JOURNAL OF MULTIVARIATE ANALYSIS 48, 1-30 (1994)

On Statistical Information of Extreme Order Statistics, Local Extreme Value Alternatives, and Poisson Point Processes

A. JANSSEN

Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

AND

F. MAROHN

Katholische Universität Eichstätt, Eichstätt, Germany

The aim of the present paper is to clarify the rôle of extreme order statistics in general statistical models. This is done within the general setup of statistical experiments in LeCam's sense. Under the assumption of monotone likelihood ratios, we prove that a sequence of experiments is asymptotically Gaussian if, and only if, a fixed number of extremes asymptotically does not contain any information. In other words: A fixed number of extremes asymptotically contains information iff the Poisson part of the limit experiment is non-trivial. Suggested by this result, we propose a new extreme value model given by local alternatives. The local structure is described by introducing the space of extreme value tangents. It turns out that under local alternatives a new class of extreme value distributions appears as limit distributions. Moreover, explicit representations of the Poisson limit experiments via Poisson point processes are found. As a concrete example nonparametric tests for Fréchet type distributions against stochastically larger alternatives are treated. We find asymptotically optimal tests within certain threshold models.

1. INTRODUCTION AND NOTATION

The present paper aims to clarify the rôle of extreme observations within i.i.d. models of real valued random variables in the asymptotic setting. The starting points of our investigation are the recent monographs of Reiss

Received August 28, 1990; revised March 18, 1993.

MSC 1991 subject classifications: primary 62B10, 62G30, secondary 60G55.

Key words and phrases: Extreme order statistics, statistical information, Gaussian experiments, Poisson experiments, Poisson point processes, local extreme value alternatives, extreme value tangents, Fréchet distribution, intensity model, hazard rate model.

0047-259X/94 $6.00

[19] and Resnick [21] about extreme value theory where the high mathematical standard of extreme value theory is documented. On the other hand, the asymptotic statistical theory was recently deeply influenced by the description of models given by tangent cones; see Pfanzagl and Wefelmeyer [18], which goes back to earlier work of Koshevnik and Levit [11].

The investigation of the structure of extremes and their statistical experiments has two aspects, which may be described as follows:

I. What is the asymptotic contribution of a finite number of extremes for a given arbitrary model?

II. Which kind of asymptotic models can be used for extreme value problems?

The present questions can naturally be embedded in the universal applicable theory of statistical experiments of LeCam [12]. Recall that the asymptotic statistical properties of a given sequence of rowwise i.i.d. observations are completely described by the class of infinitely divisible limit experiments $F = G \otimes P$ which can uniquely be decomposed into the product of a Gaussian experiment G and a Poisson experiment P; see LeCam [12, Chapt. 9] and Milbrodt and Strasser [17]. Within this concept, the relevance of a given portion of extreme order statistics is discussed in terms of their limit experiments.

The most popular models (such as those given by tangent cones) are asymptotically Gaussian models. In that case, a finite number of extremes can asymptotically be neglected without any loss of information. In these circumstances, the extremes are often suspected to be outliers and they may be thrown out for the sake of robustness. In Section 2, the main result, Theorem 2.2, shows that this is the only case where extremes may be cancelled. The Gaussian limit experiments are completely characterized by extremes: Under monotone likelihood ratios the limit experiment is Gaussian if, and only if, the extreme observations yield no information. The consequences of that result are twofold: Note first that extreme value problems cannot be described by asymptotically Gaussian experiments (including tangent cones) but Poisson experiments appear naturally. Second, the result gives a further foundation of extreme value theory which is always relevant except for the purely Gaussian limit case. The characterization of Gaussian limit experiments may be compared with the behaviour of central order statistics; see Example 2.3. They are, of course, non-negligible and are frequently used as rough estimators of the underlying parameter. Finally, we remark that there is a formal accordance between the present results for statistical experiments and sums of independent random variables considered earlier by Gnedenko and Kolmogorov

[3] and Loève [15]. They showed that the underlying sums are asymptotically normal if, and only if, the lower and upper extremes converge to zero in probability.

In Section 3, we show that certain Poisson limit experiments can completely be described by extremes. Note that from abstract results it is known that they can be realized by Poisson point processes; compare with the general program of LeCam [12] and Milbrodt [16]. Here, we find an explicit representation via extreme value processes of that result. The results of Section 2 suggest extreme value models given by a set of intensities of Poisson point processes, which stand for a collection of local extreme value alternatives. The local structure of alternatives is described by the space of extreme value tangents, which is introduced (Definition 3.1). In Section 4, a first statistical application of the concept of extreme value tangents is given by an example. We study tests for Fréchet type distributions against stochastically larger alternatives. They have a natural interpretation within the concept of intensity measures and hazard rates. As an application, their asymptotic optimality is discussed. In order not to disturb the main ideas of the present paper, all proofs are postponed to Section 5.

For the remainder of this section, the notation is introduced and some facts concerning statistical experiments are recalled. The reader is referred to LeCam [12], LeCam and Yang [13], Milbrodt and Strasser [17], Strasser [23], and Torgersen [25].

Let $E=(\Omega, \mathscr{A}, \{P_t : t \in T\})$ be a statistical experiment (sometimes briefly denoted by $\{P_t : t \in T\}$) Then E^n denotes the nth product experiment of E consisting of the product measures P_t^n. The restriction of E to some subset $\Theta \subset T$ is denoted by $E|_\Theta$. Let

$$\frac{dP_t}{dP_s}=\frac{dP_t}{d(P_s+P_t)}\left(\frac{dP_s}{d(P_s+P_t)}\right)^{-1} 1_{(0,\infty)}\left(\frac{dP_s}{d(P_s+P_t)}\right)+\infty 1_{\{0\}}\left(\frac{dP_s}{d(P_s+P_t)}\right)$$

denote the likelihood ratio of P_t w.r.t. P_s.

Two statistical experiments $E=(\Omega_1, \mathscr{A}_1, \{P_t : t \in T\})$ and $F=(\Omega_2, \mathscr{A}_2, \{Q_t : t \in T\})$ are called equivalent $(E \sim F)$ if the distributions of all likelihood processes of E and F coincide:

$$\mathscr{L}\left(\left(\frac{dP_t}{dP_s}\right)_{t\in T} \,\middle|\, P_s\right)=\mathscr{L}\left(\left(\frac{dQ_t}{dQ_s}\right)_{t\in T} \,\middle|\, Q_s\right), \qquad s \in T.$$

The weak convergence of classes of experiments (w.r.t. $\sim$) is defined by the weak convergence of all finite dimensional marginal distributions of the log-likelihood process. The weak convergence of E_n to E is denoted by $E_n \to E$.

By definition,

$$d(P, Q)=\left(\frac{1}{2}\int\left(\left(\frac{dP}{d(P+Q)}\right)^{1/2}-\left(\frac{dQ}{d(P+Q)}\right)^{1/2}\right)^2 d(P+Q)\right)^{1/2}$$

is the Hellinger distance and $\|P-Q\|$ is the norm of total variation between two probability measures P and Q. Recall that

$$1-d^2(P^n, Q^n)=(1-d^2(P, Q))^n \tag{1.1}$$

and that weak convergence of experiments implies convergence of the corresponding Hellinger distances.

Let $E_n=(\Omega_n, \mathscr{A}_n, \{P_n, Q_n\})$, $n\in\mathbb{N}$, be a sequence of binary experiments. A sequence $(Q_n)_n$ is called contiguous to the sequence $(P_n)_n$ if $P_n(A_n)\to 0$ implies $Q_n(A_n)\to 0$ for $A_n\in\mathscr{A}_n$. Suppose that $(E_n)_n$ is weakly convergent to some binary limit experiment $E=(\Omega, \mathscr{A}, \{P, Q\})$. By LeCam's first lemma, the sequence $(Q_n)_n$ is then contiguous to the sequence $(P_n)_n$ iff Q is absolutely continuous w.r.t. P.

An experiment of mutually absolutely continuous distributions is called Gaussian if at least one log-likelihood process is a Gaussian process. For example, the experiment $(\mathbb{R}^k, \mathscr{B}^k, \{N(t\Gamma, \Gamma): t\in\mathbb{R}^k\})$, where $N(a, \Gamma)$ denotes the normal distribution with mean a and covariance matrix Γ on $\mathbb{R}^k$, defines a Gaussian experiment which is usually called Gaussian shift on $\mathbb{R}^k$. An experiment E_0 is said to be totally uninformative if all distributions coincide for $t\in T$. Recall that convergence of a sequence of experiments $(\{P_{n,t}: t\in T\})_n$ to the totally uninformative experiment is equivalent to $\|P_{n,t_1}-P_{n,t_2}\|\to 0$ for all $t_1, t_2\in T$. In this case, no information about the unknown parameter t is available within the asymptotic setting.

A positive function L defined on some neighborhood $[x_0, \infty)$ of infinitely is regularly varying (at infinity) of index $\rho\in\mathbb{R}$ if

$$L(tx)/L(t)\to x^\rho, \qquad t\to\infty.$$

For $\rho=0$ the function L is usually called slowly varying. For the background concerning regular varying functions, we refer to the monograph by Bingham, Goldie, and Teugels [1]. The concept of regularly varying functions has proved to be a powerful tool in extreme value theory and for details we refer to Section 8.13 of [1] and to the monograph by de Haan [4].

Let S be locally compact with countable base and let $\mathscr{B}$ denote the corresponding Borel σ-field. Designate by $M(S, \mathscr{B})$ the set of all locally finite point measures defined on S. Recall that $\mu\in M(S, \mathscr{B})$ if there exists a denumerable set of points $x_i\in S$, $i\in I$, such that $\mu=\sum_{i\in I}\varepsilon_{x_i}$ and $\mu(K)<\infty$ for every compact set K. The set $M(S, \mathscr{B})$ is endowed with the σ-field

$\mathcal{M}(S, \mathcal{B})$, which is by definition the smallest σ-fied such that the projections $\mu \to \mu(B)$, $B \in \mathcal{B}$, are measurable. The space $M(S, \mathcal{B})$ is Polish in the vague topology. Moreover, the σ-field $\mathcal{M}(S, \mathcal{B})$ coincides with the Borel ($\equiv$ Baire) σ-field w.r.t. the vague topology; see Kallenberg [9]). Sometimes, we simply write $M(S)$ and $\mathcal{M}(S)$ instead of $M(S, \mathcal{B})$ and $\mathcal{M}(S, \mathcal{B})$, respectively. A point process on $(S, \mathcal{B})$ is a measurable mapping N on some measurable space $(\Omega, \mathcal{A})$ into $M(S, \mathcal{B})$. An excellent introduction to the theory of point processes is the recent monograph by Reiss [20].

