The Maximum Likelihood Estimator of the Extreme Value Index

Chen Zhou

Tinbergen Institute, Erasmus University Rotterdam

January 22, 2008

Introduction: Maximum Likelihood Estimator (1)

The extreme value condition:
Define the exceed distribution function as

$$F_t(x) := P(X \le t + x | X > t) = \frac{F(t+x) - F(t)}{1 - F(t)}.$$

Then $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t \to x^*} F_t(x\sigma(t)) = H_\gamma(x) := 1 - (1 + \gamma x)^{-1/\gamma},$$

for all $1 + \gamma x > 0$, where x^* is the right endpoint of F, σ is a positive function and H_γ is the Generalized Pareto Distribution)

Peak over threshold method:
Smith (1987) proposed to find a maximum likelihood estimator for γ and σ by fitting the GPD to the tail of F. The consistency and asymptotic normality is proved for $\gamma > -1/2$ under some extra conditions.

Introduction: Maximum Likelihood Estimator (2)

Taking high order statistic as the threshold:

Denote $X_{n,1} \leq X_{n,2} \leq \cdots \leq X_{n,n}$ as the order statistics of an i.i.d. sample $X_1, \cdots, X_n$. Drees, Ferreira and de Haan (2004) proposed to use the order statistic $X_{n,n-k}$ ($k \to \infty$ and $k/n \to 0$ as $n \to \infty$) as the threshold in the maximum likelihood procedure.

Likelihood equations

The maximum likelihood estimators satisfy (when $\gamma \neq 0$):

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k} \log\left(1+\frac{\gamma}{\sigma}(X_{n,n-i+1}-X_{n,n-k})\right) &= \gamma \\ \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1+(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})} &= \frac{1}{\gamma+1}\end{aligned}$$

Introduction: First order and second order conditions

First order condition:

An alternative way to present the extreme value condition is via $U(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}$. $F \in D(G_\gamma)$ is equivalent to

$$\lim_{t\to\infty} \frac{U(tx) - U(t)}{a(t)} = \frac{x^\gamma - 1}{\gamma},$$

for all $x > 0$, where a is a positive function. Note that $\sigma(t) = a(U^{\leftarrow}(t))$.

Second order condition:

The second order condition characterizes the speed of convergence in the first order condition as follows.

$$\lim_{t\to\infty} \frac{\frac{U(tx)-U(t)}{a(t)} - \frac{x^\gamma-1}{\gamma}}{A(t)} = \Psi(x).$$

Introduction: Asymptotic behavior of estimators

Estimation of the extreme value index γ

Estimators: Hill, Pickands, Moment, etc.

Properties: Consistency and Asymptotic Normality
First order condition $\Rightarrow$ consistency
Second order condition $\Rightarrow$ asymptotic normality

Introduction: Maximum Likelihood Estimator (3)

Asymptotic normality (Drees, Ferreira and de Haan (2004)):

Suppose the sequence k satisfies the condition that as $n \to \infty$, $k \to \infty$, $k/n \to 0$ and $k^{1/2}A(n/k) \to \lambda$ for some real constant λ. With the second order condition, for $\gamma > -1/2$, for any solution of the likelihood equations satisfying

$$\left| \frac{\widehat{\gamma}_n}{\widehat{\sigma}_n / a(n/k)} - \gamma \right| = O_p(k^{-1/2}) \quad \text{and} \quad \log \frac{\widehat{\sigma}_n}{a(n/k)} = O_p(1),$$

the asymptotic normality holds as follows:

$$\sqrt{k}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} W_1,$$

$$\sqrt{k} \left(\frac{\widehat{\sigma}_n}{a(n/k)} - 1 \right) \xrightarrow{d} W_2,$$

where (W_1, W_2) follows a 2-dimensional normal distribution.

Research Question (1)

Under only the first order condition, without the second order condition

1) Are the likelihood equations always solvable? (existence)

2) Is the maximum likelihood estimator consistent?

With the second order condition

3) Does a solution satisfying the requirements in Drees, Ferreira and de Haan (2004) exist?

