0.1 Asymptotics of the Hill-Weissman Estimator

In this section we consider the special case of extreme value theory based on the Type I or Fréchet limit. Gnedenko [2] showed that a limit of the form

$$F^n(a_n x) \quad \rightarrow \quad \Phi_\alpha(x) \ = \ \exp\left(-cx^{-\alpha}\right), \ x \geq 0, \ \alpha > 0, \ c > 0, \tag{1}$$

holds if $1-F(x)$ is regularly varying with index α, and in that case a_n may without loss of generality be taken as the solution of $F(a_n) = 1 - 1/n$, and $c = 1$. Note that in this case, there is no location parameter to the distribution ($b_n = 0$), but for statistical purposes, it makes sense to retain c as well as α as an unknown parameter.

In this case, Weissman's representation [11, 12] for the asymptotic joint distribution of the k largest order statistics $m_1 \geq m_2 \geq \ldots m_k$ reduces to

$$L(\alpha, c \mid m_1, ..., m_k) \quad = \quad \prod_{i=1}^{k} \left(c\alpha m_i^{-\alpha-1}\right) \cdot \exp\left(-cm_k^{-\alpha}\right). \tag{2}$$

where the notation is intended to indicate that we are thinking of (2) as a likelihood function for the parameters α and c. The dependence on $m_1, ..., m_k$ will be omitted in many of the formulas. Taking logarithms, we want to minimize

$$\ell(\alpha, c) \ = \ -\log L(\alpha, c) \ = \ -k\log\alpha - k\log c + (\alpha - 1)\sum_{i=1}^{k}\log m_i + cm_k^{-\alpha}.$$

It is quickly established that this expression is minimized when $\alpha = \hat{\alpha}, \ c = \hat{c}$ where

$$\hat{\alpha} \ = \ \left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{m_i}{m_k}\right)^{-1}, \ \hat{c} \ = \ km_k^{\hat{\alpha}}. \tag{3}$$

Note, in particular, the simple direct formula for the estimator of α. The derivation is the same as that in [11], but that paper did it for the equivalent case where the limit distribution is Gumbel (the Fréchet model is turned into the Gumbel model by taking logarithms of the observations).

An alternative, even simpler, derivation of an equivalent result was given by Hill [9]. Hill assumed, in effect, that the relationship $1 - F(x) = cx^{-\alpha}$ is exact for $x \geq u$, for some known threshold u, but that $F(x)$ is unspecified for $x < u$. If data $X_1, ..., X_n$ are ordered so that $X_1 \geq X_2 \geq \ldots X_k > u \geq X_{k+1} \geq \ldots X_n$ then the likelihood function is

$$L(\alpha, c \mid X_1, ..., X_n) \quad = \quad \prod_{i=1}^{k}\left(\alpha c X_i^{-\alpha-1}\right) \cdot \left(1 - cu^{-\alpha}\right)^{n-k}$$

Taking logarithms and minimizing with respect to first c and then α leads to

$$\hat{\alpha} \ = \ \left(\frac{1}{k}\sum_{i=1}^{k}\log\frac{X_i}{u}\right)^{-1}, \ \hat{c} \ = \ \frac{k}{n}u^{\hat{\alpha}}. \tag{4}$$

Note, in particular, the similarity of the two estimators of α: in effect, the role of the threshold u in (4) is replaced by the kth largest order statistic in (3). (The different estimators of c arise because of different definitions: (3) uses the limit distribution for sample maxima whereas (4) assumes the same functional form directly for the individual observations. The two definitions differ by a factor of n, which is reflected in the estimates.)

