ACSC/STAT 4703, Actuarial Models II

FALL 2022
Toby Kenney

Homework Sheet 2

Model Solutions

Basic Questions

1. *An insurance company has the following portfolio of liability insurance policies:*

Type of product	*Number*	*Probability of claim*	*mean claim (millions)*	*standard deviation (millions)*
Electrical	*2,800*	*0.00125*	*\$1.3*	*\$33.1*
Health	*4,300*	*0.00374*	*\$2.2*	*\$21.6*
Other	*3,700*	*0.00072*	*\$0.8*	*\$24.5*

They model aggregate losses using a Pareto distribution. Calculate the cost of reinsuring losses above \$10,000,000, if there is a 30% loading on the reinsurance premium.

We calculate the mean and variance of the aggregate loss:

Type of product	$\mathbb{E}(N)$	Var(N) of claim	mean aggregate loss (millions)	var aggregate loss (trillions)
Electrical	3.500	3.495625	4.55	3840.542606253840.4760625
Health	16.082	16.02185332	35.3804	7580.763690077578.14370069
Other	2.664	2.66208192	2.1312	1600.769732431600.75868429
Total			42.0616	13022.0760287

Using a Pareto approximation, the method of moments gives the following parameters (with θ in millions):

$$\frac{\theta}{\alpha-1} = 42.0616$$
$$\frac{\theta^2\alpha}{(\alpha-1)^2(\alpha-2)} = 13022.076028713019.3784475$$
$$\frac{\alpha-2}{\alpha} = \frac{42.0616^2}{13022.0760287} = 0.135859919007$$
$$\alpha = \frac{2}{1-0.135859919007} = 2.31443957292$$
$$\theta = 42.0616 \times 1.31443957292 = 55.2874315403$$

The expected reinsurance payment in millions is therefore

$$\int_{10}^{\infty}\left(\frac{55.2874315403}{55.2874315403+x}\right)^{2.31443957292}dx = \int_{10+55.2874315403}^{\infty} 55.2874315403^{2.31443957292}u^{-2.31443957292}\,du$$
$$= 55.2874315403^{2.31443957292}\left[-\frac{u^{-1.31443957292}}{1.31443957292}\right]_{65.2874315403}^{\infty}$$
$$= \frac{55.2874315403^{2.31443957292}}{1.31443957292(65.2874315403)^{1.31443957292}}$$
$$= 33.8048557108$$

With a 30% loading The premium is therefore $33,804,855.7108 \times 1.3 =$ \$43,946,312.42.

2. *An insurance company is modelling claim data as following a Weibull distribution with $\tau = 0.4$. It collects the following sample of claims:*

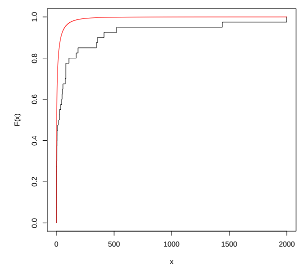
   0.1 0.2 0.2 0.4 0.4 0.8 0.8 1.1 1.5 1.5 1.8 2.5 3.2 3.3 3.4 4.1 4.8 5.1 10.7 20.1 25.3 26.1 36.2 45.8 48.7 50.5 56.3 75.6 81.5 81.5 81.6 108.2 170.8 186.9 345.9 356.0 412.7 523.1 1439.0 1998.8

   ```
   X<- c( 0.1, 0.2, 0.2, 0.4, 0.4, 0.8, 0.8, 1.1, 1.5, 1.5
   1.8, 2.5, 3.2, 3.3, 3.4, 4.1, 4.8, 5.1, 10.7, 20.1
   25.3, 26.1, 36.2, 45.8, 48.7, 50.5, 56.3, 75.6, 81.5, 81.5
   81.6,108.2,170.8,186.9,345.9,356.0,412.7,523.1,1439.0,1998.8)
   ```

   The MLE for θ is 4.657714. Graphically compare this empirical distribution with the best Weibull distribution with $\tau = 0.4$. Include the following plots:

