LOCAL ASYMPTOTIC NORMALITY IN EXTREME VALUE INDEX ESTIMATION

Frank Marohn

Mathematisch-Geographische Fakultät, Katholische Universität Eichstätt,
85071 Eichstätt, Germany

Abstract. This paper deals with the estimation of the extreme value index in local extreme value models. We establish local asymptotic normality (LAN) under certain extreme value alternatives. It turns out that the central sequence occurring in the LAN expansion of the likelihood process is up to a rescaling procedure the Hill estimator. The central sequence plays a crucial role for the construction of asymptotic optimal statistical procedures. In particular, the Hill estimator is asymptotically minimax.

Key words and phrases: Extreme value distribution, extreme value index, domain of attraction, generalized Pareto distribution, Hill estimator, local asymptotic normality, central sequence, asymptotic efficiency.

1. Introduction

Problems in extreme value statistics are mainly concerned with the upper tail of a distribution function, which belongs to some (parametric or semiparametric) class of probability measures. A common view is that only large observations contain the relevant statistical information on the upper tail. One possibility to define observations to be large is to consider the k-largest order statistics $X_{n-k+1:n}, \dots, X_{n:n}$, where $X_{1:n} \le \cdots \le X_{n:n}$ denote the pertaining order statistics of the canonical projections $X_j : \mathbb{R}^n \to \mathbb{R}$ on the j-th coordinate, $j = 1, \dots, n$.

Since one has to extrapolate outside the range of the observations, regularity conditions for the upper tail of the distribution have to be made. A standard assumption in extreme value statistics is that the underlying and unknown distribution function F belongs to the (weak) domain of attraction of an extreme value distribution (EVD) G, i.e.

$$(1.1) \qquad P\{a_n^{-1}(X_{n:n} - b_n) \le x\} = F^n(a_n x + b_n) \to G(x), \qquad x \in \mathbb{R}$$

for some normalizing constants $a_n > 0$, $b_n \in \mathbb{R}$. Note that in our notation we do not distinguish between a distribution function and its pertaining distribution.

Up to a location and scale parameter, G has to be one of the following EVDs

$$(1.2) \qquad G_\beta(x) = \begin{cases} \exp(-x^{-1/\beta}), & x > 0, & \text{if } \beta > 0 & \text{"Fréchet"} \\ \exp(-(-x)^{-1/\beta}), & x < 0, & \text{if } \beta < 0 & \text{"Weibull"} \\ \exp(-e^{-x}), & x \in \mathbb{R}, & \text{if } \beta = 0 & \text{"Gumbel"} \end{cases}$$

(see e.g. Galambos (1987)). Condition (1.1) is a fairly weak one. Most common text book distributions satisfy this condition.

A major problem in extreme value statistics is to estimate the shape parameter β (also called extreme value index or tail index for $\beta > 0$). There is a rich literature about the estimation of β based on intermediate and extreme order statistics $X_{n-k_n+1:n}, \ldots, X_{n:n}$, i.e. $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$, see e.g. Hill (1975), Pickands (1975), Dekkers et al. (1989), among many others.

Recently, this estimation problem has been treated in a local set-up in the articles of Falk (1995a,b) and Wei (1995). This means, roughly speaking, the study of the distributions of the rescaled deviations $\delta_n^{-1}(\hat{\beta}_n - (\beta_0 + \vartheta \delta_n))$ under $\beta_0 + \vartheta \delta_n$, $\vartheta \in \mathbb{R}$, where $(\hat{\beta}_n)_n$ is a sequence of estimators, β_0 is a fixed parameter, and $(\delta_n)_n$ is a suitable chosen normalizing sequence tending to zero. The quantity ϑ is usually called the "local parameter". The local approach, which was originally suggested by LeCam (1953), has turned out to be a very general and fruitful concept in asymptotic statistics. The most important case, which is applicable to many situations, is the so called *local asymptotic normality* (LAN)-case introduced by LeCam (1960). In the LAN case, asymptotically optimal statistical procedures are well known. For certain extreme value models the LAN situation holds as it was shown by the papers mentioned above. We also establish LAN in this paper but propose different models. For the readers' convenience, the notation LAN is recalled at the end of this section.

The local approach in extreme value statistics leads to interesting and partially unexpected results. For example, consider the Pareto distribution

$$W_\beta(x) = 1 - x^{-1/\beta}, \qquad x \ge 1, \ \beta > 0,$$

which belongs to the Fréchet domain of attraction. Then, asymptotically, the maximum likelihood estimator based on the k_n-largest order statistics is given by

$$(1.3) \qquad \hat{\beta}_{n,k_n}^H = \frac{1}{k_n - 1} \sum_{j=1}^{k_n} \log \frac{X_{n-j+1:n}}{X_{n-k_n+1:n}}$$

which is the Hill estimator (see e.g. Reiss (1989), Section 9.5). Using conditional techniques, it is readily seen that

$$\mathcal{L}(\sqrt{k_n}(\hat{\beta}_{n,k_n}^H - \beta_0) \mid W_{\beta_0}^n) \to N(0, \beta_0^2) \qquad \text{weakly.}$$

By $\mathcal{L}(X \mid P)$ we denote the distribution of a random element X under a probability measure P.

In the class of Pareto distributions $(W_\beta)_{\beta>0}$, the Hill estimator is outperformed by the simple estimator

$$(1.4) \qquad \tilde{\beta}_{n,k_n} := \frac{\log(X_{n-k_n+1:n})}{\log(n/k_n)}$$

since

$$\mathcal{L}(\sqrt{k_n}\log(n/k_n)(\tilde{\beta}_{n,k_n} - \beta_0) \mid W_{\beta_0}^n) \to N(0, \beta_0^2) \quad \text{weakly.}$$

This follows at once from asymptotic normality results of intermediate order statistics (Falk (1989), Theorem 2.1). But more can be said. If one chooses the local extreme value alternatives $\beta_n(\vartheta) := \beta_0 + \vartheta/(\sqrt{k_n}\log(n/k_n))$, then LAN can be shown with the central sequence $\tilde{\beta}_{n,k_n}$ (Falk (1995b), Theorem 2.3). This shows that $\tilde{\beta}_{n,k_n}$ is not only asymptotically normal but asymptotically efficient in the LAN-sense. In particular, the k_n-largest order statistic contains asymptotically all the relevant information on the tail index β, i.e. is asymptotically sufficient (see Strasser (1985), Theorem 81.4, cf. with Falk (1995b), Theorem 2.4). Hence, the Hill estimator is not efficient even in the class of Pareto distributions (the opposite conclusion from Smith (1987), p. 1176).

On the other hand, Theorem 3.2 in Falk (1995b) shows that the central sequence depends again only on the k_n-largest order statistic if a scale parameter is added. This can only hold if the local model in consideration leads asymptotically to a degenerate situation (see Remark 4 in this paper). We overcome this lack by choosing different local extreme value alternatives (thereby treating the scale parameter also as a nuisance parameter).

In the remainder of this section we recall the notion of LAN, thereby introducing some further notations used in this work. For details the reader is referred to LeCam (1986), LeCam and Yang (1990), and Strasser (1985). Denote by $N(\mu, \Gamma)$ the normal distribution with expectation $\mu \in \mathbb{R}^k$ and positive definite covariance matrix Γ. By $\langle \cdot, \cdot \rangle_\Gamma$ we denote the inner product on $\mathbb{R}^k$ defined by $\langle s, t \rangle_\Gamma = s'\Gamma t$, $s, t \in \mathbb{R}^k$. The norm induced by $\langle \cdot, \cdot \rangle_\Gamma$ is denoted by $\| \ \|_\Gamma$. The statistical experiment $(\mathbb{R}^k, \mathcal{B}^k, \{N(\vartheta, \Gamma^{-1}) : \vartheta \in \mathbb{R}^k\})$ is called Gaussian shift on $(\mathbb{R}^k, \langle \cdot, \cdot \rangle_\Gamma)$. Recall that (by definition) $N(0, \Gamma^{-1})$ is just the standard normal distribution on $(\mathbb{R}^k, \langle \cdot, \cdot \rangle_\Gamma)$.