2. Extremes of Asymptotically Gaussian Models

In this section, let $P_{n,\vartheta}$ always denote continuous distributions of $\mathbb{R}$, and let $X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$ denote the order statistics of the canonical projections $X_i\colon \mathbb{R}^n \to \mathbb{R}$ on the ith coordinate. The consideration below deals with the k-dimensional lower and upper extremes given by

$$W_{n,k} = (X_{1:n}, ..., X_{k:n}) \qquad \text{and} \qquad Z_{n,k} = (X_{n+1-k:n}, ..., X_{n:n}). \quad (2.1)$$

The associated statistical experiments are abbreviated by

$$E_{j:n, j \leqslant k} = (\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(W_{n,k} \mid P^n_{n,\vartheta}) : \vartheta \in \Theta_n\})$$

and $E_{n+1-j:n, j \leqslant k}$, respectively, with $W_{n,k}$ replaced by $Z_{n,k}$, where $\Theta_n \subset \mathbb{R}$ denotes a suitable parameter set with $1_{\Theta_n} \to 1_\Theta$ for some parameter set Θ. Throughout, we assume that $0 \in \Theta$. The following two theorems show that under standard regularity assumptions asymptotic Gaussian models are not appropriate for extreme value problems. On the other hand, Theorem 2.2 is essential for extreme value theory. Note that whenever the limit experiment is not Gaussian then the extremes yield a non-negligible asymptotic contribution.

Asymptotic normality is often derived under the standard assumption of L^q-differentiability (at 0) of a given family $(P_\vartheta)_\vartheta$ for some $1 \leqslant q \leqslant 2$, which is satisfied if there exists some $g \in L^q(P_0)$, called the q-derivative of the likelihood ratio, with

$$q\left(\left(\frac{dP_\vartheta}{dP_0}\right)^{1/q} - 1\right)\Big/\vartheta \to g$$

in $L^q(P_0)$ as $\vartheta \to 0$ and

$$P_\vartheta\left(\left\{\frac{dP_\vartheta}{d(P_0 + P_\vartheta)} = 1\right\}\right) = o(|\vartheta|^q)$$

as $\vartheta \to 0$. By $L^q(P_0)$ we denote the space of L^q-integrable functions w.r.t. P_0.

Recall that a family $(P_\vartheta)_\vartheta$ is called stochastically increasing if $P_{\vartheta_1}((x, \infty)) \geqslant P_{\vartheta_2}((x, \infty))$ for all $x \in \mathbb{R}$ and each pair $\vartheta_1 > \vartheta_2$. Before we state our results, note that the totally uninformative experiment E_0 contains no information about the unknown parameter ϑ.

2.1. THEOREM. *Assume that the sequence of experiments* $(\mathbb{R}^n, \mathscr{B}^n, \{P^n_{n,\vartheta} : \vartheta \in \Theta_n\})$ *is weakly convergent to some Gaussian limit experiment G. Either let*

(a) $P_{n,\vartheta} = P_{n^{-1/q}\vartheta}$ *arise from an L^q-differentiable family* $(P_\vartheta)_\vartheta$ *(at zero) for* $1 \leqslant q \leqslant 2$, *or let*

(b) $(P_{n,\vartheta})_\vartheta$ *be stochastically increasing for each n.*

Then for each $k \geqslant 1$ *the extreme value experiments* $E_{j:n, j \leqslant k}$ *and* $E_{n+1-j:n, j \leqslant k}$ *are weakly convergent to the totally uninformative experiment* E_0.

Under slightly stronger conditions also the converse of Theorem 2.1 can be obtained. In that case, Gaussian limit experiments are completely characterized by the extremes. Recall that a family $(P_\vartheta)_\vartheta$ on $\mathbb{R}$ has monotone likelihood ratio in x, if for each pair $\vartheta_1 < \vartheta_2$

$$\frac{dP_{\vartheta_2}}{dP_{\vartheta_1}}(x) = h_{\vartheta_1, \vartheta_2}(x) \tag{2.2}$$

holds $P_{\vartheta_1} + P_{\vartheta_2}$-almost everywhere for some nondecreasing function $h_{\vartheta_1, \vartheta_2} \colon \mathbb{R} \to [0, \infty]$. It is well known that (2.2) implies that $(P_\vartheta)_\vartheta$ is stochastically increasing; cf. Witting [26, p. 214]. The boundedness assumption of (2.3) below is standard in asymptotic theory; cf. [17, p. 40].

2.2. THEOREM. *Assume that* $(P_{n,\vartheta})_{\vartheta \in \Theta_n}$ *admits monotone likelihood ratios in x for each* $n \in \mathbb{N}$. *Let the n-fold product experiment* $E^n = (\mathbb{R}^n, \mathscr{B}^n, \{P^n_{n,\vartheta} : \vartheta \in \Theta_n\})$ *be bounded, that is,*

$$\limsup_{n \to \infty} n\, d^2(P_{n,\vartheta_1}, P_{n,\vartheta_2}) < \infty \qquad \text{for all } \vartheta_1, \vartheta_2. \tag{2.3}$$

Let E^n be weakly convergent to some limit experiment E. For fixed $k \in \mathbb{N}$ *the following assertions* (a)–(c) *are equivalent:*

(a) *E is Gaussian.*

(b) *For each* $\vartheta \in \Theta$

$$\left(\log \frac{dP_{n,\vartheta}}{dP_{n,0}}(X_{i:n})\right)_{i \in \{1, \ldots, k, n+1-k, \ldots, n\}} \to 0$$

in $P^n_{n,0}$- *and* $P^n_{n,\vartheta}$-*probability.*

(c) $E_{j:n, j \leqslant k} \to E_0$ *and* $E_{n+1-j:n, j \leqslant k} \to E_0$ *weakly as* $n \to \infty$.

In the case of central order statistics the situation is completely different. The experiments of central order statistics are in general not negligible and often again asymptotically Gaussian, as can be seen from the next example.

2.3. EXAMPLE. Let F denote a distribution function on $\mathbb{R}$ with absolutely continuous Lebesgue density f and finite Fisher information $I_f = \int (f'(x))^2/f(x)\,dx < \infty$. Consider the stochastically increasing location family P_ϑ with distribution functions $F(\cdot - \vartheta)$.

Whenever $f(F^{-1}(q)) > 0$, $0 < q < 1$, the experiment of central order statistics

$$(\mathbb{R}, \mathscr{B}, \{\mathscr{L}(X_{[nq]:n} \mid P^n_{n^{-1/2}\vartheta}) : \vartheta \in \mathbb{R}\})$$

converges weakly to the one-dimensional Gaussian shift $(\mathbb{R}, \mathscr{B}, \{N(\vartheta\sigma^2, \sigma^2) : \vartheta \in \mathbb{R}\})$ with the variance

$$\sigma^2 = f^2(F^{-1}(q))(q(1-q))^{-1}.$$

The proof follows from Theorem 4.1.4 of Reiss [19], where the convergence of

$$n^{1/2} f(F^{-1}(q))(X_{[nq]:n} - F^{-1}(q))$$

to $N(\vartheta f(F^{-1}(q)), q(1-q))$ under $P^n_{n^{-1/2}\vartheta}$ is proved w.r.t. the variational distance.

Finally, we discuss the median $X_{[n/2]:n}$ $(q = \frac{1}{2})$ for 0-symmetric densities f. Its Fisher efficiency is given by $1 \geqslant \rho := \sigma^2/I_f = 4f^2(0)/I_f$. The inequality $4f^2(0) \leqslant I_f$ is well known and can be proved directly by the Cauchy–Schwarz inequality. The discussion of the equality sign proves that the median is asymptotically efficient (w.r.t. the Fisher efficiency), that is, $\rho = 1$, if and only if $f(x) = \exp(-|x|)/2$ is the density of the double exponential distribution.

3. Extreme Value Alternatives and Poisson Point Processes

As a conclusion of Section 2, we now study the precise influence of the (upper) extremes when the limit experiment is not Gaussian or only a portion of upper extremes is observable. It turns out that in various cases we find an explicit representation of the corresponding limit experiment by Poisson point processes. Let us first summarize some frequently applied approaches in extreme value theory.

Practical problems are often concerned with the upper tails of the underlying distributions. Here, we may think about the insurance mathematics, where extremely large claim sizes are most important. Another example, where naturally extremes appear, is the flood of the ocean or of rivers. These examples have the common feature that the interesting statistical information is located in the tails of the distributions and the shape of the rest of the distribution may be suppressed. There are two traditional methods in extreme value theory which take care of this effect.

1. For a given sample size n often only the k_n largest order statistics are taken into account (or even the k_n largest data points are only observable), where $X_{n+1-k_n:n}, ..., X_{n:n}$ is a relatively small portion of order statistics.

2. Often only the exceedances $Y_i = X_i 1_{[d, \infty)}(X_i)$ of X_i over a nonrandom threshold d are considered. In this case, the statistician has the idea that the effects of interest lie behind the level of size d.

In the sequel, we motivate a new local consideration in extreme value theory. Assume that a nonparametric model depends on a family $\mathscr{P}$ of extreme value distributions or related ones on $[0, \infty)$. We are interested in the performance of given procedures at a distribution $P_0 \in \mathscr{P}$. In practice, ad hoc methods relying on preliminary estimates $\hat{P}_n$ bring us into a neighborhood of P_0, where the accuracy of the approximation depends on the sample size n. At this stage, we again refer to the classical Gaussian situation; see Pfanzagl and Wefelmeyer [18]. The investigation of local parameters, expressed in terms of the tangent cone, leads for $n \to \infty$ to Gaussian shifts and provides there an adequate tool for treating asymptotic problems. By the results of Section 2, this approach fails in extreme value theory. Motivated by the points 1 and 2 above, we introduce now the following local model. The direction of deviation from P_0 is described by the set of all "smooth" functions $h \geq 0$ such that

$$\frac{dP_\vartheta}{dP_0}(x) = h(\vartheta x) + r(x, \vartheta), \qquad \vartheta \in \Theta \subset [0, \infty), 0 \in \Theta$$

leads to a curve of distributions P_ϑ which belong to $\mathscr{P}$ locally at zero. Among other conditions, we require $h(x) \to 1$ as $x \downarrow 0$, and that the remainder term $r(x, \vartheta)$ be sufficiently small as $\vartheta \downarrow 0$. Also, one may think about related models where $h(\vartheta x)$ is replaced by $h(\vartheta \varphi(x))$ given by a scale transformation φ and $\mathscr{L}(\varphi \mid P_0)$ satisfies the regularity conditions of our extreme value model. The present model has the following meaning. For ϑ near zero the difference between P_0 and P_ϑ is more and more located in the tails. Following the motivation above, the tail effects of the curve are described by the function h, which can be considered as an extreme value version of a tangent function or an influence function. The local structure of the family $\mathscr{P}$ at P_0 is specified below in Definition 3.1.