   (a) Comparisons of $F(x)$ and $F^*(x)$

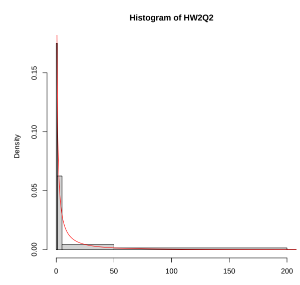
```
x<-(0:100000)/50 # The sample is slightly heavy-tailled. We could restrict the range to e
theta<-4.657714
tau<-0.4
FnX<-rowMeans(x%*%t(rep(1,40))>rep(1,100001)%*%t(HW2Q2))
plot(x,FnX,type='l',xlab="x",ylab="F(x)")
FX<-1-exp(-(x/theta)^tau)
points(x,FX,type='l',col="red")
```



(b) Comparisons of $f(x)$ and $f^*(x)$

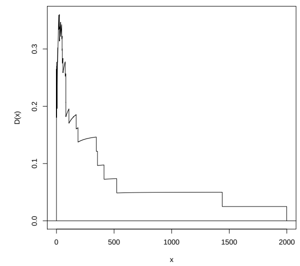
```
fx<-tau*x^(tau-1)/theta^tau*exp(-(x/theta)^tau)

hist(HW2Q2, probability=TRUE, breaks=c(0,1,5,50,200,2000), xlim=c(0,200))
  #These breaks produce a fairly smooth curve. Other choices are possible.
# I've limited the plot to 0-200 to better show the results.
points(x,fx,type='l',col="red")
```



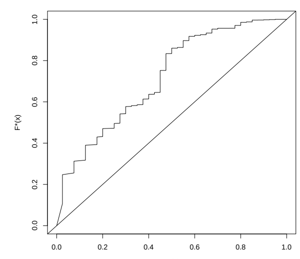
(c) A plot of $D(x)$ against x.

```
plot(x,FX-FnX,type='l',xlab="x",ylab="D(x)")
abline(h=0)
```



(d) *A p-p plot of $F(x)$ against $F^*(x)$.*

```
plot(FnX,FX, type='l', xlab=expression(F[n](x)), ylab="F*(x)", ylim=c(0,1))
abline(0,1)
```



3. *For the data in Question 2, calculate the following test statistics for the goodness of fit of the Weibull distribution with $\tau = 0.4$ and $\theta = 4.657714$:*

   (a) *The Kolmogorov-Smirnov test.*

X	$F_n(X^-)$	$F_n(X^+)$	$F^*(X)$	$-D(X^-)$	$D(X^+)$
0.1	0.000	0.025	0.1935753	0.19357535	−0.1781425
0.2	0.025	0.050	0.2471470	0.22214699	−0.1971470
0.2	0.050	0.075	0.2471470	0.19714699	−0.1721470
0.4	0.075	0.100	0.3124281	0.23742810	−0.2124281
0.4	0.100	0.125	0.3124281	0.21242810	−0.1874281
0.8	0.125	0.150	0.3899858	0.26498576	−0.2399858
0.8	0.150	0.175	0.3899858	0.23998576	−0.2149858
1.1	0.175	0.200	0.4296015	0.25460153	−0.2296015
1.5	0.200	0.225	0.4703696	0.27036958	−0.2453696
1.5	0.225	0.250	0.4703696	0.24536958	−0.2203696
1.8	0.250	0.275	0.4952336	0.24523360	−0.2202336
2.5	0.275	0.300	0.5414394	0.26643944	−0.2414394
3.2	0.300	0.325	0.5770830	0.27708296	−0.2520830
3.3	0.325	0.350	0.5815665	0.25656649	−0.2315665
3.4	0.350	0.375	0.5859230	0.23592296	−0.2109230
4.1	0.375	0.400	0.6133612	0.23836118	−0.2133612
4.8	0.400	0.425	0.6365484	0.23654842	−0.2115484
5.1	0.425	0.450	0.6454666	0.22046659	−0.1954666
10.7	0.450	0.475	0.7520955	0.30209546	−0.2770955
20.1	0.475	0.500	0.8338347	0.35883467	−0.3338347
25.3	0.500	0.525	0.8602348	**0.36023481**	−0.3352348
26.1	0.525	0.550	0.8636388	0.33863885	−0.3136388
36.2	0.550	0.575	0.8967896	0.34678955	−0.3217896
45.8	0.575	0.600	0.9175066	0.34250664	−0.3175066
48.7	0.600	0.625	0.9224684	0.32246835	−0.2974684
50.5	0.625	0.650	0.9253140	0.30031400	−0.2753140
56.3	0.650	0.675	0.9334486	0.28344864	−0.2584486
75.6	0.675	0.700	0.9525878	0.27758776	−0.2525878
81.5	0.700	0.725	0.9567998	0.25679985	−0.2317998
81.5	0.725	0.750	0.9567998	0.23179985	−0.2067998
81.6	0.750	0.775	0.9568664	0.20686639	−0.1818664
108.2	0.775	0.800	0.9703716	0.19537155	−0.1703716
170.8	0.800	0.825	0.9853603	0.18536033	−0.1603603
186.9	0.825	0.850	0.9874621	0.16246207	−0.1374621
345.9	0.850	0.875	0.9963080	0.14630800	−0.1213080
356.0	0.875	0.900	0.9965399	0.12153986	−0.09653986
412.7	0.900	0.925	0.9975497	0.09754965	−0.07254965
523.1	0.925	0.950	0.9986524	0.07365242	−0.04865242
1439.0	0.950	0.975	0.9999502	0.04995019	−0.02495019
1998.8	0.975	1.000	0.9999876	0.02498761	0.00001239

