Bounding a good region

Xufei Wang

February 7, 2017
Suppose a continuous random variable $X$ follows power law distribution when $X > x_{min}$ and $\alpha > 1$, then we have

$$f(x) = Cx^{-\alpha}, \quad x \in (x_{min}, \infty).$$

Let $Y = \log X$ and $y_{min} = \log x_{min}$, then the conditional distribution of $Y - y_{min} | Y > y_{min}$ follows $\text{Exp}(\alpha - 1)$. 
Different estimators

Hill estimator
Suppose we have observation $X_{(1)} \geq X_{(k)} \geq X_{(n)}$, then

$$H_{k,n} = 1 + \frac{1}{\frac{1}{k} \sum_{i=1}^{k} (y - y_{min})}.$$ 

QQ estimator
We have

$$1 - F(y|y > y_{min}) = e^{- (\alpha - 1)(y - y_{min})}$$

so we can get an estimator for $\frac{1}{1-\alpha}$ if we perform linear regression for

$$\left[ \log \left( \frac{i}{k+1} \right), Y(i) \right], \quad 1 \leq i \leq k.$$
An example

Suppose $X_1, \ldots, X_{1000}$ follows power law distribution with $\alpha = 2.5$, by Hill estimator we get $\alpha = 2.522$ and by QQ estimator we get $\alpha = 2.599$. If we plot the log-log plot of the histogram of $X$, with the true density function.
weighted regression

beta=1, alpha=2.401

beta=0.9, alpha=2.499

beta=0.8, alpha=2.524

beta=0.7, alpha=2.584

beta=0.6, alpha=2.706

beta=0.5, alpha=2.885
real data

By Hill estimator we get $\alpha = 1.977$ and by QQ estimator we get $\alpha = 1.951$.
weighted regression

beta=1, alpha=1.657

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy
nbeta=0.9, alpha=1.914

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy
nbeta=0.8, alpha=1.981

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy
nbeta=0.7, alpha=2.066

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy
nbeta=0.6, alpha=2.21

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy
nbeta=0.5, alpha=2.421

counts
100 200 500 2000 5000 20000
1 2 5 10 20 50
energy