# STOR 836 p. 55

Also $\mathrm{var}(\log Y_u) \to \frac{1}{\alpha^2}$ as $u \to \infty$

hence $\mathrm{var}\left(\frac{1}{\hat{\alpha}}\right) \sim \frac{1}{k\alpha^2}$

But $k \sim ncu^{-\alpha}$ so

$$\mathrm{var}\left(\frac{1}{\hat{\alpha}}\right) \approx \frac{1}{\alpha^2 ncu^{-\alpha}}$$

Overall:

$$MSE \approx \frac{Au^{\alpha}}{n} + B^2 u^{-2\beta}$$

where $A = \frac{1}{\alpha^2 c}$, $B = \frac{d \cdot \beta}{\alpha(\alpha+\beta)}$

Minimize when $u = \left(\frac{2\beta B^2 n}{\alpha A}\right)^{\frac{1}{\alpha+2\beta}}$

and then

$$MSE = \frac{B^2(\alpha+2\beta)}{\alpha}\left(\frac{2\beta B^2 n}{\alpha A}\right)^{-2\beta/(\alpha+2\beta)}$$

Key result $MSE = O\left(n^{-2\beta/(\alpha+2\beta)}\right)$

Hill's estimator: fix $u$, assume $1-F(x) = cx^{-\alpha}$, $x>u$

MLE: $\hat{\alpha} = \left\{ \frac{1}{k} \sum_{i=1}^{k} \log\left(\frac{X_i}{u}\right) \right\}^{-1}$, $\hat{c} = \frac{k}{n} u^{\hat{\alpha}}$

Here $X_1 > X_2 > \dots > X_k > u > X_{k+1} > \dots > X_n$ are ordered sample from $F$.

Alternative (Weissman): same but fix $k$ first, set $u = X_{k+1}$.

Usual focus is on $\alpha$ but we could also consider
$1-\hat{F}(x) = \hat{c}x^{-\hat{\alpha}}$, any $x>u$

or fix $1-F(x)=p$, $\hat{x}_p = \left(\frac{\hat{c}}{p}\right)^{1/\hat{\alpha}}$.

Alternative representation: let $Y_u = \frac{X}{u}$ conditioned on $X>u$.

For $\overset{y>1}{\cancel{y>0}}$, $P(Y_u > y) = \frac{1-F(uy)}{1-F(u)}$

Assume $1-F(x) = cx^{-\alpha}\{1 + dx^{-\beta} + o(x^{-\beta})\}$ $x\to\infty$

Then $\frac{1-F(uy)}{1-F(u)} = y^{-\alpha}\{1 + du^{-\beta}(y^{-\beta}-1) + o(u^{-\beta})\}$

$u\to\infty$, fixed $\underset{y>1}{\cancel{y>0}}$

Density $\alpha y^{-\alpha-1} + d u^{-\beta}\{(\alpha+\beta)y^{-\alpha-\beta-1} - \alpha y^{\alpha-1}\} + o(u^{-\beta})$

Note integral $\int_1^\infty (\log y)^k y^{-\alpha-1} dy = \alpha^{-k-1}\Gamma(k+1)$ $= k!$ if $k$ integer

So $E \log Y_u = \frac{1}{\alpha} + d u^{-\beta}\left(\frac{1}{\alpha+\beta} - \frac{1}{\alpha}\right) + o(u^{-\beta})$

$E \log^2 Y_u = \frac{2}{\alpha^2}$ etc.

So $E\left(\frac{1}{\hat\alpha}\right) = \frac{1}{\alpha} - d u^{-\beta}\cdot\frac{\beta}{\alpha(\alpha+\beta)} + o(u^{-\beta})$

$\text{var}\left(\frac{1}{\hat\alpha}\right) \approx \frac{1}{k\alpha^2}$

But $k \approx ncu^{-\alpha}$ so

Bias of $\frac{1}{\hat\alpha} \approx -\frac{d\beta}{\alpha(\alpha+\beta)} u^{-\beta}$, $\text{Var} \approx \frac{1}{\alpha^2 ncu^{-\alpha}}$

$\text{MSE} \approx \frac{Au^\alpha}{n} + B^2 u^{-2\beta}$ where $A = \frac{1}{\alpha^2 c}$, $B = \frac{d\cdot\beta}{\alpha(\alpha+\beta)}$.