In the following we restrict ourselves to absolutely continuous distributions P_0 on $\mathbb{R}_+$ with distribution function F_0 and density f_0 w.r.t. the Lebesgue measure λ such that the von Mises condition

$$\lim_{x \to \infty} \frac{x f_0(x)}{1 - F_0(x)} = \alpha, \qquad \alpha > 0 \tag{3.1}$$

holds and f_0 is bounded on each intervall $[y, \infty)$, $y > 0$. Condition (3.1) implies that P_0 belongs to the max-domain of attraction of the Fréchet distribution with shape parameter α

$$G_{1,\alpha}(x) = \exp(-x^{-\alpha})\, 1_{(0,\infty)}(x), \qquad \alpha > 0, \tag{3.2}$$

that means, there exist constants $\delta_n > 0$, $\gamma_n \in \mathbb{R}$, such that $\mathscr{L}(\delta_n(X_{n:n} - \gamma_n) \mid P_0^n) \to G_{1,\alpha}$ weakly. Moreover, one can choose $\delta_n := 1/F_0^{-1}(1 - 1/n)$ and $b_n = 0$, where F_0^{-1} denotes the generalized inverse of F_0, see, e.g., De Haan [4]. Note that in our notation we do not distinguish between a distribution and its distribution function.

3.1. Definition (Extreme value tangent space). Let P_0 fulfill the von Mises condition (3.1) and let $\mathscr{P}$ be a family of probability measures with $P_0 \in \mathscr{P}$. Denote by Ψ the set of measurable functions h: $[0, \infty) \to [0, \infty)$ with

$$\lim_{x \downarrow 0} h(x) = 1 \tag{3.3}$$

and

$$\int_0^\infty h(x) \min\{1, x^{-(1+\alpha)+\varepsilon}\}\, dx < \infty \tag{3.4}$$

for some $0 < \varepsilon < \alpha$.

The extreme value tangent space $T_\Psi(P_0, \mathscr{P})$ is the set of functions $h \in \Psi$ such that a curve

$$\frac{dP_\vartheta}{dP_0}(x) = h(\vartheta x) + r(x, \vartheta), \qquad \vartheta \in \Theta \subset [0, \infty), 0 \in \Theta \tag{3.5}$$

and some $\eta > 0$ exist with $P_\vartheta \in \mathscr{P}$ for $0 < \vartheta < \eta$, where the remainder term r in (3.5) fulfills the regularity conditions

(i) $r(\delta_n^{-1} x, \delta_n \vartheta) \to 0$ for λ almost all $x > 0$ and

(ii) $\int_{x/\delta_n}^\infty |r(z, \delta_n \vartheta)| f_0(z)\, dz = o(n^{-1})$, for each $x > 0$

and for each $\vartheta > 0$ as $n \to \infty$, where $\delta_n := 1/F_0^{-1}(1 - 1/n)$.

Note that the function $x \to h(\vartheta x)$ also belongs to the tangent space for $\vartheta \geqslant 0$ if $h \in T_\Psi(P_0, \mathscr{P})$. The relevance of the normalizing sequence $(\delta_n)_n$ is discussed in (3.11) below. The next example has a natural application for testing problems; see Section 4.

3.2. EXAMPLE. Consider the Fréchet distribution P_0 of index $\alpha>0$ with distribution function $G_{1,\alpha}$; see (3.2). Let $\mathscr{P}$ denote the set of all distributions which are stochastically larger than P_0. The tangent space $T_\Psi(P_0, \mathscr{P})$ consists of all functions $h \in \Psi$ satisfying

$$\alpha \int_x^\infty h(z)\, z^{-(1+\alpha)}\, dz \geqslant x^{-\alpha} \qquad \text{for all} \quad x>0. \tag{3.6}$$

Let $P_\vartheta \in \mathscr{P}$ be a curve of the form (3.5). Then Theorem 3.5 below implies (3.6). Conversely, for all $h \in \Psi$ satisfying (3.6), a curve P_ϑ exists which fulfills (i) and (ii) of Defintion 3.1. Straightforward calculations show that one can choose the curve P_ϑ represented by

$$P_\vartheta((-\infty, x]) = \exp\left(-\alpha \int_x^\infty h(\vartheta z)\, z^{-(1+\alpha)}\, dz\right), \qquad x>0. \tag{3.7}$$

The model (3.7) has a very interesting interpretation in terms of intensity rates and hazard rates if in the latter case the transformation $x \to \varphi(x) := 1/x$, $x>0$, is applied. Let f_ϑ denote the Lebesgue density of P_ϑ. Then the intensity rate is defined by

$$\frac{f_\vartheta(x)}{P_\vartheta((-\infty, x])} = h(\vartheta x)\, \alpha x^{-(1+\alpha)}, \qquad x>0$$

and the intensity ratio is just

$$\frac{f_\vartheta(x)}{P_\vartheta((-\infty, x])} \left(\frac{f_0(x)}{P_0((-\infty, x])}\right)^{-1} = h(\vartheta x). \tag{3.8}$$

The family $(\tilde{P}_\vartheta)_\vartheta := (\mathscr{L}(\varphi \mid P_\vartheta))_\vartheta$ yields lifetime distributions of Weibull type with

$$\tilde{P}_\vartheta((-\infty, x]) = 1 - \exp\left(-\alpha \int_0^x h(\vartheta/z)\, z^{\alpha-1}\, dz\right),$$

where now the hazard rate $\tilde{\lambda}_\vartheta$ of $\tilde{P}_\vartheta$ satisfies

$$\tilde{\lambda}_\vartheta(x) := \frac{\tilde{f}_\vartheta(x)}{1-\tilde{P}_\vartheta((-\infty, x])} = h(\vartheta/x)\, \alpha x^{\alpha-1}, \qquad x>0.$$

The model (3.8) is thus equivalent to the lifetime model with the ratio of hazard rates

$$\tilde{\lambda}_{\vartheta}(x)/\tilde{\lambda}_0(x) = h(\vartheta/x). \tag{3.9}$$

Further examples can be obtained from the following lemma, which is concerned with distributions of exponential family type.

3.3. LEMMA. *For a given direction $h \in \Psi$ define*

$$\frac{dP_{\vartheta}}{dP_0}(x) = c(h(\vartheta \cdot))\, h(\vartheta x) = h(\vartheta x) + r(x, \vartheta) \tag{3.10}$$

with $1/c(h(\vartheta \cdot)) = \int h(\vartheta x)\, dP_0(x)$. *Then, the conditions* (i) *and* (ii) *of Definition* 3.1 *are satisfied.*

3.4. EXAMPLES. (a) If we take $h(x) = \exp(-x)$, the family $(P_{\vartheta})_{\vartheta}$ (3.10) gives an exponential family. Since extreme value distributions often have heavy tails, P_0 usually does not lie in the domain of attraction of $N(0, 1)$ and Gaussian limit experiments cannot be expected.

(b) Assume that $(P_{\vartheta})_{\vartheta}$ denotes a family given by (3.10) and some function h. For $d > 0$ define a new family of distributions $(P_{\vartheta})_{\vartheta \geqslant 0}$ by $\tilde{P}_0 = P_0$ and

$$\begin{aligned}\frac{d\tilde{P}_{\vartheta}}{dP_0}(x) &= \tilde{c}(\vartheta)(1_{[0, d)}(\vartheta x) + h(\vartheta x)\, 1_{[d, \infty)}(\vartheta x))\\ &= 1_{[0, d)}(\vartheta x) + h(\vartheta x)\, 1_{[d, \infty)}(\vartheta x) + r(x, \vartheta).\end{aligned}$$

For sufficiently large n this model can be used to obtain a threshold model at size d/δ_n, where the interesting effects occur in the tails.

The local considerations now require a rescaling procedure with $\delta_n \downarrow 0$, such that the experiment

$$(\mathbb{R}^{k_n}, \mathscr{B}^{k_n}, \{\mathscr{L}(Z_{n, k_n} \mid P^n_{\delta_n \vartheta}) : \vartheta \in \Theta\}), \tag{3.11}$$

given by the relevant part of the order statistics (2.1), is convergent to some non-trivial limit experiment, where now ϑ denotes a local parameter. The rescaling procedure is necessary in order to compensate the influence of an increasing sample size n. It turns out that the asymptotic behaviour of this experiment is completely determined by the function h which justifies the investigation of extreme value tangents. For these reasons we introduce the following nonparametric structure model (3.12).

For a given set $\Psi_0 \subset \Psi$ with $h \equiv 1 \in \Psi_0$ define

$$\frac{dP_h}{dP_1}(x) = c(h)\, h(x), \qquad h \in \Psi_0 \tag{3.12}$$

with normalizing factor $c(h) = (\int h\, dP_1)^{-1}$, where, for obvious reasons, we prefer the notation P_1 instead of P_0. Our nonparametric structure model is then $\{P_h : h \in \Psi_0\}$. If $\Psi_0 = \{h(\vartheta \cdot) : \vartheta \in \Theta\}$ we obtain the model (3.10).