Consider a statistical experiment $E = (\Omega, \mathcal{A}, \{P_\vartheta : \vartheta \in \Theta\})$, where Θ is an open subset of $\mathbb{R}^k$. In the iid set-up we embed E into a sequence of localized experiments $E_{n,\vartheta_0} = (\Omega^n, \mathcal{A}^n, \{P_{\vartheta_0 + \delta_n t}^n : t \in T_n(\vartheta_0)\})$ with localization point $\vartheta_0 \in \Theta$, where $\delta_n \downarrow 0$ is a rescaling rate and $T_n(\vartheta_0) = \{t \in \mathbb{R}^k : \vartheta_0 + \delta_n t \in \Theta\}$. Note that $T_n(\vartheta_0) \to \mathbb{R}^k$ as $n \to \infty$. Recall that the rescaling procedure is necessary in order to get non-degenerate limit experiments. The sequence $(E_{n,\vartheta_0})_n$ is called local asymptotic normal (in ϑ_0) if it converges weakly to a Gaussian shift, equipped with a suitable inner product which is allowed to depend on ϑ_0. To be more precise, we have *local asymptotic normality* (LAN) of $(E_{n,\vartheta_0})_n$ if the likelihood ratio with base ϑ_0 admits the asymptotic expansion

$$\frac{dP_{\vartheta_0 + \delta_n t}^n}{dP_{\vartheta_0}^n} = \exp\left(\langle t, Z_{n,\vartheta_0} \rangle_{\Gamma(\vartheta_0)} - \frac{1}{2}\|t\|_{\Gamma(\vartheta_0)}^2 + o_{P_{\vartheta_0}^n}(1) \right), \quad t \in \mathbb{R},$$

where $Z_{n,\vartheta_0} : \Omega^n \to \mathbb{R}$ satisfies

$$\mathcal{L}(Z_{n,\vartheta_0} \mid P^n_{\vartheta_0}) \to_{n\to\infty} N(0, \Gamma(\vartheta_0)^{-1}).$$

The notation $o_P(1)$ indicates a stochastic remainder term which converges to zero in probability as $n \to \infty$. The sequence $(Z_{n,\vartheta_0})_n$ plays a crucial role in testing and estimation theory and is called *central sequence*. For the sake of completeness we remark that the quadratic term in the LAN expansion can also be random which leads to the so called LAMN (locally asymptotically mixed normal)-case, see e.g. Jeganathan (1982), and that Gaussian shifts can be defined on infinite dimensional Hilbert spaces (i.e. tangent spaces) which becomes important in nonparametric statistics, see Pfanzagl and Wefelmeyer (1982).

The paper is organized as follows. In Section 2 we specify our local models and present the results. In Section 3 we discuss some statistical consequences and compare the results with related papers known in the literature. The proofs are postponed to Section 4.

2. Local extreme value models

In the following we exclude the Gumbel case ($\beta = 0$) in our model but allow, in contrast to Hill, also negative values of the shape parameter. Moreover, we restrict ourselves to certain (semiparametric) classes of distribution functions, which belong to the domain of attraction of an EVD (1.2). To be more precise: We consider δ-neighborhoods of generalized Pareto distributions (GPDs) W_β, given by

$$(2.1) \quad W_\beta(x) := \begin{cases} 1 - x^{-1/\beta}, & x \geq 1, & \beta > 0 & \text{``Pareto''} \\ 1 - (-x)^{-1/\beta}, & x \in [-1, 0] & \beta < 0 & \text{``Uniform etc.''} \end{cases}$$

Denote by $\omega(F) := \sup\{x \in \mathbb{R} : F(x) < 1\} \in (-\infty, \infty]$ the right endpoint of a distribution function F and w_β the Lebesgue density of W_β. We say that F belongs to a δ-neighborhood of a GPD for some fixed constants $\delta > 0$, $D > 0$, iff $\omega(F) = \omega(W_\beta)$ for some β and F is differentiable with derivative f on $(x_0, \omega(F))$ for some $x_0 = x_0(\beta, f) < \omega(F)$ such that

$$\left| \frac{f(x)}{w_\beta(x)} - 1 \right| \leq D(1 - W_\beta(x))^\delta, \quad x \in (x_0, \omega(W_\beta))$$

(briefly $F \in \mathcal{Q}_{\delta,D}(W_\beta)$). Note that $\omega(W_\beta) = \infty$ for $\beta > 0$ and $\omega(W_\beta) = 0$ for $\beta < 0$. In particular, the EVDs (1.2) belong to a δ-neighborhood for $\delta = 1$. The importance of the GPDs is explained by the fact, that F belongs to the domain of attraction of an EVD if, and only if, the upper tail of F can be approximated in a suitable sense by the upper tail of a GPD (Balkema and de Haan (1974), Pickands (1975)). For a review of the basic role played by δ-neighborhoods of GPDs in extreme value statistics we refer to Chapter 2 of Falk *et al.* (1994).

Let F be a distribution function, whose upper tail belongs to some parametric family, that is, we assume that

$$F(x) - F_\beta(x), \quad x \geq x_0(\beta), \quad \beta \neq 0$$

for some unknown point $x_0(\beta)$, where $(F_\beta)_\beta$ is a parametric family of distribution functions. A natural choice are the GPD (2.1), i.e.

$$(2.2) \qquad F_\beta(x) = W_\beta(x), \qquad x \geq x_0(\beta), \quad \beta \neq 0$$

or, more generally,

$$(2.3) \qquad F_\beta \in \mathcal{Q}_{\delta,D}(W_\beta).$$

Introducing a scale parameter $\sigma > 0$ our starting model is

$$F_{\beta,\sigma}(x) := F_\beta(x/\sigma), \qquad x \geq \sigma x_0(\beta)$$

with F_β as in (2.2) or (2.3).

The asymptotic setting requires a rescaling procedure, i.e. local alternatives, in order to get non-degenerate limit experiments. Considering first the scale parameter as a nuisance parameter, we choose the following local alternatives ($\beta_0 \neq 0, \vartheta \neq 0$)

$$(2.4) \qquad \begin{aligned} \beta_n(\vartheta) &= \beta_0 + \vartheta k_n^{-1/2} \\ \sigma_n(\vartheta) &= \sigma_0(n/k_n)^{\beta_0-\beta_n(\vartheta)} = \sigma_0 \exp(-\vartheta k_n^{-1/2} \log(n/k_n)), \end{aligned}$$

where $(k_n)_n$ is any sequence of positive integers such that $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$. We remark that in certain threshold models the shape and scale alternatives (2.4) are independent in the sense that they lead to a diagonal Fisher information matrix (for details see Marohn (1995)). Our local model is then the statistical experiment

$$E_{n,k_n,\beta_0} := (\mathbb{R}^{k_n}, \mathcal{B}^{k_n}, \{\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_n(\vartheta),\sigma_n(\vartheta)}) : \vartheta \in \Theta_n\}),$$

where

$$\begin{aligned} \Theta_n &:= \begin{cases} \{\vartheta \in \mathbb{R} : \beta_n(\vartheta) > 0\} & \text{if} \quad \beta_0 > 0 \\ \{\vartheta \in \mathbb{R} : \beta_n(\vartheta) < 0\} & \text{if} \quad \beta_0 < 0 \end{cases} \\ &= \begin{cases} (-\beta_0 k_n^{1/2}, \infty) & \text{if} \quad \beta_0 > 0 \\ (-\infty, -\beta_0 k_n^{1/2}) & \text{if} \quad \beta_0 < 0. \end{cases} \end{aligned}$$

As in Falk (1995b) we allow that the point $x_0(\beta_n(\vartheta))$ tends to the upper endpoint $\omega(W_{\beta_0})$ as n increases, but not too fast, since we require

$$(2.5) \qquad \limsup_{n\to\infty}(x_0(\beta_n(\vartheta)))^{1/\beta_0}\frac{k_n}{n} < 1, \qquad \vartheta \in \mathbb{R} \setminus \{0\}.$$

Note that nothing is said on the convergence of $\sigma_n(\vartheta)$. We have $\sigma_n(\vartheta) \to \sigma_0$ as $n \to \infty$ if, and only if, $k_n^{-1/2} \log n \to 0$ as $n \to \infty$. But the convergence of $\sigma_n(\vartheta)$ is

actually not needed to establish LAN, as a careful study of the proof of Theorem 2.1 shows.

Before we treat δ-neighborhoods we consider the model (2.2), where the upper tail of F_β coincides with the tail of a GPD. The following result shows that the central sequence is up to a rescaling procedure just the Hill estimator.

THEOREM 2.1. (LAN of $(E_{n,k_n,\beta_0})_n$)   *Consider the model* (2.2). *Suppose that the sequence* $(k_n)_n$ *satisfies condition* (2.5). *Then the likelihood ratio with base* (β_0, σ_0) *admits the expansion*

$$\frac{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_n(\vartheta),\sigma_n(\vartheta)})}{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0,\sigma_0})}$$

$$= \exp\left(\langle \vartheta, Z_{n,k_n,\beta_0} \rangle_{\beta_0^{-2}} - \frac{1}{2}\|\vartheta\|^2_{\beta_0^{-2}} + r_n(\vartheta) \right)$$

with the central sequence

$$Z_{n,k_n,\beta_0}(X_{n-k_n+1:n}, \ldots, X_{n:n}) = \sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0),$$

where $\hat{\beta}^H_{n,k_n}$ *is the Hill estimator* (1.3) *and the remainder term* r_n *satisfies for every compact set* $K \subset \mathbb{R}$ *and* $\varepsilon > 0$

$$\lim_{n \to \infty} \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0,\sigma_0})\left\{ \sup_{\vartheta \in K} |r_n(\vartheta)| > \varepsilon \right\} = 0.$$

Observe that the central sequence is invariant under scale transformations. The Theorem states the weak convergence of E_{n,k_n,β_0} to the Gaussian shift on $(\mathbb{R}, \langle \cdot, \cdot \rangle_{\beta_0^{-2}})$. Since the remainder term is asymptotically vanishing uniformly on compact sets, the sequence $(E_{n,k_n,\beta_0})_n$ is equicontinuous (Strasser (1985), Theorem 80.13).