The estimator $\hat{\alpha}$ in (4) is widely known as *Hill's estimator* but in the present section, to emphasize the close similarity with Weissman's [12] result, we shall call it the *Hill-Weissman estimator.*

In order to develop some asymptotics for this estimator, we assume an expansion of the form

$$1 - F(x) \quad = \quad cx^{-\alpha}\left\{1 + dx^{-\beta} + o(x^{-\beta})\right\}, \ x \to \infty. \qquad (5)$$

In general, the assumption (5) may be replaced by an assumption of *second-order regular variation* which allows the terms with $x^{-\alpha}$ and $x^{-\beta}$ to be replaced by general regularly varying functions; see in particular [3] for a survey of this theory and its applications (including the present one). This, in turn, is a special case of the general sceond-order regular variation theory of [4], For the present discussion, we mke the simpler assumption (5) which is sufficient for most practical applications, and easier to manipulate.

Our focus will be on the condition distribution of X given $X > u$, for some high threshold u. Let $Y_u = X/u$. Then the conditional probability $P\{Y_u > y \mid Y_u > 1\}$ is represented as

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

so, assuming it is valid to differentiate term by term, we calculate the density as

$$f_{Y_u}(y) \quad = \quad \alpha y^{-\alpha-1} + du^{-\beta}\left\{(\alpha+\beta)y^{-\alpha-\beta-1} - \alpha y^{-\alpha-1}\right\} + o(u^{-\beta}).$$

We note integrals of the form

$$\int_1^\infty (\log y)^k y^{-\alpha-1} dy \quad = \quad \alpha^{-k-1} k!$$

where we shall mainly be interested in the cases $k = 1$ and 2 but for non-integer k the same formula holds with $k!$ replaced by $\Gamma(k+1)$. We therefore deduce

$$\mathrm{E}(\log Y_u)^k \quad = \quad \alpha^{-k} k! + du^{-\beta} k! \left\{(\alpha+\beta)^{-k} - \alpha^{-k}\right\} = o(u^{-\beta}). \qquad (6)$$

Now let's consider the bias and variance of $\frac{1}{\hat{\alpha}} = \frac{1}{k}\sum_{i=1}^k \log \frac{X_i}{u}$ as an estimator of $\frac{1}{\alpha}$, where k is the number of exceedances of u. Since $\mathrm{E}(\log Y_u) = \frac{1}{\alpha} - du^{-\beta}\frac{\beta}{\alpha(\alpha+\beta)} + o(u^{-\beta})$, we deduce

$$\text{Bias of } \frac{1}{\hat{\alpha}} \quad \approx \quad -du^{-\beta}\frac{\beta}{\alpha(\alpha+\beta)}.$$

However, we also have from the $k=1$ and $k=2$ cases of (6) that $\mathrm{Var}(\log Y_u) \to \frac{1}{\alpha^2}$ as $u \to \infty$ and hence the variance of $\frac{1}{\alpha}$ is asymptotically $\frac{1}{k\alpha^2}$. However if the whole sample is of size n, and k is the random number of exceedances of u, we have $k \sim ncu^{-\alpha}$. Therefore, in large samples we have

$$\text{Variance of } \frac{1}{\hat{\alpha}} \approx \frac{1}{\alpha^2 ncu^{-\alpha}}.$$

Combining the espressions for bias and variance, and writing mean squared error (MSE) for the sum of squared bias and variance, we deduce

$$\text{MSE of } \frac{1}{\hat{\alpha}} \approx \frac{Au^{\alpha}}{n} + B^2 u^{-2\beta}$$

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

This asymptotic MSE is minimized with

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

which in turn leads to an asymptotic MSE of

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

The most important consequence of this is that the MSE is of $O\left(n^{-2\beta/(\alpha+2\beta)}\right)$ as $n \to \infty$, which could be arbitrarily slow for very small β but is of $O(n^{-1})$ as $\beta \to \infty$ — this makes sense, because in thaty limit the $cx^{-\alpha}$ result is exact and we are back in the original case considered by Hill.