In the following, we point out that typically Poisson experiments occur as limit experiments of the structure model (3.12), which can be identified by Poisson point processes. Let

$$w(z) = (\log G_{1,\alpha}(z))' = \alpha z^{-(1+\alpha)} 1_{(0,\infty)}(z)$$

and define for $k \geqslant 1$ the probability measure $Q_{k,h}$ by

$$\frac{dQ_{k,h}}{d\lambda^k}(x_k, ..., x_1) = \exp\left(-\int_{x_k}^{\infty} w(z)\, h(z)\, dz \right) \prod_{j=1}^{k} w(x_j)\, h(x_j)\, 1_{A_k}(x_k, ..., x_1)$$

with

$$A_k = \{(y_k, ..., y_1) \in \mathbb{R}^k : 0 \leqslant y_k \leqslant \cdots \leqslant y_1\};$$

cf. Lemma 5.3. Note that for $k = 1$ and $h \equiv 1$ we obtain the Fréchet density. Under the regularity assumptions (i) and (ii) of Definition 3.1, we now derive the limit distribution of a fixed number $k_n = k$ of upper extremes $Z_{n,k}$ (2.1) under local alternatives (3.5). Increasing portions k_n of upper extremes are discussed later. For $\vartheta = 0$ the following result is known from uniform convergence results of extremes (Falk [2], Sweeting [24]; see also De Haan and Resnick [5]).

3.5. THEOREM. *Suppose that P_0 fulfills the von Mises condition* (3.1). *Consider a curve* (3.5) *with $h \in \Psi$, where the remainder term satisfies* (i) *and* (ii) *of Definition* 3.1. *Then*

$$\| \mathcal{L}(\delta_n Z_{n,k} \mid P^n_{\delta_n \vartheta}) - Q_{k, h(\vartheta \cdot)} \| \to 0$$

for $n \to \infty$ and $\vartheta > 0$.

From Lemma 3.3 we know that $(P_{h(\delta_n \vartheta \cdot)})_\vartheta$ satisfies the regularity assumptions (i) and (ii) of Definition 3.1 and thus Theorem 3.5 implies the following result.

3.6. COROLLARY. *Let* $\{P_h : h \in \Psi_0\}$ *be the structure model* (3.12). *Then, the experiments induced by the k largest extremes*

$$\begin{aligned} E_{n+1-j:n, j \leqslant k} &= E_{n+1-j:n, j \leqslant k}(\Psi_0) \\ &= (\mathbb{R}^k, \mathscr{B}^k, \{\mathscr{L}(Z_{n,k} \mid P^n_{h(\delta_n \cdot)}) : h \in \Psi_0\}) \end{aligned} \tag{3.13}$$

converge weakly to

$$E_k = E_k(\Psi_0) = (\mathbb{R}^k, \mathscr{B}^k, \{Q_{k,h} : h \in \Psi_0\}). \tag{3.14}$$

Next, we study the limit experiment of the sequence $(E_{n+1-j:n, j \leqslant k})_n$ in more detail. For $h \in \Psi$ define

$$\nu_h([x, \infty)) = \int_x^\infty h(z)\, w(z)\, dz$$

and let ψ_h denote the inverse of ν_h given by

$$\psi_h(u) = \sup\{t : \nu_h([t, \infty)) \geqslant u\}, \qquad u > 0.$$

Note that $\psi_1(u) = u^{-1/\alpha}$. Let $(\eta_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. standard exponential random variables and

$$S_k = \sum_{i=1}^k \eta_i. \tag{3.15}$$

By Lemma 5.3, we see that

$$Q_{k,h} = \mathscr{L}((\psi_h(S_j))_{j \leqslant k}). \tag{3.16}$$

The limit experiment (3.14) can now be rewritten in terms of point processes. By ε_x we denote the Dirac measure in x, that is, $\varepsilon_x(B) = 1_B(x)$. Then

$$N_{k,h} = \sum_{j=1}^k \varepsilon_{\psi_h(S_j)}$$

denotes the point process corresponding to $Q_{k,h}$ and let

$$N_h = \sum_{j \geqslant 1} \varepsilon_{\psi_h(S_j)}$$

be the full point process. It is easy to see that N_h is a Poisson point process with intensity measure ν_h (see, e.g., Resnick [21, Proposition 3.7]). We call $N_{k,h}$ the k-limited point process of N_h. Obviously

$$(M((0, \infty)), \mathcal{M}((0, \infty)), \{\mathcal{L}(N_{k,h}) : h \in \Psi_0\})$$

is equivalent to E_k; see Lemma 5.3.

In the next step we briefly discuss the asymptotic behaviour of the experiments given by Z_{n,k_n} where the sequence k_n tends to infinity slowly enough. It is natural to consider first the limit of $E_{n,k}$ as $n \to \infty$ and then the limit $k \to \infty$. The subsequent lemma gives the limit experiment E_∞ of $(E_k)_{k \in \mathbb{N}}$. In conclusion, we see that there exist sequences $k_n \to \infty$ (increasing slowly enough) such that E_∞ is a weak accumulation point of $E_{n+1-j:n, j \leq k_n}$ given by (3.13). For concrete examples we will see which kind of sequences k_n yield convergence to the limit experiment E_∞; see Example 3.8 below.

3.7. Lemma. *The weak limit experiment of* $(E_k)_{k \in \mathbb{N}}$ *is given by*

$$E_\infty = E_\infty(\Psi_0) = (M((0, \infty)), \mathcal{M}((0, \infty)), \{\mathcal{L}(N_h) : h \in \Psi_0\}).$$

Remarks. (a) The parameter h is just the density of ν_h w.r.t. ν_1. Thus the tangent space has a quite natural interpretation in terms of the intensity measures $(\nu_h)_{h \in \Psi_0}$ of the limit experiment; see also (3.8) and (3.9). For practical purposes extreme value models can now be established by the investigation of structure models (3.12) given by a relevant subfamily Ψ_0 of tangent functions of intensity measures. Note that in the case of Example 3.2 the corresponding intensities of stochastically larger alternatives are just those ν_h with $\nu_h([x, \infty)) \geq \nu_1([x, \infty))$ for each $x > 0$. Statistical inference of Poisson point processes can be found in the books by Karr [10] and Reiss [20].

(b) Recall from Karr [10, Proposition 6.14], that $\mathcal{L}(N_h)$ is absolutely continuous w.r.t. $\mathcal{L}(N_1)$ iff

$$\int_0^\infty (1 - h^{1/2}(x))^2 \, x^{-(1+\alpha)} \, dx < \infty.$$

Otherwise the distributions are mutually singular. Note that each tangent $h \in \Psi$ with $h(x) = 1$ for $0 < x < d$ satisfies that condition. On the other hand, there exists $h \in \Psi$ such that $\mathcal{L}(N_h)$ and $\mathcal{L}(N_1)$ become singular. However, the distributions $Q_{k,h}$ and $Q_{k,1}$ are always non-singular for each $k \in \mathbb{N}$.

The present concept includes various parametric experiments which were earlier considered in the literature. A collection of them is given in our next example. Special attention is given to the experiments (3.13) with increasing portion of extremes $k_n \to \infty$. In Example 3.8(b) and (c) (under the restriction $\alpha < 2$) it is already known that the limit experiment of $E_{n+1-j:n, j \leqslant k_n}$ is E_∞ for each sequence $k_n \to \infty$ (even for $k_n = n$) under mild regularity conditions concerning Ψ_0; see [6, 7]. In that case, a finite number of upper extremes is approximately sufficient.

3.8. EXAMPLES. (a) Our first example deals with a threshold model with $d > 0$. Introduce the transformation

$$\xi(h)(x) = 1_{[0, d)}(x) + h(x)\, 1_{[d, \infty)}(x)$$

on Ψ and consider the structure model $(P_{\xi(h)})_{h \in \Psi}$. If we take first $\delta_n \downarrow 0$ and then $k \to \infty$ we arrive at the limit experiment

$$(M((0, \infty)), \mathscr{M}((0, \infty)), \{\mathscr{L}(N_{\xi(h)}) : h \in \Psi\}), \tag{3.17}$$

which is a subexperiment of $E_\infty(\Psi)$. We now identify (3.17) with an experiment of truncated point processes. Throughout let $N(\cdot \cap [t, \infty))$ denote the restriction of a point process $N(\cdot)$ on $[t, \infty)$. For $0 < t < d$ the truncated version of (3.17) relative to $[t, \infty)$ has the likelihood ratio

$$\frac{dN_{\xi(h)}(\cdot \cap [t, \infty))}{dN_1(\cdot \cap [t, \infty))}(\mu) = \exp\left(\int_d^\infty \log h(x)\, d\mu(x) + \int_d^\infty (1 - h(x))\, w(x)\, dx\right) \tag{3.18}$$

(see Karr [10, Theorem 2.31] and Reiss [20, Theorem 3.1.1], for the density formula). Note that the likelihood (3.18) is independent of t and its distribution coincides with the likelihood distribution of

$$(M((0, \infty)), \mathscr{M}((0, \infty)), \{\mathscr{L}(N_h(\cdot \cap [d, \infty))) : h \in \Psi\}). \tag{3.19}$$

If we now let t tend to zero the next theorem implies that the experiments (3.17) and (3.19) are equivalent. Thus also the asymptotic threshold limit experiments are embedded in our approach.

(b) In Janssen [6] the extremes of exponential families were investigated. The models include as a practical application the family of inverse Gaussian distribution which is used to make a statistical inference about the Wiener processes with unknown drift under inverse sampling. By Corollary 3.6, the limit experiment of the k largest extremes of Example 3.4(a) is now

$$F_k = (M((0, \infty)), \mathscr{M}((0, \infty)), \{\mathscr{L}(N_{k, \exp(-\vartheta \cdot)}) : \vartheta \geqslant 0\})$$

with

$$F_k \to F := (M((0, \infty)), \mathscr{M}((0, \infty)), \{\mathscr{L}(N_{k, \exp(-\vartheta \cdot)}) : \vartheta \geqslant 0\}).$$

For $\alpha<2$ and $k=k_n\to\infty$ as $n\to\infty$ the occurring limit experiment of $E_{n+1-j:n,j\leqslant k_n}$ is equivalent to some exponential family $F'=\{Q_\vartheta:\vartheta\geqslant 0\}$ with

$$\frac{dQ_\vartheta}{dQ_0}(x)=c(\vartheta)\exp(-\vartheta x),$$

where Q_0 is a one-sided stable distribution with index α. We see that F and F' are equivalent. It can be shown for $0<\alpha<1$ that the relation between F and F' is given by the sufficient statistic

$$M((0,\infty))\ni\mu\to\int_0^\infty x\,d\mu(x). \tag{3.20}$$

In the case $1\leqslant\alpha\leqslant 2$ the mapping (3.20) must be centered.

