*1,250*	*1,624*	*1,815*	*1,860*		
2014	*1,325*	*1,512*	*1,685*			
2015	*1,471*	*1,536*				
2016	*2,036*					

Using the average for calculating loss development factors, esimate the total reserve needed for payments to be made in 2018 using.

We calculate the following loss development factors:

Development year	Loss Development Factor
1/0	$\frac{1}{5}\left(\frac{1022}{751}+\frac{1297}{1337}+\frac{1624}{1250}+\frac{1512}{1325}+\frac{1536}{1471}\right)=1.16309083475$
2/1	$\frac{1}{4}\left(\frac{1448}{1022}+\frac{1460}{1297}+\frac{1815}{1624}+\frac{1685}{1512}\right)=1.19363330156$
3/2	$\frac{1}{3}\left(\frac{1133}{1448}+\frac{1537}{1460}+\frac{1860}{1815}\right)=0.95333055933$
4/3	$\frac{1}{2}\left(\frac{1473}{1133}+\frac{1679}{1537}\right)=1.19623801482$
5/4	$\frac{1493}{1473}=1.01357773252$

(a) The loss development triangle method

Using the loss development triangle method, the cumulative payments up to 2017 are

$$1679\ast1.01357773252+1860\ast1.19623801482+1685\ast0.95333055933+1536\ast1.19363330156+2036\ast1.16309083$$

The cumulative payments up to 2018 are

$$1679\ast1.01357773252+1860\ast1.19623801482\ast1.01357773252+1685\ast0.95333055933\ast1.19623801482+1536\ast1.$$

The payments to be made in 2018 are therefore $10453.0443718-9734.63540369=\$718,408.97$.

(b) The Bornhuetter-Fergusson method. The expected loss ratio is 0.76 and the earned premiums in each year are given in the following table:

Year	*Earned Premiums (000's)*
2011	*1943*
2012	*2430*
2013	*2623*
2014	*2804*
2015	*3356*
2016	*3673*

Under the Bornhuetter-Fergusson method, the proportion of total payments in each year is given by:

Development year	Proportion of total payments
0	$\frac{1}{1.01357773252\times1.19623801482\times0.95333055933\times1.19363330156\times1.16309083475} = 0.623156693912$
1	$\frac{1}{1.01357773252\times1.19623801482\times0.95333055933\times1.19363330156} - \frac{}{1.01357773252\times1.19623801482\times0.9}$
2	$\frac{1}{1.01357773252\times1.19623801482\times0.95333055933} - \frac{1}{1.01357773252\times1.19623801482\times0.95333055933\times1.1}$
3	$\frac{1}{1.01357773252\times1.19623801482} - \frac{1}{1.01357773252\times1.19623801482\times0.95333055933} = -0.04037517$
4	$\frac{1}{1.01357773252} - \frac{1}{1.01357773252\times1.19623801482} = 0.161848426438$
5	$1 - \frac{1}{1.01357773252} = 0.013395847289$

This gives us:

Year	Expected total losses	Expected losses in 2018
2013	1993.48	$1993.48 \times 0.013395847289 = 26.7043536537$
2014	2131.04	$2131.04 \times 0.161848426438 = 344.905470676$
2015	2550.56	$2550.56 \times -0.040375175282 = -102.979307067$
2016	2791.48	$2791.48 \times 0.140343062249 = 391.764851407$

The total reserves needed for payments in 2018 are therefore $26.7043536537 + 344.905470676 - 102.979307067 + 391.764851407 = \$660,395.37$.

18. *An actuary is reviewing the following loss development triangles:*

No. of closed claims

Acc. Year	*Development Year 0*	*1*	*2*	*3*	*Ult.*
2013	*482*	*481*	*579*	*636*	*660*
2014	*672*	*677*	*786*		*802*
2015	*657*	*734*			*823*
2016	*745*				*963*

Total paid losses on closed claims (000's)

Acc. Year	*Development Year 0*	*1*	*2*	*3*
2013	*1176*	*1163*	*1284*	*1372*
2014	*1130*	*1356*	*1292*	
2015	*1409*	*1507*		
2016	*2262*			

(a) Calculate tables of percentage of claims closed and cumulative average losses.

Percentage of closed claims

Acc. Year	Development Year 0	1	2	3	Ult.
2013	73.03	72.88	87.73	96.36	
2014	83.79	84.41	98.00		
2015	79.83	89.19			
2016	77.36				

Average paid losses per claim on closed claims (000's)

Acc. Year	Development Year 0	1	2	3
2013	2,440	2,418	2,218	2,157
2014	1,682	2,003	1,644	
2015	2,145	2,053		
2016	3,036			

(b) Adjust the total paid losses to use the current disposal rate.

Acc. Year	Development Year 0	1	2	3
2013	1246	1423	1434	1372
2014	1043	1433	1292	
2015	1365	1507		
2016	2262			