On Statistical Information of Extreme Order Statistics

von

Dipl.-Math. Frank Marohn

Vom Fachbereich Mathematik
der Universität-Gesamthochschule Siegen
zur Verleihung des akademischen Grades

Doktor der Naturwissenschaften
(Dr. rer. nat.)

genehmigte Dissertation

Siegen 1990

Tag der Promotion: 30.01.1990

Dekan: Prof. Dr. W. Hein
Referent: Prof. Dr. R.-D. Reiss
Koreferent: Prof. Dr. A. Janssen

CONTENTS

1 INTRODUCTION

This treatise is concerned with the statistical information of extreme order statistics in certain parametric models. Section 1.1 is a chapter-by-chapter overview of the contents of this paper. Section 1.2 contains some basic definitions concerning statistical experiments. In Section 1.3, we recall the main ideas concerning the comparison of statistical experiments.

1.1 Summary

In the asymptotic theory of statistics, local as well as global results were established. Take, for example, *LeCam's* local and global asymptotic bounds for risk functions of estimates or the local and global asymptotic normality of statistical experiments. In the second chapter of the present paper, we formulate a global version of the local result of *Janssen* and *Reiss* (1988). We adopt their notation.

Before specifying our model, we give a short motivation. Assume that $X_1, ..., X_n$ are the (random) lifetimes of n aggregates, and the lifetime distribution is unknown. In practice, one observes the failure times $X_{1:n} \leq X_{2:n} \leq \cdots$ (so called "Typ II-censored data"). For obvious reasons, it often makes no sense to wait until the last aggregat has failed. So one has to come to a statistical decision based on the first $k\,(< n)$ observations. Since $X_{k+1:n}, ..., X_{n:n}$ are not observed, we suffer a loss of information. For many lifetime models, the smallest observations turn out to be very informative. Consider the following particular case: The Weibull distribution with shape parameter $a > -1$ (for $a = 0$ we get the exponential distribution) is an important lifetime distribution, and a large body of literature on statistical models has evolved for it (see, for example, the book by *Lawless* (1982). For the exponential distribution with unknown location parameter the minimum $X_{1:n}$ contains all the information.

In the following, we study location models of Weibull type distributions "near" the exponential distribution. It turns out that the k smallest observations are important. In order to give a guide how to choose k if one accepts a given loss of information, one has to calculate the loss of information. This mathematical problem is treated within the deficiency concept of statistical experiments in the sense of *LeCam*.

The starting point is a location family P_t, $t \in \mathbb{R}$, with Lebesgue density $f_t(x) = f(x-t)$, where f is of Weibull type; that is f has a representation

$$f(x) = \begin{cases} x^a r(x) & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

where r varies slowly at zero.

It is assumed that the shape parameter $a > -1$ is known to the statistician. If, for example, $r(x) = (1+a)exp(-x^{1+a})$ we get the Weibull density. We get the generalized Pareto density of type II if $r(x) = (1+a)1_{(0,1)}(x)$.

Now let $X_1, \dots, X_n$ be i.i.d. random variables with common distribution P_t and let $X_{1:n}, \dots, X_{n:n}$ denote the pertaining order statistics. It is well known that the order statistic $(X_{1:n}, \dots, X_{n:n})$ is sufficient, i.e. it contains all the information about the unknown parameter. We reduce the number of order statistics to the $k(n)$ lower extremes $X_{1:n}, \dots, X_{k(n):n}$ and calculate upper bounds for the loss of information. These calculations will be carried out within the framework of deficiency of statistical experiments. We restrict ourselves to the case $-1 < a < 1$. Notice that for $a \geq 1$ the LAN condition holds. The latter case is examined in Chapter 4.

We consider the statistical experiments E_n, $E_{n,k}$, G_k and G which arise out of this context and which were already introduced by *Janssen* and *Reiss* (1988). In Section 2.1, we collect their definitions. In Section 2.2, we establish an upper bound of the deficiency between E_n and $E_{n,k}$ using the Markov kernel criterion. Section 2.3 is concerned with the asymptotic information contained in the k smallest order statistics. Theorem 2.3.3 shows the *global sufficiency* of the $k(n)$ smallest order statistics. The comparison of the four experiments within the deficiency concept will be carried out in Section 2.4. As a suprising result, we find that $\Delta(E_n, G) \to 0$ and $\Delta(E_{n,k(n)}, G) \to 0$ as $n \to \infty$ and $k(n) \to \infty$ (Theorem 2.4.1). This strengthens a result of *Janssen* (1989 b).

Chapter 3 is concerned with Weibull type densities having a compact support $[0, b]$ where 0 and b are singularity points. In contrast to the densities considered in the previous chapter, a second singularity occurs at the right endpoint. Because of this fact, we must take the upper extremes into consideration. In Section 3.1, we define these densities and introduce the statistical experiments $\tilde{E}_n$, $\tilde{E}_{n,k_1,k_2}$, $\tilde{G}_1 \otimes \tilde{G}_2$, and $\tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}$. Sections 3.2 to 3.4 are organized similar to Sections 2.2 to 2.4. In particular, we show that the $k_1(n)$ lower and $k_2(n)$ upper extremes are *asymptotically global sufficient* (Theorem 3.3.2). This extends a result of *Weiss* (1979).

Chapter 4 is concerned with Gaussian sequences of statistical experiments. We examine the borderline case of Weibull type densities with shape parameter $a = 1$ in Section 4.1. Here, we are in the almost regular LAN situation. If $k(n) = o(n)$ tends to infinity at a sufficiently fast rate, then one may conjecture that the $k(n)$ lower order statistics are sufficient. Theorem 4.1.1 gives the following solution: If $\lim_{n\to\infty} k(n)/log(n) = 1$, then the $k(n)$ lower order statistics $(X_{1:n}, \dots, X_{k(n):n})$ are asymptotically sufficient. Conversely, if $\lim_{n\to\infty} k(n)/log(n) = 0$, the statistic $(X_{k(n)+1:n}, \dots, X_{n-k(n):n})$ turns out to be asymptotically sufficient. Section 4.2 provides a proof of the fact that in the regular LAN-case (L_2-differentiability), a fixed number of extremes does not contain any information (Theorem 4.2.1). A characterization theorem is established in Section 4.3. We show that under monotone likelihood ratios, a sequence of experiments has a non-*Gaussian* limit if and only if a fixed number of extreme order statistics, asymptotically, contains information. This result is suggested by a well-known criterion in the theory of sums of i.i.d. random variables.

Chapter 5 is devoted to point processes. In Section 5.1, we recall the definition of a point process and introduce some notations used in later sections. In Section 5.2, we show that in the i.i.d. case the original experiment and the corresponding point process experiment are equivalent (Corollary 5.2.2). Corollary 5.2.3 states that the loss of information due to a reduction of order statistics in the original experiment is the same as in the corresponding point process experiment. Empirical point processes of Weibull type samples are concerned in Section 5.3. It is in this section that we combine the results of Chapter 2 and Section 5.2.

I wish to thank *Prof. R.-D. Reiss* for drawing my attention to the problem of global sufficiency. I am also grateful to *Prof. A. Janssen* for valueable suggestions concerning Chapter 4.

1.2 Statistical Experiments

Here we list some basic definitions and recall some facts concerning statistical experiments. For the remainder of Section 1.2, we refer to *Milbrodt* and *Strasser* (1985) and *Strasser* (1985 a).

1.2.1 Definition. Let $T \neq \emptyset$ be an arbitrary set. A *statistical experiment* for the parameter set T is a triple

$$E = (\Omega, \mathcal{A}, \{P_t : t \in T\}) \tag{1.2.1}$$

where $(\Omega, \mathcal{A})$ is a sample space and $\{P_t : t \in T\}$ is a family of probability measures. If T contains exactly two points, i.e. $T = \{t_1, t_2\}$, then E is called a *binary experiment.*

The experiment E is said to be *homogeneous*, if the measures P_t, $t \in T$, are mutually equivalent, i.e. $P_s \ll P_t$ and $P_t \ll P_s$ for $s, t \in T$.