0.2 Extension to the GPD

The above calculation was relatively straightforward because of the explicit closed form of the estimator. In most cases of interest (for example, estimating the two-parameter GPD or the three-parameter GEV distribution), there is no closed form estimator and the MLE is obtained by solving the likelihood equations. In such case, we may in principle proceed as follows. Suppose the negative log likelihood function based on n observations is $\ell_n(\theta)$ for some multidimensional parameter θ whose true value we shall write θ_0. Also write $\hat{\theta}_n$ for the MLE. The Taylor expansion

$$\nabla \ell_n(\hat{\theta}_n) - \nabla \ell_n(\theta_0) \approx \left(\hat{\theta}_n - \theta_0\right)^T \nabla^2 \ell_n(\theta_0)$$

leads to the approximation

$$\hat{\theta}_n - \theta_0 \approx -\left(\nabla^2 \ell_n(\theta_0)\right)^{-1} \nabla \ell_n(\theta_0).$$

Now suppose that as $n \to \infty$, $n^{-1}\nabla^2 \ell_n(\theta_0) \xrightarrow{p} J$ (the Fisher information matrix) and $n^{-1}\nabla \ell_n(\theta_0) \xrightarrow{p}$ $\mathbf{b}$ (bias due to model misspecification; if the model is correctly specified, $\mathbf{b} = \mathbf{0}$). Then for $\hat{\theta}_n$ we have, for large n,

$$\text{Bias} \;\approx\; J^{-1}\mathbf{b},\ \text{Covariance Matrix} \;\approx\; n^{-1}J^{-1}. \tag{7}$$

Now let's apply this to the case of the GPD, again under the assumption that the true distribution satisfies (5). Note that in the case where $1 - F(x) = cx^{-\alpha}$ is exact, we have

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

so the two forms are identical if $\sigma = \frac{u}{\alpha}$, $\xi = \frac{1}{\alpha}$. From now on, we treat these as the "true" GPD parameter values in this case.

In this model, the Fisher information matrix [10] is

$$J \;=\; \begin{pmatrix} \frac{1}{\sigma^2(1+2\xi)} & \frac{1}{\sigma(1+\xi)(1+2\xi)} \\ \frac{1}{\sigma(1+\xi)(1+2\xi)} & \frac{2}{(1+\xi)(1+2\xi)} \end{pmatrix}$$

provided $1 + 2\xi > 0$, and hence

$$J^{-1} \;=\; (1+\xi)\begin{pmatrix} 2\sigma^2 & -\sigma \\ -\sigma & (1+\xi) \end{pmatrix}$$

Now let's compute the $\mathbf{b}$ term in (7). The log likelihood for a single observation is

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

Hence,

$$\begin{aligned} \sigma\frac{\partial \ell}{\partial \sigma} &= -\frac{1}{\xi} + \left(\frac{1}{\xi} + 1\right)\left(1 + \xi\frac{y}{\sigma}\right)^{-1}, \\ \frac{\partial \ell}{\partial \xi} &= -\frac{1}{\xi^2}\log\left(1 + \xi\frac{y}{\sigma}\right) + \frac{1}{\xi}\left(\frac{1}{\xi} + 1\right)\left\{1 - \left(1 + \xi\frac{y}{\sigma}\right)^{-1}\right\}. \end{aligned}$$

To calculate $\mathbf{b}$, we need to find expressions for the expected values of these terms.

To recast in the notation of Section 0.1, we first make the substitutions $\sigma = \frac{u}{\alpha}$, $\xi = \frac{1}{\alpha}$, and also that if y denotes the excess over the threshold u, then $y = u(Y_u - 1)$ and so $1 + \xi\frac{y}{\sigma} = Y_u$. Also, by the same reasoning as led to (6)

$$\mathrm{E}\left(Y_u^{-1}\right) \;=\; \frac{\alpha}{\alpha+1} + du^{-\beta}\cdot\frac{\beta}{(\alpha+1)(\alpha+\beta+1)} + o(u^{-\beta}).$$

