The Annals of Statistics
1998, Vol. 26, No. 2, 692–718

LOCAL ASYMPTOTIC NORMALITY OF TRUNCATED EMPIRICAL PROCESSES

BY MICHAEL FALK

Katholische Universität Eichstätt

Given n iid copies $X_1, \dots, X_n$ of a random element X in some arbitrary measurable space S, we are only interested in those observations that fall into some subset D having but a small probability of occurrence. It is assumed that the distribution P_X of X belongs on D to a parametric family $P_X(\cdot \cap D) = P_\vartheta$, $\vartheta \in \Theta \subset \mathbb{R}^d$. Nonlinear regression analysis and the peaks-over-threshold (POT) approach in extreme value analysis are prominent examples. For the POT approach on $S = \mathbb{R}$ and P_ϑ being a generalized Pareto distribution, it is known that the complete information about the underlying parameter ϑ_0 is asymptotically contained in the number $\tau(n)$ of observations in D among $X_1, \dots, X_n$, but not in their actual values. This result is formulated in terms of local asymptotic normality of the log-likelihood ratio of the point process of exceedances with $\tau(n)$ being the central sequence.

In this paper we establish a necessary and sufficient condition such that $\tau(n)$ has this property for a general truncated empirical process in an arbitrary sample space and for an arbitrary parametric family. The known results are then consequences of this result. We can, moreover, characterize the influence of the actual observations in D on the central sequence, if this condition is violated.

Immediate applications are asymptotically optimal tests for testing ϑ_0 and, if $\Theta \subset \mathbb{R}$, asymptotic efficiency of the ML-estimator $\hat{\vartheta}_n$ satisfying $P_{\hat{\vartheta}_n}(D) = \tau(n)/n$, where these statistics are based on $\tau(n)$ only.

0. Introduction. Let $X_1, \dots, X_n$ be independent copies of a random element (re)X with values in some measurable space $(S, \mathscr{D})$, whose distribution P_X belongs locally to some parametric family. Precisely, there exists $D \in \mathscr{D}$ such that

$$(M) \qquad P_X(\cdot \cap D) = P_\vartheta(\cdot \cap D),$$

where $\{P_\vartheta: \vartheta \in \Theta\}$ is a family of distributions, parametrized by ϑ from some parameter space $\Theta \subset \mathbb{R}^d$. Typical examples are regression analysis or density estimation, in which case the set D is located in the center of the distribution of X, or in extreme value theory, where $D = D_n$ is located at the border and usually shrinks with increasing sample size n; see Section 1.3 in Falk, Hüsler and Reiss (1994) for details.

A statistical analysis concerning the underlying parameter ϑ, such as the estimation of conditional distributions in regression analysis or of extreme

Received October 1996; revised August 1997.

AMS 1991 *subject classifications*. Primary 62F05, 62F12; secondary 60G55.

Key words and phrases. Truncated empirical point process, log-likelihood ratio, local asymptotic normality, central sequence, optimal tests, *ML*-estimator, efficiency.

quantiles in extreme value theory, has obviously to be based on these observations Y_j among $X_1, \dots, X_n$ that fall into the set D. A mathematically convenient way to describe these observations Y_j is by means of truncated empirical point processes. First, we identify any point $x \in S$ with the pertaining Dirac measure $\varepsilon_x(B) = 1$ if $x \in B \in \mathscr{D}$ and 0 otherwise. We thus identify a re X_i with the random Dirac measure ε_{X_i}. A mathematically precise representation of those observations Y_j among $X_1, \dots, X_n$ that fall into D is then the truncated empirical point process

$$\begin{aligned} N_n^D(B) &:= \sum_{j=1}^{n} \varepsilon_{X_j}(B)\varepsilon_{X_j}(D) \\ &= \sum_{j=1}^{n} \varepsilon_{X_j}(B \cap D) = \sum_{j=1}^{\tau(n)} \varepsilon_{Y_j}(B), \qquad B \in \mathscr{D}, \end{aligned}$$

where the number $\tau(n) := N_n^D(D)$ of observations in D is binomial $B(n, P_X(D))$-distributed. The process N_n^D is a re in the set $\mathbb{M} := \{\mu = \sum_{j=1}^{n} \varepsilon_{x_j}: x_1, \dots, x_n \in S,\ n = 0, 1, 2, \dots\}$ of finite point measures on $(S, \mathscr{D})$, equipped with the smallest σ-field $\mathscr{M}$ such that for any $B \in \mathscr{D}$ the projection

$$\pi_B: \mathbb{M} \to \{0, 1, 2, \dots\}, \qquad \pi_B(\mu) := \mu(B)$$

is measurable. As such, N_n^D is called a *point process*. For technical details we refer to Section 1.1 of Reiss (1993).

The following crucial representation result for N_n^D is well known [see, e.g., Theorem 1.4.1 in Reiss (1993)]. We state it here explicitly for easier reference. By $\mathscr{L}(Z)$ we denote the distribution of a re Z.

LEMMA 0.1. *Suppose that* $0 < P_X(D) < 1$. *Then we have*

$$\mathscr{L}(N_n^D) = \mathscr{L}\left(\sum_{j=1}^{\tau(n)} \varepsilon_{Y_j}\right) = \mathscr{L}\left(\sum_{j=1}^{K(n)} \varepsilon_{W_j}\right),$$

where $K(n)$ *is* $B(n, P_X(D))$*-distributed*, $W_1, W_2, \dots, W_n$ *are iid re's with common distribution*

$$P_W(B) := P_X(B \cap D)/P_X(D)$$

and $K(n)$ *and the vector* $(W_1, \dots, W_n)$ *are independent.*

By the preceding lemma we can handle those observations $Y_1, \dots, Y_{\tau(n)}$ among $X_1, \dots, X_n$ that fall into D, as a set of iid re's, whose common distribution is the conditional distribution of X, given $X \in D$; these are independent of their number $\tau(n)$, which is $B(n, P_X(D))$-distributed.

If $(S, \mathscr{D}) = (\mathbb{R}, \mathbb{B})$, where $\mathbb{B}$ is the Borel σ-field on $\mathbb{R}$, and $D = D_n = [t_n, \infty]$ satisfies $P_X(D) \to 0$, $nP_X(D) \to \infty$ as $n \to \infty$, then N_n^D describes the exceedances among $X_1, \dots, X_n \in \mathbb{R}$ over some high threshold $t_n \in \mathbb{R}$, with the expected number $nP_X(D)$ of observations above t_n tending to infinity as the sample size n increases. Lemma 0.1 formalizes in this case the *peaks-over-*

threshold method (POT), which has become a quite popular alternative in recent years to the order statistics approach in extreme value statistics for the analysis of tails of a distribution. Different from the order statistics approach, the POT approach is universally applicable and not restricted to $\mathbb{R}^1$ with its natural ordering; see, for example, Section 2.3 in Falk, Hüsler and Reiss (1994) and the literature cited therein.

Suppose now that we want to test the hypothesis $\vartheta_0 = 0$ in model (M) against some alternative $\vartheta_n \neq 0$, where we assume that $\Theta \subset \mathbb{R}^d$ contains 0 as an inner point. If we suppose that P_{ϑ_n} is dominated on D by P_0, then $\mathscr{L}_{\vartheta_n}(N_n^D)$ is dominated by $\mathscr{L}_0(N_n^D)$ [see, e.g., Theorem 3.1.2 in Reiss (1993)] and thus, the Neyman–Pearson lemma implies that with $u \in \mathbb{R}$,

$$\varphi(N_n^D) := 1_{(u,\infty)}\Bigl(\log\bigl\{d\mathscr{L}_{\vartheta_n}(N_n^D)/d\mathscr{L}_0(N_n^D)\bigr\}(N_n^D)\Bigr)$$

is the most powerful test for $\vartheta = 0$ against ϑ_n of level $E_0(\varphi(N_n^D))$, based on N_n^D. We index expectations, distributions and so on by the underlying parameter. By 1_A we denote the indicator function of a set A, that is, $1_A(x) = \varepsilon_x(A)$.

If ϑ_n now converges to 0 as the sample size n increases, such that for some $\xi \in \mathbb{R}$ the log-likelihood ratio satisfies the expansion

$$(L) \qquad \begin{aligned} L_n &:= \log\bigl\{d\mathscr{L}_{\vartheta_n}(N_n^D)/d\mathscr{L}_0(N_n^D)\bigr\}(N_n^D) \\ &= \xi Z_{(n)} - \xi^2/2 + o_{P_0}(1), \end{aligned}$$

where for $n \to \infty$

$$Z_{(n)} \to_{\mathscr{D}_0} N(0,1),$$

then L_n is called *local asymptotic normal* (LAN) with *central sequence* $Z_{(n)}$, $n \in \mathbb{N}$. By $\to_{\mathscr{D}_\vartheta}$ we denote weak convergence with underlying parameter ϑ. This asymptotic expansion enables us to apply the powerful general LAN theory to the above particular testing problem [cf. LeCam (1986), Strasser (1985), LeCam and Yang (1990) for the general theory; for applications in estimation problems we refer to the books by Ibragimov and Hasminskii (1981) and Pfanzagl (1994). A very readable introduction to both estimation and testing is in Chapter 8 of Anderson, Borgan, Gill and Keiding (1992)]. We give in the following a brief introduction to basic concepts of the theory, leading to asymptotically optimal tests and efficient estimators.

Denote for $\alpha \in (0,1)$ by $u_\alpha = \Phi^{-1}(1-\alpha)$ the $(1-\alpha)$ quantile of the standard normal distribution function (df) Φ. The LAN expansion (L) then implies that

$$\varphi_{n,\mathrm{opt}}(N_n^D) := 1_{(u_\alpha,\infty)}\bigl(\mathrm{sign}(\xi) Z_{(n)}\bigr)$$

is an asymptotically optimal level α-test for testing $\vartheta = 0$ against ϑ_n. By LeCam's first lemma [LeCam (1965), named by Hájek (1962); see, e.g., Theorem 1 in Chapter 3 of LeCam and Yang (1990)], expansion (L) implies moreover that the alternatives $\mathscr{L}_{\vartheta_n}(N_n^D)$ are contiguous with respect to $\mathscr{L}_0(N_n^D)$, yielding that (L) is also valid under the alternative ϑ_n

$$L_n = \xi Z_{(n)} - \xi^2/2 + o_{P_{\vartheta_n}}(1),$$

where the central sequence now satisfies

$$Z_{(n)} \to_{\mathscr{D}_{\vartheta_n}} N(\xi, 1).$$

This result in turn provides immediately the asymptotic power of $\varphi_{n,\mathrm{opt}}(N_n^D)$

$$\lim_{n\to\infty} E_{\vartheta_n}\left(\varphi_{n,\mathrm{opt}}\left(N_n^D\right)\right) = 1 - \Phi(u_\alpha - |\xi|).$$

