ACSC/STAT 4703, Actuarial Models II

FALL 2024

Toby Kenney

Homework Sheet 4

Due: Thursday 10th October: 11:30

Basic Questions

1. The file `HW4_data1.txt` contains 200 i.i.d. samples of a random variable. An insurer is trying to model this random variable as following an inverse exponential distribution, as suggested by data sets from earlier years. Graphically compare this empirical distribution with the best inverse exponential distribution. From the data, they find that the MLE for θ is $\theta = 71.1524$. Include the following plots:
   (a) Comparisons of $F(x)$ and $F^*(x)$
   (b) Comparisons of $f(x)$ and $f^*(x)$
   (c) A plot of $D(x)$ against x.
   (d) A p-p plot of $F(x)$ against $F^*(x)$.

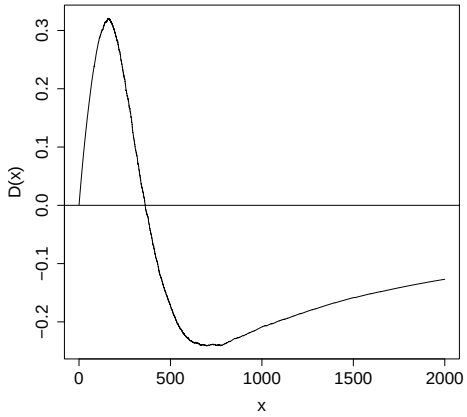
2. For the data in `HW4_data1.txt`, calculate the following test statistics for the goodness of fit of the Inverse exponential distribution with θ estimated by MLE:
   (a) The Kolmogorov-Smirnov test.
   (b) The Anderson-Darling test.
   (c) The chi-square test, dividing into the intervals 0–50,50–100,100–300 and more than 300.

3. For the data in `HW4_data1.txt`, perform a likelihood ratio test to determine whether an inverse exponential distribution, or an inverse transformed gamma distribution with α, τ and θ freely estimated is a better fit for the data. [For the inverse transformed gamma distribution, the MLE is $\alpha = 8.603$, $\tau = 0.4237$ and $\theta = 13615$.]

4. For the data in `HW4_data1.txt`, use AIC and BIC to choose between an inverse exponential distribution and an inverse Pareto distribution. [The MLE for the inverse Pareto distribution is $\tau = 0.925$ and $\theta = 129.9$.]

Standard Questions

5. An insurance company collects a sample of 2741 past claims, and attempts to fit a distribution to the claims. Based on experience with other claims, the actuary believes that a Pareto distribution may be appropriate to model these claims. She fits the MLE parameters $\alpha = 0.9085505$ and $\theta = 230.6825$ and constructs the following plot $D(x) = F^*(x) - F_n(x)$ for this distribution and data.



(a) How many data points in the sample were less than 240?

(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.

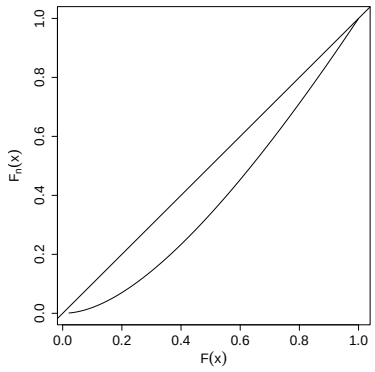
(iii) The Pareto distribution assigns too much probability to tail values and too little probability to central values.

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

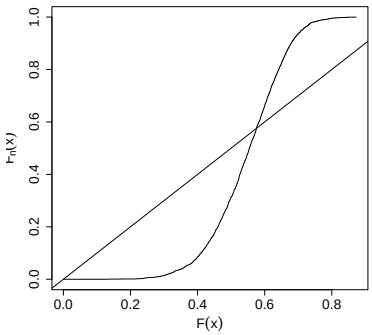
Justify your answer.

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

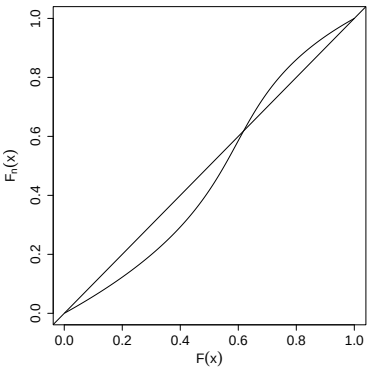
(i)


$F_n(x)$
0.0
0.2
0.4
0.6
0.8
1.0
F(x)

(ii)


$F_n(x)$
0.0
0.2
0.4
0.6
0.8
1.0
F(x)

(iii)


$F_n(x)$
0.0
0.2
0.4
0.6
0.8
1.0
F(x)