Remark 1.   From the discussion given in the previous section we already know that in a pure shape parameter model, i.e. the scale parameter is dropped, the Hill estimator is no longer the central sequence which depends only on the k_n-largest order statistic (Falk (1995b), Theorem 2.3). In this case the rate of convergence of the local shape alternatives must be of a higher order than $1/\sqrt{k_n}$ which is intuitively clear. If one adds a scale parameter as a nuisance parameter and cancel out the arising scale effects, the Hill estimator becomes relevant.

Remark 2.   Wei (1995) considered the conditional distribution of the k_n largest order statistics $X_{n-k_n+1:n}, \ldots, X_{n:n}$ given $X_{n-k_n:n}$, where the k_n largest order statistics are distributed according to the local alternatives $\beta_0 + t/\sqrt{k_n}$, but $X_{n-k_n:n}$ itself follows the hypothesis β_0. Thereby, the underlying distribution function w.r.t. β_0 belongs to the domain of attraction of the Fréchet distribution. Using this local model LAN was shown and efficiency of the proposed conditional

maximum likelihood estimator was verified. Wei motivated her conditional view by the fact that LAN cannot hold for the joint distribution of the k_n-largest order statistics under L_2-differentiability, since the score function w.r.t. the joint distribution of the k_n-largest order statistics normalized by $1/\sqrt{k_n}$ is unbounded in probability and tends therefore not a normal law. But extending the model by a scale parameter σ, i.e. making the model less informative, it is possible to get LAN with the rate $1/\sqrt{k_n}$ as Theorem 2.1 shows.

Next we show that Theorem 2.1 remains true for the model (2.3), that is, the condition that the upper tail of a distribution function F coincides with the upper tail of a GPD (2.1) is replaced by the condition that it is in a δ-neighborhood of some GPD, provided $(k_n)_n$ tends to infinity sufficiently slow.

COROLLARY 2.1. *Suppose that $(F_\beta)_\beta$ is a parametric family such that $F_\beta \in \mathcal{Q}_{\delta,D}(W_\beta)$ and that the sequence of positive integers $(k_n)_n$ satisfies the conditions (2.5) and*

$$\lim_{n \to \infty} k_n (k_n/n)^\delta = 0.$$

Then the LAN expansion of Theorem 2.1 holds.

If the scale parameter is no longer considered to be a nuisance parameter we choose the shape alternatives

$$\beta_n(\vartheta) = \beta_0 + \vartheta k_n^{-1/2}$$

as before and the curves of scale alternatives

$$(2.6) \qquad \sigma_n(\vartheta, \xi) = \sigma_n(\vartheta) \left(1 + \frac{\xi k_n^{-1/2}}{\beta_0 \sigma_0} \right)^{\beta_n(\vartheta)}$$

$$= \sigma_0 (n/k_n)^{\beta_0 - \beta_n(\vartheta)} (1 + \xi k_n^{-1/2}/(\beta_0 \sigma_0))^{\beta_n(\vartheta)}$$

with $\sigma_n(\vartheta)$ as in (2.4), where $(k_n)_n$ is any sequence of integers such that $k_n \to \infty$, $k_n/n \to 0$, and $k_n^{-1/2} \log n \to 0$ as n tends to infinity. Note that the local parametrization of the scale parameter depends on the shape alternatives and that $\sigma_n(\vartheta, \xi) \to \sigma_0$ as the sample size n increases, i.e. $\sigma_n(\vartheta, \xi)$ is a local alternative.

Again, we allow the sequence $x_0(\beta_n(\vartheta))$ to tend to $\omega(W_{\beta_0})$, but not too fast, since we require again that condition (2.5) holds. Our local model is then

$$E_{n,k_n,\beta_0,\sigma_0}$$
$$:= (\mathbb{R}^{k_n}, \mathcal{B}^{k_n}, \{\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_n(\vartheta),\sigma_n(\vartheta,\xi)}) : (\vartheta, \xi) \in \Theta_n \times \Xi_n\})$$

with Θ_n as before and

$$\Xi_n := \{\zeta \subset \mathbb{R} : 1 + \zeta k_n^{-1/2}/(\beta_0 \sigma_0) > 0\}$$
$$= \begin{cases} (-\beta_0 \sigma_0 k_n^{1/2}, \infty), & \text{if } \beta_0 > 0 \\ (\ \infty,\ \ \beta_0 \sigma_0 k_n^{1/2}), & \text{if } \beta_0 < 0. \end{cases}$$

The next theorem shows, that the sequence $(E_{n,k_n,\beta_0,\sigma_0})_n$ converges weakly to the two-dimensional Gaussian shift on $(\mathbb{R}^2, \langle \cdot, \cdot \rangle_\Gamma)$ with covariance matrix

$$\Gamma = \Gamma(\beta_0, \sigma_0) = \begin{pmatrix} \beta_0^{-2} & 0 \\ 0 & \beta_0^{-2}\sigma_0^{-2} \end{pmatrix}.$$

In order not to overload the present paper we omit the proof. It uses similar arguments as given in the proof of Theorem 2.1 (for details see Marohn (1995)). Note that the asymptotic independence of the central sequences $Z_{1,n}$ and $Z_{2,n}$ given below is an easy consequence of the fact that the spacings of an ordered iid standard exponential sample are independent (see e.g. Reiss (1989), Theorem 1.6.1).

THEOREM 2.2. (LAN of $(E_{n,k_n,\beta_0,\sigma_0})_n$) *Consider the model (2.2) and suppose that condition (2.5) is satisfied. Then the likelihood ratio with base (β_0, σ_0) admits the expansion*

$$\frac{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_n(\vartheta),\sigma_n(\vartheta,\xi)})}{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0,\sigma_0})}$$

$$= \exp\left(\langle (\vartheta,\xi)^T, (Z_{1,n}, Z_{2,n})^T \rangle_\Gamma - \frac{1}{2}\|(\vartheta,\xi)^T\|_\Gamma^2 + r_n(\vartheta,\xi) \right)$$

with the central sequence

$$\begin{pmatrix} Z_{1,n}(X_{n-k_n+1:n}, \ldots, X_{n:n}) = \sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0) \\ Z_{2,n}(X_{n-k_n+1:n}) = \sqrt{k_n}\left(\left(\frac{k_n}{n}\right)^{\beta_0} (|X_{n-k_n+1:n}| - \sigma_0) \right) \end{pmatrix},$$

where the remainder term $r_n(\vartheta,\xi)$ satisfies for every compact set $K \subset \mathbb{R}^2$ and $\varepsilon > 0$

$$\lim_{n\to\infty} \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0,\sigma_0})\left\{ \sup_{(\vartheta,\xi)\in K} |r_n(\vartheta,\xi)| > \varepsilon \right\} = 0.$$

If, in addition, $(k_n)_n$ satisfies $\lim_{n\to\infty} k_n(k_n/n)^\delta = 0$ then the LAN expansion remains true for δ-neighborhoods (2.3).

Note that the first component of the central sequence is the central Z_{n,k_n,β_0} sequence of Theorem 2.1.

Remark 3. If the shape parameter β_0 is known, i.e. $\vartheta = 0$, then we get a scale model with the local scale alternatives $\sigma_n(0,\xi)$ and central sequence $Z_{2,n}$. In this case, the scale alternatives are of the order $O(1/\sqrt{k_n})$. As in the pure shape parameter model the central sequence depends only on the k_n-largest order

statistic (see Remark 1). This shows that these two parameters —shape and scale—act asymptotically very closely. If $\xi = 0$, then $\sigma_n(\vartheta,\xi) = \sigma_n(\vartheta)$ and we are in the situation of Theorem 2.1, where the scale parameter is treated as a nuisance parameter.

Remark 4. In Falk (1995*b*) the local alternatives

$$\tilde{\beta}_n(\vartheta) := \beta_0 - \vartheta\beta_0 k_n^{-1/2}/\log(n/k_n)$$
$$\tilde{\sigma}_n(\xi) := 1 - \xi k_n^{-1/2}/\beta_0$$

were proposed (where in our notation $F_{1/\beta,1/\sigma}$ is considered instead of $F_{\beta,\sigma}$). Using these parametrizations the expansion of the log-likelihood ratio

$$(\vartheta + \xi)Z_{2,n} - \frac{1}{2}(\vartheta + \xi)^2 + o_{F_{\beta_0,1}^n}$$

with $Z_{2,n}$ as above (with $\sigma_0 = 1$) was established, which shows the weak convergence to a Gaussian experiment $G = (\mathbb{R}, \mathcal{B}, \{N(\psi(\vartheta,\xi),1) : (\vartheta,\xi) \in \mathbb{R}^2\})$ with $\psi(\vartheta,\xi) = \vartheta + \xi$ (in the sense of Definition 2.3 of Milbrodt and Strasser (1985)). This is not a Gaussian shift. The covariance of the Gaussian process $X(\vartheta,\xi) := (\vartheta + \xi)Z - (\vartheta + \xi)^2/2$ with Z $N(0,1)$-distributed is

$$\mathrm{Cov}(X(\vartheta_1,\xi_1), X(\vartheta_2,\xi_2)) = (\vartheta_1 + \xi_1)(\vartheta_2 + \xi_2)$$
$$= (\vartheta_1,\xi_1)A(\vartheta_2,\xi_2)^T, \qquad A := \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

which is bilinear, no doubt, but the matrix A is not positive definite.