Note that $\varphi_{n,\mathrm{opt}}(N_n^D)$ is obviously asymptotically optimal uniformly for $\xi > 0$ and for $\xi < 0$. Expansion (L) is also crucial for the estimation problem as we will explicate in the following. Suppose for simplicity that $\Theta \subset \mathbb{R}$ and assume that (L) is true with the parameter 0 replaced by ϑ_0 and $\vartheta_n = \vartheta_n(\xi) = \vartheta_0 + \xi\delta_n$, where $\delta_n \to_{n\to\infty} 0$. Then (L) reads

$$L_n = L_n(\vartheta_n) = \delta_n^{-1}(\vartheta_n - \vartheta_0)Z_{(n)} - \delta_n^{-2}(\vartheta_n - \vartheta_0)^2/2 + o_{P_{\vartheta_0}}(1).$$

We denote now by $\hat{\vartheta}_n$ the ML-estimator of ϑ_0 maximizing $L_n = L_n(\vartheta)$ and assume its consistency. By neglecting the remainder term, the maximization

$$L_n\left(\hat{\vartheta}_n\right) = \max_{\vartheta} L_n(\vartheta)$$

is then roughly equivalent to maximizing the leading terms on the right-hand side of the above expansion

$$\begin{aligned}
&\delta_n^{-1}\left(\hat{\vartheta}_n - \vartheta_0\right)Z_{(n)} - \delta_n^{-2}\left(\hat{\vartheta}_n - \vartheta_0\right)^2/2 \\
&\quad = \max_{\vartheta}\left(\delta_n^{-1}(\vartheta - \vartheta_0)Z_{(n)} - \delta_n^{-2}(\vartheta - \vartheta_0)^2/2\right) \\
&\quad \Leftrightarrow \delta_n^{-1}\left(\hat{\vartheta}_n - \vartheta_0\right) = Z_{(n)}.
\end{aligned}$$

We thus have, by LAN and the implications of LeCam's first lemma,

$$\delta_n^{-1}\left(\hat{\vartheta}_n - \vartheta_0\right) \to_{\mathscr{D}_{\vartheta_0}} N(0,1), \qquad \delta_n^{-1}\left(\hat{\vartheta}_n - \vartheta_n\right) \to_{\mathscr{D}_{\vartheta_n}} N(0,1).$$

The estimator $\hat{\vartheta}_n$ is therefore asymptotically unbiased under ϑ_0 as well as under ϑ_n. Let $\hat{T}_n$ be another asymptotically unbiased estimator of ϑ_0, precisely,

$$\delta_n^{-1}\left(\hat{T}_n - \vartheta_0\right) \to_{\mathscr{D}_{\vartheta_0}} Q_{\vartheta_0}, \qquad \delta_n^{-1}\left(\hat{T}_n - \vartheta_n\right) \to_{\mathscr{D}_{\vartheta_n}} Q_{\vartheta_0},$$

where Q_{ϑ_0} is some probability measure on $(\mathbb{R}, \mathbb{B})$, independent of ξ. In this case $\hat{T}_n$ is called *regular*. Then Hájek's convolution theorem [Hájek (1970)] implies that Q_{ϑ_0} is the convolution of $N(0,1)$ with a further probability measure and thus, the variance of the limiting distribution of T_n exceeds that of $\hat{\vartheta}_n$, which is one. Due to this minimal property, $\hat{\vartheta}_n$ is called (asymptotically) *efficient* within the class of regular estimators.

The preceding considerations indicate that a powerful statistical machinery starts, as soon as LAN has been established, and that the central sequence plays a key role in testing as well as in estimation problems. In Falk (1995a) this expansion could be verified for the POT method mentioned above, with the aim of testing the extreme value index in certain neighborhoods of generalized Pareto distributions. In this case we assume that

$X_1, X_2, \dots$ are independent copies of a random variable (rv) X on $\mathbb{R}$, whose df F has a generalized Pareto upper tail. Precisely, there exists an (unknown) point x_0 with $F(x_0) < 1$ such that

$$F(x) = G_\beta(x), \qquad x \geq x_0,$$

where G_β is *a generalized Pareto distribution* (GPD), that is,

$$G_\beta(x) = \begin{cases} 1 - x^{-\beta}, & x \geq 1, \quad \text{if } \beta > 0, \\ 1 - |x|^{-\beta}, & -1 \leq x \leq 0, \quad \text{if } \beta < 0, \\ 1 - \exp(-x), & x \geq 0, \quad \beta = 0. \end{cases}$$

The assumption that the upper tail of F coincides with that of a GPD is supported by results from Balkema and de Haan (1974), Pickands (1975) and Rychlik (1992), which roughly imply that the exceedances in the sample $X_1, \dots, X_n$ over a high threshold t_n have a nondegenerate weak limit distribution iff the upper tail of F is close to that of a GPD. This leads to the definition of δ-neighborhoods of GPDs, which have been extensively studied in recent years [cf. Chapter 2 in Falk, Hüsler and Reiss (1994) and the literature cited therein].

It was observed in Falk (1995a) that in this particular model the *number* $\tau(n)$ of exceedances alone provides the central sequence, not only in a pure shape parameter problem but also if a scale parameter is added. In view of (L) and the pertaining considerations, this means that the complete information in the testing and estimation problem is contained only in the number of exceedances but not in their actual values, as one would expect. Best unbiased estimators for the parameter of a Pareto distribution with unknown shape and scale parameters for a *fixed* sample size are given, for example, in Saksena and Johnson (1984). On the other hand it was observed in Falk (1995b), using the order statistics approach, or in Marohn (1995), that the exceedances enter the central sequence if a location parameter is added.

These observations were the motivation for the present paper, in which we will clarify the contribution of the number $\tau(n)$ and the observations Y_j in D to the central sequence for general truncated empirical processes N_n^D. We can establish in Theorem 1.1 a necessary and sufficient condition for the family $\{P_\vartheta\colon \vartheta \in \Theta\}$, satisfied, for example, for a normal location and scale family, such that $\tau(n)$ alone provides the central sequence, and we can characterize the influence of the observations Y_j on the central sequence. The results in Falk (1995a, b) are then consequences of Theorem 1.1. In Section 3 we apply the LAN theory to establish asymptotically optimal estimators of the underlying parameter which are based on $\tau(n)$.

The LAN of certain Poisson point process models is, for example, established in Proposition 3.25 of Karr (1991) and of marked point processes in Nishiyama (1995). Höpfner and Jacod (1994) and Marohn (1995) established LAN of two-parameter point processes of exceedances. For LAN of closely related counting process models and for excellent bibliographic remarks, we refer to Chapter 8 of Andersen, Burgan, Bill and Keiding (1992). The LAN of extreme order statistics in certain order statistics models were established by Falk (1995b), Marohn (1995) and Wei (1995). A characterization of LAN within iid models of real valued random variables, in terms of the information contained asymptotically in an arbitrary fixed number of extreme order statistics was proved by Janssen and Marohn (1994).

1. Main result. Suppose that the model (M) is satisfied with $\Theta \subset \mathbb{R}^d$ and 0 being an inner point of Θ. Suppose further that the distributions P_ϑ, $\vartheta \in \Theta$, are dominated on $D = D_n$ by some σ-finite measure μ such that the densities $f_\vartheta := dP_\vartheta / d\mu$ satisfy for ϑ near 0 the expansion

$$f_\vartheta = f_0\left(1 + \langle \vartheta, g \rangle + \|\vartheta\|^2 h_\vartheta\right) \tag{1}$$

on D, where the function $g = (g_1, \dots, g_d)$ satisfies $\int_D \|g\|^2 \, dP_0 < \infty$. By $\langle \cdot, \cdot \rangle$ we denote the usual inner product of $\mathbb{R}^d$ and by $\|\vartheta\| := \langle \vartheta, \vartheta \rangle^{1/2}$ the usual norm. Note that each of the functions f_ϑ, g, h_ϑ in expansion (1) is defined on $D = D_n$ and thus, they actually depend on the sample size n as well, that is,

$$f_\vartheta = f_{n,\vartheta}, \qquad g = g_n, \qquad h_\vartheta = h_{n,\vartheta},$$

so that only the parameter space Θ is kept fixed. But for the sake of a clear presentation we omit in the following the dependence of these quantities on the sample size n.

Define the alternatives

$$\begin{aligned} \vartheta_n &:= \vartheta_n(\xi_1, \dots, \xi_n) := (\vartheta_{ni})_{i=1}^d \\ &:= \left(\frac{\xi_i P_0^{1/2}(D)}{n^{1/2} \int_D g_i \, dP_0} \right)_{i=1}^d \end{aligned} \tag{2}$$

where $\xi_1, \dots, \xi_d \in \mathbb{R}$ are fixed. We suppose $\vartheta_n \to 0$ as $n \to \infty$ and $\sup_n \int_D \|g\|^2 \, dP_0 < \infty$ as well as $\sup_n \int_D h_{\vartheta_n}^4 dP_0 < \infty$. The proof of Theorem 1.1 shows that the alternatives ϑ_{ni} can be replaced by any sequence ϑ'_{ni} such that

$$\vartheta'_{ni} / \vartheta_{ni} \to 1 \text{ as } n \to \infty, \qquad 1 \le i \le d, \tag{3}$$

provided $h_{\vartheta'_n}$ satisfies the above integrability conditions. We suppose

$$P_0(D) \to 0, \qquad nP_0(D) \to \infty \text{ as } n \to \infty$$

and the crucial condition

$$\frac{P_0(D) \int_D g_i^2 \, dP_0}{\left(\int_D g_i \, dP_0\right)^2} \to 1 + c_i \text{ as } n \to \infty, \qquad 1 \le i \le d, \tag{4}$$

where $c_i \geq 0$, $1 \leq i \leq d$, are finite constants. We assume further

$$\frac{1}{P_0(D)} \int_D \left(\frac{P_0(D) g_j}{\int_D g_j \, dP_0} - 1 \right) \left(\frac{P_0(D) g_k}{\int_D g_k \, dP_0} - 1 \right) dP_0 \to 0 \quad \text{as } n \to \infty, \qquad i \leq j \neq k \leq d, \tag{5}$$

which is, by the Cauchy–Schwarz inequality, automatically satisfied, if (4) holds with c_j or c_k equal to 0. Recall that an observation Y among those random elements X_i, which actually fall into D, has by Lemma 0.1 under the parameter $\vartheta = 0$ the distribution $P_0(\cdot \cap D)/P_0(D)$. Condition (4) then reads

$$\frac{P_0(D) \int_D g_i^2 \, dP}{\left(\int_D g_i \, dP_0 \right)^2} - 1 = \frac{E(Y^2)}{E(Y)^2} - 1 = \frac{\operatorname{Var}(Y)}{E(Y)^2} \to c_i \qquad \text{as } n \to \infty$$

and (5) becomes, for $j \neq k$,

$$\begin{aligned} &\frac{1}{P_0(D)} \int_D \left(\frac{P_0(D) g_j}{\int_D g_j \, dP_0} - 1 \right) \left(\frac{P_0(D) g_k}{\int_D g_k \, dP_0} - 1 \right) dP_0 \\ &\quad = \operatorname{Cov}\left(\frac{g_j(Y)}{E(g_j(Y))}, \frac{g_k(Y)}{E(g_j(Y))} \right) \to 0 \qquad \text{as } n \to \infty. \end{aligned}$$

For those indices i such that $c_i > 0$ in (4) we require in addition,

$$\frac{P_0^2(D)}{n} \frac{\int_D g_i^4 \, dP_0}{\left(\int_D g_i \, dP_0 \right)^4} \to 0 \qquad \text{as } n \to \infty, \tag{6}$$

which is the condition

$$\frac{1}{nP_0(D)} \frac{E(g_i^4(Y))}{E(g_i(Y))^4} \to 0 \qquad \text{as } n \to \infty.$$

Finally we assume that

$$nP_0^{1/2}(D) \|\vartheta_n\|^4 \to 0 \qquad \text{as } n \to \infty. \tag{7}$$

If the remainder term h_ϑ in expansion (1) is uniformly bounded for ϑ near 0 and $x \in D$, then it is sufficient to require

$$nP_0(D) \|\vartheta_n\|^4 \to 0 \qquad \text{as } n \to \infty. \tag{7$'$}$$

The preceding conditions are discussed in the sequel. Condition (4) and its connection to the Fisher information is particularly discussed in Section 3. Now we are ready to state our main result.