Answers of question 1) and 2)

Theorem. *Suppose the first order condition holds for the extreme value index $\gamma > -1$. If the sequence $k = k(n)$ satisfies $k(n) \to \infty$, $k(n)/n \to 0$, and $k(n)/(\log\log n)^5 \to \infty$, then*

$$P(\{\textit{The MLE does not exist for infinitely many } n\}) = 0.$$

Or, equivalently,

$$P(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{\textit{The MLE exists for sample size } n\}) = 1.$$

On this probability 1 set, there exists an random integer N, such that for any sample size $n > N$, there is a suitable solution of the likelihood equations, $(\widehat{\gamma}_n, \widehat{\sigma}_n)$, satisfying

$$\widehat{\gamma}_n \xrightarrow{a.s.} \gamma \quad \textit{and} \quad \frac{\widehat{\sigma}_n}{a(n/k)} \xrightarrow{a.s.} 1$$

as $n \to \infty$.

Idea of Grimshaw's numerical solution (1)

Grimshaw's numerical method on solving the likelihood equations:

Recall the likelihood equations

$$\begin{aligned}\frac{1}{k}\sum_{i=1}^{k}\log\left(1+\frac{\gamma}{\sigma}(X_{n,n-i+1}-X_{n,n-k})\right) &= \gamma\\ \frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+(\gamma/\sigma)(X_{n,n-i+1}-X_{n,n-k})} &= \frac{1}{\gamma+1}.\end{aligned}$$

With the notation $Y_i := X_{n,n-i+1} - X_{n,n-k}$, we can derive that,

$$\left(\frac{1}{k}\sum_{i=1}^{k}\log\left(1+\frac{\gamma}{\sigma}Y_i\right)+1\right)\cdot\frac{1}{k}\sum_{i=1}^{k}\frac{1}{1+(\gamma/\sigma)Y_i}=1.$$

Idea of Grimshaw's numerical solution (2)

Denote

$$\begin{aligned} f_n(t) &= \frac{1}{k}\sum_{i=1}^{k} \log(1 + tY_i) + 1, \\ g_n(t) &= \frac{1}{k}\sum_{i=1}^{k} \frac{1}{1 + tY_i}, \\ h_n(t) &= f_n(t)g_n(t) - 1. \end{aligned}$$

Then, it is clear that any root $(\widehat{\gamma}_n, \widehat{\sigma}_n)$ of the likelihood equations satisfies $h_n(\widehat{\gamma}_n/\widehat{\sigma}_n) = 0$.

The maximum likelihood estimator can be calculated in the following procedure:
(1) find the root t_n^* of $h_n(t) = 0$;
(2) $\widehat{\gamma}_n = f_n(t_n^*) - 1$;
(3) $\widehat{\sigma}_n = \widehat{\gamma}_n / t_n^*$.

Idea of proof (1) ($\gamma > 0$)

We try to approximate the root $t_n^* = \widehat{\gamma}_n / \widehat{\sigma}_n$. We use an approximate solution $\frac{\gamma}{\sigma(X_{n,n-k})} = \frac{1}{X_{n,n-k}}$, but disturb it as $t_n^{(\delta)} = \frac{1+\delta}{X_{n,n-k}}$ for $\delta \in (-1/2, 1/2)$. It can be proved that

$$\begin{aligned} f_n(t_n^{(\delta)}) &\xrightarrow{a.s.} f(\delta) := 1 + \int_0^1 \log((1+\delta)t^{-\gamma} - \delta)dt, \\ g_n(t_n^{(\delta)}) &\xrightarrow{a.s.} g(\delta) := \int_0^1 \frac{dt}{(1+\delta)t^{-\gamma} - \delta}, \\ h_n(t_n^{(\delta)}) &\xrightarrow{a.s.} h(\delta) := f(\delta)g(\delta) - 1. \end{aligned}$$

We can calculate that

$$f(0) = \gamma + 1, \;\; g(0) = \frac{1}{\gamma+1}, \;\; h(0) = 0$$

and

$$h'(0) = f(0)g'(0) + g(0)f'(0) = -\frac{\gamma^3}{(\gamma+1)^2(2\gamma+1)} < 0.$$

Idea of proof (2) ($\gamma > 0$)

Now we can choose a suitable $\delta > 0$ such that $h(\delta) < 0 < h(-\delta)$. Then, $h_n(t_n^{(\delta)}) < 0 < h_n(t_n^{(-\delta)})$ for sufficient large n. Considering the continuity of the function h_n, the existence of the root is proved. Furthermore, the root t_n^* is in between $t_n^{(-\delta)}$ and $t_n^{(\delta)}$.

Since f_n is an increasing function and $f(\delta)$ is continuous at 0. we have

$$f_n(t_n^*) \xrightarrow{a.s.} f(0) = \gamma + 1,$$

i.e.

$$\widehat{\gamma}_n = f_n(t_n^*) - 1 \xrightarrow{a.s.} \gamma.$$

The consistency of $\widehat{\sigma}_n$ can be similarly proved.