In particular, we have

$$\|\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F_{\tilde{\beta}_n(\vartheta),\tilde{\sigma}_n(-\vartheta)}^n) - \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F_{\beta_0,1}^n)\|$$
$$\to 0$$

as $n \to \infty$, where $\|P - Q\| := \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$ denotes the variational distance between probability measures P and Q defined on some measurable space $(\Omega, \mathcal{A})$. Consequently, differences of the shape parameter are completely cancelled out by the scale transformation $\tilde{\sigma}_n(-\vartheta)$. The parametrization of the scale parameter does not depend on the parametrization of the shape parameter. Theorem 2.2 shows that the scale effects dominate in the local model proposed by Falk (1995*b*). The two-dimensional starting problem is approximated by an one-dimensional limit experiment. Note that the rate of convergence of the shape alternatives $\tilde{\beta}_n(\vartheta)$ is too high compared with $\beta_n(\vartheta)$.

3. Discussion

The Hill estimator has been extensively studied in the literature. Asymptotic consistency results are given e.g. in Mason (1982), Deheuvels *et al.* (1988), and de Haan (1994). An asymptotic normality result was first shown by Hall (1982). Refinements can be found in Davis and Resnick (1984), Hall and Welsh (1984, 1985), Csörgő and Mason (1985), Csörgő *et al.* (1985), Häusler and Teugels (1985), Beirlant and Teugels (1987, 1989), de Haan (1994) among others. Asymptotic behaviour of the Hill estimator based on exceedances over certain thresholds instead of order statistics can be found e.g. in Goldie and Smith (1987) and Smith (1987).

Some of these papers are addressed to the optimal choice of k_n. In the present (unbiased) situation it is desirable to choose k_n as large as possible. But the condition $k_n(k_n/n)^\delta \to 0$, $n \to \infty$, in Corollary 2.1 rules out an optimal choice. In practice one should take $k_n = cn^{\delta/(1+\delta)}$ for some appropriate chosen constant $c > 0$. This is a general feature: If an optimal rate k_n^* exists (in a certain sense) one has a limiting centered normal distribution whenever $k_n = o(k_n^*)$ and a non-centered one (but usually with a smaller variance) for k_n^* (for a further discussion see Csörgő *et al.* (1987)). It was shown by Hall (1982) that the optimal rate is given by $k_n^* = n^{2\delta/(2\delta+1)}$. Csörgő *et al.* (1985) generalizes the Hill estimator. Their estimate achieves the optimal rate as the Hill estimator but has better asymptotic performance by minimizing the mean square error. As long as optimal rates are concerned, optimality is focussed with the bias-variance trade-off. In this paper optimality means having an optimal property within a certain class of (asymptotically unbiased) estimators. Thereby, for any reasonable comparison estimators must be based on the same number of order statistics.

In view of the importance of the Hill estimator, we will now discuss some statistical consequences in estimation and testing theory in detail and tackle the problem of joint estimation of the shape and scale parameter. Throughout we assume that $(k_n)_n$ satisfies the conditions of the previous section.

LAN-efficiency of the Hill estimator

(a) *Fisher efficiency*

The results of Section 2 show not only the asymptotic normality of the Hill estimator $\hat\beta_{n,k_n}^H$, i.e.

$$\mathcal{L}(\sqrt{k_n}(\hat\beta_{n,k_n}^H - \beta_0) \mid F_{\beta_0,\sigma_0}^n) \to N(0,\beta_0^2)$$

weakly, but also the asymptotic efficiency in the sense of Fisher. That the Hill estimator is also consistent for $\beta < 0$ should not surprise too much. If X_i is W_β-distributed for $\beta < 0$ then $Z_i := -1/X_i$ is $W_{-\beta}$-distributed and $\hat\beta_{n,k_n}^H(Z_{n-k_n+1:n}, \ldots, Z_{n:n}) = -\hat\beta_{n,k_n}^H(X_{n-k_n+1:n}, \ldots, X_{n:n})$. Before we say some words on the definition of Fisher's efficiency, we remark that

$$\mathcal{L}(\sqrt{k_n}(\hat\beta_{n,k_n}^H - \beta_n(\vartheta) \mid \Gamma_{\beta_n(\vartheta),\sigma_n(\vartheta)}^n) \to N(0,\beta_0^2)$$

weakly, which is a well known consequence of LeCam's First Lemma (see e.g. LeCam and Yang (1990), Theorem 1, Chapter 3). Note that the limit distribution is independent of the local parameter ϑ.

Asymptotic efficiency in the sense of Fisher means that the variance of the weak limit distribution of the rescaled estimation errors $\sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0)$ attains the lower bound of the variance β_0^2, $\beta_0 \neq 0$, within the class of regular estimators based on the k_n-largest order statistics. A sequence of estimators based on the k_n-largest order statistics $T_{n,k_n} = T_{n,k_n}(X_{n-k_n+1:n}, \ldots, X_{n:n})$ is called regular, if the limit distribution of the rescaled estimation errors $\sqrt{k_n}(T_{n,k_n} - \beta_n(\vartheta))$ under $F^n_{\beta_n(\vartheta),\sigma_n(\vartheta)}$ is independent of ϑ, i.e.

$$\mathcal{L}(\sqrt{k_n}(T_{n,k_n} - \beta_n(\vartheta)) \mid F^n_{\beta_n(\vartheta),\sigma_n(\vartheta)}) \rightarrow Q_{\beta_0}$$

weakly. In this case a probability measure R_{β_0} on $(\mathbb{R}, \mathcal{B})$ exists such that

$$Q_{\beta_0} = N(0, \beta_0^2) * R_{\beta_0},$$

where $*$ denotes the convolution. This is the asymptotic Convolution Theorem due to Hájek (1970), cf. Strasser (1985), Remark 83.14 (see also Section 8.4 of Pfanzagl (1994) for an extensive discussion). Hence, the lower bound of the variance in the limit distribution of a regular sequence of estimators, based on the k_n-largest order statistics, is β_0^2.

This lower bound is not achieved by other estimators known in the literature, e.g. the moment estimator (Dekkers et al. (1989)), the Pickands estimator (1975) and its refinements (Drees (1995), Falk (1994a)). This shows the superior of the Hill estimator for the model in consideration. But at least two points should be stressed if one compares these estimators simply by means of the variances of their limit distributions.

First, in our model we explicitly assume that $\beta \neq 0$, an assumption which is not made for the other competitive estimators which have a limiting centered normal distribution also under $\beta = 0$, where the von Mises parametrization of the class of EVD (1.2) is used. But LAN does not hold for $\beta_0 \neq 0$ and $\beta_0 = 0$ simultaneously (see Marohn (1995), Theorem 3.2.1). This can be seen as a consequence of the fact that $\beta_0 = 0$ is some kind of singularity point in the von Mises parametrization.

Second, we assume that the location parameter is known. The Hill estimator is not location invariant in contrast to the other ones. At the present we do not know whether LAN remains to be true if a location parameter μ is added (as a nuisance parameter). But if LAN holds, then Theorem 3.3 in Falk (1995b) (see also Marohn (1995), Theorem 5.1.1) suggests that the central sequence is no longer given by the Hill estimator. If μ is known, then the Hill estimator occurring in the LAN-expansion has to be replaced by

$$\frac{1}{k_n - 1} \sum_{j=1}^{k_n-1} \log \left| \frac{\mu - X_{n-j+1:n}}{\mu - X_{n-k_n+1:n}} \right|.$$

For $\beta < 0$ an unknown location parameter means that the upper endpoint of the distribution function—which is equal to μ—is not known. In this case, an apparent idea is to replace μ by the largest observation $X_{n:n}$ which yields the estimator

$$\frac{1}{k_n - 1} \sum_{j=2}^{k_n-1} \log \frac{X_{n:n} - X_{n-j+1:n}}{X_{n:n} - X_{n-k_n+1:n}}.$$

This estimator was proposed by Falk (1995c) if $\beta < -1/2$ is known.

(b) *Wolfowitz efficiency*

The Hill estimator $\hat{\beta}^H_{n,k_n}$ is also asymptotically efficient in the sense of Wolfowitz (1965). He suggested to compare the quality of estimators by the degree of their "concentration" about the true value of the parameter, that means, covering probabilities come up in the Wolfowitz definition of efficiency.