The *log-likelihood process* of E with base $s \in T$ is the process

$$(log(\frac{dP_t}{dP_s}))_{t \in T} \tag{1.2.2}$$

defined on $(\Omega, \mathcal{A}, P_s)$.

1.2.2 Definition. A homogeneous experiment $E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$ is called *Gaussian* if at least one log-likelihood process is a Gaussian process.

For example, the statistical experiment

$$(\mathbb{R}, \mathcal{B}, \{N(t\sigma^2, \sigma^2) : t \in \mathbb{R}\})$$

is Gaussian, and is called *Gaussian shift* on $\mathbb{R}$, where $N(\mu, \sigma^2)$ denotes the normal distribution with expectation μ and variance σ^2.

Every log-likelihood process of a Gaussian experiment is a Gaussian process.

1.2.3 Definition. Two statistical experiments $E = (\Omega_1, \mathcal{A}_1, \{P_t : t \in T\})$ and $F = (\Omega_2, \mathcal{A}_2, \{Q_t : t \in T\})$ are called *equivalent* (briefly $E \sim F$) if

$$\mathcal{L}((log(\frac{dP_t}{dP_s}))_{t \in T} | P_s) = \mathcal{L}((log(\frac{dQ_t}{dQ_s}))_{t \in T} | Q_s) \tag{1.2.3}$$

for every $s \in T$.

Denote by $\mathcal{E}(T)$ the collection of all experiments for the parameter space T ($\mathcal{E}(T)$ is not a set!). The elements of the quotient set $\mathcal{E}(T)/\sim$ are called *experiment types*.

We denote the *product experiment* of E and F by

$$E \otimes F = (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \{P_t \otimes Q_t : t \in T\}).$$

The n-fold product experiment of $E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$ is denoted by

$$E^n = (\Omega^n, \mathcal{A}^n, \{P_t^n : t \in T\})$$

where $(\Omega^n, \mathcal{A}^n)$ denotes the n-fold product space of $(\Omega, \mathcal{A})$ and P_t^n denotes the n-fold product measure of P_t.

1.2.4 Definition. A sequence of experiments

$$E_n = (\Omega_n, \mathcal{A}_n, \{P_{nt} : t \in T\}),\ n \in \mathbb{N},$$

converges weakly to $E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$ (briefly $E_n \to E$) if for every finite subset α of T and for every $s \in \alpha$

$$\mathcal{L}((log(\frac{dP_{nt}}{dP_{ns}}))_{t\in\alpha}|P_{ns}) \longrightarrow \mathcal{L}((log(\frac{dP_t}{dP_s}))_{t\in\alpha}|P_s) \text{ weakly.} \tag{1.2.4}$$

The most important class of limit experiments are the *Gaussian*.

1.2.5 Definition. A sequence of statistical experiments $E_n = (\Omega_n, \mathcal{A}_n, \{P_{nt} : t \in T\})$ is called *Gaussian sequence* if the pertaining sequence of product experiments $E_n^n = (\Omega_n^n, \mathcal{A}_n^n, \{P_{nt}^n : t \in T\})$ has only Gaussian accumulation points.

A sequence $(E_n)_{n\in\mathbb{N}} \in \mathcal{E}(\mathbb{R})$ is called *asymptotically normal* if it converges weakly to a Gaussian shift (on $\mathbb{R}$).

1.2.6 Definition. Let P_n and Q_n be probability measures on $(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$. The sequence $(Q_n)_{n\in\mathbb{N}}$ is *contiguous* to the sequence $(P_n)_{n\in\mathbb{N}}$ if

$$P_n(A_n) \longrightarrow 0,\ A_n \in \mathcal{A}_n, \text{ implies } Q_n(A_n) \longrightarrow 0. \tag{1.2.5}$$

1.2.7 Definition. A sequence of experiments $E_n = (\Omega_n, \mathcal{A}_n, \{P_{nt} : t \in T\})$ is *contiguous* if for every pair $(s,t) \in T \times T$ the sequence of probability measures $(P_{ns})_{n\in\mathbb{N}}$ and $(P_{nt})_{n\in\mathbb{N}}$ are mutually contiguous.

The meaning of contiguity is the following: The weak convergence of a contiguous sequence of experiments is equivalent to the weak convergence of one particular log-likelihood process.

If $E_n \to E$ weakly, then E is homogeneous iff $(E_n)_{n\in\mathbb{N}}$ is contiguous.

Let us denote by

$$\|P - Q\| := \sup_{A\in\mathcal{A}} |P(A) - Q(A)| \tag{1.2.6}$$

the variational distance between two probability measures P and Q, defined on some measurable space $(\Omega, \mathcal{A})$.

1.2.8 Definition. A sequence of experiments $E_n = (\Omega_n, \mathcal{A}_n, \{P_{nt} : t \in T\})$ is said to be *infinitesimal* if for every $s, t \in T$

$$\lim_{n \to \infty} \|P_{ns} - P_{nt}\| = 0. \tag{1.2.7}$$

Intuitively speaking, the assumption of infinitesimality guarantees that the influence of a single measure is asymptotically negligible.

1.3 Comparison of Statistical Experiments

When investigating a complicated statistical experiment, it is sometimes useful to construct another experiment which is close to the original one but is of statistically simple nature. The usual way to obtain an approximating experiment is to embed the original experiment into a sequence of experiments and to expand the log-likelihood function. As one is normally more interested in approximations than in limit theorems, one has to estimate the distance of the two experiments. A natural quantity for comparing two experiments is the deficiency distance of *LeCam*. It is based on the comparison of risk functions available in the two experiments.

The theory of ***comparison of statistical experiments*** was initiated by the papers of *Blackwell* (1951) and *LeCam* (1964). For an excellent full depth treatment of this topic, the reader should consult the monograph by *Strasser* (1985 a).

Let $T \neq \emptyset$ be an arbitrary set and consider the experiments

$$E = (\Omega_1, \mathcal{A}_1, \{P_t : t \in T\})$$

and

$$F = (\Omega_2, \mathcal{A}_2, \{Q_t : t \in T\}).$$

Moreover, we consider decision problems (T, D, W) consisting of a topological space D and a bounded, continuous loss function $W : T \times D \to \mathbb{R}$. If D contains two elements, one speaks of a ***testing*** problem. Let $\epsilon \geq 0$. E is called ***ϵ-deficient*** w.r.t. F, if for every decision problem (T, D, W) and for every decision function ρ_2 in F there exists a decision function ρ_1 in E such that for every $t \in T$ the following inequality between the risk functions is valid:

$$(1.3.1) \quad \int_{\Omega_1} \int_D W(t,x)\rho_1(\omega_1, dx)\, dP_t(\omega_1)$$
$$\leq \int_{\Omega_2} \int_D W(t,x)\rho_2(\omega_2, dx)\, dQ_t(\omega_2) + \epsilon\, sup_{x \in D}|W(t,x)|.$$

The *deficiency* of E w.r.t. F is the number

$$(1.3.2) \qquad \delta(E,F) = inf\{\epsilon : E \text{ is } \epsilon - \text{deficient w.r.t. } F\}.$$

E is called *more informative* than F, if $\delta(E,F) = 0$. The *deficiency* between E and F is the symmetrical quantity

$$(1.3.3) \qquad \Delta(E,F) = max\{\delta(E,F), \delta(F,E)\}.$$

E and F are *equivalent* iff $\Delta(E,F) = 0$.

The deficiency is a pseudodistance on the collection $\mathcal{E}(T)$ of all experiments for the parameter space T. Moreover, $(\mathcal{E}(T)/\sim, \Delta)$ is a complete metric space.

A sequence of experiments $E_n \in \mathcal{E}(T)$ converges (in the strong sense) to an experiment $E \in \mathcal{E}(T)$, if $\Delta(E_n, E) \longrightarrow 0$ as $n \to \infty$. The sequence $(E_n)_n$ converges *weakly* to E iff $\Delta_\alpha(E_n, E) \longrightarrow 0$ as $n \to \infty$ for every finite subset α of T. With Δ_α we denote the restriction to the parameter set α.