The consideration of exponential families has further aspects. If we make use of the arguments of Theorem 3.9 below, we see that the truncated version of F w.r.t. $[t,\infty)$, $t>0$,

$$N^{(t)}=\sum_{j\geqslant 1}\varepsilon_{\psi_h(S_j)}1_{[t,\infty)}(\psi_h(S_j))$$

is again an exponential family whenever $h(x)=\exp(-x)\,1_{(0,\infty)}(x)$ and $\mu\to\int_t^\infty x\,d\mu$ is a sufficient statistic.

(c) Lifetime location models of Weibull type with density

$$f(x)=(1+a)\,x^a\exp(-x^{1+a})\,1_{(0,\infty)}(x) \tag{3.21}$$

and shape parameter $a\in(-1,1)$ yield a limit experiment G given by the family of point processes

$$\sum_{j\geqslant 1}\varepsilon_{S_j^{1/(1+a)}+\vartheta},\qquad \vartheta\geqslant 0$$

(see Janssen and Reiss [8]). From $G_{1,\alpha}(x)$ with $\alpha=1+a$ we arrive at (3.21) by using the transformations $x\to 1/x$. This model is also contained in our approach. Straightforward calculations show that for $a\neq 0$ G is equivalent to the sub-experiment

$$(M((0,\infty)),\mathscr{M}((0,\infty)),\{\mathscr{L}(N_{h_\vartheta}:\vartheta\geqslant 0\})$$

given by

$$h_\vartheta(x)=h(\vartheta x),\qquad h(x):=(1-x)^a\,1_{(0,1)}(x)$$

and the corresponding intensity measures

$$\frac{d\nu_\vartheta}{d\lambda}(x) = h_\vartheta(x)(1+a)\, x^{-(2+a)} 1_{(0,\infty)}(x).$$

Note that the corresponding ψ_ϑ-functions have the form

$$\psi_\vartheta(x) = (x^{1/(1+a)} + \vartheta)^{-1}.$$

Statistical applications for lifetime tests making use of the limit experiment G are contained in Janssen and Mason [7]. For instance, this method proves that related score tests for survival times have certain optimality properties. It should be remarked that G is a stable Poisson experiment in the sense of Strasser [22].

The experiment $E_\infty(\Psi)$ appears as the limit experiment of the truncated experiments $\{\mathscr{L}(N_h(\cdot \cap [t, \infty))) : h \in \Psi\}$, as the following theorem shows.

3.9. Theorem. *The experiment*

$$(M((0, \infty)), \mathscr{M}((0, \infty)), \{\mathscr{L}(N_h(\cdot \cap [t, \infty))) : h \in \Psi\})$$

converges weakly to $E_\infty(\Psi)$ *whenever* $t \downarrow 0$.

Our last result concerns the limit experiment of the threshold model given in Example 3.8(a). Under additional assumptions on the function h the number of extremes k_n may tend to infinity at any rate.

3.10. Theorem. *Consider the structure model* (3.12) *with parameter space*

$$\Psi_1 := \{h \in \Psi : h(x) = 1, x \in [0, d), \textit{for some } d > 0\}.$$

Then, for each sequence $k_n \uparrow \infty$, $k_n \leqslant n$, *we have*

$$E_{n+1-j:n,\, j \leqslant k_n}(\Psi_1) \to E_\infty(\Psi_1) \tag{3.22}$$

weakly.

4. Testing Extreme Value Hypothesis

The results of the preceding section are applied for testing the Fréchet distribution $G_{1,\alpha}$ against stochastically larger alternatives; see Example 3.2. As a consequence of the convergence of extreme value experiments the performance of statistical procedures can then be compared along curves of alternatives (3.5). Recall from LeCam [12] and Strasser [23] that, whenever a (non-degenerate) limit experiment exists, the statistician has (only) to solve the underlying decision problem for the limit experiment and the lower bounds (for power functions, risk functions, etc.) of the limit experiment yield lower bounds for the sequence of experiments. So far, the extreme value models are embedded in the general asymptotic decision theory.

However, in contrast to the case of local asymptotic normality it cannot be expected that the risk bounds of the limit experiments can be attained by the underlying procedures in general. Recently, LeCam [14] discussed lower bounds for Poisson experiments. Here we study a concrete example showing which type of asymptotic results can be expected.

As in Example 3.2 denote by $\mathscr{P}$ the class of distributions which are stochastically larger than $G_{1,\alpha}$. We consider the testing problem

$$H_0 = \{G_{1,\alpha}\} \qquad \text{against} \quad H_1 = \mathscr{P} \backslash H_0$$

at sample size n, where we assume that the tail index $\alpha > 0$ is known. In addition, we assume that the relevant differences between H_0 and H_1 lie behind a given threshold d/δ_n, $d > 0$, where $\delta_n = (\log(n/(n-1)))^{1/\alpha}$. Throughout, the following two models are compared.

1. Only $X_{n+1-k:n}, ..., X_{n:n}$ are available for fixed $k \in \mathbb{N}$.

2. An increasing sequence of order statistics $(X_{n+1-j:n})_{j \leqslant k_n}$, $k_n \uparrow \infty$ (including $k_n = n$), is available.

Our model is a structure model (3.12) with $P_1 = G_{1,\alpha}$ and parameter space

$$\Psi_2 = \{h \in \Psi : h 1_{[0,d)} = 1 \text{ and } h \text{ fulfills (3.6)}\}.$$

Within this setup, now let $1 \leqslant h_0 \in \Psi_2$, $h_0 \neq 1$, be a fixed "tangent." Then the structure of the likelihood ratio suggests the test statistic

$$T_{k_n,n} = \int_{\max\{d,\,\delta_n X_{n+1-k_n:n}\}}^{\infty} w(z)(1-h_0(z))\,dz + \sum_{j=1}^{k_n} (\log h_0(\delta_n X_{n+1-j:n}))\, 1_{[d,\infty)}(\delta_n X_{n+1-j:n}).$$

Then two asymptotic tests are proposed (for the situations 1 and 2 above):

$$\varphi_{1,n} = \begin{cases} 1 \\ 0 \end{cases} T_{k,n} \begin{matrix} \geqslant \\ < \end{matrix} c_1$$

$$\varphi_{2,n} = \begin{cases} 1 \\ 0 \end{cases} T_{k_n,n} \begin{matrix} \geqslant \\ < \end{matrix} c_2. \qquad (4.1)$$

Introduce the limit distributions under $G_{1,\alpha}$,

$$\mu_1 = \mathscr{L}\left(\int_{\max\{d, S_k^{-1/\alpha}\}}^{\infty} w(z)(1-h_0(z))\,dz + \sum_{j=1}^{k} \log h_0(S_j^{-1/\alpha})\right)$$

and

$$\mu_2 = \mathscr{L}\left(\int_{d}^{\infty} w(z)(1-h_0(z))\,dz + \sum_{j=1}^{\infty} \log h_0(S_j^{-1/\alpha})\right).$$

Then $(\varphi_{1,n})_n$ and $(\varphi_{2,n})_n$ are test sequences of asymptotic level α (not to be confused with the notation of the tail index α), whenever $\mu_i([c_i, \infty)) \leqslant \alpha$ for $i = 1, 2$. These tests are asymptotic Neyman–Pearson tests for

$$\mathscr{L}(Z_{n,r} \mid P_1^n) \qquad \text{against} \quad \mathscr{L}(Z_{n,r} \mid P^n_{h_0(\delta_n \cdot)}), \qquad r \in \{k, k_n\}$$

if c_i are continuity points of the distribution functions of μ_i and $\mu_i([c_i, \infty)) = \alpha$ holds. According to Example 3.2, this is an asymptotically optimal test for alternatives given by the ratio of intensities $h_0(\delta_n \cdot)$ (3.8).

A proper choice of h_0 leads to further relevant tests. We obtain shape alternatives if we choose

$$h_\beta(x) = 1_{[0,d)}(x) + (\beta/\alpha)\, x^{\alpha-\beta}\, 1_{[d,\infty)}(x), \qquad 0 < \beta < \alpha.$$

Then the tail $P_{h_\beta}([x, \infty)) = G_{1,\beta}([x, \infty))$ is for $x \geqslant d$ just the tail of the Fréchet distribution with shape parameter β. Since

$$\log h_\beta(x) = (\log(\beta/\alpha) + (\alpha-\beta)\log x)\; 1_{[d,\infty)}(x)$$

it is convenient to introduce test statistics

$$T_{k_n,n} = \sum_{j=1}^{k_n} (\log(\delta_n X_{n+1-j:n}))\; 1_{[d,\infty)}(\delta_n X_{n+1-j:n}),$$

which are independent of β! Note that the limit distribution of $T_{k_n,n}$ under H_0 is

$$\mathscr{L}\left(\sum_{j=1}^{\infty} \log(S_j^{-1/\alpha})\; 1_{[d,\infty)}(S_j^{-1/\alpha})\right)$$

whenever $k_n \to \infty$ as $n \to \infty$. If the critical value is taken as the $(1-\alpha)$-quantile of that distribution we arrive at tests which are asymptotically equivalent to $\varphi_{2,n}$. These tests are asymptotically optimal tests for

$$\mathscr{L}(Z_{n,k_n} \mid G^n_{1,\alpha}) \qquad \text{against} \quad \mathscr{L}(Z_{n,k_n} \mid P^n_{h_\beta(\delta_n \cdot)})$$

for arbitrary $0 < \beta < \alpha$.

We obtain scale alternatives if we choose $h_\sigma(x) = 1_{[0,d)}(x) + \sigma^\alpha 1_{[d,\infty)}(x)$, $\sigma > 1$. In this case the asymptotic optimal test statistic depends only on the number of exceedances.

Remarks. (a) The assertion above remains valid if $G_{1,\alpha}$ is substituted by P_0, where P_0 fulfills the von Mises condition (3.1).

(b) Check that h^{ϑ} belongs to Ψ, whenever $h \in \Psi$ and $0 < \vartheta < 1$ (use Hölder's inequality). Thus $\{N_{h^{\vartheta}} : \vartheta \in [0, 1]\}$ defines an exponential family which is a subfamily of the limit experiment. For this reason the asymptotic optimality of the test $\varphi_{1,n}$ and $\varphi_{2,n}$ in (4.1) carries over to alternatives specified by $h_0^{\vartheta}(\delta_n \cdot)$, $0 < \vartheta \leqslant 1$.

5. Proofs

The proofs require some technical preparations. First we derive the likelihood ratio of the extremes (2.1).