Let $T_{n,k_n} = T_{n,k_n}(X_{n-k_n+1:n}, \ldots, X_{n:n})$ be any regular sequence of estimators. For the bounded loss functions

$$l_s = 1 - 1_{[-s,s]}, \quad s > 0,$$

we get covering probabilities

$$F^n_{\beta_0,\sigma_0}\{\sqrt{k_n}(T_{n,k_n} - \beta_0) \in [-s,s]\} = 1 - \int l_s(\sqrt{k_n}(T_{n,k_n} - \beta_0))dF^n_{\beta_0,\sigma_0}.$$

By Hájek's Convolution Theorem, l_s is continuous Q_{β_0}-almost everywhere and therefore we have weak convergence

$$\lim_{n\to\infty} \int l_s(\sqrt{k_n}(T_{n,k_n} - \beta_0))dF^n_{\beta_0,\sigma_0} = \int l_s dN(0,\beta_0^2) * R_{\beta_0}.$$

Since l_s is subconvex (i.e. $l_s(x) = l_s(-x)$, $\{l_s \le a\}$ is convex, $a \ge 0$) the Lemma of Anderson (see e.g. Strasser (1985), Lemma 38.21) states that

$$\int l_s dN(0,\beta_0^2) * R_{\beta_0} \ge \int l_s dN(0,\beta_0^2).$$

Consequently, we arrive at the asymptotic upper bound

$$\lim_{n\to\infty} F^n_{\beta_0,\sigma_0}\{\sqrt{k_n}(T_{n,k_n} - \beta_0) \subset [-s,s]\} \le 1 - \int l_s dN(0,\beta_0^2)$$

which is attained by $(\hat{\beta}^H_{n,k_n})_n$. Hence, $(\hat{\beta}^H_{n,k_n})_n$ is also optimal in the sense of maximizing covering probabilities (cf. Reiss (1989), Section 9.5). For a further discussion of these two efficiency concepts the reader is referred to the book by Ibragimov and Has'minskii (1981), Section 11 of Chapter II.

(c) *Minimax property*

In extension to (b) the central sequence (i.e. the Hill estimator) is asymptotically minimax. Suppose that l is a bounded and continuous loss function. If

$\tilde{T}_{n,k_n} = \tilde{T}_{n,k_n}(X_{n-k_n+1:n}, \ldots, X_{n:n})$ is any sequence of estimates (not necessarily unbiased and normal distributed) for the deviations $\vartheta = \sqrt{k_n}(\beta_n(\vartheta) - \beta_0)$, then

$$\liminf_{n \to \infty} \sup_{\vartheta \in K} \int l(\vartheta - \tilde{T}_{n,k_n}) dF^n_{\beta_n(\vartheta), \sigma_n(\vartheta)} \geq \int l \, dN(0, \beta_0^2)$$

for compact sets $K \subset \mathbb{R}$. This is a special case of Hájek-LeCam's asymptotic minimax bound for risk functions. Since $(E_{n,k_n,\beta_0})_n$ is equicontinuous, the asymptotic lower bound is achieved by the central sequence Z_{n,k_n,β_0}, i.e.

$$\lim_{n \to \infty} \sup_{\vartheta \subset K} \int l(\vartheta - \sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0)) dF^n_{\beta_n(\vartheta), \sigma_n(\vartheta)} = \int l \, dN(0, \beta_0^2).$$

A similar result holds within the class of asymptotically median unbiased estimators (Pfanzagl (1970), cf. Strasser (1985), Section 83). For unbounded and discontinuous loss functions see the discussion 83.7 given in Strasser (1985).

Application in testing theory

Suppose that $F_\beta \in \mathcal{Q}_{\delta,D}(W_\beta)$ and that one is interested in the two sided testing problem $\beta = \beta_0$ against $\beta \neq \beta_0$, considering the scale parameter as a nuisance parameter. Based on the k_n-largest observations, we embed this testing problem into the asymptotic testing problem

$$\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0, \sigma_0})$$
$$\text{against} \{ \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_n(\vartheta), \sigma_n(\vartheta)}) : \vartheta \in \Theta_n \setminus \{0\} \}.$$

Denote by $u_\alpha = \Phi^{-1}(\alpha)$ the α-quantile of the standard normal distribution function Φ. The asymptotically optimal test of level α is then given by

$$\varphi_n^* = \begin{cases} 1, & \text{if } |\sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0)| > |\beta_0| u_{1-\alpha/2} \\ 0, & \text{if } |\sqrt{k_n}(\hat{\beta}^H_{n,k_n} - \beta_0)| < |\beta_0| u_{1-\alpha/2} \end{cases}$$

i.e. $(\varphi_n^*)_n$ attains asymptotically the upper bound for power functions

$$\lim_{n \to \infty} E_{\beta_n(\vartheta), \sigma_n(\vartheta)} \varphi_n^* = \Phi\left(u_{\alpha/2} + \frac{\vartheta}{|\beta_0|}\right) + \Phi\left(u_{\alpha/2} - \frac{\vartheta}{|\beta_0|}\right)$$

(see e.g. Strasser (1985), Section 82). Note that $(-\infty, -|\beta_0| u_{1-\alpha/2}) \cup (|\beta_0| u_{1-\alpha/2}, \infty)$ is an asymptotic critical region for φ_n^*. The proposed test φ_n^* is in addition a level-α-test, if $u_{1-\alpha/2}$ is replaced by the $(1-\alpha/2)$-quantile of the (exact) distribution of the test statistic $\sqrt{k_n}(\hat{\beta}^H_{n,k_n}/\beta_0 - 1)$. For example, in the case of GPDs the distribution of the test statistic is of gamma type (see the Proof of Theorem 2.1).

Joint estimation of (β, σ)

Next we want to discuss the problem of joint estimation of (β, σ) in the shape-scale model. Suppose that β_0 is known (i.e. $\vartheta = 0$). Then the efficient estimator of σ

$$\hat{\sigma}_{n,\beta_0} := \left(\frac{k_n}{n}\right)^{\beta_0} |X_{n-k_n+1:n}|$$

depends on the shape parameter. Next, we want to replace β_0 by an estimator $\hat{\beta}_n$ in order to get an asymptotically efficient estimator $(\hat{\beta}^H_{n,k_n}, \hat{\sigma}_{n,\hat{\beta}_n})$ for the shape-scale parameter. First, we consider an estimator $\hat{\beta}_n$ which depends only on the k_n-largest order statistic $X_{n-k_n+1:n}$, since in this case the estimator $\hat{\sigma}_{n,\hat{\beta}_n}$ is asymptotically independent of the Hill estimator. The only reasonable choice is $\hat{\beta}_n = \log|X_{n-k_n+1:n}|/\log(n/k_n)$ (cf. (1.4)), which is a consistent estimator. But we have $\hat{\sigma}_{n,\hat{\beta}_n} = 1$. Of course, such a result is not surprising, since the k_n-largest order statistic cannot estimate the scale and shape parameter simultaneously in a satisfactory way. So we have to take into account also the larger order statistics in order to get an asymptotically efficient estimator of the shape-scale parameter, if at all. We already know that the whole information on the unknown shape parameter is (asymptotically) contained in the k_n-largest order statistics $X_{n-k_n+1:n}, \ldots, X_{n:n}$. Recall that a central sequence is asymptotically sufficient, see Strasser (1985), Section 81. But replacing β_0 by the Hill estimator $\hat{\beta}^H_{n,k_n}$, the estimator $(\hat{\beta}^H_{n,k_n}, \hat{\sigma}_{n,\hat{\beta}^H_{n,k_n}})$ is not asymptotically efficient for (β_0, σ_0), since its limit distribution is degenerate. First, observe that the difference $\sigma_n(\vartheta, \xi) - \sigma_0$ is of order $O(k_n^{-1/2} \log(n/k_n))$. Straightforward calculations show the degeneracy property

$$\frac{k_n^{1/2}}{\log(n/k_n)} (\hat{\sigma}_{n,\hat{\beta}^H_{n,k_n}} - \sigma_0) = \sigma_0 Z_{1,n} + o_{F^n_{\beta_0,\sigma_0}}(1).$$

The considerations above indicate that there exists no asymptotically efficient estimator for (β_0, σ_0) independent of (β_0, σ_0).

A degeneracy property in extreme value statistics was also recognized by Falk (1994b). It was shown that a suitable data transformation reduces the estimation of extreme quantiles of the unknown distribution function (which is closely related to estimation of the shape parameter) to the problem of estimating the location and scale parameter of a certain exponential family. The proposed estimators are close to our ones. (The reader should be aware that "degeneracy property" means here that the asymptotic distribution of the joint estimators in consideration is degenerate, while Remark 5 is concerned with a degenerate limit model.)