Minimize the MSE:

$u = \left(\frac{2\beta B^2 n}{\alpha A}\right)^{1/(\alpha+2\beta)}$, $\text{MSE} = \frac{B^2(\alpha+2\beta)}{\alpha}\cdot\left(\frac{2\beta B^2 n}{\alpha A}\right)^{-2\beta/(\alpha+2\beta)}$.

Key point: $\text{MSE} \propto n^{-2\beta/(\alpha+2\beta)}$

Alternative approach: use GPD

If $1-F(x) = cx^{-\alpha}$ is exact, then

$$\frac{1-F(u+y)}{1-F(u)} = \left(1+\frac{y}{u}\right)^{-\alpha} = \left(1+\frac{\xi y}{\sigma}\right)^{-1/\xi}$$

identical if $\sigma = \frac{u}{\alpha}$, $\xi = \frac{1}{\alpha}$ } Treat as "true" parameters for GPD

Fisher info matrix $J$, $J^{-1} = (1+\xi)\begin{pmatrix} 2\sigma^2 & -\sigma \\ -\sigma & 1+\xi \end{pmatrix}$

so $\mathrm{var}(\hat{\xi}) \approx \frac{(1+\xi)^2}{k} = \left(\frac{\alpha+1}{\alpha}\right)^2 \cdot \frac{1}{k} > \frac{1}{\alpha^2 k}$

worse than Hill est$^n$.

But what about the bias?

Side calculation: bias in MLE and related estimators

Suppose we have a model $X_{ni} \sim f_n(x; \underset{\sim}{\theta}_n)$

and an estimator $\hat{\underset{\sim}{\theta}}_n$ defined by

$$\sum_{i=1}^{k_n} \underset{\sim}{T}(X_{ni}; \hat{\underset{\sim}{\theta}}_n) = 0, \quad \text{some function } \underset{\sim}{T}, \quad \underset{\sim}{T} = (T_1, \ldots, T_p)$$

~~where $E_{f_n}\{$~~

Suppose model $X_i \sim f(X_i; \underset{\sim}{\theta}_0)$, $\underset{\sim}{\theta} \in \Theta$
↑ some par. space

"True value" $\underset{\sim}{\theta} = \underset{\sim}{\theta}_0$.
($p$ dimensional)

Suppose estimating equation

$$\sum_{i=1}^{k} \underset{\sim}{T}(X_i, \hat{\underset{\sim}{\theta}}_k) = 0$$

where $\int \underset{\sim}{T}(x; \underset{\sim}{\theta}_0) = 0$

Also define a $p \times p$ matrix $W(X, \underset{\sim}{\theta})$ by

$$w_{rs}(X; \underset{\sim}{\theta}) = \frac{\partial T_r(X; \underset{\sim}{\theta})}{\partial \theta_s}$$

$1 \leq r \leq p, \quad 1 \leq s \leq p.$

In case $\underset{\sim}{T}(x; \underset{\sim}{\theta}) = -\nabla_\theta \log f(x; \theta)$ this is standard MLE and $W$ is observed info matrix (for a single obs.)

Under regularity conditions

$$0 = \sum_{i=1}^{k_n} \underset{\sim}{T}(X_i; \hat{\theta}_n) = \sum_{i=1}^{k_n} \{ \underset{\sim}{T}(X_i; \underset{\sim}{\theta}_0) + W(X_i, \theta_n^*)(\hat{\theta}_n - \underset{\sim}{\theta}_0) \}$$

$$\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0 \approx -\left\{\sum_i W(X_i;\underset{\sim}{\theta}_0)\right\}^{-1}\left\{\sum_i \underset{\sim}{T}(X_i;\underset{\sim}{\theta}_0)\right\}$$

If $E\,W(X_i;\underset{\sim}{\theta}_0) = J_0$, cov $\underset{\sim}{T}(X_i;\underset{\sim}{\theta}_0) = C_0$ then

$$\sqrt{k_n}\,(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \approx N\left[0, J_0^{-1} C_0 J_0^{-1}\right]$$

"information sandwich formula"