so the Kolmogorov-Smirnov statistic is 0.36023481

(b) The Anderson-Darling test.

The Anderson-Darling statisic is given by

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

We calculate this for our data set

```
Asq<-40*(sum(((40:1)/40)^2*(c(0,(-(HW2Q2[1:39]/theta)^tau))
-(-(HW2Q2/theta)^tau)))+
sum(((1:40)/40)^2*(c(log(1-exp(-(HW2Q2[2:40]/theta)^tau)),0)-log(1-exp(-(HW2Q2/theta)^t
```

This gives the Anderson-Darling statistic as 17.37052

(c) The chi-square test, dividing into the intervals 0–1,1–5,5–50 and more than 50.

We have the following table:

Interval	O	E	$\frac{(O-E)^2}{E}$
$[0,1)$	7	$40\times\left(1-\exp\left(-4.657714^{-0.4}\right)\right) = 16.6998423561$	$\frac{(7-16.6998423561)^2}{16.6998423561} = 5.63400179037$
$[1,5)$	10	$40\times\left(\exp\left(-4.657714^{-0.4}\right) - \exp\left(-\left(\frac{5}{4.657714}\right)^{0.4}\right)\right) = 9.00232342856$	$\frac{(10-9.00232342856)^2}{9.00232342856} = 0.110566849669$
$[5,50)$	8	$40\times\left(\exp\left(-\left(\frac{5}{4.657714}\right)^{0.4}\right) - \exp\left(-\left(\frac{50}{4.657714}\right)^{0.4}\right)\right) = 11.2794472265$	$\frac{(8-11.2794472265)^2}{11.2794472265} = 0.953484146469$
$[50,\infty)$	15	$40\times\exp\left(-\left(\frac{50}{4.657714}\right)^{0.4}\right) = 3.0183869889$	$\frac{(15-3.0183869889)^2}{3.0183869889} = 47.5615124488$
Total			54.2595652353

The Chi-squared statistic is 54.2595652353.

4. *For the data in Question 2, perform a likelihood ratio test to determine whether a Weibull distribution with fixed $\tau = 0.4$, or a transformed gamma distribution with α and τ freely estimated is a better fit for the data. [For the transformed gamma distribution, the MLE is $\alpha = 0.0273$, $\tau = 7.415$ and $\theta = 2045.683$.]*

The log-likelihood is given by

$$\sum_{i=1}^{40}\log(\tau) + \tau\alpha(\log(x_i) - \log(\theta)) - \left(\frac{x_i}{\theta}\right)^{\tau} - \log(x_i) - \log(\Gamma(\alpha))$$

We calculate this for the two parameter values

```
lltrg<-function(alpha,theta,tau,X){
40*log(tau)-40*alpha*tau*log(theta)+alpha*tau*sum(log(X))-sum((X/theta)^tau)-sum(log(X))-
}

llGeneral<-lltrg(0.0273,2045.683,7.415,HW2Q2)
llalpha1tau0_4<-lltrg(1,4.657714,0.4,HW2Q2)
```

Gives the log-likelihoods -209.9763 and -225.3924 respectively. Thus the log-likelihood ratio is $2(-209.9763-(-225.3924)) = 30.8322$. This is compared to a chi-squared distribution with two degrees of freedom, so the critical value, at the 5% significance level, is 5.991465, so we reject $\alpha = 1, \tau = 0.4$.