Höpfner and Jacod (1994) showed a degeneracy property of the maximum likelihood estimator $(\hat{\alpha}_n, \hat{\xi}_n)$ (see their Corollary 2.1), where the local model is given by a Poisson point process with intensity

$$\alpha \xi x^{-(1+\alpha)} 1_{(0,\infty)}(t) 1_{(0,\infty)}(x) dt dx, \qquad \alpha > 0, \ \xi > 0,$$

which is at stage n observed in a certain window D_n. Under suitable chosen local alternatives (similar to (2.6)) and growing conditions on D_n they showed LAN.

We remark that the intensity above does not arise from a scale family of Pareto distributions.

It was recognized by Höpfner (1997) that the degeneracy property of the estimator $(\hat{\beta}_{n,k_n}^{H}, \hat{\sigma}_{n,\hat{\beta}_{n,k_n}^{H}})$ turns out to be an optimality property with respect to certain sequences of contiguous alternatives. We consider the one-dimensional local model at (β_0, σ_0) with the alternatives $\beta_n(\vartheta)$ and $\sigma_n(\vartheta, 0) = \sigma_n(\vartheta)$. The following assertion is a consequence of LeCam's Third Lemma (LeCam (1986), Chapter 6, Proposition 5) and Hájek's Convolution Theorem (Hájek (1970)), for details we refer to Höpfner (1997).

Optimality property of $(\hat{\beta}_{n,k_n}, \hat{\sigma}_{n,\hat{\beta}_{n,k_n}^{H}})$:

(i) For $\vartheta \in \mathbb{R}$ the estimator $(\hat{\beta}_{n,k_n}^{H}, \hat{\sigma}_{n,\hat{\beta}_{n,k_n}^{H}})$ satisfies

$$\mathcal{L}\left(\left(\sqrt{k_n}(\hat{\beta}_{n,k_n}^{H} - \beta_n(\vartheta)), \frac{\sqrt{k_n}}{\sigma_0 \log(n/k_n)}(\hat{\sigma}_{n,\hat{\beta}_{n,k_n}^{H}} - \sigma_n(\vartheta))\right) \mid F_{\beta_n(\vartheta),\sigma_n(\vartheta)}^{n}\right)$$
$$\to \mathcal{L}(Z, Z),$$

where Z is $N(0, \beta_0^2)$-distributed.

(ii) If $(\hat{\beta}_n, \hat{\sigma}_n)$ is any estimator of (β, σ) such that for $\vartheta \in \mathbb{R}$

$$\mathcal{L}\left(\left(\sqrt{k_n}(\hat{\beta}_n - \beta_n(\vartheta)), \frac{\sqrt{k_n}}{\sigma_0 \log(n/k_n)}(\hat{\sigma}_n - \sigma_n(\vartheta))\right) \mid F_{\beta_n(\vartheta),\sigma_n(\vartheta)}^{n}\right) \to \mathcal{L}(U, V),$$

where $\mathcal{L}(U, V)$ is independent of ϑ, then $\operatorname{Var}(U) \geq \beta_0^2$, $\operatorname{Var}(V) \geq \beta_0^2$, and one has $\mathcal{L}(U) = \mathcal{L}(V) = N(0, \beta_0^2)$ if, and only if, $\mathcal{L}(U, V)$ is degenerate: $\mathcal{L}(U, V) = \mathcal{L}(Z, Z)$.

Concluding remark. Hill's estimator becomes relevant for estimating the extreme value index if scale effects but no location effects occur. A characteristic and nice property of the Hill estimator is to be scale invariant—a property which is meaningless if no scale alternatives are involved in the model. Hence, the results of this paper clarify the role of this popular estimator.

4. Proofs

PROOF OF THEOREM 2.1. Define $Y_{n-j+1:n} = \beta_0^{-1} \log(|X_{n-j+1:n}|/\sigma_0)$, $1 \leq j \leq n$. Then

(4.1) $k_n^{1/2}(Y_{n-k_n+1:n} - \log(n/k_n)) \to_{n\to\infty} N(0, 1)$

weakly under F_{β_0,σ_0}^{n} which, in turn, implies $Y_{n-k_n+1:n}/\log(n/k_n) \to_{n\to\infty} 1$ in F_{β_0,σ_0}^{n}-probability. The asymptotic normality of the intermediate order statistic (4.1) follows at once from Theorem 2.1 in Falk (1989). Note that $\beta_0^{-1} \log(|X|/\sigma_0)$ is standard exponential if X is distributed according to W_{β_0,σ_0}.

Denote by $w_{\beta,\sigma}$ the density of $W_{\beta,\sigma}$. From the density formula of order statistics (see e.g. (1.4.8) in Reiss (1989)) we obtain

$$\log \frac{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F_{\beta_n(\vartheta),\sigma_n(\vartheta)}^{n})}{d\mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F_{\beta_0,\sigma_0}^{n})}(y_1, \ldots, y_{k_n})$$

$$= \sum_{j=1}^{k_n} \log \frac{w_{\beta_n(\vartheta),\sigma_n(\vartheta)}(y_j)}{w_{\beta_0,\sigma_0}(y_j)} + (n - k_n) \log \frac{W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(y_1)}{W_{\beta_0,\sigma_0}(y_1)}$$

provided $\sigma_0 x_0(\beta_n(\vartheta)) \le y_1 \le \cdots \le y_{k_n}$. But the probability of this event converges to one

$$F_{\beta_0,\sigma_0}^n\{X_{n-k_n+1:n} \ge \sigma_0 x_0(\beta_n(\vartheta))\} \to_{n\to\infty} 1$$

which is an easy consequence of (2.5) and (4.1). In the following we assume that $X_{n-k_n+1:n} \ge \sigma_0 x_0(\beta_n(\vartheta))$. We have

$$\frac{W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n}) - W_{\beta_0,\sigma_0}(X_{n-k_n+1:n})}{W_{\beta_0,\sigma_0}(X_{n-k_n+1:n})}$$

$$= \frac{\exp(-Y_{n-k_n+1:n})}{1 - \exp(-Y_{n-k_n+1:n})}$$
$$\times \left\{ 1 - \exp\left(\left(1 - \frac{\beta_0}{\beta_n(\vartheta)}\right) Y_{n-k_n+1:n} + \frac{1}{\beta_n(\vartheta)} \log \frac{\sigma_n(\vartheta)}{\sigma_0}\right)\right\}$$

$$= \frac{\exp(-Y_{n-k_n+1:n})}{1 - \exp(-Y_{n-k_n+1:n})}$$
$$\times \left\{ 1 - \exp\left(\frac{1}{\beta_n(\vartheta)} \vartheta k_n^{-1/2}(Y_{n-k_n+1:n} - \log(n/k_n))\right)\right\}$$

$$= \frac{1}{n} \frac{\exp(-(Y_{n-k_n+1:n} - \log(n/k_n)))}{1 - (k_n/n)\exp(-(Y_{n-k_n+1:n} - \log(n/k_n)))}$$
$$\times \left\{ -\frac{1}{\beta_n(\vartheta)} \vartheta k_n^{1/2}(Y_{n-k_n+1:n} - \log(n/k_n)) + O_{F_{\beta_0,\sigma_0}^n}(k_n^{-1})\right\}$$

$$= O_{F_{\beta_0,\sigma_0}^n}(n^{-1})$$

by (4.1). Hence, by Taylor expansion, we conclude that

$$(4.2) \qquad (n - k_n) \log \frac{W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n})}{W_{\beta_0,\sigma_0}(X_{n-k_n+1:n})}$$
$$= -\frac{1}{\beta_n(\vartheta)} \vartheta k_n^{1/2}(Y_{n-k_n+1:n} - \log(n/k_n)) + o_{F_{\beta_0,\sigma_0}^n}(1).$$

Furthermore,

$$(4.3) \qquad \sum_{j=1}^{k_n} \log \frac{w_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})}$$

$$= \sum_{j=1}^{k_n} \log \left\{ (|X_{n-j+1:n}|/\sigma_0)^{1/\beta_0 - 1/\beta_n(\vartheta)} \frac{\beta_0}{\beta_n(\vartheta)} \left(\frac{\sigma_n(\vartheta)}{\sigma_0}\right)^{1/\beta_n(\vartheta)}\right\}$$

$$= \sum_{j=1}^{k_n} \left\{ \frac{\beta_n(\vartheta) - \beta_0}{\beta_n(\vartheta)} Y_{n-j+1:n} + \frac{1}{\beta_n(\vartheta)} \log \frac{\sigma_n(\vartheta)}{\sigma_0} - \log \frac{\beta_n(\vartheta)}{\beta_0}\right\}$$

$$- \sum_{j=1}^{k_n} \left\{ \frac{1}{\beta_n(\vartheta)} \vartheta k_n^{-1/2}(Y_{n-j+1:n} - \log(n/k_n))\right.$$

$$\left. - \vartheta \beta_0^{-1} k_n^{-1/2} + \frac{1}{2}\vartheta^2 \beta_0^{-2} k_n^{-1} + O(k_n^{-3/2})\right\}.$$