5.1. Lemma. *Consider distributions P_i with continuous distribution functions F_i for $i = 0, 1$. Then for $1 \leqslant k \leqslant n$*

$$\frac{d\mathscr{L}(W_{n,k} \mid P_1^n)}{d\mathscr{L}(W_{n,k} \mid P_0^n)}(x_1, ..., x_k) = \prod_{i=1}^{k} \frac{dP_1}{dP_0}(x_i) \left(\frac{1 - F_1(x_k)}{1 - F_0(x_k)}\right)^{n-k} \tag{5.1}$$

for $x_1 < \cdots < x_k$ and zero otherwise.

Proof. It is known that $\mathscr{L}(W_{n,k} \mid P_i^n)$, $i = 0, 1$, has the P_i^k density

$$\frac{d\mathscr{L}(W_{n,k} \mid P_i^n)}{dP_i^k}(x_1, ..., x_k) = \frac{n!}{(n-k)!}(1 - F_i(x_k))^{n-k} \tag{5.2}$$

if $x_1 < \cdots < x_k$ and the density is equal to zero otherwise; see, e.g., Reiss [19, Theorem 1.5.2]. If P_1 is P_0 absolutely continuous (5.1) is an easy consequence of (5.2). In general, the density is first calculated for $P_2 := (P_0 + P_1)/2$. Then the expression $(dP_1/dP_2)/(dP_0/dP_2)$ gives the desired formula (5.1). ∎

A similar formula holds for upper extremes; replace $(x_1, ..., x_k)$ by $(x_k, ..., x_1)$ and $1 - F_i$ by F_i.

Proof of Theorem 2.1. Using contiguity it is sufficient to consider the likelihood process with basis 0. In the sequel, the proof is carried out for the lower extremes. According to (5.1), we must show that

$$\sum_{i=1}^{k} \log \frac{dP_{n,\vartheta}}{dP_{n,0}}(X_{i:n}) + \log\left(\frac{1 - F_{n,\vartheta}(X_{k:n})}{1 - F_{n,0}(X_{k:n})}\right)^{n-k} \to 0 \tag{5.3}$$

in $P_{n,0}^n$-probability, where $F_{n,\vartheta}$ denotes the distribution function of $P_{n,\vartheta}$. By means of Hellinger distances (see (1.1)), the expression $nd^2(P_{n,\vartheta}, P_{n,0})$

is convergent and (2.3) holds. Thus Theorem (6.3) of Milbrodt and Strasser [17] can be applied to $Y_{ni} := \log(dP_{n,\vartheta}/dP_{n,0})(X_i)$. Hence, for $\varepsilon \geqslant 0$

$$nP^n_{n,0}\{|Y_{n1}| \geqslant \varepsilon\} \to 0 \tag{5.4}$$

(recall that X_i is by definition the canonical projection) and thus

$$Y_{n,1:n} \to 0 \quad \text{and} \quad Y_{n,n:n} \to 0 \tag{5.5}$$

in $P^n_{n,0}$-probability for the order statistics of Y_{ni}. Since

$$Y_{n,1:n} \leqslant \log \frac{dP_{n,\vartheta}}{dP_{n,0}}(X_{i:n}) \leqslant Y_{n,n:n} \tag{5.6}$$

the first term in (5.3) vanishes as $n \to \infty$. A Taylor expansion now shows that the proof of (5.3) is complete whenever

$$(n-k)(F_{n,\vartheta}(X_{k:n}) - F_{n,0}(X_{k:n})) \to 0 \tag{5.7}$$

in $P^n_{n,0}$-probability. The verification of (5.7) will be done separately for the cases (a) and (b) of Theorem 2.1.

Case (a). For $x, y \in [0, 1]$ the mean value theorem yields for $q \geqslant 1$

$$|x-y| \leqslant q\,|x^{1/q} - y^{1/q}|. \tag{5.8}$$

Let $\mu_n = P_0 + P_{n^{-1/q}\vartheta}$. Hölder's inequality together with (5.8) implies

$$\begin{aligned} &|n(F_{n,\vartheta}(X_{k:n}) - F_{n,0}(X_{k:n})| \\ &\leqslant nq \int_{(-\infty, X_{k:n}]} \left| \left(\frac{dP_{n^{-1/q}\vartheta}}{d\mu_n}\right)^{1/q} - \left(\frac{dP_0}{d\mu_n}\right)^{1/q} \right| d\mu_n \\ &\leqslant \left(\int_{(-\infty, X_{k:n}]} \left| q\left(\left(\frac{dP_{n^{-1/q}\vartheta}}{d\mu_n}\right)^{1/q} - \left(\frac{dP_0}{d\mu_n}\right)^{1/q}\right)\Big/ n^{-1/q} \right|^q d\mu_n \right)^{1/q} \\ &\quad \times (n\mu_n(-\infty, X_{k:n}])^{1-1/q}. \end{aligned} \tag{5.9}$$

Next, we remark that

$$n\mu_n((-\infty, X_{k:n}]) = n(P_{n,0}((-\infty, X_{k:n}]) + P_{n,\vartheta}((-\infty, X_{k:n}]))$$

is stochastically bounded under $P^n_{n,0}$. The boundedness of $nP_{n,0}((-\infty, X_{k:n}])$ is immediate, since the expression coincides in distribution with the order statistic $U_{k:n}$ of n independent random variables which are uniformly distributed in the unit interval. By the same arguments $nP_{n,\vartheta}((-\infty, X_{k:n}])$ is stochastically bounded under $P^n_{n,\vartheta}$. Contiguity gives the result also under

$P^n_{n,0}$. It remains to show that the first factor on the right-hand side of (5.9) converges in P^n_0-probability to zero. Note that L^q-differentiability yields

$$\left(\int_{(-\infty,\, X_{k:n}]} \left| q\left(\left(\frac{dP_{n^{-1/q}\vartheta}}{d\mu_n}\right)^{1/q} - \left(\frac{dP_0}{d\mu_n}\right)^{1/q}\right)\Big/ n^{-1/q} \right|^q d\mu_n\right)^{1/q}$$
$$\leqslant \left(\int_{(-\infty,\, X_{k:n}]} |g|^q \, dP_0\right)^{1/q} + o_{P^n_0}(1). \tag{5.10}$$

Since $X_{k:n}$ converges to the lower endpoint of the support of P_0, the dominated convergence theorem implies that (5.10) converges in $P^n_{n,0}$-probability to zero.

Case (b). For stochastically increasing families the convergence of (5.7) can be derived by the following arguments: Let $\vartheta > 0$. By the weak sequential compactness of statistical experiments (for finite parameter sets) it is sufficient to show that each weak accumulation point $F = \{P, Q\}$ of $E_{j:n, j \leqslant k}|_{\{0,\vartheta\}}$ is totally uninformative. Assume that convergence holds along a subsequence $(n_j)_j$. Taking into account the arguments (5.4)–(5.6) above, we obtain

$$\mathscr{L}\left((n_j - k) \log \frac{1 - F_{n_j,\vartheta}(X_{k:n_j})}{1 - F_{n_j,0}(X_{k:n_j})} \,\Big|\, P_0^{n_j}\right) \to \nu_0 := \mathscr{L}\left(\log \frac{dQ}{dP} \,\Big|\, P\right) \tag{5.11}$$

in distribution. Since $F_{n,\vartheta} \leqslant F_{n,0}$ we have $\nu_0([0, \infty]) = 1$. On the other hand, the definition of ν_0 implies $\int \exp(x) \, d\nu_0(x) \leqslant 1$. Thus $\nu_0 = \varepsilon_0$ and $F = E_0$ follows.

In the case $\vartheta < 0$ the rôle of 0 and ϑ can be interchanged. In connection with upper extremes the random variables $-X_i$ can be regarded. An obvious modification of the present proof yields the result for stochastically decreasing families. ∎

The following auxiliary result is crucial for the proof of Theorem 2.2; its proof is elementary.

5.2. Lemma. *Let $Y_n, Z_n: (\Omega_n, \mathscr{A}_n, P_n) \to \mathbb{R}$ denote random variables with $Z_n \geqslant Y_n$. Assume that $Z_n \to Z$ and $Y_n \to Z$ converge in distribution to some random variable Z defined on some probability space $(\Omega, \mathscr{A}, P)$. Then $Z_n - Y_n \to 0$ in P_n-probability.*

Proof of Theorem 2.2. The proof is devided into two steps.

I. Here, we show the equivalence of the assertions (b) and (c) for stochastically increasing families $(P_{n,\vartheta})_\vartheta$. Note that the implication (b) $\Rightarrow$ (c) is given in part (b) of the preceding proof. Assume now that the assertion (c) holds. Again, it is only necessary to treat the lower extremes and by induction it remains to prove that

$$\sum_{i=1}^{k} \log \frac{dP_{n,\vartheta}}{dP_{n,0}}(X_{i:n}) \to 0$$

in $P_{n,0}^n$- and $P_{n,\vartheta}^n$-probability. By our assumptions (5.3) converges to zero in $P_{n,0}^n$- and $P_{n,\vartheta}^n$-probability. The proof now reduces by showing that (5.7) holds under $P_{n,0}^n$ and $P_{n,\vartheta}^n$. To this end, we recall from extreme value theory that under $P_{n,\vartheta}^n$

$$nF_{n,\vartheta}(X_{k:n}) \to \sum_{i=1}^{k} \eta_i,$$

where η_i are i.i.d. exponential random variables with mean 1. Since $X_{k:n}$ contains asymptotically no information about ϑ, the assertion (5.7) also holds under $P_{n,0}^n$.

Define

$$Y_n = nF_{n,\vartheta}(X_{k:n}) \qquad \text{and} \qquad Z_n = nF_{n,0}(X_{k:n}).$$

Then $Z_n \geqslant Y_n$ for $\vartheta > 0$ and $Z_n \leqslant Y_n$ whenever $\vartheta < 0$. Thus Lemma 5.2 yields $Z_n - Y_n \to 0$ in $P_{n,0}^n$-probability. Again, $Z_n - Y_n$ contains no information about ϑ and thus the convergence holds also under $P_{n,\vartheta}^n$. Hence, the assertion (5.7) is proved.