THEOREM 1.1. *Assume the above conditions. Then we have under* $\vartheta = 0$ *for any* $(\xi_1, \dots, \xi_d) \in \mathbb{R}^d$,

$$\log\left\{\frac{d\mathscr{L}_{\vartheta_n}(N_n^D)}{d\mathscr{L}_0(N_n^D)}\right\}(N_n^D)$$

$$= \left(\sum_{i=1}^d \xi_i\right)\frac{\tau(n) - nP_0(D)}{(nP_0(D))^{1/2}} - \frac{1}{2}\left(\sum_{i=1}^d \xi_i\right)^2$$

$$+ (nP_0(D))^{-1/2}\sum_{j=1}^{\tau(n)}\sum_{i:\, c_i>0}\xi_i\left(\frac{P_0(D)}{\int_D g_i\, dP_0}g_i(Y_j) - 1\right)$$

$$-\frac{1}{2}\sum_{i=1}^d \xi_i^2 c_i + o_{P_0}(1)$$

$$\to_{\mathscr{D}} N\left(-\frac{1}{2}\left(\left(\sum_{i=1}^d \xi_i\right)^2 + \sum_{i=1}^d \xi_i^2 c_i\right), \left(\sum_{i=1}^d \xi_i\right)^2 + \sum_{i=1}^d \xi_i^2 c_i\right),$$

where $Y_1, \dots, Y_{\tau(n)}$ *are those rvs in the original sample that fall into the set* D, *and they are independent of the number* $\tau(n)$.

The preceding result shows that the deviations c_i from 0 in (4) determine the central sequence in Theorem 1.1, and the term

$$(nP_0(D))^{-1/2}\sum_{j=1}^{\tau(n)}\xi_i\left(\left(P_0(D)\Big/\int_D g_i\, dP_0\right)g_i(Y_j) - 1\right) - \xi_i^2 c_i/2$$

contributes to the central sequence iff $c_i > 0$. The number $\tau(n)$ constitutes therefore the central sequence alone iff $c_i = 0$ for $1 \le i \le d$. The examples in Section 2 show that each case can occur.

REMARK 1.2. As mentioned before, (4) is actually a condition on the asymptotic variance of the conditional distribution of $g_i(X)$, given $X \in D$, since

$$\frac{P_0(D)\int_D g_i^2\, dP_0}{(\int_D g_i\, dP_0)^2} - 1 = \frac{\mathrm{Var}_0(g_i(Y_j))}{E_0(g_i(Y_j))^2} \ge 0.$$

A possible limit c_i in (4) is therefore necessarily nonnegative. Conditions (4) with $c_i = 0$ and (7) are automatically satisfied under appropriate local regularity conditions on g and h_ϑ in expansion (1), if $D = D_n$ is located in the center of the distribution P_0. Suppose for simplicity that P_ϑ, $\vartheta \in \Theta \subset \mathbb{R}^d$, are distributions on $\mathbb{R}$. Fix $x_0 \in \mathbb{R}$ and assume that expansion (1) holds for x near x_0 and ϑ close to 0, where f_0 is continuous at x_0 with $f_0(x_0) > 0$. Put for $n \in \mathbb{N}$,

$$D = D_n = \left[x_0 - \frac{1}{2na_n f_0(x_0)},\ x_0 + \frac{1}{2na_n f_0(x_0)}\right],$$

where $0 < a_n \to_{n\to\infty} 0$, $na_n \to_{n\to\infty} \infty$. Then we have

$$P_0(D) = \frac{1}{na_n}(1 + o(1))$$

and

$$\int_D g_i P_0(dx) = \frac{g_i(x_0)}{na_n}(1 + o(1)), \qquad 1 \le i \le d,$$

provided the function g_i in expansion (1) is continuous at x_0 with $g_i(x_0) > 0$ for $1 \le i \le d$. In this case, condition (4) is obviously satisfied with $c_i = 0$. We can choose, moreover,

$$\vartheta_{ni} := \xi_i a_n^{1/2}/g_i(x_0), \qquad 1 \le i \le d.$$

If we assume in addition that the remainder term $h_\vartheta(x)$ in expansion (1) is uniformly bounded for ϑ near 0 and x near x_0, then

$$nP_0(D)\vartheta_{ni}^4 = O(a_n),$$

that is, condition (7′) is automatically satisfied. The assertion of Theorem 1.1 now follows.

Proof of Theorem 1.1. From Example 3.1.2 in Reiss (1993) we obtain that for ϑ near zero, $\mathscr{L}_\vartheta(N_n^D)$ has the $\mathscr{L}_0(N_n^D)$ density

$$\frac{d\mathscr{L}_\vartheta(N_n^D)}{d\mathscr{L}_0(N_n^D)}(\mu) = \left(\prod_{i=1}^{\mu(D)} \frac{f_\vartheta(x_i)}{f_0(x_i)} \frac{P_0(D)}{P_\vartheta(D)}\right)\left(\frac{P_\vartheta(D)}{P_0(D)}\right)^{\mu(D)}\left(\frac{1 - P_\vartheta(D)}{1 - P_0(D)}\right)^{n-\mu(D)}$$

if $\mu = \sum_{i=1}^{\mu(D)} \varepsilon_{x_i}$ and $0 \le \mu(D) \le n$. Consequently,

$$\begin{aligned} &\log\left\{\frac{d\mathscr{L}_\vartheta(N_n^D)}{d\mathscr{L}_0(N_n^D)}\right\}(N_n^D) \\ &\quad = \int \log \frac{f_\vartheta(x)}{f_0(x)} \frac{P_0(D)}{P_\vartheta(D)} N_n^D(dx) \qquad (8) \\ &\qquad + \tau(n)\log\left(\frac{P_\vartheta(D)}{P_0(D)}\right) + (n - \tau(n))\log\left(\frac{1 - P_\vartheta(D)}{1 - P_0(D)}\right). \end{aligned}$$

A suitable version of the central limit theorem implies the following.
Fact 1.

$$(\tau(n) - nP_0(D))/(nP_0(D))^{1/2} \to_{\mathscr{D}_0} N(0,1).$$

From expansion (1) we obtain the following.

Fact 2.

$$\begin{aligned}
P_{\vartheta_n}(D) - P_0(D) &= \int_D f_{\vartheta_n} - f_0 \, d\mu \\
&= \int_D \langle \vartheta_n, g \rangle + \|\vartheta_n\|^2 h_{\vartheta_n} \, dP_0 \\
&= n^{-1/2} P_0^{1/2}(D) \sum_{i=1}^{d} \xi_i + \|\vartheta_n\|^2 \int_D h_{\vartheta_n} \, dP_0 \\
&= n^{-1/2} P_0^{1/2}(D) \sum_{i=1}^{d} \xi_i + O\big(\|\vartheta_n\|^2 P_0^{3/4}(D)\big) \\
&= n^{-1/2} P_0^{1/2}(D) \left(\sum_{i=1}^{d} \xi_i + o(1) \right)
\end{aligned}$$

by the Cauchy-Schwarz inequality and condition (7).

Fact 3.

$$\frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)} = \big(nP_0(D)\big)^{-1/2} \left(\sum_{i=1}^{d} \xi_i + o(1) \right) = o(1)$$

by Fact 2 and the condition $nP_0(D) \to \infty$ as $n \to \infty$. By making use of Facts 1–3 we will next show that

$$\begin{aligned}
&\tau(n) \log\left(\frac{P_{\vartheta_n}(D)}{P_0(D)} \right) + (n - \tau(n)) \log\left(\frac{1 - P_{\vartheta_n}(D)}{1 - P_0(D)} \right) \\
&\quad = \left(\sum_{i=1}^{d} \xi_i \right) \big(nP_0(D)\big)^{-1/2} \big(\tau(n) - nP_0(D)\big) \\
&\qquad - \frac{1}{2} \left(\sum_{i=1}^{d} \xi_i \right)^2 + o_{P_0}(1).
\end{aligned} \tag{9}$$

The Taylor expansion $\log(1 + \varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$ for $\varepsilon \to 0$ implies

$$\begin{aligned}
&\tau(n) \log\left(\frac{P_{\vartheta_n}(D)}{P_0(D)} \right) + (n - \tau(n)) \log\left(\frac{1 - P_{\vartheta_n}(D)}{1 - P_0(D)} \right) \\
&\quad = \tau(n) \left(\frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)} - \frac{1}{2} \left(\frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)} \right)^2 \right. \\
&\qquad\qquad \left. + O\left(\left| \frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)} \right|^3 \right) \right) \\
&\qquad + (n - \tau(n)) \left(\frac{P_0(D) - P_{\vartheta_n}(D)}{1 - P_0(D)} + O\big(|P_0(D) - P_{\vartheta_n}(D)|^2 \big) \right)
\end{aligned}$$

$$= \tau(n) \frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)} + (n - \tau(n)) \frac{P_0(D) - P_{\vartheta_n}(D)}{1 - P_0(D)}$$

$$- \frac{\tau(n)}{2} \frac{\left(\sum_{i=1}^d \xi_i + o(1)\right)^2}{nP_0(D)} + o_{P_0}(1)$$

$$= \left(P_{\vartheta_n}(D) - P_0(D)\right)\left(\frac{\tau(n)}{P_0(D)} - \frac{n - \tau(n)}{1 - P_0(D)}\right) - \frac{\left(\sum_{i=1}^d \xi_i\right)^2}{2} + o_{P_0}(1)$$

$$= \frac{P_{\vartheta_n}(D) - P_0(D)}{P_0(D)(1 - P_0(D))}(\tau(n) - nP_0(D)) - \frac{\left(\sum_{i=1}^d \xi_i\right)^2}{2} + o_{P_0}(1)$$

$$= \left(nP_0(D)\right)^{-1/2}\left(\sum_{i=1}^d \xi_i + o(1)\right)(\tau(n) - nP_0(D))$$

$$- \frac{1}{2}\left(\sum_{i=1}^d \xi_i\right)^2 + o_{P_0}(1)$$

by utilizing Facts 1–3. This proves representation (9). In order to establish Theorem 1.1, it remains to prove

$$\int \log\left\{\frac{f_{\vartheta_n}(x)}{f_0(x)} \frac{P_0(D)}{P_{\vartheta_n}(D)}\right\} N_n^D(dx)$$

$$(10) \qquad = \left(nP_0(D)\right)^{-1/2} \sum_{j=1}^{\tau(n)} \sum_{i:\, c_i > 0} \xi_i \left(\frac{P_0(D)}{\int_D g_i \, dP_0} g_i(Y_j) - 1\right)$$

$$- \sum_{i=1}^d \xi_i^2 c_i/2 + o_{P_0}(1).$$

For n large enough, we have the following expansion on $D \cap \{f_0 > 0\}$.