Adding (4.2) and (4.3) we arrive at the expansion

$$\frac{1}{\beta_n(\vartheta)}\vartheta k_n^{-1/2}\sum_{j=1}^{k_n}(Y_{n-j+1:n}-Y_{n-k_n+1:n}-1)$$

$$+\vartheta k_n^{1/2}\left(\frac{1}{\beta_n(\vartheta)}-\frac{1}{\beta_0}\right)+\frac{1}{2}\vartheta^2\beta_0^{-2}+o_{F_{\beta_0,\sigma_0}^n}(1)$$

$$=\vartheta\beta_0^{-1}k_n^{-1/2}\sum_{j=1}^{k_n}(Y_{n-j+1:n}-Y_{n-k_n+1:n}-1)$$

$$-\frac{1}{2}\vartheta^2\beta_0^{-2}+o_{F_{\beta_0,\sigma_0}^n}(1).$$

Hence, the central sequence is given by

$$\beta_0 k_n^{-1/2}\sum_{j=1}^{k_n}(\beta_0^{-1}\log|X_{n-j+1:n}|-\beta_0^{-1}\log|X_{n-k_n+1:n}|-1)$$

$$=(k_n-1)^{1/2}\left\{\left(\frac{1}{k_n-1}\sum_{j=1}^{k_n-1}\log|X_{n-k_n+1:n}|\right)-\log|X_{n-k_n+1:n}|-\beta_0\right\}$$

$$+o_{F_{\beta_0,\sigma_0}^n}(1)$$

$$=k_n^{1/2}(\hat{\beta}_{n,k_n}^H(X_{n-k_n+1:n},\ldots,X_{n:n})-\beta_0)+o_{F_{\beta_0,\sigma_0}^n}(1).$$

The LAN expansion is shown if

$$(4.4)\qquad k_n^{-1/2}\sum_{j=1}^{k_n}(Y_{n-j+1:n}-Y_{n-k_n+1:n}-1)\to_{n\to\infty}N(0,1)$$

weakly under F_{β_0,σ_0}^n. Since, by assumption, the underlying distribution function is ultimately a GPD, we conclude from the first part of Theorem 5.4.5 in Reiss (1989) that the left-hand side of (4.4) is stochastically equivalent to $k_n^{-1/2}\sum_{j=1}^{k_n}(\eta_{n-j+1:n}-\eta_{n-k_n+1:n}-1)$, where $\eta_1,\eta_2,\ldots$ are iid from a standard exponential distribution. It was observed by Dekkers et al. (1989) (see their first part of Lemma 3.4) that this sum is asymptotically standard normal. Note that we have equality in distribution to $k_n^{-1/2}\sum_{j=1}^{k_n-1}(\eta_j-1)$, which is a consequence of the well-known behaviour of order statistics of an iid sample of standard exponential random variables (see e.g. Reiss (1989), Theorem 1.6.1).

Since, by (4.1),

$$\sup_{\vartheta\in K}\left|1-\exp\left(\frac{1}{\beta_n(\vartheta)}\vartheta k_n^{-1/2}(Y_{n-k_n+1:n}-\log(n/k_n))\right)\right|=o_{F_{\beta_0,\sigma_0}^n}(1),$$

$\sup_{\vartheta\in K}\vartheta/\beta_n(\vartheta)$ is bounded for n large, and

$$\sup_{\vartheta\in K}\left|\log\frac{\beta_n(\vartheta)}{\beta_0}-\left(\frac{\vartheta k_n^{-1/2}}{\beta_0}-\frac{1}{2}\frac{\vartheta^2 k_n^{-1}}{\beta_0^2}\right)\right|$$

$$\leq\frac{1}{3}\frac{k_n^{-3/2}}{|\beta_0^3|}\sup_{\vartheta\in K}|\vartheta^3|\to_{n\to\infty}0$$

the assertion for the remainder term follows and the proof is complete. $\square$

For the Proof of Corollary 2.1 we need the following lemma, which can easily be established.

LEMMA 4.1.   *Suppose $F \in \mathcal{Q}_{\delta,D}(W_\beta)$. Then*

$$\left| \frac{1 - F(x)}{1 - W_\beta(x)} - 1 \right| \le C(1 - W_\beta(x))^\delta, \quad x \in (x_0, \omega(W_\beta)),$$

where the constant $C = D/(1 + \delta)$ is independent of β.

PROOF OF COROLLARY 2.1.   First, the uniform convergence of extremes (Reiss (1989), Corollary 5.5.5) implies

$$\| \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid F^n_{\beta_0,\sigma_0}) - \mathcal{L}((X_{n-k_n+1:n}, \ldots, X_{n:n}) \mid W^n_{\beta_0,\sigma_0}) \|$$
$$= O((k_n^{1/2}(k_n/n)^\delta + k_n/n).$$

Hence, from the Proof of Theorem 2.1 we conclude that

$$\sum_{j=1}^{k_n} \log \frac{w_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})} + (n - k_n) \log \frac{W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n})}{W_{\beta_0,\sigma_0}(X_{n-k_n+1:n})}$$

admits the LAN expansion of the Theorem w.r.t. $F^n_{\beta_0,\sigma_0}$. So it remains to show

$$(4.5) \qquad (n - k_n) \log \frac{F_{\beta_0,\sigma_0}(X_{n-k_n+1:n})}{W_{\beta_0,\sigma_0}(X_{n-k_n+1:n})} \to_{n\to\infty} 0$$

$$(4.6) \qquad (n - k_n) \log \frac{F_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n})}{W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n})} \to_{n\to\infty} 0$$

and

$$(4.7) \qquad \sum_{j=1}^{k_n} \log \frac{f_{\beta_0,\sigma_0}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})} \to_{n\to\infty} 0$$

$$(4.8) \qquad \sum_{j=1}^{k_n} \log \frac{f_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})}{w_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})} \to_{n\to\infty} 0$$

in $W^n_{\beta_0,\sigma_0}$-probability.

Taking Lemma 4.1 into account, we have for $x \ge x_0(\beta)$

$$\left| \frac{F_{\beta,\sigma}(x)}{W_{\beta,\sigma}(x)} - 1 \right| = \left| \frac{1 - W_{\beta,\sigma}(x) - (1 - F_{\beta,\sigma}(x))}{W_{\beta,\sigma}(x)} \right|$$

$$= \left| 1 - \frac{1 - F_{\beta,\sigma}(x)}{1 - W_{\beta,\sigma}(x)} \right| \frac{1 - W_{\beta,\sigma}(x)}{W_{\beta,\sigma}(x)}$$

$$\le \frac{C(1 - W_{\beta,\sigma}(x))^{1+\delta}}{W_{\beta,\sigma}(x)}.$$

With $Y_{n-j+1:n}$ defined as in the Proof of Theorem 2.1 we have

$$
\begin{aligned}
(n - k_n)&(1 - W_{\beta_0,\sigma_0}(X_{n-k_n+1:n}))^{1+\delta} \\
&= (n - k_n)\exp(-(1+\delta)Y_{n-k_n+1:n}) \\
&= \frac{n - k_n}{n}k_n \left(\frac{k_n}{n}\right)^\delta \exp(-(1+\delta)(Y_{n-k_n+1:n} - \log(n/k_n))) \to_{n\to\infty} 0
\end{aligned}
$$

in $W^n_{\beta_0,\sigma_0}$-probability. This shows (4.5) by the Taylor expansion $\log(1+x) = x + O(x^2)$, $x \to 0$. Similarly, (4.6) follows from

$$
\begin{aligned}
(n - k_n)&(1 - W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-k_n+1:n}))^{1+\delta} \\
&= (n - k_n)\exp(-(1+\delta)Y_{n-k_n+1:n} \\
&\qquad\qquad + (1+\delta)\vartheta\beta_n(\vartheta)^{-1}k_n^{-1/2}(Y_{n-k_n+1:n} - \log(n/k_n))) \\
&= \frac{n - k_n}{n}k_n \left(\frac{k_n}{n}\right)^\delta \exp(-(1+\delta)(Y_{n-k_n+1:n} - \log(n/k_n))) \\
&\qquad \times \exp((1+\delta)\vartheta\beta_n(\vartheta)^{-1}k_n^{-1/2}(Y_{n-k_n+1:n} - \log(n/k_n))) \to_{n\to\infty} 0
\end{aligned}
$$

in $W^n_{\beta_0,\sigma_0}$-probability.