Now, the famous *randomization criterion* due to *LeCam* (1964) says that

$$\delta(E,F) = \inf_M \sup_{t \in T} \|Q_t - MP_t\| \tag{1.3.4}$$

where the infimum is taken over all *transitions* (*stochastic operators*) from the $L-$space $L(E)$ of the experiment E to the $L-$space $L(F)$ of the experiment F. In the case of equal sample spaces we get $\Delta(E,F) \le \sup_{t\in T} \|P_t - Q_t\|$, since the identity defines a transition.

If, in addition, E is a dominated experiment and F is such that Ω_2 is Polish (i.e. metrizable as complete, separable metric space) and $\mathcal{A}_2$ is the Borel $\sigma-$field, then the *Markov kernel criterion* holds, i.e. the infimum in (1.3.4) can be taken over all Markov kernels $K : \mathcal{A}_2 \times \Omega_1 \to [0,1]$ from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$, where by definition

$$KP_t(\cdot) = \int K(\cdot|\omega_1)\, dP_t(\omega_1). \tag{1.3.5}$$

If $KP_t = Q_t$ for some kernel K then E is more informative than F. In this case, E is also called *Blackwell sufficient* or *exhaustive* for F.

In general, it is not possible to calculate the deficiency; some exceptional cases may be found in the paper of *Torgersen* (1972). Due to the Markov kernel criterion, an upper bound of $\delta(E,F)$ is obtained by $\sup_{t\in T} \|Q_t - K^* P_t\|$ where K^* is some appropriate Markov kernel. The choice of the kernel K^* is crucial.

The relation between sufficiency and deficiency is the following. Let $X : \Omega_1 \longrightarrow \Omega_2$ be any $\mathcal{A}_1, \mathcal{A}_2$-measurable map with $\mathcal{L}(X|P_t) = Q_t$. Clearly, E is more informative than F, since the transition

$$K(A_2|\omega_1) := 1_{A_2} \circ X(\omega_1)$$

satisfies $KP_t = Q_t$. If X is a sufficient statistic, then the statistical experiments E and F are equivalent. If, in addition, Ω_2 is Polish, then F is even Blackwell sufficient for E: First, choose a conditional distribution $P(\cdot|X) \in \bigcap_{t\in T} P_t(\cdot|X)$ which is independent of the parameter t. Second, find a version $\tilde{P}(\cdot|X)$ of $P(\cdot|X)$ which is the regular conditional distribution. The kernel

$$\tilde{K}(A_1|\omega_2) = \tilde{P}(A_1|X = \omega_2)$$

satisfies $\tilde{K}Q_t = P_t$. For more details concerning Blackwell sufficiency see *Heyer* (1982).

An experiment $(\Omega, \mathcal{A}, \{P_t : t \in T\})$ is said to be "totally uninformative" or "trivial" if all measures P_t are equal (see *Heyer* (1982), p. 240, *LeCam* (1986), p. 19). Such an experiment is equivalent to $E_0 = (\mathbb{R}, \mathcal{B}, \{\epsilon_0\})$ (E_0 considered as an element of $\mathcal{E}(T)$). The totally uninformative experiment is the weakest element of $\mathcal{E}(T)$ in the sense that

$$\delta(E, E_0) = 0 \qquad \forall E \in \mathcal{E}(T).$$

The transition $K(B|\omega) = 1_B(0)$, $B \in \mathcal{B}$, satisfies

$$\epsilon_0(B) = \int K(B|\omega)\, dP_t(\omega), \quad t \in T.$$

Now, we recall the definition of asymptotic sufficiency, see *Strasser* (1985 a), p. 422, 423. Let $E_n = (\Omega_n, \mathcal{A}_n, \{P_{nt} : t \in T\})$ be a sequence of statistical experiments, where T is a Hilbert space with $0 \in T$. Let X_n be an $\mathcal{A}_n$-measurable random variable, and let $\mathcal{F}(\Omega_n, \mathcal{A}_n)$ denote the set of all critical functions defined on $(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$.

The sequence $(X_n)_n$ is *asymptotically sufficient* (for E_n) if

$$\lim_{n\to\infty} \Big| \int_{\Omega_n} \varphi_n\, dP_{nt} - \int_{\Omega_n} E_{P_{n0}}(\varphi_n|X_n)\, dP_{nt} \Big| = 0 \tag{1.3.6}$$

for every $t \in T$ and for every sequence of critical functions $\varphi_n \in \mathcal{F}(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$.

The sequence $(X_n)_n$ is *asymptotically sufficient uniformly on compact subsets of* T if

$$\lim_{n\to\infty} \sup_{t\in K} \Big| \int_{\Omega_n} \varphi_n\, dP_{nt} - \int_{\Omega_n} E_{P_{n0}}(\varphi_n|X_n)\, dP_{nt} \Big| = 0 \tag{1.3.7}$$

for every compact set $K \subseteq T$, and for every sequence of critical functions $\varphi_n \in \mathcal{F}(\Omega_n, \mathcal{A}_n)$, $n \in \mathbb{N}$.

Definition (1.3.7) is *LeCam's* definition of asymptotic sufficiency, see the paper of *LeCam* (1956) and the book by *LeCam* (1986), Theorem 1, p. 177.

In the last years, many papers have been concerned with the comparison of statistical experiments. Besides the previously-mentioned papers and the famous paper of *Torgersen* (1970) we quote the following ones:

Helgeland (1982) studied the increase in statistical information by adding independent observations, where the underlying model is a 1-parametric exponential family.

Mammen (1983, 1986) deduced upper bounds for the gain of information due to additional independent observations for experiments which fulfill general dimensionality conditions. Exponential families were also studied.

One dependent case can be found in the article of *Lindqvist* (1984), where homogeneous Markov chains $(X_n)_n$ are concerned. Here the starting value X_0 is concerned to be the unknown parameter and the loss of information of X_n is investigated if n tends to infinity.

The deficiency of one shift experiment relative to another on infinite dimensional Banach spaces was studied by *Luschgy* (1987).

Janssen (1989 a) was the first to recognize the importance of extreme order statistics for exponential families. It was shown that the extreme order statistics asymptotically contain all the information.

Approximate sufficiency of sparse order statistics (also in nonparametric models) was investigated in the book by *Reiss* (1989).

The deficiency concept is a mathematically rigorous one. However, *Lehmann* (1989) in his article *Comparing location models* came to the conclusion that the requirements for an experiment to be more informative than another are "too strong to hold in many situations in which intuition suggests that one experiment is more informative than another". One natural approach to weaken the requirements is to define the deficiency not for all decision problems but only for some (appropriate) class of decision problems. This approach will not be discussed is this paper.

2 LOCATION EXPERIMENTS OF WEIBULL TYPE

The four sections of this chapter concern location families of Weibull type densities.

2.1 Definition and Notations

Let P_t, $t \in \mathbb{R}$, be a family of probability measures with Lebesgue density $f_t(x) = f(x-t)$, where f is of Weibull type; that is, f has a representation

$$f(x) = \begin{cases} x^a r(x) & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \tag{2.1.1}$$

where $a > -1$ and r varies slowly at zero. Our starting point is the statistical experiment

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \{P^n_{\delta_n t} : t \in \mathbb{R}\}\right) \tag{2.1.2}$$

where the normalizing sequence $(\delta_n)_{n\in\mathbb{N}}$ will be explained in Section 2.3. (Throughout of this paper, the dependence of the shape parameter a is suppressed in the notations since a is held constant.)

The second experiment is

$$E_{n,k} = \left(\mathbb{R}^k, \mathcal{B}^k, \{V_{n,k,t} : t \in \mathbb{R}\}\right) \tag{2.1.3}$$

where

$$V_{n,k,t} = \mathcal{L}\left(\delta_n^{-1}(X_{1:n}, \ldots, X_{k:n}) | P^n_{\delta_n t}\right). \tag{2.1.4}$$

Notice that $\delta(E_n, E_{n,k}) = 0$, i.e. E_n is more informative than $E_{n,k}$.

Finally, we consider the experiments G and G_k which occur as limit experiments of E_n and $E_{n,k}$, respectively. Let $(Y_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of standard exponential random variables and denote

$$S_m = \sum_{i=1}^{m} Y_i \tag{2.1.5}$$

the m-th partial sum.