In MLE case, $J_0 = C_0$ and RHS is $N[0, J_0^{-1}]$

Now suppose model is not exact but

$\frac{1}{k_n}\sum_{i=1}^{k_n} T(X_i;\underset{\sim}{\theta}_0) \approx b_n$ where $b_n$ is small as $n \to \infty$

then

$$\cancel{k^{-\frac{1}{2}}(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) \sim N[k_n^{-1/2}\, \underset{\sim}{b}_k, C_0]}$$

$$\frac{1}{\sqrt{k_n}}\sum_{1}^{k_n} T(X_i;\theta_0) \sim N\left[\frac{1}{\sqrt{k_n}}\underset{\sim}{b}_n, C_0(\underset{\sim}{\theta}_0)\right]$$

(approx)

hence $\sqrt{k_n}\,(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0) + k_n^{-1/2} J_0^{-1}\underset{\sim}{b}_n \to N\left[0, J_0^{-1} C_0 J_0^{-1}\right]$

If $k_n^{-1/2} \underset{\sim}{b}_n \to \underset{\sim}{c}$ for some vector $\underset{\sim}{c}$,

$$\sqrt{k_n}\left(\hat{\underset{\sim}{\theta}}_n - \underset{\sim}{\theta}_0\right) \xrightarrow{d} N\left[-J_0^{-1}\underset{\sim}{c},\ J_0^{-1} C_0 J_0^{-1}\right].$$

Application to GPD:

$$\ell(\sigma,\xi) = \log\sigma + \left(\frac{1}{\xi}+1\right)\log\left(1+\frac{\xi y}{\sigma}\right)$$

$$\sigma\frac{\partial\ell}{\partial\sigma} = -\frac{1}{\xi} + \left(\frac{1}{\xi}+1\right)\left(1+\frac{\xi y}{\sigma}\right)^{-1}$$

$$\frac{\partial\ell}{\partial\xi} = -\frac{1}{\xi^2}\log\left(1+\frac{\xi y}{\sigma}\right) + \frac{1}{\xi}\left(\frac{1}{\xi}+1\right)\left\{1-\left(1+\frac{\xi y}{\sigma}\right)^{-1}\right\}$$

set $y = u(Y_u - 1)$

$$\cancel{EY_u^{-1} = \int_1^\infty \left\{\alpha y^{-\alpha-2} + du^{-\beta}\left\{(\alpha+\beta)y^{-\alpha-\beta-1} - \alpha y^{-\alpha-1}\right\}dy\right.}$$

$$EY_u^{-1} = \int_1^\infty \left[\alpha y^{-\alpha-2} + du^{-\beta}\left\{(\alpha+\beta)y^{-\alpha-\beta-2} - \alpha y^{-\alpha-2}\right\}\right]dy + o(u^{-\beta})$$

But $\int_1^\infty y^{-\alpha-2}\,dy = \left[-\frac{y^{-\alpha-1}}{\alpha+1}\right]_1^\infty = \frac{1}{\alpha+1}$ etc.

$$EY_u^{-1} = \frac{\alpha}{\alpha+1} + du^{-\beta}\left(\frac{\alpha+\beta}{\alpha+\beta+1} - \frac{\alpha}{\alpha+1}\right) + o(u^{-\beta}) \text{ etc.}$$

STOR834 p.62

Ultimately, bias of $\hat{\xi} \approx du^{-\beta} \frac{(\alpha+1)\beta(1-\beta)}{\alpha(\alpha+\beta)(\alpha+\beta+1)}$

and var $\sim \frac{(\alpha+1)^2}{\alpha^2 ncu^{-\alpha}}$

Calculate optimal threshold for GPD and Hill estimators separately

Ratio of optimal MSE is

$$\left| \frac{(1-\beta)(\alpha+1)}{\alpha(\alpha+\beta)(\alpha+\beta+1)} \right|^{2\alpha/(\alpha+2\beta)} \cdot |\alpha+1|^{4\beta/(\alpha+2\beta)}$$

Fig 5.1 p.63 : shows this curve as a fn. of $\beta$ for various $\alpha$.

Conclusion: in every case MSE favors GPD for small $\beta$ (critical case)

Hall & Weissman (1997)

Distribution function $F$, interest in $\bar{F} = 1 - F$

Some model leads to $\bar{F}_\theta(x)$ in terms of parameters $\theta$.