5. *For the data in Question 2, use AIC and BIC to choose between a Weibull distribution with* $\tau = 0.4$ *for the data and a Pareto distribution. [The MLE for the Pareto distribution is* $\alpha = 0.4108$, $\theta = 2.112$.*]*

   The log-likelihood for the Pareto distribution is

$$\sum_{i=1}^{40} \log(\alpha) + \alpha\log(\theta) - (\alpha+1)\log(\theta + x_i)$$

   We substitute the MLE for α and θ to calculate the log-likelihood:

   This gives the log-likelihood as -202.8737

   The AIC for the Weibul with $\tau = 0.4$ is $-225.3924 - 1 = -226.3924$, and the BIC is $-225.3924 - \frac{1}{2}\log(40) = -227.236839727$

   For the Pareto distribution, the AIC is $-202.8737 - 2 = -204.8737$ and the BIC is $-202.8737 - \log(40) = -206.562579454$. Thus the Pareto distribution is prefered by both AIC and BIC.

Standard Questions

6. *An auto insurer divides insureds into three categories: new drivers; average drivers; and good drivers. The number of claims made by an individual follows a negative binomial distribution with parameters* r *and* β*. The insurance company has the following portfolio of policies.*

Category	*Number insured*	r	β *of claim*	*mean claim*	*standard deviation*
new driver	*1,300*	*0.3*	*22.6*	*\$639*	*\$250,340*
average driver	*6,422*	*3.2*	*0.8*	*\$1,430*	*\$180,460*
safe driver	*1,105*	*0.9*	*1.8*	*\$1,100*	*\$105,660*

The insurance company models the aggregate losses as following an Pareto distribution with the correct mean and variance. It charges a 10% loading on its premiums. It wants to buy stop-loss reinsurance for its policies. The reinsurance company uses the same Pareto distribution to model aggregate losses and sets its premium at 125% of expected payments on the policy. The insurer is willing to pay 30% of its total premiums towards reinsurance. What attachment point can it set to achieve this?

We calculate the expectation and variance of aggregate claims

Category	$\mathbb{E}(N)$	$\mathrm{Var}(N)$	$\mathbb{E}(S)$ (millions)	$\mathrm{Var}(S)$ (trillions)
new driver	8814.00	208010.400	5.632146	552.459333913
average driver	16440.32	29592.576	23.5096576	535.4528776
safe driver	1790.10	5012.280	1.969110	19.9908049864
Total			31.1109136	1107.9030165

Thus the total premium is $31.1109136 \times 1.1 = \$34.22200496$ million. The insurer is therefore willing to pay $34.22200496 \times 0.3 = \$10.266601488$ million in reinsurance premiums. Setting the aggregate mean and variance as the mean and variance of a Pareto distribution gives

$$\begin{aligned}
\frac{\theta}{\alpha-1} &= 31.1109136\\
\frac{\alpha\theta}{(\alpha-1)^2(\alpha-2)} &= 1107.9030165\\
\frac{\alpha-2}{\alpha} &= \frac{31.1109136^2}{1107.9030165} = 0.873622447644\\
\alpha &= \frac{2}{1-0.873622447644} = 15.8255953112\\
\theta &= 14.8255953112 \times 31.1109136 = 461.237814795
\end{aligned}$$

For reinsurance with attachment point a, the expected payment is

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

Thus we need to solve

$$1.25\theta^\alpha \frac{(a+\theta)^{1-\alpha}}{(\alpha-1)} = 10.266601488$$

$$(a+\theta)^{1-\alpha} = \frac{(\alpha-1)10.266601488}{1.25\theta^\alpha}$$

$$= \frac{14.8255953112 \times 10.266601488}{1.25 \times 461.237814795^{15.8255953112}}$$

