ASRM569 HW3

Minhao Zhang

September 2020

Question One

Proof:
QQ-plot coordinates (HTE-type: Pareto, LTE-type: Weibull ($tau > 1$))

Proof

Q1 ① HTE-type: Pareto $F(x)=1-x^{-\alpha}$ $(x>1, \alpha>0)$ $x^{-\alpha}=1-F(x)$

$\ln x = -\frac{1}{\alpha}\ln[1-F(x)]$ So $\ln[Q_\alpha(p)] = \frac{1}{\alpha}[-\ln(1-p)] = \frac{1}{\alpha}\ln[Q_1(p)]$ for $p\in(0,1)$

Let $\hat{Q}_n(p) = x_{i,n}$ for $\frac{i-1}{n}<p\le\frac{i}{n}$

coordinate: $(\log[Q_1(p)], \log[\hat{Q}_n(p)]) = (-\log(1-p_{i,n}), \log(x_{i,n}))$

② LTE-type: Weibull (tau>1) $F(x)=1-e^{-\lambda x^r}$ $(x>0;\lambda>0, r>1)$

$e^{-\lambda x^r}=1-F(x)$ $x^r = -\frac{1}{\lambda}\ln(1-F(x))$ $r\ln x = \ln(\frac{1}{\lambda}) + \ln[-\ln(1-F(x))]$

$\ln x = \frac{\ln(\frac{1}{\lambda})}{r} + \frac{1}{r}\ln[-\ln(1-F(x))]$

So $\ln[Q_{r,\lambda}(p)] = \frac{\ln(\frac{1}{\lambda})}{r} + \frac{1}{r}\ln[-\ln(1-p)] = \frac{1}{r}\ln[Q_{1,1}(p)] + \frac{\ln(\frac{1}{\lambda})}{r}$ for $p\in(0,1)$

Let $\hat{Q}_n(p) = x_{i,n}$ for $\frac{i-1}{n}<p\le\frac{i}{n}$

coordinate: $(\log[Q_{1,1}(p)], \log[\hat{Q}_n(p)]) = (\log(-\log(1-p_{i,n})), \log(x_{i,n}))$

Figure 1: Proof

Question Two

Create QQ-plot and mean excess plot without using the existing package and verify Figure 1.8.

R codes

(github link: https://github.com/zhangminhao00/569HW.git)

```r
# Pareto
set.seed(569)
n = 50000
data = sort(rpareto(n,shape=3,scale=10))

p_values = seq(1/(n+1), n/(n+1), by=1/(n+1))
theoretical_Q = -log(1-p_values)
empirical_Q = quantile(data, probs=p_values)

plot(theoretical_Q, empirical_Q,
     main=c("Exponential QQ-plot","\nPareto"),
     xlab="-log(1-p)",
     ylab="Q(p)")

e=numeric(n)
K=1:(n-1)
e[n-K] = cumsum(data[n-K+1])/K - data[n-K]

plot(data[K],e[K],xlab=bquote(X["n-k,n"]), ylab=bquote(e["k,n"]),
     main=c("Mean excess plot","\nPareto"))



# Weibull (tau>1)
set.seed(569)
n = 50000
data = sort(rweibull(n,shape=3,scale=10))

p_values = seq(1/(n+1), n/(n+1), by=1/(n+1))
theoretical_Q = -log(1-p_values)
empirical_Q = quantile(data, probs=p_values)

plot(theoretical_Q, empirical_Q,
     main=c("Exponential QQ-plot","\nWeibull (tau>1)"),
     xlab="-log(1-p)",
     ylab="Q(p)")

e=numeric(n)
K=1:(n-1)
e[n-K] = cumsum(data[n-K+1])/K - data[n-K]

plot(data[K],e[K],xlab=bquote(X["n-k,n"]), ylab=bquote(e["k,n"]),
     main=c("Mean excess plot","\nWeibull (tau>1)"))
```

Figure 2: R codes

Outcomes

Pareto:

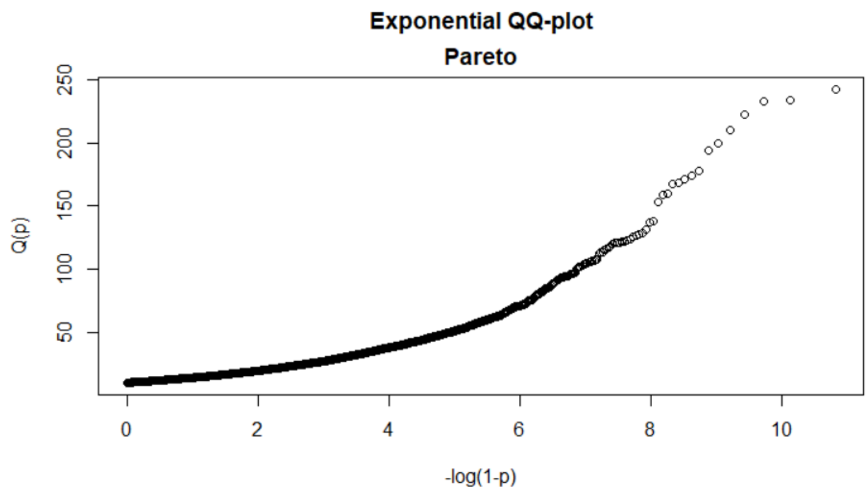

Figure 3: Exponential QQ-plot (Pareto)