Fact 4.

$$\frac{f_{\vartheta_n}}{f_0} \frac{P_0(D)}{P_{\vartheta_n}(D)} - 1$$

$$= \frac{\left(1 + \langle \vartheta_n, g \rangle + \|\vartheta_n\|^2 h_{\vartheta_n}\right) P_0(D) - P_{\vartheta_n}(D)}{P_{\vartheta_n}(D)}$$

$$= \frac{P_0(D)}{P_{\vartheta_n}(D)}\left(\left\langle \vartheta_n, g - \frac{\int_D g \, dP_0}{P_0(D)} \right\rangle + \|\vartheta_n\|^2 \left(h_{\vartheta_n} - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)}\right)\right)$$

$$= \frac{P_0(D)}{P_{\vartheta_n}(D)}\left(\sum_{i=1}^d \vartheta_{ni}\left(g_i - \frac{\int_D g_i \, dP_0}{P_0(D)}\right) + \|\vartheta_n\|^2 \left(h_{\vartheta_n} - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)}\right)\right).$$

In the next step we will show that we can assume, for $1 \le j \le \tau(n)$,

$$\sum_{i=1}^{d} \left| \vartheta_{ni} \left(g_i(Y_j) - \frac{\int_D g_i \, dP_0}{P_0(D)} \right) \right| + \|\vartheta_n\|^2 \left| h_{\vartheta_n}(Y_j) - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)} \right| \le \varepsilon_n \tag{11}$$

for some sequence $0 < \varepsilon_n \to 0$ as $n \to \infty$.

Fact 5.

$$\begin{aligned}
P_0 & \left\{ \|\vartheta_n\|^2 \left| h_{\vartheta_n}(Y_j) - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)} \right| \ge \varepsilon_n \text{ for some } 1 \le j \le \tau(n) \right\} \\
&= \sum_{k=1}^{n} P_0 \left\{ \|\vartheta_n\|^2 \left| h_{\vartheta_n}(Y_j) - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)} \right| \ge \varepsilon_n \text{ for some } 1 \le j \le k \right\} \\
&\qquad \times P_0\{\tau(n) = k\} \\
&\le P_0 \left\{ \|\vartheta_n\|^2 \left| h_{\vartheta_n}(Y_1) - \frac{\int_D h_{\vartheta_n} \, dP_0}{P_0(D)} \right| \ge \varepsilon_n \right\} \sum_{k=1}^{n} k P_0\{\tau(n) = k\} \\
&\le \varepsilon_n^{-2} \|\vartheta_n\|^4 E_0\left(h_{\vartheta_n}^2(Y_1)\right) E_0(\tau(n)) \\
&= \varepsilon_n^{-2} n \|\vartheta_n\|^4 \int_D h_{\vartheta_n}^2 \, dP_0 \\
&\le \varepsilon_n^{-2} n P_0^{1/2}(D) \|\vartheta_n\|^4 \left(\int_D h_{\vartheta_n}^4 \, dP_0 \right)^{1/2} \to 0 \qquad \text{as } n \to \infty
\end{aligned}$$

by the independence of $\tau(n)$, $Y_1, Y_2, \ldots,$ the Chebyshev and the Cauchy inequalities and condition (7), if $\varepsilon_n > 0$ converges slowly enough to zero. By repeating the above arguments, we have Fact 6 for those indices i such that condition (4) is satisfied with $c_i = 0$.

Fact 6.

$$\begin{aligned}
P_0 & \left\{ \left| \vartheta_{ni} \left(g_i(Y_j) - \frac{\int_D g \, dP_0}{P_0(D)} \right) \right| \ge \varepsilon_n \text{ for some } 1 \le j \le \tau(n) \right\} \\
&\le \varepsilon_n^{-2} n \vartheta_{ni}^2 \int_D \left(g_i - \frac{\int_D g_i \, dP_0}{P_0(D)} \right)^2 dP_0 \\
&= O\left(\varepsilon_n^{-2} \left(\frac{P_0(D) \int_D g_i^2 \, dP_0}{(\int_D g_i \, dP_0)^2} - 1 \right) \right) \to 0 \qquad \text{as } n \to \infty
\end{aligned}$$

if $0 \le \varepsilon_n \to 0$ as $n \to \infty$ slowly enough. Equally, we have Fact 7 for those indices i such that condition (4) is satisfied with $c_i > 0$.

Fact 7.

$$P_0\left\{\left|\vartheta_{ni}\left(g_i(Y_j) - \frac{\int_D g_i\, dP_0}{P_0(D)}\right)\right| \ge \varepsilon_n \text{ for some } 1 \le j \le \tau(n)\right\}$$
$$\le \varepsilon_n^{-4} n P_0(D)\vartheta_{ni}^4 E_0\left(\left(g_i(Y_1) - \frac{\int_D g_i\, dP_0}{P_0(D)}\right)^4\right)$$
$$= O\left(\frac{P_0^2(D)}{\varepsilon_n^4 n}\frac{\int_D g_i^4\, dP_0}{\left(\int_D g_i\, dP_0\right)^4}\right) \to 0 \qquad \text{as } n \to \infty$$

if $0 < \varepsilon_n \to 0$ as $n \to \infty$ slowly enough. This follows from condition (6).

Facts 5–7 imply that (11) holds on a set, whose probability converges to 1, that is, we will assume in the following that (11) holds. Put now for $j = 1, 2, \dots,$

$$R_{nj} := \sum_{i=1}^{d} \vartheta_{ni}\left(g_i(Y_j) - \frac{\int_D g_i\, dP_0}{P_0(D)}\right)$$

and

$$S_{nj} := \|\vartheta_n\|^2\left(h_{\vartheta_n}(Y_i) - \frac{\int_D h_{\vartheta_n}\, dP_0}{P_0(D)}\right).$$

Then we have by Fact 4 if n is large,

$$\int \log\left\{\frac{f_{\vartheta_n}(x)}{f_0(x)}\frac{P_0(D)}{P_{\vartheta_n}(D)}\right\} N_n^D(dx) = \sum_{j=1}^{\tau(n)} \log\left\{1 + \frac{P_0(D)}{P_{\vartheta_n}(D)}(R_{nj} + S_{nj})\right\}, \tag{12}$$

where we can assume by Fact 2 and (11) that $(P_0(D)/P_{\vartheta_n}(D))|R_{nj} + S_{nj}| \le \varepsilon_n$ for $1 \le j \le \tau(n)$ with $0 < \varepsilon_n \to 0$ as $n \to \infty$. Taylor's formula implies the expansion $\log(1 + x\varepsilon_n) = x\varepsilon_n - ((x\varepsilon_n)^2/2)(1 + o(1))$ uniformly for $|x| \le 1$ and thus,

$$\begin{aligned}
&\sum_{j=1}^{\tau(n)} \log\left\{1 + \frac{P_0(D)}{P_{\vartheta_n}(D)}(R_{nj} + S_{nj})\right\}\\
&\quad= \frac{P_0(D)}{P_{\vartheta_n}(D)}\sum_{j=1}^{\tau(n)}(R_{nj} + S_{nj})\\
&\qquad - \frac{1}{2}\left(\frac{P_0(D)}{P_{\vartheta_n}(D)}\right)^2\sum_{j=1}^{\tau(n)}(R_{nj} + S_{nj})^2(1 + o(1)).
\end{aligned} \tag{13}$$

A suitable version of the central limit theorem implies the following.

Fact 8.

$$\frac{1}{(nP_0(D))^{1/2}}\sum_{j=1}^{\tau(n)}\sum_{i=1}^{d}\xi_i\left(\frac{P_0(D)}{\int_D g_i\, dP_0}g_i(Y_j) - 1\right) \to_{\mathscr{D}_0} N\left(0, \sum_{i=1}^{d}\xi_i^2 c_i\right)$$

by conditions (4) and (5).

Fact (8) yields

$$
\begin{aligned}
&\frac{P_0(D)}{P_{\vartheta_n}(D)} \sum_{j=1}^{\tau(n)} R_{nj} \\
&\quad = (1+o(1)) \sum_{j=1}^{\tau(n)} \sum_{i=1}^{d} \vartheta_{ni}\left(g_i(Y_j) - \frac{\int_D g_i \, dP_0}{P_0(D)}\right) \\
&\quad = (1+o(1)) \frac{1}{(nP_0(D))^{1/2}} \sum_{j=1}^{\tau(n)} \sum_{i=1}^{d} \xi_i \left(\frac{P_0(D)}{\int_D g_i \, dP_0} g_i(Y_j) - 1\right) \\
&\quad = \frac{1}{(nP_0(D))^{1/2}} \sum_{j=1}^{\tau(n)} \sum_{i=1}^{d} \xi_i \left(\frac{P_0(D)}{\int_D g_i \, dP_0} g_i(Y_j) - 1\right) + o_{P_0}(1).
\end{aligned} \tag{14}
$$

Conditions (4) and (6) imply that in P_0 probability we have the following.