Define

$$Q_{k,t} = \mathcal{L}\left((S_m^{1/(1+a)} + t)_{m\leq k}\right) \tag{2.1.6}$$

and

$$Q_t = \mathcal{L}\left((S_m^{1/(1+a)} + t)_{m\in\mathbb{N}}\right). \tag{2.1.7}$$

Then

$$G_k = \left(\mathbb{R}^k, \mathcal{B}^k, \{Q_{k,t} : t \in \mathbb{R}\}\right) \tag{2.1.8}$$

and

$$G = \left(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \{Q_t : t \in \mathbb{R}\}\right). \tag{2.1.9}$$

The comparison of these experiments is carried out in Section 2.4, according to the following diagram:

$$\begin{array}{ccc} E_n & \overset{(2.4.9)}{\longleftrightarrow} & E_{n,k} \\ (2.4.21)\;\updownarrow & & \updownarrow\;(2.4.20) \\ \mathrm{G} & \underset{(2.4.16)}{\longleftrightarrow} & G_k \end{array}$$

2.2 Upper Bound of the Deficiency between E_n and $E_{n,k}$

Let $-1 < a < 1$ and let P_t be defined as in Section 2.1. Denote with

$$K_t^{(n,k)}(\cdot|\underline{x}) \tag{2.2.1}$$

the conditional distribution of $(X_{1:n}, \ldots, X_{n:n})$ given $(X_{1:n}, \ldots, X_{k:n}) = (x_1, \ldots, x_k)$ $=: \underline{x}$ under the parameter t. It is well-known (see *Reiss* (1989), Theorem 1.8.1) that

$$K_t^{(n,k)}(\cdot|\underline{x}) = \epsilon_{x_1} \times \cdots \times \epsilon_{x_k} \times \mathcal{L}(Y_{1:n-k}, \ldots, Y_{n-k:n-k}) \tag{2.2.2}$$

where ϵ_y denotes the Dirac measure at y and the Y_i, $i \in \{1, \ldots, n-k\}$, are i.i.d. random variables with common distribution P_{t,x_k} (the truncation of P_t on the left at x_k).

Let F denote the distribution function of P_0. If $F(x_k - t) < 1$ then the distribution P_{t,x_k} has the Lebesgue density

$$f_{t,x_k} = \frac{f_t}{1 - F(x_k - t)} 1_{[x_k,\infty)}. \tag{2.2.3}$$

To obtain an upper bound of $\Delta(E_n, E_{n,k})$ we choose a kernel of the following type:

$$K_{\hat{\kappa}(X_{1:n},\ldots,X_{k:n})}^{(n,k)}(\cdot|X_{1:n}, \ldots, X_{k:n}) \tag{2.2.4}$$

where $\hat{\kappa}$ is an appropriate estimator of the unknown parameter. *Janssen* and *Reiss* (1988) considered the kernel $K_0^{(n,k)}$ for their local treatment of the problem on a right neighborhood of 0. In our situation, a plausible choice of $\hat{\kappa}$ will be the minimum, i.e.

$$\hat{\kappa}(X_{1:n}, \ldots, X_{k:n}) = X_{1:n}. \tag{2.2.5}$$

Using the kernel

$$K_{x_1}^{(n,k(n))}(\cdot|\underline{x}) \tag{2.2.6}$$

we will be able to verify the global sufficiency of the $k(n)$ smallest order statistics. The use of kernels under an estimated parameter has turned out to be very successful in order to establish bounds of the deficiency, see e.g. *Helgeland* (1982), *Mammen* (1986) and *Weiss* (1979). The upper bound of $\Delta(E_n, E_{n,k})$ will depend on three auxiliary functions h, g, and ψ, cf. *Janssen* and *Reiss* (1988),

$$\begin{aligned} h(y) &= y^{\frac{a}{2}} - (y-1)^{\frac{a}{2}},\ y \geq 1, \\ g(x) &= \int_1^\infty \left((y^{\frac{a}{2}} r^{\frac{1}{2}}(xy) - (y-1)^{\frac{a}{2}} r^{\frac{1}{2}}(x(y-1)))/(r^{\frac{1}{2}}(x)) - h(y)\right)^2 dy, \\ \psi(z) &= \int_z^\infty h^2(y)\, dy,\ z \geq 1. \end{aligned} \tag{2.2.7}$$

We note that $h \in L_2(\lambda)$ and $\psi \equiv 0$ for $a = 0$.

2.2.1 Theorem. *For $k \in \{1,\dots,n\}$, $t \in \mathbb{R}$, and $\epsilon > 0$ such that $k/n \le F(\epsilon) < 1$, the following inequality holds:*

$$(2.2.8)\quad \|\mathcal{L}((X_{1:n},\dots,X_{n:n})|P_t^n) - K_{x_1}^{(n,k)}\mathcal{L}((X_{1:n},\dots,X_{k:n})|P_t^n)\| \le (1-F(\epsilon))^{-\frac{1}{2}}(n-k)^{\frac{1}{2}} I_{1,n}(I_{2,n}+I_{3,n,k}) + R_{n,k}$$

where

$$I_{1,n}^2 = \int_{(0,\epsilon)} r(x_1)x_1^{a+1}\, d\mathcal{L}(X_{1:n}|P_0^n)(x_1),$$

$$I_{2,n}^2 = \int_{(0,\epsilon)} g(x_1)\, d\mathcal{L}(X_{1:n}|P_0^n)(x_1),$$

$$I_{3,n,k}^2 = \iint_{(0,\epsilon)\times(0,\epsilon)} \psi(\frac{x_k}{x_1})\, d\mathcal{L}((X_{1:n},X_{k:n})|P_0^n)(x_1,x_k),$$

and

$$R_{n,k} = exp(-n(F(\epsilon)-\frac{k}{n})^2/3).$$

We remark that the right-hand side of (2.2.8) is independent of the parameter t!

Proof of Theorem 2.2.1. For the sake of convenience, let us abbreviate the left-hand side of (2.2.8) by $\rho(n,k,t)$. Similar to the proof of Theorem (2.8) of *Janssen* and *Reiss* (1988), we obtain

$$\begin{aligned}
&\rho(n,k,t)\\
&= \sup_{B\in\mathcal{B}^n}\left|\int (K_t^{(n,k)}(B|\underline{x}) - K_{x_1}^{(n,k)}(B|\underline{x}))\, d\mathcal{L}((X_{1:n},\dots,X_{k:n})|P_t^n)(\underline{x})\right|\\
&\le \int \sup_{B\in\mathcal{B}^n}\left|K_t^{(n,k)}(B|\underline{x}) - K_{x_1}^{(n,k)}(B|\underline{x})\right| d\mathcal{L}((X_{1:n},\dots,X_{k:n})|P_t^n)(\underline{x})\\
&= \iint \|P_{t,x_k}^{n-k} - P_{x_1,x_k}^{n-k}\|\, d\mathcal{L}((X_{1:n},X_{k:n})|P_t^n)(x_1,x_k)\\
&\le \int_t^{\epsilon+t}\int_t^{\epsilon+t} \|P_{t,x_k}^{n-k} - P_{x_1,x_k}^{n-k}\|\, d\mathcal{L}((X_{1:n},X_{k:n})|P_t^n)(x_1,x_k) + P_t^n\{X_{k:n} > \epsilon+t\}\\
&\le \sqrt{2(n-k)}\int_t^{\epsilon+t}\int_t^{\epsilon+t} H(P_{t,x_k},P_{x_1,x_k})\, d\mathcal{L}((X_{1:n},X_{k:n})|P_t^n)(x_1,x_k) + P_t^n\{X_{k:n} > \epsilon+t\}.
\end{aligned}$$

The last step follows from the inequality

$$H(P^m,Q^m) \le \sqrt{2m}H(P,Q)$$

where

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

denotes the *Hellinger distance* between the probability measures P and Q.