We now calculate the expectations of $\sigma\frac{\partial \ell}{\partial \sigma}$ and $\frac{\partial \ell}{\partial \xi}$, respectively, to be

$$-\alpha + (\alpha+1)\left\{\frac{\alpha}{\alpha+1} + du^{-\beta}\cdot\frac{\beta}{(\alpha+1)(\alpha+\beta+1)} + o(u^{-\beta})\right\} = du^{-\beta}\cdot\frac{\beta}{\alpha+\beta+1} + o(u^{-\beta})$$

and

$$\begin{aligned} & -\alpha^2\left\{\frac{1}{\alpha} - du^{-\beta}\frac{\beta}{\alpha(\alpha+\beta)}\right\} + \alpha(\alpha+1)\left\{\frac{1}{\alpha+1} - du^{-\beta}\frac{\beta}{(\alpha+1)(\alpha+\beta+1)}\right\} + o(u^{-\beta}) \\ = \;& du^{-\beta}\cdot\frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} + o(u^{-\beta}). \end{aligned}$$

Therefore, we conclude

$$\begin{aligned} \mathbf{b} &\sim du^{-\beta}\begin{pmatrix} \frac{1}{\sigma}\frac{\beta}{\alpha+\beta+1} \\ \frac{\alpha\beta}{(\alpha+\beta)(\alpha+\beta+1)} \end{pmatrix}, \\ J^{-1}\mathbf{b} &\sim du^{-\beta}\frac{(\alpha+1)\beta}{\alpha(\alpha+\beta)(\alpha+\beta+1)}\begin{pmatrix} \sigma(\alpha+2\beta) \\ 1-\beta \end{pmatrix}. \end{aligned}$$

Focussing on the second entries in these vectors, we deduce that $\hat{\xi}$ has asymptotic bias

$$du^{-\beta}\frac{(\alpha+1)\beta(1-\beta)}{\alpha(\alpha+\beta)(\alpha+\beta+1)}$$

and asymptotic variance (based on $k \approx ncu^{-\alpha}$ exceedances of the threshold

$$\frac{1}{k}\left(\frac{\alpha+1}{\alpha}\right)^2 \sim \frac{(\alpha+1)^2}{\alpha^2 ncu^{-\alpha}}.$$

0.2.1 Comparisons with the Hill-Weissman Estimator

For the Hill-Weissman estimator, we deduced that the bias was asymptotically $Bu^{-\beta}$, variance Au^{α}/n, with $B = -d\beta/(\alpha(\alpha+\beta)),\ A = 1/(\alpha^2 c)$.

For the GPD estimator, we get asymptotic bias $B'u^{-\beta}$, asymptotic variance $A'u^{\alpha}/n$, where $B‘ = d\beta(1-\beta)(\alpha+1)/(\alpha(\alpha+\beta)(\alpha+\beta+1)$.

The optimal MSE is proportional to

$$|B|^{2\alpha/(\alpha+2\beta)}A^{2\beta/(\alpha+2\beta)}$$

Therefore, the ratio of the optimal MSE for the GPD estimator to that of the Hill-Weissman estimator is

$$\left|\frac{B'}{B}\right|^{2\alpha/(\alpha+2\beta)}\left|\frac{A'}{A}\right|^{2\beta/(\alpha+2\beta)} = \left|\frac{(1-\beta)(\alpha+1)}{\alpha(\alpha+\beta)(\alpha+\beta+1)}\right|^{2\alpha/(\alpha+2\beta)}|\alpha+1|^{4\beta/(\alpha+2\beta)}$$

See Figure 1.