Fact 9.

$$
\frac{1}{nP_0(D)} \sum_{j=1}^{\tau(n)} \left(\sum_{i=1}^{d} \xi_i \left(\frac{P_0(D)}{\int_D g_i \, dP_0} g_i(Y_j) - 1 \right)\right)^2 \to \sum_{i=1}^{d} \xi_i^2 c_i \qquad \text{as } n \to \infty.
$$

From Fact 9 we obtain

$$
\frac{P_0(D)}{P_{\vartheta_n}(D)} \sum_{j=1}^{\tau(n)} R_{nj}^2 = \sum_{i=1}^{d} \xi_i^2 c_i + o_{P_0}(1). \tag{15}
$$

Fact 10.

$$
E_0\left(\left(\sum_{j=1}^{\tau(n)} S_{nj}\right)^2\right) = E_0(S_{n1}^2) E_0(\tau(n)) = O\left(nP_0^{1/2}(D) \|\vartheta_n\|^4\right) = o(1)
$$

by condition (7). Fact (10) implies

$$
\sum_{j=1}^{\tau(n)} S_{nj} = o_{P_0}(1) \tag{16}
$$

and

$$
\sum_{j=1}^{\tau(n)} S_{nj}^2 = o_P(1). \tag{17}
$$

From (12)–(17) we obtain now

$$
\begin{aligned}
&\int \log\left\{\frac{f_{\vartheta_n}(x)}{f_0(x)} \frac{P_0(D)}{P_{\vartheta_n}(D)}\right\} N_n^D(dx) \\
&\quad = \frac{P_0(D)}{P_{\vartheta_n}(D)} \sum_{j=1}^{\tau(n)} (R_{nj} + S_{nj}) - \frac{1}{2}\left(\frac{P_0(D)}{P_{\vartheta_n}(D)}\right)^2 \sum_{j=1}^{\tau(n)} R_{nj}^2 \\
&\qquad - \frac{1}{2} \sum_{j=1}^{\tau(n)} \left(2R_{nj}S_{nj} + S_{nj}^2\right)(1+o(1))
\end{aligned}
$$

$$= \frac{1}{(nP_0(D))^{1/2}} \sum_{j=1}^{\tau(n)} \sum_{i:\, c_i>0}^{d} \xi_i \left(\frac{P_0(D)}{\int_D g_i \, dP_0} g_i(Y_j) - 1 \right) - \frac{1}{2} \sum_{i=1}^{d} \xi_i^2 c_i + o_P(1)$$

by using the Cauchy–Schwarz inequality. This proves (10) and completes therefore the proof of Theorem 1.1. □

2. Examples. In this section we provide various examples, where conditions (4)–(7) are satisfied, as well as an example from regression analysis, where condition (4) is violated.

EXAMPLE 2.1 (GPD with shape parameter). Let $\{G_\beta \colon \beta \in \mathbb{R}\}$ be again the class of generalized Pareto distributions as defined in the introduction. Fix $\beta_0 \in \mathbb{R}$ and put $P_\vartheta := G_{\beta_0+\vartheta}$, $\vartheta \in \mathbb{R}$.

(i) *The case* $\beta_0 \neq 0$. In this case we have for ϑ near 0 on the support of G_{β_0},

$$f_\vartheta = f_0\left(1 + \vartheta g + \vartheta^2 h_\vartheta\right),$$

where

$$g(x) = \frac{1}{\beta_0} - \log(|x|)$$

and

$$h_\vartheta(x) = O\left(\log^2(|x|)|x|^{|\vartheta|} + 1\right).$$

If we put now $t_n := G_{\beta_0}^{-1}(1-(na_n)^{-1})$ and $D = D_n = [t_n, \omega(G_{\beta_0})]$, where $0 < a_n \to 0$ as $n \to \infty$, $na_n \to \infty$ as $n \to \infty$, then $P_0(D) = 1/(na_n)$,

$$\int_D g \, dP_0 = -\frac{1}{\beta_0} \frac{\log(na_n)}{na_n}$$

and thus,

$$\vartheta_n = \xi \frac{P_0^{1/2}(D)}{n^{1/2} \int_D g \, dP_0} = -\beta_0 \xi \frac{a_n^{1/2}}{\log(na_n)}.$$

Condition (4) is satisfied here with $c = 0$,

$$P_0(D) \frac{\int_D g^2 \, dP_0}{\left(\int_D g \, dP_0\right)^2} = 1 + O\left(\frac{1}{\log(na_n)} \right)$$

and thus Theorem 1.1 implies that under P_0,

$$\begin{aligned} &\log\left\{d\mathscr{L}_{\vartheta_n}(N_n^D)/d\mathscr{L}_0(N_n^D)\right\}(N_n^D) \\ &\quad = \xi a_n^{-1/2}\left(\tau(n) - a_n^{-1}\right) - \xi^2/2 + o_{P_0}(1) \to_{\mathscr{D}_0} N(-\xi^2/2, \xi^2), \end{aligned} \tag{18}$$

which is Theorem 1.3 in Falk (1995a).

(ii) *The case* $\beta_0 = 0$. In this case we reparametrize the class $\{G_\beta\colon \beta \in \mathbb{R}\}$ and put for $\beta \neq 0$,

$$H_\beta(x) := 1 - (1 + \beta x)^{-1/\beta}, \qquad 0 < (1 + \beta x)^{-1/\beta} \le 1,$$

which is the von Mises parametrization of the family of GPDs. Interpret $H_0(x) = \lim_{\beta \to 0} H_\beta(x) = 1 - \exp(-x) = G_0(x)$, $x \ge 0$. Note that H_β has support $(0, -1/\beta)$ if $\beta < 0$ and $(0, \infty)$ if $\beta > 0$. Furthermore, H_β can be obtained from $G_{1/\beta}$ by the identity

$$H_\beta(x) = \begin{cases} G_{1/\beta}(1 + \beta x), & \text{if } \beta > 0, \\ G_{1/\beta}(-(1 + \beta x)), & \text{if } \beta < 0. \end{cases}$$

Denote by f_β the density of H_β, that is,

$$f_\beta(x) = (1 + \beta x)^{-(1/\beta) - 1} \quad \text{for } \begin{cases} x \ge 0, & \text{if } \beta \ge 0, \\ 0 \le x \le -1/\beta, & \text{if } \beta < 0. \end{cases}$$

Then we have for x in the support of H_β by using iterated Taylor expansions,

$$f_\beta(x) = f_0(x)\big(\beta g(x) + \beta^2 h_\beta(x)\big),$$

where

$$g(x) = \frac{x^2}{2} - x, \qquad x \ge 0.$$

If we put now $t_n := \log(na_n)$ and $D = [t_n, t_n a_n^{-1/4})$, where again $0 < a_n \to 0$, $na_n \to \infty$, then $P_0(D) = (na_n)^{-1}(1 + o(1))$,

$$\int_D g\, dP_0 = \frac{\log^2(na_n)}{2na_n}(1 + o(1))$$

and thus we can choose

$$\vartheta_n = 2\xi \frac{a_n^{1/2}}{\log^2(na_n)}.$$

Note that the functions $h_{\vartheta_n}(x)$ are uniformly bounded for $n \in \mathbb{N}$ and $x \in D$, and that the probability that an observation among $X_1, \dots, X_n$ exceeds $t_n a_n^{-1/4}$ converges to zero under $\vartheta = 0$ as well as under $\vartheta = \vartheta_n$ if n increases. The crucial condition (4) is satisfied here with $c = 0$ and thus Theorem 1.1 implies essentially Theorem 1.1 in Falk (1995a) with $\tau(n)$ yielding the central sequence.

EXAMPLE 2.2 (The normal case). Let P_ϑ with $\vartheta = (\mu, \sigma)$ be the normal distribution on $\mathbb{R}$ with mean μ and variance $(1 + \sigma)^{-2}$, that is, $P_\vartheta = N(\mu, (1 + \sigma)^{-2})$, $\vartheta \in \Theta := \mathbb{R} \times (-1, \infty)$. Put $D = D_n := [\Phi^{-1}(1 - 1/(na_n)), \infty)$, where Φ denotes the standard normal df and $a_n > 0$ satisfies $a_n \to 0$, $na_n \to \infty$ as $n \to \infty$. A Taylor expansion of exp at zero implies, for $x \in \mathbb{R}$ and ϑ near 0,

$$\frac{f_\vartheta(x)}{f_0(x)} = 1 + \langle \vartheta, g(x) \rangle + \|\vartheta\|^2 h_\vartheta(x),$$

where

$$g_1(x) = x, \qquad g_2(x) = 1 - x^2$$

and

$$h_\vartheta(x) = O\big(\exp(\|\vartheta\| x^2)(x^4 + 1)\big).$$

Expansion (1) is therefore satisfied. Denote by $\varphi(x) = f_0(x) = \Phi'(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, $x \in \mathbb{R}$, the standard normal density and put $t_n := \Phi^{-1}(1 - 1/(na_n))$. Integration by parts yields

$$\int_D g_1 \, dN(0,1) = \varphi(t_n), \qquad \int_D g_1^2 \, dN(0,1) = t_n \varphi(t_n) + 1/(na_n)$$

and

$$\int_D g_2 \, dN(0,1) = t_n \varphi(t_n),$$

$$\int_D g_2^2 \, dN(0,1) = \big(t_n^3 + 9t_n\big)\varphi(t_n) + 10/(na_n).$$

From the asymptotic equivalence $\varphi(x) \sim x(1 - \Phi(x))$ as $x \to \infty$, we obtain therefore

$$N(0,1)(D)\frac{\int_D g_i^2 \, dN(0,1)}{\big(\int_D g_i \, dN(0,1)\big)^2} \to 1 \qquad \text{as } n \to \infty, \qquad i = 1, 2.$$

Condition (4) is therefore also satisfied with $c_1 = c_2 = 0$ and hence, the assertion of Theorem 1.1 is valid for the family $\{P_\vartheta\colon \vartheta \in \Theta\} = \{N(\mu, (1+\sigma)^{-2})\colon (\mu, \sigma) \in \mathbb{R} \times (-1, \infty)\}$, where we can choose the alternatives $\vartheta_n := (\vartheta_{ni}(\xi_i))_{i=1}^2$ as

$$\vartheta_{n1}(\xi_1) = \xi_1 \frac{a_n^{1/2}}{t_n}, \qquad \vartheta_{n2}(\xi_2) = \xi_2 \frac{a_n^{1/2}}{t_n^2}.$$

Note that the normal distributions do *not* belong to any δ-neighborhood of a GPD and thus the preceding example is a counterexample, showing that the assertion of Theorem 1.1 with D vanishing in the tail and $\tau(n)$ yielding the central sequence, is *not* only valid for distributions from a δ neighborhood of a GPD as was conjectured in Falk (1995a). For the definition of these classes of distributions and a discussion of their significance in extreme value theory, we refer to Sections 2.2 and 2.4 in Falk, Hüsler and Reiss (1994).