In the following, let $(x_1,x_k)\in(t,\epsilon+t)\times(t,\epsilon+t)$. Then we find that $F(x_k-x_1)\le F(x_k-t)\le F(\epsilon)<1$. Since $x+y\ge 2\sqrt{xy}$ for $x,y\ge 0$, we obtain

$$\begin{aligned}
2H^2(P_{t,x_k},P_{x_1,x_k}) &= \int\left(f_{t,x_k}^{1/2}(y)-f_{x_1,x_k}^{1/2}(y)\right)^2 dy\\
&= 2\Big\{1-\left((1-F(x_k-t))(1-F(x_k-x_1))\right)^{-1/2}\int_{x_k}^{\infty} f^{1/2}(y-t)\,f^{1/2}(y-x_1)\,dy\Big\}\\
&= 2\Big\{1-\left((1-F(x_k-t))(1-F(x_k-x_1))\right)^{-1/2}\\
&\quad\times\Big[(1-F(x_k-t)+1-F(x_k-x_1))/2-\frac{1}{2}\int_{x_k}^{\infty}\left(f^{1/2}(y-t)-f^{1/2}(y-x_1)\right)^2 dy\Big]\Big\}\\
&\le (1-F(\epsilon))^{-1}\int_{x_k}^{\infty}\left(f^{1/2}(y-t)-f^{1/2}(y-x_1)\right)^2 dy\\
&= (1-F(\epsilon))^{-1}\int_{x_k-t}^{\infty}\left(f^{1/2}(y)-f^{1/2}(y-(x_1-t))\right)^2 dy,
\end{aligned}$$

where the last step follows by substituting y by $y+t$. Combining these results, we get

$$\begin{aligned}
\rho(n,k,t)\le(1-F(\epsilon))^{-\frac{1}{2}}(n-k)^{\frac{1}{2}}\int_t^{\epsilon+t}\int_t^{\epsilon+t}\tilde{d}(x_1-t,x_k-t)\,d\mathcal{L}\big((X_{1:n},X_{k:n})|P_t^n\big)(x_1,x_k)\\
+\mathcal{L}\big(X_{k:n}|P_t^n\big)\big((\epsilon+t,\infty)\big)
\end{aligned}$$

where

$$\tilde{d}^2(x_1-t,x_k-t)=\int\left(f^{\frac{1}{2}}(y)-f^{\frac{1}{2}}(y-(x_1-t))\right)^2 1_{(x_k-t,\infty)}(y)\,dy.$$

Since $\{P_t\}$ is a location family, we obtain

(2.2.9)

$$\begin{aligned}
\rho(n,k,t)\le(1-F(\epsilon))^{-\frac{1}{2}}(n-k)^{\frac{1}{2}}\iint_{(0,\epsilon)\times(0,\epsilon)}\tilde{d}(x_1,x_k)\,d\mathcal{L}\big((X_{1:n},X_{k:n})|P_0^n\big)(x_1,x_k)\\
+P_0^n\{X_{k:n}>\epsilon\}.
\end{aligned}$$

Using the exponential bound for order statistics (see *Reiss* (1989), Lemma 3.3.1) and the quantile transformation we get

$$P_0^n\{X_{k:n} > \epsilon\} \le exp(-n(F(\epsilon) - \frac{k}{n})^2/3) \tag{2.2.10}$$

for $k/n \le F(\epsilon)$ (cf. Lemma 2.8 of *Janssen* and *Reiss* (1988)). By substituting y by $x_1 y$ and applying the Minkowski inequality, we obtain

(2.2.11)

$$\begin{aligned}
\tilde{d}(x_1, x_k) &= x_1^{\frac{1}{2}} \Big(\int_{x_k/x_1}^{\infty} \big(f^{\frac{1}{2}}(x_1 y) - f^{\frac{1}{2}}(x_1(y-1))\big)^2 \, dy \Big)^{\frac{1}{2}} \\
&= r^{\frac{1}{2}}(x_1)\, x_1^{\frac{1+a}{2}} \Big(\int_{x_k/x_1}^{\infty} \big(\frac{y^{\frac{a}{2}} r^{\frac{1}{2}}(x_1 y) - (y-1)^{\frac{a}{2}} r^{\frac{1}{2}}(x_1(y-1))}{r^{1/2}(x_1)}\big)^2 \, dy \Big)^{\frac{1}{2}} \\
&\le r^{\frac{1}{2}}(x_1)\, x_1^{\frac{1+a}{2}} \Big(\int_{x_k/x_1}^{\infty} \big(\frac{y^{\frac{a}{2}} r^{\frac{1}{2}}(x_1 y) - (y-1)^{\frac{a}{2}} r^{\frac{1}{2}}(x_1(y-1))}{r^{1/2}(x_1)} - h(y)\big)^2 \, dy \Big)^{\frac{1}{2}} \\
&\quad + r^{\frac{1}{2}}(x_1)\, x_1^{\frac{1+a}{2}} \Big(\int_{x_k/x_1}^{\infty} h^2(y) \, dy \Big)^{\frac{1}{2}}.
\end{aligned}$$

Now, using the Cauchy Schwarz inequality, combining (2.2.9) – (2.2.11), and taking into account the definition of g and ψ, the proof is completed. ∎

It is obvious that using the kernel $K_0^{(n,k)}$ it is not possible to establish an upper bound of the deficiency independent of the parameter. Because of this, one has to restrict the parameter space to compact sets.

Let us denote the right-hand side of (2.2.8) by $D(n,k)$. Notice that $D(n,k)$ is an upper bound of the deficiency between E_n and $E_{n,k}$, i.e.

$$\Delta(E_n, E_{n,k}) \le D(n,k). \tag{2.2.12}$$

2.3 The Asymptotic Information Contained in the k Smallest Order Statistics

Once again, let F denote the distribution function of P_0. We recall that for the normalizing sequence $\delta_n = F^{-1}(\frac{1}{n})$ occuring in E_n and $E_{n,k}$, we have (see *Bingham et al.* (1987), Theorem 1.5.12)

$$\delta_n = n^{-1/(1+a)} L(\frac{1}{n}) \tag{2.3.1}$$

for some slowly varying function L.

From the theory of regular variation it is known that densities of type (2.1.1) fulfill the *von Mises* condition

$$\lim_{x \to 0} \frac{x f(x)}{F(x)} = 1 + a \tag{2.3.2}$$

(see *Bingham et al.* (1987), Proposition 1.5.10). Condition (2.3.2) implies that F belongs to the strong domain of attraction (see *Falk* (1985) or *Sweeting* (1985)), i.e. convergence of the extremes holds w.r.t. the variational distance

$$\|\mathcal{L}(\delta_n^{-1}(X_{1:n}, \ldots, X_{k:n})|P_0^n) - \mathcal{L}(S_1^{1/(1+a)}, \ldots, S_k^{1/(1+a)})\| \to 0 \text{ as } n \to \infty. \tag{2.3.3}$$

We will assume that for some $\epsilon > 0$

$$r \text{ is bounded on } (0, \epsilon) \text{ and } \liminf_{x \downarrow 0} r(x) > 0, \tag{2.3.4}$$

or

$$r \text{ is decreasing on } (0, \epsilon). \tag{2.3.4'}$$

Note that $r(x) > 0$ for $x \in (0, x_0)$ and for some x_0.

In addition, we assume that the following condition of *Janssen* and *Mason* (1989) is fulfilled. First, it is assumed that

$$\int_{\epsilon}^{\infty} \left(f^{1/2}(x-t) - f^{1/2}(x)\right)^2 dx = o\left(t^{1+a} r(t)\right) \ (t \downarrow 0) \tag{2.3.5 (i)}$$

for each $\epsilon > 0$.

Condition (2.3.5) (i) says that the Hellinger distance is mainly determined by the local behaviour of the density at the singularity 0. Moreover, no other singularities of higher order occur.

Now, let

$$r(x) = l(x) r_1(x) = l(x) exp\left(\int_x^{x_0} \frac{b(u)}{u}\, du\right)$$

for some $x_0 > 0$, l measurable and $l(x) \to c \in (0, \infty)$, $b(x) \to 0$, as $x \to 0$, be the *Karamata representation* of the slowly varying function r (see e.g. *Bingham et al.* (1987), Theorem 1.3.1) It is known that r_1 is absolutely continuous on $(0, x_0)$ with $x r_1'(x)/r_1(x) = b(x)$ *a.e.* (see *Bingham et al.* (1987), p. 15).