Estimator $\hat{\theta}(t)$, tuning parameter $t$.

"Pulling back": replace $(n, x)$ by $(m, y)$

Approximate $\bar{F}$ by $\hat{\bar{F}}$ empirical cdf

then estimate $t$

Criteria: $D_1(t; n, x) = E\{\bar{F}_{\hat{\theta}(t)}(x) - \bar{F}(x)\}^2$

$$D_2(t; n, p) = D_1(t; n, \bar{F}^{-1}(p))$$
$$= E\left[\bar{F}_{\hat{\theta}(t)}\{\bar{F}^{-1}(p)\} - p\right]^2$$

Restrict to $\bar{F}_\theta(x) = \theta_1 x^{-\theta_2}$, $\theta_1 > 0$, $\theta_2 > 0$

Relabel so $\bar{F}_{\alpha,c} = cx^{-\alpha}$

$\hat{\alpha}, \hat{c}$ : Hill's est$^r$, tuning parameter $k$.

They argue: preserve $\frac{\log x}{\log n} = \frac{\log y}{\log m}$

Assume $\bar{F}(x) = cx^{-\alpha}\{1 + dx^{-\beta} + o(x^{-\beta})\}$

Assume $x \to \infty$, $n \to \infty$ s.t. $\frac{\log x}{\log n}$ is bounded away from 0 and $\infty$.

Thm!
$$E\left[\{\bar{F}_{\hat{\alpha},\hat{c}}(x) - \bar{F}(x)\}^2\{\bar{F}_{\alpha,c}(x)\}^{-2}\right]$$
$$= D_1(k; n, x)\{\bar{F}_{\alpha,c}(x)\}^{2}$$
$$= E\Big[\delta_1\left(\tfrac{k}{n}\right) - \delta_2(x) + \tfrac{1}{\sqrt{k}}Y_1 + \alpha f\left(\tfrac{k}{n}, x\right)\left\{\tfrac{Y_2}{\sqrt{k}} - a\left(\tfrac{k}{n}\right)^{\beta/\alpha}\right\}\Big]^2$$
+ smaller order terms.

$\delta_1\left(\frac{k}{n}\right) \sim dc^{-\beta/\alpha}\left(\frac{k}{n}\right)^{\beta/\alpha}$ $\quad \delta_2(x) \sim dx^{-\beta}$

**Bootstrap** Choose $m << n$

Resample ~~$\mathbb{X}^* = (X$~~

$\mathfrak{X}^* = (X^*_1, \ldots, X^*_m)$

$\hat{\alpha}^*, \hat{c}^*$ Hill est$^n$ based on $(X^*_1, \ldots, X^*_m)$

$$\hat{D}_1 = E' \{ \bar{F}_{\hat{\theta}^*(k)}(y) - \hat{\bar{F}}(y) \}^2$$

$E'$ : expectation ~~consistent~~ cond. on original data

Let $k(m,y)$ minimize $D_1(k; m, y)$

$\hat{k}(m,y)$ minimize $\hat{D}_1(k; m, y)$

Assume $y = O(n^{\frac{1}{\alpha+2\beta}} - \varepsilon)$

Theorem 3.1: for any $\varepsilon > 0$, $\lambda > 0$,

$$P\left\{ \frac{\hat{k}(m,y) - k_0(m,y)}{k_0(m,y)} > \varepsilon \right\} = O(n^{-\lambda})$$

STOR834 p.66

Objective is to find $\hat{k}$ so that

$$P\left[\frac{\hat{k}-k_0}{k_0} > \varepsilon\right] = o(\eta) \qquad (\ddagger)$$

where $\eta = n^{-2\beta/(\alpha+2\beta)} (\log n)^2 x^{-2\alpha}$

They argue:

$$\log \hat{k}(m,y) = \log C_1 + \gamma \log m + \xi(m,y)$$

smaller order

Regress $\log \hat{k}(m,y)$ on $\log m$, get estimators $\hat{C}_1$, $\hat{\gamma}$, set

$$\hat{k}(n,x) = \hat{C}_1 n^{\hat{\gamma}}$$

Then, $\hat{k}$ satisfies $(\ddagger)$.