EXAMPLE 2.3 (Pareto with shape, scale and location parameter). Fix $\beta_0 > 0$ and denote by P_ϑ with $\vartheta := (\beta, \sigma, \mu) \in (-\beta_0, \infty) \times (-1, \infty) \times \mathbb{R}$ the Pareto distribution with shape, scale and location parameter $\beta_0 + \beta, 1 + \sigma$ and μ, that is, P_ϑ has the Lebesgue density

$$f_\vartheta(x) = (\beta_0 + \beta)(1 + \sigma)^{\beta_0 + \beta}(x - \mu)^{-(\beta_0 + \beta) - 1} \quad \text{if } x - \mu \geq 1 + \sigma$$

and zero elsewhere. Then we have for ϑ near 0 and $x \geq x_0 > 1$,

$$\frac{f_\vartheta(x)}{f_0(x)} = 1 + \beta g_1(x) + \sigma g_2(x) + \mu g_3(x) + \|\vartheta\|^2 h_\vartheta(x),$$

where

$$g_1(x) := \frac{1}{\beta_0} - \log(x), \qquad g_2(x) := \beta_0, \qquad g_3(x) := \frac{\beta_0 + 1}{x}$$

and

$$h_\vartheta(x) = O\big(\log^2(x)\, x^{|\vartheta|} + 1\big).$$

Put $t_n := (na_n)^{1/\beta_0}$ and $D = D_n := [t_n, \infty)$, where again $0 < a_n \to 0$ and $na_n \to \infty$ as $n \to \infty$. Then we have, for n large enough, $P_0(D) = 1/(na_n)$ and

$$\int_D g_1\, dP_0 = -\frac{\log(na_n)}{\beta_0 na_n}, \qquad \int_D g_2\, dP_0 = \frac{\beta_0}{na_n}, \qquad \int_D g_3\, dP_0 = \frac{\beta_0}{(na_n)^{1+1/\beta_0}}$$

and condition (4) is satisfied with

$$c_i = \begin{cases} 0, & \text{if } i = 1, \\ 0, & \text{if } i = 2, \\ \dfrac{1}{\beta_0(\beta_0 + 2)}, & \text{if } i = 3. \end{cases}$$

Thus, we can choose

$$\vartheta_{n1} := -\xi_1 \beta_0 \frac{a_n^{1/2}}{\log(na_n)}, \qquad \vartheta_{n2} := \xi_2 \frac{a_n^{1/2}}{\beta_0}, \qquad \vartheta_{n3} := \xi_3 \frac{(na_n)^{1/\beta_0} a_n^{1/2}}{\beta_0}$$

and obtain from Theorem 1.1 under P_0, if $nP_0^{1/2}(D)\|\vartheta_n\|^4 = (n/a_n)^{1/2}\|\vartheta_n\|^4 = o(1)$,

$$\begin{aligned}
&\log\big\{d\mathscr{L}_{\vartheta_n}(N_n^D)/d\mathscr{L}_0(N_n^D)\big\}(N_n^D) \\
&\quad = (\xi_1 + \xi_2 + \xi_3) a_n^{1/2}\big(\tau(n) - a_n^{-1}\big) - (\xi_1 + \xi_2 + \xi_3)^2/2 \\
&\qquad + a_n^{1/2} \sum_{j=1}^{\tau(n)} \xi_3 \left(\frac{\beta_0 + 1}{\beta_0} \frac{(na_n)^{1/\beta_0}}{Y_j} - 1 \right) - \frac{\xi_3^2}{2\beta_0(\beta_0 + 2)} + o_{P_0}(1) \\
&\quad \to_{\mathscr{D}_0} N\Bigg(-\frac{(\xi_1 + \xi_2 + \xi_3)^2}{2} - \frac{\xi_3^2}{2\beta_0(\beta_0 + 2)}, \\
&\qquad\qquad (\xi_1 + \xi_2 + \xi_2)^2 + \frac{\xi_3^2}{\beta_0(\beta_0 + 2)} \Bigg).
\end{aligned}$$

This result is essentially Theorem 2.3 in Falk (1995b), which is formulated for a random threshold t_n being the kth largest order statistic in the sample; see also Section 3.

The central sequence in this example is therefore the sum of two conditionally independent terms based on the number $\tau(n)$ *and* the rvs Y_j in D, with the Y_j carrying information *only* about the location parameter μ, and $\tau(n)$ containing *all* the information about the underlying shape and scale parameters β and σ and a part of that about μ.

We conclude Section 2 by adding a counterexample, in which condition (4) is violated and Theorem 1.1 cannot be applied.

EXAMPLE 2.4 (Counterexample; regression analysis). Fix $y_0, a_0, b_0 \in \mathbb{R}$ and suppose that $X = (Y, Z) \in \mathbb{R}^2$, where Y has a continuous density f near y_0 with $f(y_0) > 0$. Assume that under $\vartheta = (\vartheta_1, \vartheta_2) \in \mathbb{R}^2$ and for y near y_0, the conditional distribution of Z, given $Y = y$, is $N(a_0 + \vartheta_1 + (b_0 + \vartheta_2)(y - y_0), 1)$, that is, the regression function $E_\vartheta(Z|Y = y) = a_0 + \vartheta_1 + (b_0 + \vartheta_2)(y - y_0)$ is linear in y. Precisely, P_ϑ has the Lebesgue density

$$p_\vartheta(y, z) := f(y) q_\vartheta(z|y)$$

for y near y_0 and $z \in \mathbb{R}$, where

$$q_\vartheta(z|y) = \frac{1}{(2\pi)^{1/2}} \exp\left(-\frac{1}{2}\left(z - \left(a_0 + \vartheta_1 + (b_0 + \vartheta_2)(y - y_0)\right)\right)^2\right).$$

In this case we have on $D = D_n = [y_0 - a_n, y_0 + a_n] \times \mathbb{R}$, where $0 < a_n \to 0$ as $n \to \infty$, if n is large,

$$\begin{aligned} \frac{p_\vartheta(y, z)}{p_0(y, z)} &= \frac{q_\vartheta(z|y)}{q_0(z|y)} \\ &= 1 + \vartheta_1 g_1(y, z) + \vartheta_2 g_2(y, z) + \|\vartheta\|^2 h_\vartheta(y, z), \end{aligned}$$

where

$$\begin{aligned} g_1(y, z) &= z - (a_0 + b_0(y - y_0)), \\ g_2(y, z) &= (y - y_0)\left(z - (a_0 + b_0(y - y_0))\right) \end{aligned}$$

and h_ϑ satisfies the condition of Theorem 1.1 But now $P_0(D) \sim 2f(y_0)a_n$ and

$$\int_D g_1 \, dP_0 = 0, \qquad \int_D g_1^2 \, dP_0 = \int_{y_0 - a_n}^{y_0 + a_n} f(y) \, dy \sim 2f(y_0)a_n,$$

$$\int_D g_2 \, dP_0 = 0, \qquad \int_D g_2^2 \, dP_0 = \int_{y_0 - a_n}^{y_0 + a_n} f(y)(y - y_0)^2 \, dy \sim \tfrac{2}{3} f(y_0) a_n^3,$$

that is, condition (4) is violated and hence, Theorem 1.1 cannot be applied to this regression model. But this counterexample is in accordance with the results in Falk and Marohn (1993), where LAN expansions of loglikelihood ratios of truncated empirical processes in regression models were computed. In this case, the observations in D contribute roughly to the central sequence, but not their number; see Corollary 1.4 in Falk and Marohn (1993) for details.

3. Efficient estimators based on $\tau(n)$. Suppose for notational simplicity that $\Theta \subset \mathbb{R}^1$ and that expansion (1) holds with 0 replaced by ϑ_0, that is,

$$f_\vartheta = f_{\vartheta_0}\left(1 + (\vartheta - \vartheta_0) g_{\vartheta_0} + (\vartheta - \vartheta_0)^2 h_{\vartheta, \vartheta_0}\right).$$

Under appropriate regularity conditions on the family of functions $h_{\vartheta, \vartheta_0}$ we have for $\vartheta \neq \vartheta_0$ on D,

$$\frac{f_\vartheta - f_{\vartheta_0}}{\vartheta - \vartheta_0} = f_{\vartheta_0} g_{\vartheta_0} + (\vartheta - \vartheta_0) f_{\vartheta_0} h_{\vartheta, \vartheta_0} \to f_{\vartheta_0} g_{\vartheta_0} \text{ as } \vartheta \to \vartheta_0$$

and thus

$$\frac{\partial}{\partial \vartheta} \log\left(p_{\vartheta_0}(x)\right) = \frac{1}{p_{\vartheta_0}(x)} \frac{\partial}{\partial \vartheta} p_{\vartheta_0}(x) = g_{\vartheta_0}(x) - \frac{1}{P_{\vartheta_0}(D)} \int_D g_{\vartheta_0}\, dP_0$$

for $x \in D$ such that $f_{\vartheta_0}(x) > 0$, where

$$p_{\vartheta_0}(x) := f_{\vartheta_0}(x) / P_{\vartheta_0}(D), \qquad x \in D,$$

and 0 elsewhere, is a μ-density of Y_j. The Fisher information of the family p_ϑ is therefore

$$\begin{aligned}
I(\vartheta_0) &= \frac{1}{P_{\vartheta_0}(D)} \int_D \left(\frac{\partial}{\partial \vartheta} \log(p_{\vartheta_0})\right)^2 dP_{\vartheta_0} \\
&= \frac{1}{P_{\vartheta_0}(D)} \int_D g_{\vartheta_0}^2\, dP_{\vartheta_0} - \frac{1}{P_{\vartheta_0}^2(D)} \left(\int_D g_{\vartheta_0}\, dP_{\vartheta_0}\right)^2 \\
&= \frac{1}{P_{\vartheta_0}^2(D)} \left(\int_D g_{\vartheta_0}\, dP_{\vartheta_0}\right)^2 \left(\frac{P_{\vartheta_0}(D) \int_D g_{\vartheta_0}^2\, dP_0}{\left(\int_D g_{\vartheta_0}\, dP_0\right)^2} - 1\right) \\
&=: \frac{1}{P_{\vartheta_0}^2(D)} \left(\int_D g_{\vartheta_0}\, dP_{\vartheta_0}\right)^2 c_n;
\end{aligned}$$

recall that $D = D_n$ depends on n. The Cramér–Rao inequality together with the fact that $\tau(n), Y_1, Y_2, \dots,$ are independent rvs, suggest that the variance of any unbiased estimator $T(Y_1, \dots, Y_{\tau(n)})$ of ϑ_0, based on $Y_1, \dots, Y_{\tau(n)}$, satisfies

$$\begin{aligned}
E_{\vartheta_0}\left(\left(T(Y_1, \dots, Y_{\tau(n)}) - \vartheta_0\right)^2\right) &\geq \frac{1}{E_{\vartheta_0}(\tau(n)) I(\vartheta_0)} = \frac{1}{n P_{\vartheta_0}(D) I(\vartheta_0)} \\
&= \frac{P_{\vartheta_0}(D)}{n\left(\int_D g_{\vartheta_0}\, dP_{\vartheta_0}\right)^2} \frac{1}{c_n} =: \frac{\delta_n^2}{c_n},
\end{aligned}$$

where $\delta_n = P_{\vartheta_0}^{1/2}(D) / (n^{1/2} \int_D g_{\vartheta_0}\, dP_{\vartheta_0})$ is the rate at which the alternative ϑ_n in (2) converges to 0.

If condition (4) is therefore satisfied with $c = 0$, that is, $c_n \to 0$ as $n \to \infty$, then, roughly there exists no unbiased estimator of ϑ_0 based on $Y_1, \dots, Y_{\tau(n)}$, whose standard deviation attains the rate δ_n, at which, in the testing problem, the number $\tau(n)$ can still separate between the null hypothesis ϑ_0 and the alternative $\vartheta_n \neq \vartheta_0$. On the other hand, we will show in the following that there actually exist estimators of ϑ_0 which are based on $\tau(n)$ and attain under appropriate regularity conditions the rate ϑ_n of convergence. This can also be confirmed by the Cramér–Rao inequality as above.