In addition, l is assumed to be continuous on $[0, x_0]$ with $l(0) > 0$ and

$$\int_0^{x_0/2} \left(l^{1/2}(x+t) - l^{1/2}(x)\right)^2 x^a r_1(x)\,dx = o\left(t^{1+a} r(t)\right) \tag{2.3.5 (ii)}$$

as $t \downarrow 0$.

Under condition (2.3.5), *Janssen* and *Mason* (1989) proved that

$$\lim_{x \to 0} g(x) = 0, \tag{2.3.6}$$

see Lemma 10.13.

Moreover, we need moment convergence of the normalized sequence $\delta_n^{-1} X_{1:n}$. Concerning limit theorems for moments of extremes, we refer to *Polfeldt* (1970), p. 45, and to the book by *Resnick* (1987), Chapter 2.

2.3.1 Lemma. *Assume that*

$$\begin{aligned} E_{P_0} X_1 &< \infty \qquad \text{if } a \in (-1, 0] \\ E_{P_0} X_1^2 &< \infty \qquad \text{if } a \in (0, 1). \end{aligned} \tag{2.3.7}$$

Then

$$\limsup_{n \to \infty} \int x^{1+a}\, d\mathcal{L}(\delta_n^{-1} X_{1:n} | P_0^n)(x) \le C \in (0, \infty) \tag{2.3.8}$$

for some constant $C > 0$.

Condition (2.3.7) is always satisfied in standard cases.

Proof of Lemma 2.3.1. Let $a \in (-1, 0]$ (the case $a \in (0, 1)$ may be treated in a similar way). Since F belongs to the (weak) domain of attraction of a (min-)stable distribution, condition (2.3.7) implies that the first moment of $\delta_n^{-1} X_{1:n}$ converges to the first moment of the limiting distribution, i.e.

$$\lim_{n \to \infty} \int x\, d\mathcal{L}(\delta_n^{-1} X_{1:n} | P_0^n)(x) = \int x\, d\mathcal{L}(S_1^{1/(1+a)})(x) = C_a < \infty \tag{2.3.9}$$

(see *Resnick* (1987), Proposition 2.1).

Now the proof can be easily completed. ∎

2.3.2 Theorem. *Let* f *be a density of type* (2.1.1), *and assume that* (2.3.4) ((2.3.4′)), (2.3.5), *and* (2.3.7) *are valid. There exists a constant* $C > 0$ *such that*

$$\lim_{n \to \infty} D(n, k) \le C \left(E\psi\left(\left(\frac{S_k}{S_1}\right)^{1/(1+a)} \right) \right)^{\frac{1}{2}}. \tag{2.3.10}$$

PROOF. 1) First, let condition (2.3.4) be fulfilled. Substituting $\delta_n x$ for x, we obtain for some constant $C > 0$

$$(n-k)I_{1,n}^2 \le C(n-k)\,\delta_n^{1+a} \int\limits_{(0,\delta_n^{-1}\epsilon)} x_1^{a+1}\, d\mathcal{L}(\delta_n^{-1}X_{1:n}|P_0^n)\,(x_1).$$

Moreover,

$$\lim_{n\to\infty}(n-k)\,(\delta_n x_1)^{1+a}\, r(\delta_n x_1) = (1+a)x_1^{a+1}. \tag{2.3.11}$$

Taking into account Lemma 2.3.1 and condition (2.3.4), we obtain

$$\limsup_{n\to\infty}(n-k)^{1/2}\, I_{1,n} < \infty.$$

In the case (2.3.4′), we proceed as follows: Let $s > 0$. Then

$$(n-k)\,I_{1,n}^2 \le \int\limits_0^s (n-k)\,(\delta_n x_1)^{1+a}\, r(\delta_n x_1)\, d\mathcal{L}(\delta_n^{-1}X_{1:n}|P_0^n)(x_1)$$

$$+\,(n-k)\,\delta_n^{1+a}\, r(\delta_n s) \int\limits_s^{\delta_n^{-1}\epsilon} x_1^{a+1}\, d\mathcal{L}(\delta_n^{-1}X_{1:n}|P_0^n)\,(x_1).$$

Janssen and *Reiss* (1988) showed that

$$\lim_{n\to\infty}\sup_{0\le x_1\le s}(n-k)\,(\delta_n x_1)^{1+a}\, r(\delta_n x_1) = (1+a)\, s^{1+a}. \tag{2.3.12}$$

Now using (2.3.11), (2.3.12), and Lemma 2.3.1, we see that $\limsup_{n\to\infty}(n-k)^{1/2}I_{1,n} < \infty$ also holds under condition (2.3.4′).

2) An upper bound of $I_{2,n}$ is obtained as follows:

$$\begin{aligned} I_{2,n}^2 &= \int\limits_{(0,\delta_n^{-1}\epsilon)} g(\delta_n x_1)\, d\mathcal{L}(\delta_n^{-1}X_{1:n}|P_0^n)(x_1) \\ &\le \int\limits_{(0,\delta_n^{-1}\epsilon)} g(\delta_n x_1)\, d\mathcal{L}(S_1^{1/(1+a)})(x_1) \\ &\quad + c_2\|\mathcal{L}(\delta_n^{-1}X_{1:n}|P_0^n) - \mathcal{L}(S_1^{1/(1+a)})\| \\ &=: A_n' + c_2 B_n' \end{aligned}$$

where $c_2 := \sup_{0<x_1<\epsilon} g(x_1)$. Notice that c_2 is finite because of (2.3.6). From (2.3.3), we know that B_n' tends to zero as n tends to infinity. An application of the *dominated convergence theorem of Lebesgue* shows that A_n' also tends to zero as n tends to infinity.

3) Similar to 2), we obtain

$$\begin{aligned} I_{3,n,k}^2 &\le E\psi\big((\frac{S_k}{S_1})^{1/(1+a)}\big) \\ &\quad + \|h\|_2^2\, \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, X_{k:n})|P_0^n) - \mathcal{L}(S_1^{1/(1+a)}, S_k^{1/(1+a)})\|. \end{aligned}$$

The assertion follows from 1) - 3) and inequality (2.2.8) . ■

2.3.3 Theorem (Global Sufficiency). *Suppose that the conditions of Theorem 2.3.2 are fulfilled.*

(i) *Let $a \in (-1,1), a \neq 0$. There exists a constant $C > 0$ such that*

$$\Delta(E_n, E_{n,k}) \leq Ck^{(a-1)/(2(1+a))} + o(n^0). \tag{2.3.13}$$

(ii) *If $a = 0$ then*

$$\Delta(E_n, E_{n,1}) = o(n^0). \tag{2.3.14}$$

Remark. Theorem 2.3.3 states that the $k(n)$ lower extremes $(X_{1:n}, \dots, X_{k(n):n})$ are ***asymptotically global sufficient*** in the sense that

$$\lim_{n\to\infty} \Delta(E_n, E_{n,k(n)}) = 0 \tag{2.3.15}$$

whenever $k(n)$ with $k(n) \to \infty$ as $n \to \infty$. This has an important consequence in testing theory. Assume that $\varphi_n(X_1, \dots, X_n)$ is a test for E_n. Then

$$\varphi_{n,k}(\underline{x}) := E_{P^n_{x_1}}(\varphi_n|(X_{1:n}, \dots, X_{k:n}) = \underline{x}) = \int \varphi_n(y) K^{(n,k)}_{x_1}(dy, \underline{x}),\ \underline{x} = (x_1, \dots, x_k),$$

is a new test—based on the k smallest observations. We obtain

$$\sup_{t\in\mathbb{R}} \left| E_{P^n_t}\varphi_n - E_{P^n_t} E_{P^n_{x_1}}(\varphi_n|(X_{1:n}, \dots, X_{k(n):n}))\right| \leq \Delta(E_n, E_{n,k(n)}) \to 0$$

for $n \to \infty$ and $k(n) \to \infty$.

In the paper of *Janssen* and *Reiss* (1988), it was shown that

$$\lim_{n\to\infty} \Delta_s(E_n, E_{n,k(n)}) = 0,$$

where Δ_s indicates the restriction to the compact parameter set $[0, s]$.