II. In the second step, we prove the implication (b) $\Rightarrow$ (a) for families with monotone likelihood ratios

$$\frac{dP_{n,\vartheta}}{dP_{n,0}}(x) = h_{n,\vartheta}(x).$$

The arguments are based on the criterion (5.12) below for the convergence to Gaussian experiments. Here, we need a slight extension of Theorem (6.3) of Milbrodt and Strasser [17], given in Janssen and Mason [7, Theorem 3.1, Appendix]. Note that this theorem is applicable, since the boundedness assumption (2.3) implies the infinitesimality of $(E_n)_n$; see Lemma 5.7 in [17]. Along these lines it is sufficent to prove that for each $0 < \varepsilon < 1$

$$n(P_{n,0} + P_{n,\vartheta})\{|\log h_{n,\vartheta}| > \varepsilon\} \to 0. \tag{5.12}$$

Note that by assumption (b)

$$\begin{aligned} P_{n,0}^n\{\log h_{n,\vartheta}(X_{n:n}) > \varepsilon\} &= P_{n,0}^n\{\max_{1 \leqslant i \leqslant n} \log h_{n,\vartheta}(X_i) > \varepsilon\} \\ &= 1 - (P_{n,0}\{\log h_{n,\vartheta} \leqslant \varepsilon\})^n \to 0, \end{aligned}$$

which yields

$$nP_{n,0}\{\log h_{n,\vartheta} > \varepsilon\} \to 0 \tag{5.13}$$

for $n \to \infty$. Assertion (5.12) easily follows, since (5.13) also holds under $P_{n,\vartheta}$. The proof of Theorem 2.2 is complete. ▌

Proof of Lemma 3.3. Put

$$r(x, \vartheta) = h(\vartheta x)[c(h(\vartheta\cdot)) - 1].$$

We see that condition (i) is satisfied if

$$\lim_{\vartheta \downarrow 0} (c(h(\vartheta\cdot)))^{-1} = 1. \tag{5.14}$$

Taking into account (3.3), Fatou's lemma implies $\liminf_{\vartheta \downarrow 0}(c(h(\vartheta\cdot)))^{-1} \geqslant 1$. So it remains to show

$$\limsup_{\vartheta \downarrow 0} (c(h(\vartheta\cdot)))^{-1} \leqslant 1.$$

We split the integration into three domains whenever $0 < c\vartheta < \delta$:

$$\begin{aligned} &\int h(\vartheta x) f_0(x)\, dx \\ &\quad = \int_0^c h(\vartheta x) f_0(x)\, dx + \frac{1}{\vartheta} \int_\delta^\infty h(x) f_0(x/\vartheta)\, dx + \frac{1}{\vartheta} \int_{c\vartheta}^\delta h(x) f_0(x/\vartheta)\, dx \\ &\quad = I_1(\vartheta) + I_2(\vartheta) + I_3(\vartheta). \end{aligned}$$

For all $c > 0$, we have

$$I_1(\vartheta) \to \int_0^c f_0(x)\, dx \leqslant 1$$

as $\vartheta \downarrow 0$. Since f_0 is $-(1+\alpha)$ varying at infinity by Karamata's theorem (see Bingham *et al.* [1, Theorem 1.6.1]), we can find a constant $K = K(c)$ such that $f_0(y) \leqslant K y^{-(1+\alpha)+\varepsilon}$ for $y \geqslant c$ and (3.4) holds. Hence, for all $\delta > 0$

$$I_2(\vartheta) \leqslant \frac{1}{\vartheta} \int_\delta^\infty h(x)\, K \vartheta^{1+\alpha-\varepsilon} x^{-(1+\alpha)+\varepsilon}\, dx \to 0$$

as $\vartheta \downarrow 0$. Next, we choose δ such that $h(x) \leqslant 2$ for $x \in [0, \delta]$. Then for large c

$$
\begin{aligned}
I_3(\vartheta) &\leqslant \frac{2K}{\vartheta} \int_{c\vartheta}^{\delta} x^{-(1+\alpha)+\varepsilon} \vartheta^{1+\alpha-\varepsilon}\, dx \\
&\leqslant 2K\vartheta^{\alpha-\varepsilon} \frac{1}{\alpha-\varepsilon} (c\vartheta)^{-\alpha+\varepsilon} \\
&= \frac{2K}{\alpha-\varepsilon} c^{-\alpha+\varepsilon},
\end{aligned}
$$

which becomes arbitrarily small for large c. Combining these results, we see that (5.14) is valid.

To show the validity of condition (ii) it is sufficient to show

$$\int_{x/\delta_n}^{\infty} h(\delta_n \vartheta z) f_0(z)\, dz = O(n^{-1}). \tag{5.15}$$

The substitution z by $\delta_n^{-1} z$ yields

$$n \int_{x/\delta_n}^{\infty} h(\delta_n \vartheta z) f_0(z)\, dz = \frac{n}{\delta_n} \int_{x}^{\infty} h(\vartheta z) f_0(\delta_n^{-1} z)\, dz.$$

Since F_0 is $-\alpha$ varying and $F_0(F_0^{-1}(x)) = x$, we conclude from (3.1) that

$$n\delta_n^{-1} f_0(\delta_n^{-1} x) \to \alpha x^{-(1+\alpha)}. \tag{5.16}$$

Note that the convergence in (5.16) takes place uniformly on intervals $[a, \infty)$, $a > 0$, which is a well-known property of regularly varying functions (see Bingham *et al.* [1, Theorem 1.5.2]). Now, assertion (5.15) follows from (5.16). ∎

Proof of Theorem 3.5. Let

$$f_{n,k,h,\delta_n\vartheta} = d\mathscr{L}(\delta_n Z_{n,k} \mid P^n_{\delta_n\vartheta})/d\lambda^k.$$

By Scheffé's lemma the result is proved if

$$f_{n,k,h,\delta_n\vartheta} \to dQ_{k,h(\vartheta\cdot)}/d\lambda^k$$

as $n \to \infty$. By formula (1.4.8) of Reiss [19] we have

$$
\begin{aligned}
f_{n,k,h,\delta_n\vartheta}(x_k, ..., x_1) = F^{n-k}_{\delta_n\vartheta}(\delta_n^{-1} x_k) \prod_{j=1}^{k} (n-j+1)\, \delta_n^{-1} f_0(\delta_n^{-1} x_j) \\
\times [h(\vartheta x_j) + r(\delta_n^{-1} x_j, \delta_n \vartheta)]\, 1_{A_k}(x_k, ..., x_1),
\end{aligned}
$$

where

$$A_k = \{(y_k, ..., y_1) \in (0, \infty)^k : y_k < y_{k-1} < \cdots < y_1\}.$$

In view of (5.16) it remains to show that

$$F^{n-k}_{\delta_n \vartheta}(\delta_n^{-1} x_k) \to \exp\left(-\int_{x_k}^{\infty} w(z)\, h(\vartheta z)\, dz \right)$$

as $n \to \infty$. First, we obtain by the substitution $z \to \delta_n^{-1} z$

$$F^{n-k}_{\delta_n \vartheta}(\delta_n^{-1} x_k) = \left(1 - \int_{x_k}^{\infty} \delta_n^{-1} [h(\vartheta z) + r(\delta_n^{-1} z, \delta_n \vartheta)] f_0(\delta_n^{-1} z)\, dz \right)^{n-k}.$$

Again the uniform convergence theorem of regularly varying functions shows that

$$n\delta_n^{-1} \int_{x_k}^{\infty} h(\vartheta z) f_0(\delta_n^{-1} z)\, dz \to \int_{x_k}^{\infty} w(z)\, h(\vartheta z)\, dz.$$

Condition (ii) ensures that

$$n \int_{x_k}^{\infty} \delta_n^{-1} |r(\delta_n^{-1} z, \delta_n \vartheta)| f_0(\delta_n^{-1} z)\, dz \to 0$$

as $n \to \infty$. The proof is complete. ∎

LEMMA 5.3. (a) *Let $h \in \Psi$. Then*

$$Q_{k,h} = \mathscr{L}((\psi_h(S_j))_{j \leq k}).$$

(b) *The $(\mathscr{B}^k \cap (0, \infty)^k, \mathscr{M}((0, \infty)))$-measurable map $T_k : (0, \infty) \to M((0, \infty))$ defined by*

$$T_k(x_1, ..., x_k) := \sum_{i=1}^{k} \varepsilon_{x_i}$$

is a sufficient statistic for the family $\{Q_{k,h} : h \in \Psi\}$.

Proof. *Case* (a). First, the joint distribution $(S_1, ..., S_k)$ has the λ^k-density

$$(x_1, ..., x_k) \to e^{-x_k} 1_{A_k}(x_1, ..., x_k)$$

with $A_k = \{(y_1, ..., y_k) \in \mathbb{R}^k : 0 < y_1 < \cdots < y_k)\}$. For $h > 0$ the function ψ_h is bijective with $\psi_h^{-1}(x) = \nu_h([x, \infty)) = \int_x^\infty h(z)\, w(z)\, dz$. An application of the transformation theorem for densities yields (3.16) in the case $h > 0$. For arbitrary $h \in \Psi$ we can find a sequence $h_n \in \Psi$ with $h_n > 0$ and $h_n \downarrow h$ (for example $h_n = h 1_{\{h>0\}} + (1/n) 1_{\{h=0\}}$). By the dominated convergence theorem of Lebesgue we obtain $dQ_{k,h_n}/d\lambda \to dQ_{k,h}/d\lambda$ and $\nu_{h_n}([x, \infty)) \to \nu_h([x, \infty))$, which implies the assertion (3.16) for $h \in \Psi$.