Next, we establish (4.7) and (4.8). A Taylor expansion shows

$$
\begin{aligned}
\log\frac{f_{\beta,\sigma}(x)}{w_{\beta,\sigma}(x)} &= \log(1 + f_{\beta,\sigma}(x)/w_{\beta,\sigma}(x) - 1) \\
&= (f_{\beta,\sigma}(x)/w_{\beta,\sigma}(x) - 1) + O((1 - W_{\beta,\sigma}(x))^{2\delta}), \quad x \to \omega(W_\beta).
\end{aligned}
$$

Hence, we get

$$
\begin{aligned}
\sum_{j=1}^{k_n}&\log\frac{f_{\beta_0,\sigma_0}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})} \\
&= \sum_{j=1}^{k_n}\left(\frac{f_{\beta_0,\sigma_0}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})} - 1\right) + \sum_{j=1}^{k_n}O((1 - W_{\beta_0,\sigma_0}(X_{n-j+1:n}))^{2\delta}).
\end{aligned}
$$

Now,

$$
\begin{aligned}
\left|\sum_{j=1}^{k_n}\right.&\left.\left(\frac{f_{\beta_0,\sigma_0}(X_{n-j+1:n})}{w_{\beta_0,\sigma_0}(X_{n-j+1:n})} - 1\right)\right| \\
&\leq D\sum_{j=1}^{k_n}(1 - W_{\beta_0,\sigma_0}(X_{n-j+1:n}))^\delta \\
&\leq Dk_n(1 - W_{\beta_0,\sigma_0}(X_{n-k_n+1:n}))^\delta \\
&= Dk_n\left(\frac{k_n}{n}\right)^\delta \exp(-\delta(Y_{n-k_n+1:n} - \log(n/k_n))) \to_{n\to\infty} 0
\end{aligned}
$$

in $W^n_{\beta_0,\sigma_0}$-probability. Similarly one shows

$$\sum_{j=1}^{k_n} O((1 - W_{\beta_0,\sigma_0}(X_{n-j+1:n}))^{2\delta}) = o_{W^n_{\beta_0,\sigma_0}}(1)$$

and (4.7) is established.

Applying the same arguments we get

$$\left| \sum_{j=1}^{k_n} \left(\frac{f_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})}{w_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n})} - 1 \right) \right|$$

$$\leq k_n \left(\frac{k_n}{n} \right)^\delta \exp(-\delta(Y_{n-k_n+1:n} - \log(n/k_n)))$$

$$\times \exp(\vartheta\beta_n(\vartheta)^{-1} k_n^{1/2}(Y_{n-k_n+1:n} - \log(n/k_n))) \to_{n\to\infty} 0$$

and

$$\sum_{j=1}^{k_n} O((1 - W_{\beta_n(\vartheta),\sigma_n(\vartheta)}(X_{n-j+1:n}))^{2\delta}) \to_{n\to\infty} 0$$

in $W^n_{\beta_0,\sigma_0}$-probability which shows (4.8).

The assertion for the remainder term follows now as in the Proof of Theorem 2.1. □

Acknowledgements

REFERENCES

Balkema, A. A. and de Haan, L. (1974). Residual life time at great age, *Ann. Probab.*, **2**, 792–804.

Beirlant, J. and Teugels, J. L. (1987). The asymptotic behaviour of Hill's estimator, *Theory Probab. Appl.*, **31**, 463–469.

Beirlant, J. and Teugels, J. L. (1989). Asymptotic normality of Hill's estimator, *Extreme Value Theory* (eds. J. Hüsler and R.-D. Reiss), *Lecture Notes in Statist.*, **51**, 148–155, Springer, New York.

Csörgő, M., Horvath, L. and Révész, P. (1987). On the optimality of estimating the tail index and a naive estimator, *Austral. J. Statist.*, **29**, 166–178.

Csörgő, S. and Mason, D. M. (1985). Central limit theorems for sums of extreme value values, *Math. Proc. Cambridge Philos. Soc.*, **98**, 547–558.

Csörgő, S., Deheuvels, P. and Mason, D. (1985). Kernel estimator of the tail index of a distribution, *Ann. Statist.*, **13**, 1050–1077.

Davis, R. A. and Resnick, S. A. (1984). Tail estimates motivated by extreme value theory, *Ann. Statist.*, **12**, 1467–1487.

De Haan, L. (1994). Extreme value statistics, *Extreme Value Theory and Applications* (eds. J. Galambos *et al*), **1**, 93–122, Kluwer, Dordrecht.

Deheuvels, P., Häusler, E. and Mason, D. (1988). Almost sure convergence of the Hill estimator, *Math. Proc. Cambridge Philos. Soc.*, **104**, 371–381.

Dekkers, L. M., Einmahl, J. H. J. and De Haan, L. (1989). A moment estimator for the index of an extreme-value distribution, *Ann. Statist.*, **17**, 1833–1855.

Drees, H. (1995). Refined Pickands' estimator of the extreme value index, *Ann. Statist.*, **23**, 2059–2080.

Falk, M. (1989). A note on uniform asymptotic normality of intermediate order statistics, *Ann. Inst. Statist. Math.*, **41**, 19–29.

Falk, M. (1994a). Efficiency of convex combinations of Pickands estimator of the extreme value index, *Nonparametric Statistics*, **4**, 133–147.

Falk, M. (1994b). Extreme quantile estimation in δ-neighborhoods of generalized Pareto distributions, *Statist. Probab. Lett.*, **20**, 9–21.

Falk, M. (1995a). On testing the extreme value index via the POT-method, *Ann. Statist.*, **23**, 2013–2035.

Falk, M. (1995b). LAN of extreme order statistics, *Ann. Inst. Statist. Math.*, **47**, 693–717.

Falk, M. (1995c). Some best parameter estimates for distributions with finite endpoint, *Statistics*, **27**, 115–125.

Falk, M., Hüsler, J. and Reiss, R.-D. (1994). *Laws of Small Numbers: Extremes and Rare Events*, DMV Seminar, **23**, Birkhäuser, Basel.

Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd ed., Krieger, Malabar.

Goldie, C. M. and Smith, R. L. (1987). Slow variation with remainder theory and applications, *Quart. J. Math. Oxford Ser.* (2), **38**, 45–71.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates, *Zeit. Wahrscheinlichkeitsth.*, **14**, 323–330.

Hall, P. (1982). On some simple estimates of an exponent of regular variation, *J. Roy. Statist. Soc. Ser. B*, **44**, 37–42.

Hall, P. and Welsh, A. H. (1984). Best attainable rates of convergence for estimates of parameters of regular variation, *Ann. Statist.*, **12**, 1079–1084.

Hall, P. and Welsh, A. H. (1985). Adaptive estimates of parameters of regular variation, *Ann. Statist.*, **13**, 331–341.

Häusler, E. and Teugels, J. L. (1985). On asymptotic normality of Hill's estimator for the exponent of regular variation, *Ann. Statist.*, **13**, 743–756.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution, *Ann. Statist.*, **3**, 1163–1174.

Höpfner, R. (1997). Two comments on parameter estimation in stable processes, *Mathematical Methods of Statistics*, **6**, 125–134.

Höpfner, R. and Jacod, J. (1994). Some remarks on the joint estimation of the index and the scale parameter, for stable processe, *Asymptotic Statistics* (eds. M. Hušková and P. Mandl), 273–284, Physika, Wien.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*. Springer, Berlin.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal, *Sankhyā Ser. A*, **44**, part 2, 173–212.

LeCam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates, *University of California Publications in Statistics*, **1**, 277–330.

LeCam, L. (1960). Locally asymptotically normal families of distributions, *University of California Publications in Statistics*, **3**, 37–98.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*, Springer Series in Statistics, Sprinter, New York.

LeCam, L. and Yang, G. L. (1990). *Asymptotics in Statistics. Some Basic Concepts*, Springer Series in Statistics, Sprinter, New York.

Mason, D. (1982). Laws of large numbers for sums of extreme values, *Ann. Probab.*, **10**, 754–764.

Marohn, F. (1995). Contributions to a local treatment in extreme value statistics, Habilitation thesis, Katholische Universität Eichstätt.

Milbrodt, H. and Strasser, H. (1985). Limits of triangular arrays of experiments, *Infinitely Divisible Statistical Experiments* (eds. A. Janssen, H. Milbrodt and H. Strasser), *Lecture Notes in Statist.*, **27**, 14–54, Springer, Berlin.

Pfanzagl, J. (1970). On the asymptotic efficiency of median unbiased estimates, *Ann. Math. Statist.*, **41**, 1500–1509.

Pfanzagl, J. (1994). *Parametric Statistical Theory*, De Gruyter, Berlin.

Pfanzagl, J. and Wefelmeyer, W. (1982). Contributions to a general asymptotic statistical theory, *Lecture Notes in Statist.*, **13**, Springer, Berlin.

Pickands, J. III (1975). Statistical inference using extreme order statistics, *Ann. Statist.*, **3**, 119–131.

Reiss, R.-D. (1989). *Approximate Distributions of Order Statistics (with Applications to Nonparametric Statistics)*, Springer Series in Statistics, Sprinter, New York.

Smith, R. L. (1987). Estimating tails of probability distributions, *Ann. Statist.*, **15**, 1174–1207.

Strasser. H. (1985). *Mathematical Theory of Statistics*. De Gruyter Studies in Mathematics **7**, De Gruyter, Berlin.

Wei, X. (1995). Asymptotically efficient estimates of the index of regular variation, *Ann. Statist.*, **23**, 2036–2058.

Wolfowitz, J. (1965). Asymptotic efficiency of the maximum likelihood estimator, *Theory Probab. Appl.*, **10**, 247–260.