Answer of question 3) ($\gamma > 0$)

With second order condition and $\gamma > -1/2$, we can choose a sequence δ_n shrinking to 0 satisfying $k^{1/2}\delta_n = O_p(1)$ without destroying the inequality

$$h_n(t_n^{(\delta_n)}) < 0 < h_n(t_n^{(-\delta_n)}).$$

Since $\frac{t_n^{(0)}}{\gamma/a(n/k)} = 1 + O_p(k^{-1/2})$ and $\delta_n = O_p(k^{-1/2})$, there exits a root t_n^* of $h_n(t) = 0$ such that

$$\frac{t_n^*}{\gamma/a(n/k)} - 1 = O_p(k^{-1/2}).$$

It implies that

$$\left|\frac{\widehat{\gamma}_n}{\widehat{\sigma_n}/a(n/k)} - \gamma\right| = O_p(k^{-1/2}) \quad \text{and} \quad \log\frac{\widehat{\sigma}_n}{a(n/k)} = O_p(1),$$

i.e. there exists a solution of the likelihood equations satisfying the requirements in Drees, Ferreira and de Haan (2004).

Research Question (2)

Results up to now:

For $\gamma > -1$,
First order condition $\Rightarrow$ existence and consistency
For $\gamma > -1/2$,
Second order condition $\Rightarrow$ asymptotic normality.

New question:

How about $-1 < \gamma \leq -1/2$? Under the second order condition do we still have asymptotic normality?

Answer of new question

Yes, but extra condition on k sequence

Theorem. *Suppose the second order condition holds for the extreme value index $-1 < \gamma \leq -1/2$. If the sequence $k = k(n)$ satisfies $k \to \infty$, $k/n \to 0$ and $k^{-\gamma}A(n/k) \to 0$, then for sufficiently large n, there exist a sequence of solution $(\widehat{\gamma}_n, \widehat{\sigma}_n)$ of the likelihood equations satisfying*

$$\sqrt{k}\left(\widehat{\gamma}_n - \gamma, \frac{\widehat{\sigma}_n}{a(n/k)} - 1\right) \xrightarrow{d}, (W_1, W_2)$$

as $n \to \infty$, where $(W_1, W_2)^T$ follows a 2-dimensional normal distribution

Idea of proof ($-1 < \gamma < -1/2$)

We first prove that there exists a root t_n^* of $h_n(t) = 0$ lying in between $p_n^{(1)}$ and $p_n^{(2)}$ such that for any $\varepsilon > 0$,

$$k^{-\gamma-\varepsilon}\left(\frac{p_n^{(j)}}{\gamma/\sigma(X_{n,n-k})} - 1\right) = O_p(1) \quad \text{for } j = 1, 2.$$

By studying $f_n(p_n^{(1)}) - 1$ and $f_n(p_n^{(2)}) - 1$, we prove that

$$\sqrt{k}(\widehat{\gamma}_n - \gamma) \xrightarrow{d} W_1.$$

The proof for the scale part is different from the case $\gamma > -1/2$. We start from

$$\frac{\widehat{\sigma}_n}{a(n/k)} = \frac{\widehat{\gamma}_n/t_n^*}{a(n/k)} = \frac{\widehat{\gamma}_n}{\gamma} \cdot \frac{\gamma/\sigma(X_{n,n-k})}{t_n^*} \cdot \frac{\sigma(X_{n,n-k})}{\sigma(U(n/k))}.$$

Therefore, $\widehat{\sigma}_n/a(n/k)$ goes to 1 at speed $k^{1/2}$ with a limit distribution dominated by limit distributions of the first and third items.

Remark on asymptotic variance

When $\gamma > -1/2$, Drees, Ferreira and de Haan (2004) provided the asymptotic covariance matrix for (W_1, W_2) as

$$\begin{pmatrix} (1+\gamma)^2 & -(1+\gamma) \\ -(1+\gamma) & 1+(1+\gamma)^2 \end{pmatrix}.$$

When $-1 < \gamma \leq -1/2$, we calculate the asymptotic covariance matrix for (W_1, W_2) as

$$\begin{pmatrix} \gamma^2 & \gamma \\ \gamma & 1+\gamma^2 \end{pmatrix}.$$

They are connected at $\gamma = -1/2$, but not smoothly!

Conclusion

The maximum likelihood estimator can be used for all $\gamma > -1$ by choosing suitable high threshold.

Further question

How about $\gamma \leq -1$?

Questions and Remarks