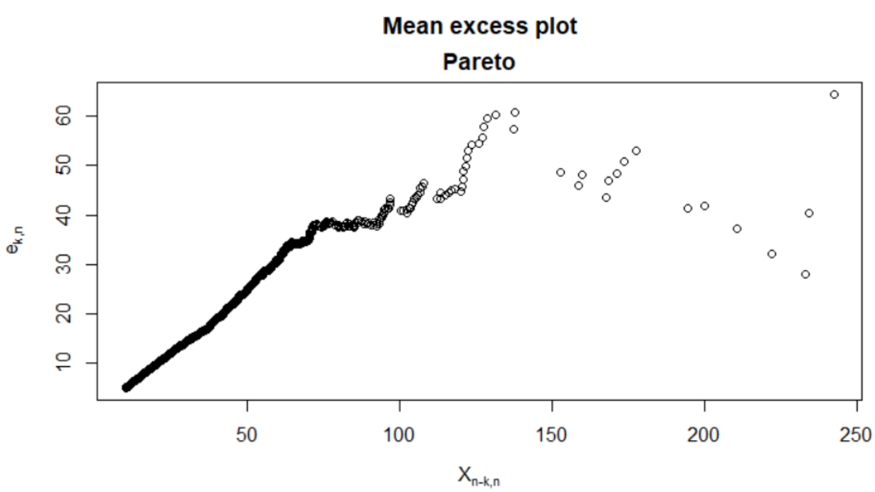

Figure 4: Mean excess plot (Pareto)

Conclusion:
The exponential QQ-plot for Pareto has a convex shape for the larger observations and the slopes continue to increase near the higher observations. Also, Pareto distribution has an increasing mean excess function. So it can be concluded that Pareto distribution is a HTE distribution.

Weibull ($tau > 1$):

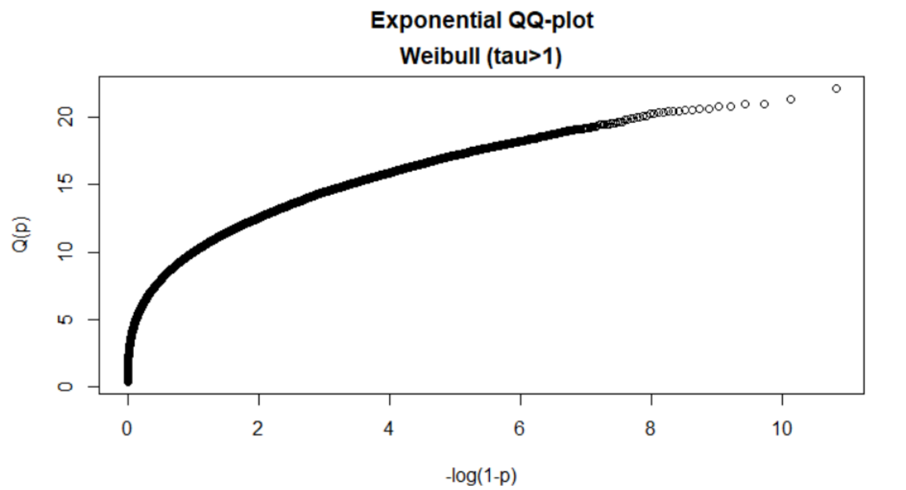

Figure 5: Exponential QQ-plot (Weibull ($tau > 1$))

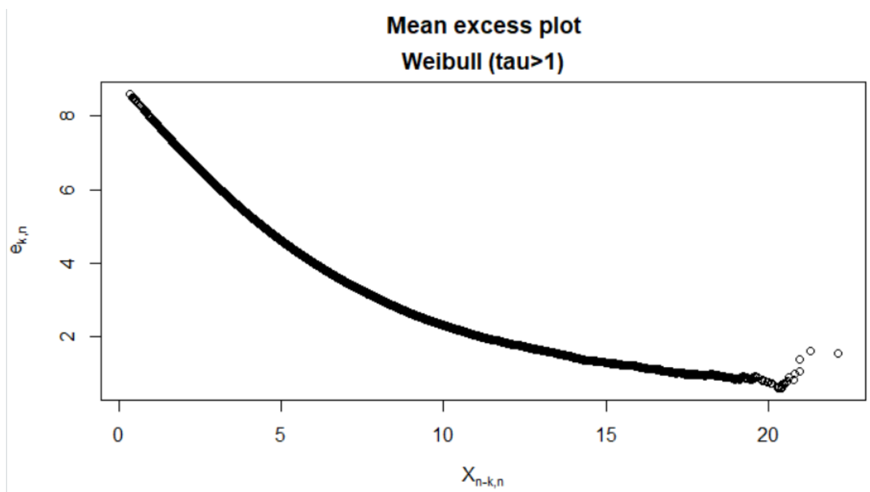

Figure 6: Mean excess plot (Weibull ($tau > 1$))

Conclusion:
The exponential QQ-plot for Weibull ($tau > 1$) has a concave shape for the larger observations and the slopes continue to decrease near the higher observations. Also, Weibull distribution ($tau > 1$) has an decreasing mean excess function. So it can be concluded that Weibull distribution ($tau > 1$) is a LTE distribution.