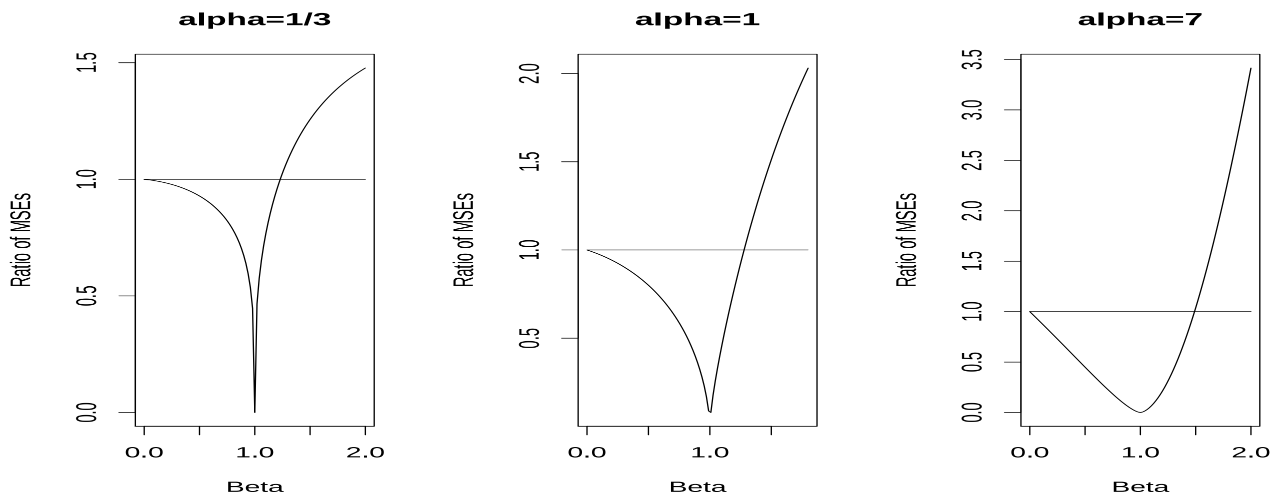

Figure 1: *Ratio of optimal mean squared error for the GPD estimator to that of the Hill-Weissman estimator, for a variety of values of α and β.*

0.3 Background References

The Hill estimator was introduced in [9] and the Weissman estimator, in its original form, in [12]. Asymptotic properties of the Hill estimator were obtained by [8, 5, 3] Optimality of the derived rate of convergence was proved by [6], and an adaptive estimator to achieve the optimal threshold was given by [7]. Many variants on the method exists, for example, [1] used a kernel-weighted version. The comparison of the two estimators was first derived in [10]. Many other authors have contributed to the theory and a more complete bibliography will be given later.

Bibliography

[1] S. Csörgő, P. Deheuvels, and D.M. Mason. Kernel estimates of the tail index of a distribution. *Annals of Statistics*, 13:1050–1077, 1985.

[2] B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44:423–453, 1943.

[3] C.M Goldie and R.L. Smith. Slow variation with remainder: A survey of the theory and its applications. *Quarterly Journal of Mathematics Oxford Series (2)*, 38:45–71, 1987.

[4] L. de Haan and U. Stadtmüller. Generalized regular variation of second order. *Journal of the Australian Mathematical Society, Series A*, 61:381–395, 1996.

[5] E. Haeusler and J.L. Teugels. On asymptotic normality of Hill's estimator for the exponent of regular variation. *Annals of Statistics*, 13:743–756, 1985.

[6] P. Hall and A.H. Welsh. Best attainable rates of convergence for estimates of parameters of regular variation. *Annals of Statistics*, 12:1079–1084, 1984.

[7] P. Hall and A.H. Welsh. Adaptive estimates of parameters of regular variation. *Annals of Statistics*, 13:331–341, 1985.

[8] Peter Hall. On some simple estimates of an exponent of regular variation. *Journal of the Royal Statistical Society, Series B*, 44:37–42, 1982.

[9] Bruce M. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3 (5):1163–1174, 1975.

[10] R.L. Smith. Estimating tails of probability distributions. *Annals of Statistics*, 15 (3):1174–1207, 1987.

[11] I. Weissman. Multivariate extremal processes generated by independent non-identically distributed random variables. *Journal of Applied Probability*, 12:477–487, 1975.

[12] I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73 (364):812–815, 1978.