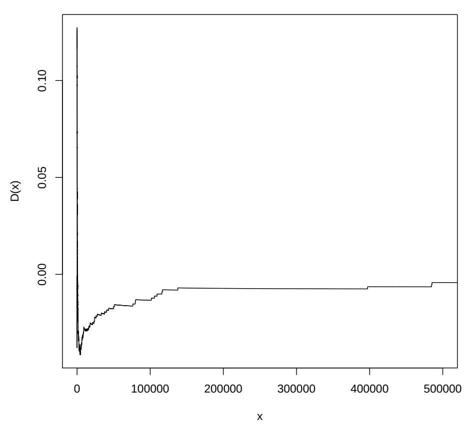
$$= 8.4592397211 \times 10^{-41}$$

$$a + 461.237814795 = (8.4592397211 \times 10^{-41})^{-\frac{1}{14.8255953112}} = 504.58954931$$

$$a = 43.351734515$$

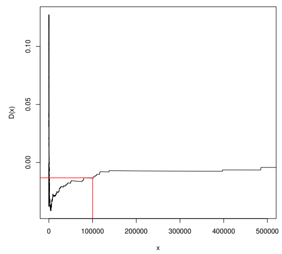
Thus attachment point is \$43,351,734.515.

7. *An insurance company collects a sample of 900 past claims, and attempts to fit a distribution to the claims. Based on experience with other claims, the company believes that an inverse Pareto distribution with $\tau = 3.1$ and $\theta = 35$ may be appropriate to model these claims. It constructs the following plot of $D(x) = F_n(x) - F^*(x)$ to compare the sample to this distribution:*



(a) How many data points in the sample were more than 100,000?

From the graph, we read $D(100000) \approx -0.0125$.



We have that $F^*(100000) = \left(\frac{100000}{100035}\right)^{3.1} = 0.998915778025$, so $F_n(100000) = F^*(100000) + D(100000) \approx 0.998915778025 - 0.0125 = 0.986415778025$, so there are $900 * (1 - 0.986415778025) = 12$ samples larger than 100,000. [In fact, there are 13 samples larger than 100,000 in the data set.]

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

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

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

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

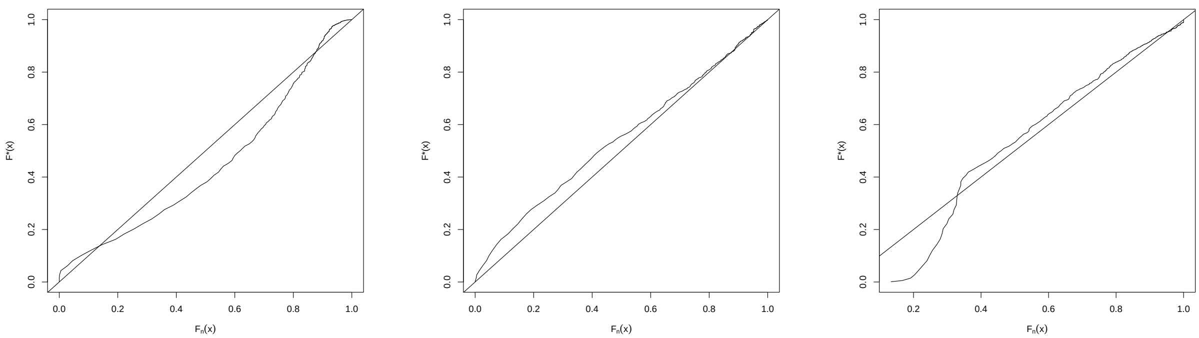
(iv) The inverse Pareto distribution assigns too much probability to central values and too little probability to tail values.

Justify your answer.

We see that $D(x)$ is positive for small values of x and negative for larger values of X. This means that $F^*(x) < F_n(x)$ for small x, and $F^*(x) > F_n(x)$ for large x. Thus the model assigns too little probability to both small values and large values, so (iv) is the best description of the fit.

(c) Which of the following plots is the p-p plot of this model on this data? Justify your answer.

(i) *(ii)* *(iii)*



Since $F^*(x) > F_n(x)$ for smaller values of x and $F^*(x) < F_n(x)$ for larger values, we expect the plot to lie above the line $y = x$ for small values of x and below for larger values of x. Only (i) shows this pattern, so (i) must be the correct plot.