Recall that $\tau(n)$ is under ϑ_0 binomial $B(n, P_{\vartheta_0}(D))$-distributed, where we assume that $P_{\vartheta_0}(D)$ is small. In this case, the binomial distribution $B(n, P_{\vartheta_0}(D))$ can reasonably be approximated by the Poisson distribution with parameter $nP_{\vartheta_0}(D)$ and, for the sake of simplicity, we assume for the moment that $\tau(n)$ actually follows a Poisson distribution. In this case, $q_\vartheta(k) := \exp(-nP_\vartheta(D))(nP_\vartheta(D))^k/k!$, $k = 0, 1, 2 \dots$ is under ϑ the density of $\tau(n)$ with respect to the counting measure on $\{0, 1, 2, \dots\}$. Under appropriate regularity conditions we then have

$$\frac{\partial}{\partial\vartheta} \log\bigl(q_{\vartheta_0}(k)\bigr) = \frac{\int_D g\, dP_{\vartheta_0}}{P_{\vartheta_0}(D)} \bigl(k - nP_{\vartheta_0}(D)\bigr).$$

The corresponding Fisher information is now

$$\begin{aligned} \tilde{I}(\vartheta_0) &= E_{\vartheta_0}\left(\left(\frac{\partial}{\partial\vartheta} \log\bigl(q_{\vartheta_0}(\tau(n))\bigr)\right)^2\right) \\ &= \left(\frac{\int_D g\, dP_{\vartheta_0}}{P_{\vartheta_0}(D)}\right)^2 E_{\vartheta_0}\Bigl(\bigl(\tau(n) - nP_{\vartheta_0}(D)\bigr)^2\Bigr) = \delta_n^{-2}. \end{aligned}$$

For $\tilde{T}(\tau(n))$, being an unbiased estimator of ϑ_0 which is based on $\tau(n)$, the Cramér–Rao inequality implies therefore

$$E_{\vartheta_0}\Bigl(\bigl(\tilde{T}(\tau(n)) - \vartheta_0\bigr)^2\Bigr) \geq \frac{1}{\tilde{I}(\vartheta_0)} = \delta_n^2.$$

This inequality indicates that in the case $c = 0$ there actually exists an estimator of ϑ_0 based on $\tau(n)$, which has smaller rate of convergence than any other (unbiased) estimator that is based on $Y_1, \dots, Y_{\tau(n)}$, that is, δ_n compared to $\delta_n / c_n^{1/2}$.

Suppose that $\hat{\vartheta}_n$ is a solution of the equation

$$P_{\hat{\vartheta}_n}(D) = \tau(n)/n. \tag{19}$$

Recall that $\tau(n)$ is under ϑ_0 binomial $B(n, P_{\vartheta_0}(D))$-distributed, and thus, $\hat{\vartheta}_n$ is the maximum likelihood estimator of ϑ_0 for the family $\{\mathscr{L}_\vartheta(\tau(n)): \vartheta \in \Theta\}$. It turns out that $\hat{\vartheta}_n$ is under appropriate conditions an asymptotically optimal estimator of the underlying parameter ϑ_0. This will be a consequence of the following result.

LEMMA 3.1. *Suppose that Θ is an open subset of $\mathbb{R}$ and that expansion* (1) *holds for every $\vartheta_0 \in \Theta$. Precisely, we assume that for every $\vartheta_0 \in \Theta$ and ϑ near ϑ_0,*

$$f_\vartheta = f_{\vartheta_0}\Big(1 + (\vartheta - \vartheta_0) g_{\vartheta_0} + (\vartheta - \vartheta_0)^2 h_{\vartheta,\vartheta_0}\Big)$$

on D, where $\int_D g_{\vartheta_0}^2 \, dP_{\vartheta_0} < \infty$ and $\sup_{|\vartheta-\vartheta_0|<\varepsilon} \int_D h_\vartheta^4 \, dP_{\vartheta_0} < \infty$ for some $0 < \varepsilon = \varepsilon(\vartheta_0)$. If $\hat{\vartheta}_n$ satisfies condition (7) *in probability, that is, $nP_{\vartheta_0}^{1/2}(D)(\hat{\vartheta}_n - \vartheta_0)^4 = o_{P_{\vartheta_0}}(1)$, $P_{\vartheta_0}(D) \to 0$ and $nP_{\vartheta_0}(D) \to \infty$ as $n \to \infty$, then*

$$\delta_n^{-1}\big(\hat{\vartheta}_n - \vartheta_0\big) = \frac{\tau(n) - nP_{\vartheta_0}(D)}{\big(nP_{\vartheta_0}(D)\big)^{1/2}} + o_{P_{\vartheta_0}}(1) \to_{\mathscr{D}_{\vartheta_0}} N(0,1).$$

PROOF. We can write, for $\hat{\vartheta}_n$ near ϑ_0,

$$\frac{\tau(n)}{n} = P_{\hat{\vartheta}_n}(D) = \int_D 1 + \big(\hat{\vartheta}_n - \vartheta_0\big) g_{\vartheta_0} + \big(\hat{\vartheta}_n - \vartheta_0\big)^2 h_{\hat{\vartheta}_n,\vartheta_0} \, dP_{\vartheta_0},$$

which implies

$$\frac{\tau(n) - nP_{\vartheta_0}(D)}{n^{1/2} P_{\vartheta_0}^{1/2}(D)} = \frac{n^{1/2}}{P_{\vartheta_0}^{1/2}(D)}\big(\hat{\vartheta}_n - \vartheta_0\big)\int_D g_{\vartheta_0} \, dP_{\vartheta_0} + \frac{n^{1/2}}{P_{\vartheta_0}^{1/2}(D)}\big(\hat{\vartheta}_n - \vartheta_0\big)^2 \int_D h_{\hat{\vartheta}_m,\vartheta_0} \, dP_{\vartheta_0}.$$

Since by the Cauchy–Schwarz inequality

$$\left|\int_D h_{\hat{\vartheta}_n,\vartheta_0} \, dP_{\vartheta_0}\right| \le P_{\vartheta_0}^{3/4}(D)\left(\int_D h_{\hat{\vartheta}_n,\vartheta_0}^4 \, dP_{\vartheta_0}\right)^{1/4},$$

the assertion of Lemma 1.2 follows from condition (7). $\square$

The estimator $\hat{\vartheta}_n$ converges therefore at the same rate δ_n at which the alternative $\vartheta_n = \vartheta_n(\xi)$ in (7) converges to ϑ_0. If $c = 0$ in condition (4), then $\delta_n^{-1}(\hat{\vartheta}_n - \vartheta_0)$ coincides asymptotically with the central sequence and thus, we have, by LeCam's first lemma and Theorem 1.1, that the expansion in Lemma 3.1 also holds under the alternative, yielding

$$\delta_n^{-1}\big(\hat{\vartheta}_n - \vartheta_n\big) \to_{\mathscr{D}_{\vartheta_n}} N(0,1).$$

Defined in (19) $\hat{\vartheta}_n$ is therefore an asymptotically efficient estimator, which attains the lower bound 1 for the limiting variance within the class of regular estimators of ϑ_0 that are based on N_n^D.

EXAMPLE 3.2 (Pareto tail). Suppose that P_ϑ, $\vartheta > 0$, is a probability measure on $\mathbb{R}$ with a standard Pareto upper tail, that is, P_ϑ has distribution function $F_\vartheta(x) = 1 - x^{-\vartheta}$, $x \ge x_0 \ge 1$, where x_0 is unknown. Choose $\vartheta_0 > 0$

and put $D = [t, \infty]$, where $t = t_n$ is sufficiently large. Then P_{ϑ_0} satisfies expansion (1) with $g_{\vartheta_0}(x) = \vartheta_0^{-1} - \log(x)$, $x \in D$ (see also Examples 2.1 and 2.3). In this case we have

$$P_{\hat{\vartheta}_n}(D) = \frac{\tau(n)}{n} \quad \Leftrightarrow \quad \hat{\vartheta}_n = -\frac{\log(\tau(n)/n)}{\log(t)}$$

and

$$\delta_n = \frac{P_{\vartheta_0}^{1/2}(D)}{n^{1/2}\int_D g_{\vartheta_0}\, dP_{\vartheta_0}} = -\frac{t^{\vartheta_0/2}}{n^{1/2}\log(t)} = -\frac{1}{\left(nP_{\vartheta_0}(D)\right)^{1/2}\log(t)}.$$

From the expansion $\log(1+\varepsilon) = \varepsilon + O(\varepsilon^2)$, $\varepsilon \to 0$, we obtain therefore

$$\begin{aligned}
\delta_n^{-1}\left(\hat{\vartheta}_n - \vartheta_0\right) &= \left(nP_{\vartheta_0}(D)\right)^{1/2}\log(t)\left(\frac{\log(\tau(n)/n)}{\log(t)} + \vartheta_0\right) \\
&= \left(nP_{\vartheta_0}(D)\right)^{1/2}\log\left(\frac{\tau(n)t^{\vartheta_0}}{n}\right) \\
&= \left(nP_{\vartheta_0}(D)\right)^{1/2}\log\left(1 + \frac{\tau(n) - nP_{\vartheta_0}(D)}{nP_{\vartheta_0}(D)}\right) \\
&= \left(nP_{\vartheta_0}(D)\right)^{1/2}\left(\frac{\tau(n) - nP_{\vartheta_0}(D)}{nP_{\vartheta_0}(D)}\right. \\
&\qquad \left. + O_{P_{\vartheta_0}}\left(\left(\frac{\tau(n) - nP_{\vartheta_0}(D)}{nP_{\vartheta_0}(D)}\right)^2\right)\right) \\
&= \frac{\tau(n) - nP_{\vartheta_0}(D)}{\left(nP_{\vartheta_0}(D)\right)^{1/2}} + o_{P_{\vartheta_0}}(1) \to_{\mathscr{D}_{P_{\vartheta_0}}} N(0,1)
\end{aligned}$$

provided $t = t_n \to \infty$ and $nP_{\vartheta_0}(D) = nt_n^{-\vartheta_0} \to \infty$ as $n \to \infty$. The condition $nP_{\vartheta_0}^{1/2}(D)(\hat{\vartheta}_n - \vartheta_0)^4 = o_{P_{\vartheta_0}}(1)$, is then automatically satisfied if $nP_{\vartheta_0}^{1/2}(D)\delta_n^4 \to 0$ as $n \to \infty$.

The estimator $\hat{\vartheta}_n = -\log(\tau(n)/n)/\log(t)$ is by the preceding remarks an asymptotically efficient estimator of the underlying shape parameter ϑ_0 within the class of regular estimators that are based on $N_n^D = \sum_{i=1}^{\tau(n)} \varepsilon_{Y_i}$. Note that $\hat{\vartheta}_n$ depends only on the number $\tau(n)$ of the exceedances Y_i over the threshold $t = t_n$ but not on their actual values.