PROOF OF THEOREM 2.3.3. (ii) is immediate from Theorem 2.3.2, since $\psi \equiv 0$ for $a = 0$. So it remains to consider the case $a \neq 0$. First observe that $(z > 1)$

$$\psi(z) = \int_z^\infty h^2(y)\, dy \leq \frac{a^2}{4(1-a)}(z-1)^{a-1}. \tag{2.3.16}$$

To verify (2.3.16) use the mean value theorem. Notice that

$$\frac{S_1}{S_k} \stackrel{d}{=} U_{1:k-1} \tag{2.3.17}$$

where the symbol $\stackrel{d}{=}$ denotes equality in distribution and $U_{1:k-1}$ is the minimum of $k-1$ i.i.d. $(0,1)$-uniform random variables ($U_{1:0} := 1$). Now let $k \geq 2$ and choose $\delta \in (\frac{1}{2}, 1)$. From (2.3.16) and (2.3.17), we deduce:

(2.3.18)
$$E\psi\left(U_{1:k-1}^{-1/(1+a)}\right) \leq \frac{a^2}{4(1-a)} \int\limits_{(0,\delta)} (u^{-1/(1+a)} - 1)^{a-1}\, d\mathcal{L}(U_{1:k-1})(u)$$
$$+ \|h\|_2^2\, P\{U_{1:k-1} > \delta\}.$$

Once more, applying the exponential bound for order statistics (compare with (2.2.10)), we see that

$$P\{U_{1:k-1} > \delta\} \leq exp\left(-(k-1)(\delta - \frac{1}{k-1})^2/3\right) \tag{2.3.19}$$

holds. A bound of the other expression on the right-hand side of (2.3.18) is obtained as follows: Substituting u by $u/(k-1)$, we get

(2.3.20)
$$\int\limits_{(0,\delta)} (u^{-1/(1+a)} - 1)^{a-1} d\mathcal{L}(U_{1:k-1})(u)$$
$$= (k-1)^{\frac{a-1}{1+a}} \int\limits_0^{(k-1)\delta} \left(u^{\frac{-1}{1+a}} - (k-1)^{\frac{-1}{1+a}}\right)^{a-1} d\mathcal{L}\left((k-1)U_{1:k-1}\right)(u)$$
$$\leq (k-1)^{\frac{a-1}{1+a}} (1 - \delta^{\frac{1}{1+a}})^{a-1} \int\limits_0^{(k-1)\delta} u^{\frac{1-a}{1+a}}\, d\mathcal{L}\left((k-1)U_{1:k-1}\right)(u)$$

where the last inequality follows from the fact that for $u \in (0, (k-1)\delta)$ we have

$$u^{-1/(1+a)} - (k-1)^{-1/(1+a)} \geq (1 - \delta^{1/(1+a)})u^{-1/(1+a)}.$$

If

$$\limsup_{k\to\infty} \int\limits_{(0,(k-1)\delta)} u^{(1-a)/(1+a)}\, d\mathcal{L}\left((k-1)U_{1:k-1}\right)(u) < \infty, \tag{2.3.21}$$

then the proof follows from Theorem 2.3.2 and (2.3.18) – (2.3.21). To verify (2.3.21), we may apply the same arguments used in the proof of Lemma 2.3.1. Notice that in case of the uniform distribution, moments of arbitrary order do exist. ■

2.4 Comparison of the Statistical Experiments $E_n, E_{n,k}, G$ and G_k

In *Janssen* (1989) (see also *Janssen* and *Reiss* (1988)) it was shown that $\Delta_s(E_n, G) \to 0$ and $\Delta_s(E_{n,k(n)}, G) \to 0$ as $n \to \infty$ and $k(n) \to \infty$, where s indicates the restriction to the compact parameter set $[0, s]$. The next theorem states that the same holds on the whole real line. This is a suprising result. Usually, one has to restrict the parameter set to compact sets, see *LeCam* (1986), Theorem 2 (Theorem of Lindae), p. 92, and Remark 2, p. 93.

2.4.1 Theorem (Strong Convergence). *Assume that the conditions of Theorem 2.3.2 are valid. Then*

$$\Delta(E_n, G) \to 0 \text{ as } n \to \infty \tag{2.4.1}$$

and

$$\Delta(E_{n,k(n)}, G) \to 0 \text{ as } n \to \infty \tag{2.4.2}$$

whenever $k(n) \leq n$ *and* $k(n) \to \infty$ *as* $n \to \infty$.

PROOF. Theorem 2.3.3 states that $\Delta(E_n, E_{n,k(n)}) \to 0$ as $n \to \infty$, $k(n) \to \infty$. Now, by omitting the index s, we see that the proof of (2.4.1) is a repetition of the proof of Lemma (5.21) in *Janssen* and *Reiss* (1988).

(2.4.2) is immediate from $\Delta(E_{n,k(n)}, G) \leq \Delta(E_{n,k(n)}, E_n) + \Delta(E_n, G)$. ∎

In the following, we establish rates of convergence. First, we proof

2.4.2 Lemma. *Let f fulfill (2.1.1) and let* $0 < \lambda < 1$. *There exists a constant* $C > 0$ *such that*

(i) *for* $a \in (-1, 1), a \neq 0$:

$$\begin{aligned} D(n,k) &\leq C\Big((k^{(a-1)/(2(1+a))} + \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, X_{k:n})|P_0^n) - \mathcal{L}(S_1^{1/(1+a)}, S_k^{1/(1+a)})\|^{\frac{1}{2}} \\ &\qquad + (\int_{(0,\delta_n^{-1}\epsilon)} g(\delta_n x_1) d\mathcal{L}(S_1^{1/(1+a)})(x_1))^{\frac{1}{2}}\Big) \end{aligned} \tag{2.4.3}$$

(ii) *for* $a = 0$:

$$D(n,1) \leq C(\int_{(0,\delta_n^{-1}\epsilon)} g(\delta_n x_1) d\mathcal{L}(S_1)(x_1))^{\frac{1}{2}} \tag{2.4.4}$$

for all $n \in \mathbb{N}$ *and* $k \leq \lambda n$.

PROOF. For $k < \lambda n$ we have

$$exp(-n(F(\epsilon) - \frac{k}{n})^2/3)) \leq exp(-k(F(\epsilon) - \lambda)^2/(3\lambda)).$$

Now (2.4.3) and (2.4.4) follow from Theorem 2.2.1 and from the arguments of the proofs of Theorems 2.3.2 and 2.3.3. ∎

The upper bounds (2.4.3) and (2.4.4) involve the term $g(\delta_n x_1)$. To establish rates of convergence one has to impose further assumptions (cf. *Janssen* and *Reiss* (1988)). Let

$$r(x) = c\,exp(\tilde{h}(x)),\ 0 < x < x_0 \tag{2.4.5}$$

where $c > 0$ and $\tilde{h}$ satisfies the condition

$$|\tilde{h}(x)| \leq Lx^{\gamma} \tag{2.4.6}$$

for some constant $L > 0$ and $\gamma > 0$. Note that $\lim_{x \to 0} r(x) = c$, and that condition (2.3.4) is fulfilled.

Under condition (2.4.5), we may choose the normalizing sequence

$$\tilde{\delta}_n = ((1+a)/c)^{1/(1+a)} n^{-1/(1+a)}. \tag{2.4.7}$$

Note that $\tilde{\delta}_n \sim \delta_n$ and that (2.3.3), (2.4.3), and (2.4.4) hold with δ_n replaced by $\tilde{\delta}_n$.

First, we treat the case $a \neq 0$.

