ACSC/STAT 4703, Actuarial Models II

FALL 2023

Toby Kenney

Homework Sheet 4

Due: Thursday 12th October: 14:30

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 a Pareto distribution with $\alpha = 9$, as suggested by data sets from earlier years. Graphically compare this empirical distribution with the best Pareto distribution with $\alpha = 9$. From the data, they find that the MLE for θ is $\theta = 52.61$. 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 $F^*(x)$.

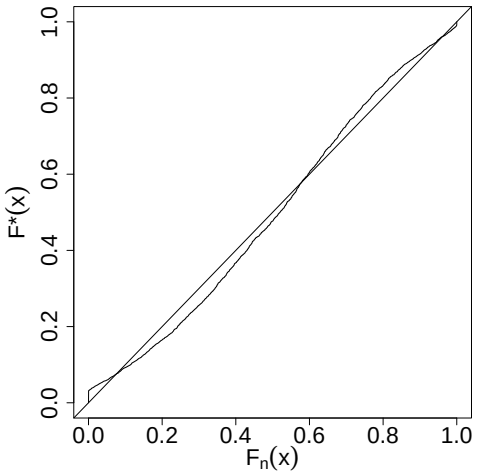
2. For the data in `HW4_data1.txt`, calculate the following test statistics for the goodness of fit of the Pareto distribution with $\alpha = 9$ and θ estimated by MLE:
   (a) The Kolmogorov-Smirnov test.
   (b) The Anderson-Darling test.
   (c) The chi-square test, dividing into the intervals 0–1,1–5,5–10 and more than 10.

3. For the data in `HW4_data1.txt`, perform a likelihood ratio test to determine whether a Pareto distribution with fixed $\alpha = 9$, or a generalised Pareto distribution with α, τ and θ freely estimated is a better fit for the data. [For the generalised Pareto distribution, the MLE is $\alpha = 5.6701$, $\tau = 1.86747$ and $\theta = 15.89494$.]

4. For the data in `HW4_data1.txt`, use AIC and BIC to choose between a Pareto distribution with $\alpha = 9$ for the data and a transformed gamma distribution. [The MLE for the transformed gamma distribution is $\alpha = 0.883801$, $\tau = 1.304570$ and $\theta = 7.650101$.]

Standard Questions

5. An insurance company collects a sample of 3,900 past claims, and attempts to fit a distribution to the claims. Based on experience with other claims, the actuary believes that a log-normal distribution may be appropriate to model these claims. She fits the MLE parameter $\mu = 0.4373128$ and $\sigma^2 = 0.3691496$ and constructs the following p-p plot of the distribution and data.



(a) How many data points in the sample were more than 2?

(b) Which of the following statements best describes the fit of the log-normal distribution to the data:

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

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

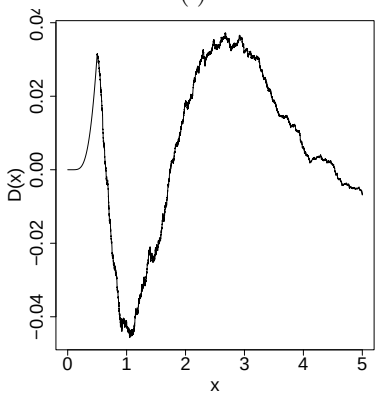
(iii) The log-normal distribution assigns too much probability to tail values and too little probability to central values.

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

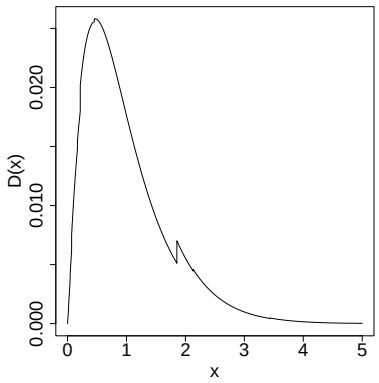
Justify your answer.

(c) Which of the following plots shows $D(x) = F^*(x) - F_n(x)$ for this model on this data? Justify your answer.

(i)



(ii)



(iii)