Case (b). Note that for $(x_k, ..., x_1) \in A_k$

$$\frac{dQ_{k,h}}{dQ_{k,1}}(x_k, ..., x_1) = \exp\left(-\int_{\phi(T_k(x_k, ..., x_1))}^{\infty} (h(z) - 1)\, w(z)\, dz \right.$$
$$\left. + \int \log h(u)\, dT_k(x_k, ..., x_1)(u)\right)$$

if we define $\phi(\mu) = \sup\{t : \mu((0, t]) = 0\}$ for $\mu \in M((0, \infty))$. The Neyman criterion for sufficient statistics implies the result. ▮

Proof of Lemma 3.7. By Lemma 5.3 we have $E_k \sim \{\mathscr{L}(N_{k,h}) : h \in \Psi_0\}$. Let $\pi_k : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^k$ denote the projection of the first k coordinates. Then

$$F_k(\Psi_0) := (\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}}, \{\mathscr{L}((\psi_h(S_j))_{j \in \mathbb{N}})|_{\pi_k^{-1}(\mathscr{B}^k)} : h \in \Psi_0\}) \sim E_k. \quad (5.17)$$

In the sequel, we apply the following well-known convergence result for experiments, which is a consequence of the application of standard martingale arguments. Let $\mathscr{F}_n$ denote an increasing sequence of sub-σ-fields which generate a σ-field $\mathscr{F}_\infty$. Then

$$\{P_\vartheta|_{\mathscr{F}_n} : \vartheta \in \Theta\} \to \{P_\vartheta|_{\mathscr{F}_\infty} : \vartheta \in \Theta\}$$

weakly. For the sequence (5.17) we have

$$F_k(\Psi_0) \to F_\infty(\Psi_0) := (\mathbb{R}^{\mathbb{N}}, \mathscr{B}^{\mathbb{N}}, \{\mathscr{L}((N_h(S_j))_{j \in \mathbb{N}}) : h \in \Psi_0\}). \quad (5.18)$$

Since $F_\infty(\Psi_0)$ is more informative than E_∞ and E_∞ is more informative that E_k, we conclude from (5.17) and (5.18) that $E_k \to E_\infty$ and $E_\infty \sim F_\infty(\Psi_0)$. ▮

Proof of Theorem 3.9. Let $\mathscr{F}_t$ denote the σ-field generated by $\mu \to \mu(B \cap [t, \infty))$ for $B \in \mathscr{B} \cap (0, \infty)$ and $t > 0$. Then $\mathscr{F}_t \uparrow \mathscr{M}(0, \infty)$ for $t \downarrow 0$. Once again, a martingale argument proves the weak convergence of the experiments. ▮

Proof of Theorem 3.10. Consider $h \in \Psi_1$ given by $h(x) = 1_{[0,d)}(x) + h(x) 1_{[d,\infty)}(x)$. Let $S_1, S_2, ...$ be as in (3.15). Then $\mathscr{L}(N_h)$ is absolutely continuous w.r.t. P_1 and we have equality in distribution under $\mathscr{L}(N_1)$:

$$\log \frac{dN_h}{dN_1} \overset{\mathscr{L}}{=} \int_d^\infty (1 - h(x))\, w(x)\, dx + \sum_{j=1}^{\infty} \log h(S_j^{-1/\alpha}). \quad (5.19)$$

Next consider the structure model (3.12). First, we show that

$$\left(\frac{dP^n_{h(\delta_n\cdot)}}{dP^n_1}\right)_{h\in\Psi_1} \to \left(\frac{dN_h}{dN_1}\right)_{h\in\Psi_1} \tag{5.20}$$

weakly as $n\to\infty$. For this purpose we show that

$$\sum_{i=1}^n \log\frac{dP_{h(\delta_n\cdot)}}{dP_1} \overset{\mathscr{L}}{=} n\log c(h(\delta_n\cdot)) + \sum_{i=1}^n \log h(\delta_n X_{i:n}) \tag{5.21}$$

converges weakly to the distribution (5.19). The convergence of the finite dimensional marginal distributions of the likelihood process (5.20) w.r.t. P_1 follows similarly by the Cramér–Wold device. Note that

$$h(\delta_n X_{n+1-j:n}) \to h(S_j^{-1/\alpha})$$

w.r.t. the variational distance for each $j\in\mathbb{N}$. Since $\delta_n X_{n+1-j_n:n}\to 0$ for each sequence $j_n\uparrow\infty$ we can find for each $\varepsilon>0$ some $k\in\mathbb{N}$ such that

$$\limsup_{n\to\infty} P(\delta_n X_{n+1-k:n}>d)\leqslant\varepsilon.$$

Since $\log h(x)=0$ for $0\leqslant x<d$ and $S_j^{-1/\alpha}\downarrow 0$ as $j\uparrow\infty$ we have

$$\sum_{i=1}^n \log h(\delta_n X_{i:n}) \to \sum_{j=1}^\infty \log h(S_j^{-1/\alpha})$$

in distribution under P_0^n.

The convergence of the $\log c(h(\delta_n\cdot))$-part of (5.21) can be established as in the proof of Lemma 3.3. Let $F_{h(\delta_n\cdot)}$ denote the distribution function of $P_{h(\delta_n\cdot)}$. Then

$$1=\int dP_{h(\delta_n\cdot)} = c(h(\delta_n\cdot))\,F_1(d/\delta_n)+1-F_{h(\delta_n\cdot)}(d/\delta_n)$$

implies

$$\log c(h(\delta_n\cdot)) = \log F_{h(\delta_n\cdot)}(d/\delta_n) - \log F_1(d/\delta_n).$$

As in the proof of Lemma 3.3 we obtain

$$n\log c(h(\delta_n\cdot)) \to -\int_d^\infty w(z)\,h(z)\,dz + \int_d^\infty w(z)\,dz$$

and (5.20) is established. Note that (5.20) implies convergence of the experiments (3.22) for the sequence $k_n=n$ since $\mathscr{L}(N_h)$ is absolutely continuous w.r.t. $\mathscr{L}(N_1)$.

Now let $k_n \leqslant n$, $k_n \uparrow \infty$, denote any sequence of integers and let Ψ_{11} be any finite subset of Ψ_1. Since $E_{n+1-j:n, j \leqslant k}(\Psi_{11})$ converges weakly to $E_k(\Psi_{11})$ we can find by Lemma 3.7 a sequence $j_n \leqslant k_n$ with

$$E_{n+1-j:n, j \leqslant j_n}(\Psi_{11}) \to E_\infty(\Psi_{11}). \tag{5.22}$$

Thus the limit experiment (5.22) is the same w.r.t. Ψ_{11} as for $k_n = n$. Since $\{P^n_{h(\delta_n \cdot)} : h \in \Psi_{11}\}$ is more informative than $E_{n+1-j:n, j \leqslant k_n}(\Psi_{11})$ and the latter experiment is more informative than that experiment based on the j_n largest extremes we have also convergence of $E_{n+1-j:n, j \leqslant k_n}(\Psi_{11})$ to $E_\infty(\Psi_{11})$. ∎

ACKNOWLEDGMENT

This paper was written while both authors were at the University of Siegen. They are grateful to R.-D. Reiss, who introduced them to extreme value theory.

REFERENCES

[1] BINGHAM, N. H., GOLDIE, C. M., AND TEUGELS, J. L. (1987). *Regular Variation.* Cambridge Univ. Press, Cambridge.

[2] FALK, M. (1986). Rates of uniform convergence of extreme order statistics. *Ann. Inst. Statist. Math. Part A* **38** 245–262.

[3] GNEDENKO, B. V., AND KOLMOGOROV, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, Reading, MA.

[4] HAAN, L. DE (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematical Centre Tracts 32. Mathematical Centre, Amsterdam.

[5] HAAN, L. DE, AND RESNICK, S. I. (1982). Local limit theorems for sample extremes. *Ann. Probab.* **10** 396–413.

[6] JANSSEN, A. (1989). The rôle of extreme order statistics for exponential families. In *Extreme Value Theory* (J. Hüsler and R.-D. Reiss, Eds.). Lecture Notes in Statistics, Vol. 51. Springer-Verlag, New York.

[7] JANSSEN, A., AND MASON, D. M. (1990). *Non-standard Rank Tests.* Lecture Notes in Statistics, Vol. 65. Springer-Verlag, New York.

[8] JANSSEN, A., AND REISS, R.-D. (1988). Comparison of location models of Weibull type samples and extreme value processes, *Probab. Theory Related Fields* **78** 273–292.

[9] KALLENBERG, O. (1986). *Random Measures*, 4th ed. Academic Press, Berlin/London.

[10] KARR, A. F. (1986). *Point Processes and Their Statistical Inference.* Dekker, New York.

[11] KOSHEVNIK, YU. A., AND LEVIT, B. YA. (1976). On a nonparametric analogue of the information matrix. *Theory Probab. Appl.* **21** 738–753.

[12] LECAM, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer Series in Statistics. Springer-Verlag, New York.

[13] LECAM, L., AND YANG, G. L. (1990). *Asymptotics in Statistics. Some Basic Concepts.* Springer Series in Statistics. Springer-Verlag, New York.

[14] LECAM, L. (1992). Sur l'estimation de l'intensité d'un processus de Poisson. Conference talk held at 24es Journées de Statistisque, Bruxelles, May 1992.

[15] LOÈVE, M. (1956). Ranking limit problem. In *Proceedings, Third Berkely Sympos. Math. Statist. Probab.*, Vol. II, pp. 177–194, Univ. of California Press, Berkeley.

[16] MILBRODT, H. (1985). Representation of Poisson experiments. In *Infinitely Divisible Statistical Experiments* (A. Janssen, H. Milbrodt, and H. Strasser, Eds.), pp. 106–123. Lecture Notes in Statistics, Vol. 27. Springer-Verlag, Berlin/Heidelberg.

[17] MILBRODT, H., AND STRASSER, H. (1985). Limits of triangular arrays of experiments. In *Infinitely Divisible Statistical Experiments* (A. Janssen, H. Milbrodt, and H. Strasser, Eds.), pp. 14–54. Lecture Notes in Statistics, Vol. 27. Springer-Verlag, Berlin/Heidelberg.

[18] PFANZAGL, J., AND WEFELMEYER, W. (1982). *Contributions to a General Asymptotic Statistical Theory*. Lecture Notes in Statistics, Vol. 13. Springer-Verlag, Berlin/Heidelberg.

[19] REISS, R.-D. (1989). *Approximate Distributions of Order Statistics (with Applications to Nonparametric Statistics)*. Springer Series in Statistics. Springer-Verlag, New York.

[20] REISS, R.-D. (1993). *A Course on Point Processes*. Springer Series in Statistics. Springer-Verlag, New York.

[21] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Applied Probability, Vol. 4. Springer-Verlag, New York.

[22] STRASSER, H. (1985). Stability and translation invariance of statistical experiments. *Probab. Math. Statist.* **5** 1–20.

[23] STRASSER, H. (1985). *Mathematical Theory of Statistics*. De Gruyter Studies in Mathematics, Vol. 7. de Gruyter, Berlin/New York.

[24] SWEETING, T. J. (1985). On domain of uniform local attraction in extreme value theory. *Ann. Probab.* **13** 196–205.

[25] TORGERSEN, E. (1991). *Comparison of Statistical Experiments*. Cambridge Univ. Press, Cambridge.

[26] WITTING, H. (1985). *Mathematische Statistik I*. Teubner, Stuttgart.