Using the order statistics approach, which is on the real line $\mathbb{R}$ a natural competitor to the POT approach for a statistical analysis of the tails of a distribution, Falk (1995b) observed an analogous phenomenon. There it was noted that

$$\tilde{\vartheta}_n := \log(n/k)/\log(X_{n-k+1:n})$$

outperforms the Hill (1975) estimator

$$\hat{\vartheta}_{n,\,\mathrm{Hill}} := \left((k-1)^{-1} \sum_{i=1}^{k-1} \log(X_{n-i+1:n}) - \log(X_{n-k+1:n}) \right)^{-1}$$

of the underlying shape parameter ϑ_0 in the above Pareto model. By $X_{1:n} \leq \cdots \leq X_{n:n}$, we denote the ordered values pertaining to the original sample $X_1, \dots, X_n$. Precisely, we have

$$\frac{k^{1/2}}{\vartheta_0} \left(\hat{\vartheta}_{n,\,\mathrm{Hill}} - \vartheta_0 \right) \to_{\mathscr{D}_{\vartheta_0}} N(0,1),$$

whereas the estimator $\tilde{\vartheta}_n$, which is based only on the kth largest order statistic $X_{n-k+1:n}$, satisfies

$$\frac{k^{1/2} \log(n/k)}{\vartheta_0} \left(\tilde{\vartheta}_n - \vartheta_0 \right) \to_{\mathscr{D}_{\beta_0}} N(0,1)$$

with $k = k(n) \to \infty$, $k/n \to 0$ as $n \to \infty$. Note that by considering $X_{n-k+1:n}$ as a *random* threshold t_n, we obtain $\tau(n) = k$ and

$$\tilde{\vartheta}_n = -\log(\tau(n)/n)/\log(t_n) = \hat{\vartheta}_n,$$

which shows the close correspondence between the POT and the order statistics approach in this case.

Adding a scale parameter σ to the Pareto model P_ϑ and reparameterizing $P_{\vartheta,\,\sigma(\vartheta)}$ via least favorable alternatives, Marohn (1995) established LAN of log-likelihood ratios in the order statistics as well as in the POT approach, with the Hill estimator now being part of the central sequence which corresponds to the shape parameter ϑ. Analogous phenomena were observed in particular for certain stable processes by Höpfner and Jacod (1994) and Höpfner (1994, 1997).

If the parameter space Θ in Lemma 1.2 is an open subset of $\mathbb{R}^d$, $d > 1$, then there is usually not a unique solution of the equation

$$P_{\hat{\vartheta}_n}(D_n) = \tau(n)/n.$$

In this case one can define $\hat{\vartheta}_n$ as an solution of the *system* of equations

$$P_{\hat{\vartheta}_n}(D(n_i)) = \tau(n_i)/n, \qquad 1 \leq i \leq d,$$

where $n_1 = n_1(n) \leq \cdots \leq n_d = n_d(n) \leq n$ are sample sizes smaller than n with corresponding subsets $D(n_i)$, $1 \leq i \leq d$. Suppose again that condition (1) is satisfied for every $\vartheta_0 \in \Theta$ with $\int \|g_{\vartheta_0}\|^2 \, dP_{\vartheta_0} < \infty$ and $\sup_{\|\vartheta - \vartheta_0\| < \varepsilon} \int h^4_{\vartheta,\,\vartheta_0} \, dP_{\vartheta_0} < \infty$ for some $0 < \varepsilon = \varepsilon(\vartheta_0)$. If $\hat{\vartheta}_n$ satisfies now condition (7) in probability, that is, $n_i P^{1/2}_{\vartheta_0}(D(n_i)) \|\hat{\vartheta}_n - \vartheta_0\|^4 = o_{P\vartheta_0}(1)$, $P_{\vartheta_0}(D(n_i)) \to 0$ and $n_i P_{\vartheta_0}(D(n_i)) \to \infty$ as $n \to \infty$, $1 \leq i \leq d$, then we obtain from the arguments in the proof of Lemma 3.1 the representation

$$\left\langle \frac{1}{\delta(n_i)}, \hat{\vartheta}_n - \vartheta_0 \right\rangle = \frac{\tau(n_i) - n_i P_{\vartheta_0}(D(n_i))}{\left(n_i P_{\vartheta_0}(D)\right)^{1/2}} + o_{P\vartheta_0}(1) \to_{\mathscr{D}_{\vartheta_0}} N(0,1),$$

$$1 \le i \le d,$$

where with $g_{\vartheta_0} = (g^1_{\vartheta_0}, \dots, g^d_{\vartheta_0})$

$$\delta(n_i) := \frac{P^{1/2}_{\vartheta_0}(D(n_i))}{n_i^{1/2}} \frac{1}{\int_{D(n_i)} g_{\vartheta_0}\, dP_{\vartheta_0}}, \qquad 1 \le i \le d,$$

and the operation $1/x$ for $x \in \mathbb{R}^d$ is meant componentwise. If the n_i are now all of the same order n, then the preceding considerations indicate that $\hat{\vartheta}_n - \vartheta_0$ is again of the precise order $\|\delta(n)\|$.

As an example, we consider again a class of distributions on the real line with upper Pareto tail, where we add now a scale parameter. Suppose therefore that with $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta := (0,\infty)^2$, the distribution P_ϑ has df $F_\vartheta(x) = 1 - (x/\vartheta_2)^{-\vartheta_1}$ for $x \ge x_0 = x_0(\vartheta_2) \ge \vartheta_2$. Put $D_n := [t_n, \infty]$. The equations

$$P_{\hat{\vartheta}_n}(D(n_i)) = \tau(n_i)/n_i, \qquad i = 1,2$$

lead to the system

$$\hat{\vartheta}^{(1)}_n\left(\log(t_{n_i}) - \log\left(\hat{\vartheta}^{(2)}_n\right)\right) = -\log(\tau(n_i)/n_i), \qquad i = 1,2,$$

with the solutions

$$\hat{\vartheta}^{(1)}_n = \frac{\log(\tau(n_2)/n_2) - \log(\tau(n_1)/n_1)}{\log(t_{n_1}) - \log(t_{n_2})}.$$

and

$$\hat{\vartheta}^{(2)}_n = \exp\left(\frac{\log(t_{n_1})\log(\tau(n_2)/n_2) - \log(t_{n_2})\log(\tau(n_1)/n_1)}{\log(\tau(n_2)/n_2) - \log(\tau(n_1)/n_1)}\right).$$

The estimator $\hat{\vartheta}^{(1)}_n$ is close to the Pickands (1975) estimator of ϑ_1 in the order statistics approach, if we replace the fixed threshold t_{n_i} by the random threshold $X_{n-k_i+1:n}$ and $\tau(n_i)/n_i$ by k_i/n, $i = 1,2$. Replacing $\tau(n_i)/n_i$ by its expectation $P_{\vartheta_0}(D(n_i)) = (t_{n_i}/\vartheta_2)^{-\vartheta_1}$, we obtain

$$\hat{\vartheta}^{(1)}_n \sim \frac{\log\left((t_{n_2}/\vartheta_2)^{-\vartheta_1}\right) - \log\left((t_{n_1}/\vartheta_2)^{-\vartheta_1}\right)}{\log(t_{n_1}) - \log(t_{n_2})} = \vartheta_1$$

and

$$\hat{\vartheta}^{(2)}_n \sim \exp\left(\frac{\log(t_{n_1})\log\left((t_{n_2}/\vartheta_2)^{-\vartheta_1}\right) - \log(t_{n_2})\log\left((t_{n_1}/\vartheta_2)^{-\vartheta_1}\right)}{\log\left((t_{n_2}/\vartheta_2)^{-\vartheta_1}\right) - \log\left((t_{n_1}/\vartheta_2)^{-\vartheta_1}\right)}\right) = \vartheta_2.$$

It is therefore easy to see that under appropriate conditions on n_i and t_{n_i}, $i = 1, 2$, the estimator $\hat{\vartheta}_n = (\hat{\vartheta}_n^{(1)}, \hat{\vartheta}_n^{(2)})$ is asymptotically consistent and $\hat{\vartheta}_n - \vartheta_0$ is asymptotically normal.

Acknowledgment. The author is indebted to Frank Marohn for valuable suggestions.

REFERENCES

ANDERSON, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1992). *Statistical Models Based on Counting Processes*. Springer, New York.

BALKEMA, A. A. and DE HAAN, L. (1974). Residual lifetime at great age. *Ann. Probab.* **2** 792–804.

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. and MAROHN, F. (1993). Asymptotically optimal tests for conditional distributions. *Ann. Statist.* **21** 45–60.

FALK, M., HÜSLER, J. and REISS, R.-D. (1994). *Laws of Small Numbers: Extreme and Rare Events*. Birkhäuser, Basel.

HÁJEK, J. (1962). Asymptotically most powerful rank order tests. *Ann. Math. Statist.* **33** 1224–1147.

HÁJEK, J. (1970). A characterisation of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* **12** 21–55.

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

HÖPFNER, R. (1994). On tail parameter estimation in certain point process models. *J. Statist. Plann. Inference*. To appear.

HÖPFNER, R. (1997). Two comments on parameter estimation in stable processes. *Math. Methods Statist.* **6** 125–134.

HÖPFNER, R. and JACOD J. (1994). Some remarks on the joint estimation of the index and the scale parameter for stable processes. In *Asymptotic Statistics. Proceedings of the Fifth Prague Symposium 1993* (P. Mandl and M. Huskova, eds.) 273–284. Physica, Heidelberg.

IBRAGIMOV, L. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation. Asymptotic Theory*. Springer, New York.

JANSSEN, A. and MAROHN, F. (1994). On statistical information of extreme order statistics, local, extreme value alternatives, and Poisson point processes. *J. Multivariate Anal.* **48** 1–30.

KARR, A. F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. Dekker, New York.

LECAM, L. (1956). On the asymptotic theory of estimation and testing hypothesis. *Proc. Third Berkeley Symp. Math. Statist. Probab.* 129–156. Univ. California Press, Berkeley.

LECAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, New York.

LECAM, L. and YANG, G. L. (1990). *Asymptotics in Statistics. Some Basic Concepts*. Springer, New York.

MAROHN, F. (1995). Contributions to a local approach in extreme value statistics. Habilitation thesis, Katholische Univ. Eichstätt.

NISHIYAMA, Y. (1995). Local asymptotic normality of a sequential model for marked point processes and its application. *Ann. Inst. Statist. Math.* **47** 195–209.

PFANZAGL, J. (1994). *Parametric Statistical Theory*. de Gruyter, Berlin.

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

REISS, R.-D. (1993). *A Course on Point Processes*. Springer, New York.

RYCHLIK, T. (1992). Weak limit theorems for stochastically largest order statistics. In *Order Statistics and Nonparametrics: Theory and Applications* (P. K. Sen and I. A. Salama, eds.) 141–154. North-Holland, Amsterdam.

SAKSENA, S. K. and JOHNSON, A. M. (1984). Best unbiased estimators for the parameters of a two-parameter Pareto distribution. *Metrika* **31** 77–83.

STRASSER, H. (1985). *Mathematical Theory of Statistics*. de Gruyter, Berlin.

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

MATHEMATISCH-GEOGRAPHISCHE FAKULTÄT
KATHOLISCHE UNIVERSITÄT EICHSTÄTT
85071 EICHSTÄTT
GERMANY
E-MAIL: michael.falk@ku-eichstaett.de