2.4.3 Theorem. *Let $a \in (-1, 1)$, $a \neq 0$. Assume in additon to (2.1.1) and (2.4.5) that f is absolutely continuous on $(0, \infty)$ and that*

$$\int_0^\infty \frac{(r'(x))^2}{r(x)} x^a\, dx < \infty. \tag{2.4.8}$$

Then for every $\lambda \in (0, 1)$ there exists a constant $C > 0$, such that for $n \in \mathbb{N}$ and $k \leq \lambda n$ the following inequality holds:

$$\begin{aligned} &\Delta(E_n, E_{n,k}) \\ &\leq C\Big((k^{(a-1)/2((1+a))} + (\frac{k}{n})^{\gamma/(2(1+a))} k^{\frac{1}{4}} + (\frac{k}{n})^{1/2} + n^{max\{a-1,-2\gamma\}/(2(1+a))} \Big). \end{aligned} \tag{2.4.9}$$

PROOF. Throughout, C denotes a generic constant which does not depend on n and $k \leq \lambda n$. Under conditions (2.4.5) and (2.4.8), we have

$$g(x) \leq Cx^{min\{1-a, 2\gamma\}}, \tag{2.4.10}$$

with g as in (2.2.7). It was shown by *Janssen* and *Reiss* (1988) that

$$\begin{aligned} \tilde{g}(x) &= \int_1^\infty \left(y^{a/2} r^{1/2}(xy) - (y-1)^{a/2} r^{1/2}(x(y-1)) - c^{1/2} h(y) \right)^2 dy \\ &= O\big(x^{min\{1-a, 2\gamma\}}\big). \end{aligned}$$

Now, assertion (2.4.10) is immediate from

$$g^{1/2}(x) \leq \frac{1}{r(x)} \tilde{g}^{1/2}(x) + |r^{1/2}(x) - c^{1/2}|\, \|h\|_{L_2(\lambda)}$$

and

$$|r^{1/2}(x) - c^{1/2}| \leq Cx^{\gamma}.$$

Thus we have

$$\int_{(0,\tilde{\delta}_n^{-1}\epsilon)} g(\tilde{\delta}_n x_1)\, d\mathcal{L}(S_1^{1/(1+a)})\,(x_1) \le C\tilde{\delta}_n^{min\{1-a,2\gamma\}}.$$

Taking account of (2.4.7), we obtain

$$\int_{(0,\tilde{\delta}_n^{-1}\epsilon)} g(\tilde{\delta}_n x_1)\, d\mathcal{L}(S_1^{1/(1+a)})\,(x_1) \le Cn^{max\{a-1,-2\gamma\}/(1+a)}. \tag{2.4.11}$$

From the proofs of Theorem 2.3.2 and Theorem 2.3.3, we know that

$$I_{3,n,k}^2 \le C\left((k-1)^{\frac{a-1}{1+a}} + \|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{1:n}, X_{k:n})|P_0^n) - \mathcal{L}(S_1^{1/(1+a)}, S_k^{1/(1+a)})\|\right). \tag{2.4.12}$$

The Corollary 5.5.5 in *Reiss* (1989) implies
(2.4.13)

$$\|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{1:n},\dots,X_{k:n})|P_0^n) - \mathcal{L}(S_1^{1/(1+a)},\dots,S_k^{1/(1+a)})\| \le C((\frac{k}{n})^{\gamma/(1+a)}k^{\frac{1}{2}} + \frac{k}{n}).$$

(Corollary 5.5.5 still holds for $c \neq 1+a$ and $c_n = \tilde{\delta}_n$. To see this, examine the proof of Corollary 5.5.5: choose $x_{0,n} = cnx_0$ and $f_n(x) = \frac{1}{cn}f(\frac{x}{cn})$). Hence, the asserted inequality is immediate from Lemma 2.4.2 and (2.4.11) – (2.4.13). ∎

Now we treat the case $a = 0$.

2.4.4 Theorem. *Let f be a density of type (2.1.1) for $a = 0$. In addition to (2.4.5), let f be absolutely continuous on $(0,\infty)$ and*

$$\int_0^\infty \frac{|f'(x)|^\eta}{f(x)^{\eta-1}}\,dx < \infty \tag{2.4.14}$$

for some $\eta \in (1,2]$. Then for $k \le \lambda n$, $\lambda \in (0,1)$, the following inequality holds:

$$\Delta(E_n, E_{n,k}) \le Cn^{(1-\eta)/2} \tag{2.4.15}$$

for some $C > 0$.

PROOF. As in *Janssen* and *Reiss* (1988) one may show $g(x) = O(x^{\eta-1})$. Now taking account of Lemma 2.4.2 (ii), we see that (2.4.15) holds. ∎

2.4.5 Theorem.

(i) *Let $a \in (-1,1), a \neq 0$. There exists a constant $C > 0$ such that*

$$\Delta(G, G_k) \le C\big(E\psi(U_{1:k-1})\big)^{\frac{1}{2}} = Ck^{(a-1)/2(1+a)}. \tag{2.4.16}$$

(ii) *Let $a = 0$. Then*

$$\Delta(G, G_1) = 0. \tag{2.4.17}$$

PROOF. From (2.3.3), we know

$$\Delta(E_{n,k}, G_k) \to 0 \text{ as } n \to \infty. \tag{2.4.18}$$

Applying the triangle inequality we obtain

$$\Delta(G, G_k) \le \Delta(G, E_n) + \Delta(E_n, E_{n,k}) + \Delta(E_{n,k}, G_k). \tag{2.4.19}$$

Taking into account Theorem 2.3.2, 2.3.3, and 2.4.1, the proof is complete. ∎

From (2.4.13), we deduce

$$\Delta(G_k, E_{n,k}) \le C\Big(\big(\frac{k}{n}\big)^{\gamma/(1+a)}k^{\frac{1}{2}} + \frac{k}{n}\Big). \tag{2.4.20}$$

Now, we are ready to establish the rate of convergence of $\Delta(E_n, G)$.

2.4.6 Theorem. *Let f be a density of type (2.1.1) with $a \in (-1,1)$ and let (2.3.7), (2.4.5), (2.4.8), and (2.4.14) be valid. Then*

$$\Delta(E_n, G) = O(n^{\beta(\gamma,a)}) \tag{2.4.21}$$

where

$$\beta(\gamma, a) = \begin{cases} \frac{a-1}{2(3+a)} & \text{for } \gamma \ge \frac{2(1+a)}{1-a},\ a \neq 0 \\ \frac{\gamma(a-1)}{(1+\gamma)4(1+a)} & \text{for } 0 < \gamma < \frac{2(1+a)}{1-a},\ a \neq 0 \\ max\{-\gamma, -\frac{1}{2}\} & \text{for } a = 0. \end{cases}$$

PROOF. The calculations are similar to those of *Janssen* and *Reiss* (1988). We start with the inequality

$$\Delta(E_n, G) \le \Delta(E_n, E_{n,k}) + \Delta(E_{n,k}, G_k) + \Delta(G_k, G).$$

Assume first $a \neq 0$. Combining (2.4.9), (2.4.20), and (2.4.16), we obtain

$$\Delta(E_n, G) \le C\Big((k^{(a-1)/(4(1+a))} + \big(\frac{k}{n}\big)^{\gamma/(2(1+a))}k^{\frac{1}{4}} + \big(\frac{k}{n}\big)^{1/2} + n^{max\{a-1,-2\gamma\}/(2(1+a))}\Big)$$

uniformly over all $k \leq \lambda n$ if $0 < \lambda < 1$.

Elementary calculations show that $n^{\gamma/(1+\gamma)}$ is the solution of the equation

$$x^{(a-1)/(4(1+a))} = \left(\frac{x}{n}\right)^{\gamma/(2(1+a))} x^{1/4}.$$

Hence we choose

$$k(n) = [n^{\gamma/(1+\gamma)}].$$

Now, $\frac{\gamma(a-1)}{(1+\gamma)4(1+a)} > -\frac{1}{2(1+\gamma)} \Leftrightarrow \gamma < \frac{2(1+a)}{1-a}$. Moreover, the relations $\frac{\gamma(a-1)}{(1+\gamma)4(1+a)} > \frac{-2\gamma}{2(1+a)}$ and $\frac{\gamma(a-1)}{(1+\gamma)4(1+a)} > \frac{a-1}{2(1+a)}$ are trivially valid.

Notice, that for $\gamma \geq \gamma_0 := \frac{2(1+a)}{1-a}$ the assertion (2.4.6) holds for γ_0 instead of γ. We obtain $\frac{\gamma_0(a-1)}{(1+\gamma_0)4(1+a)} = \frac{a-1}{2(3+a)}$. Thus, for $a \neq 0$

$$\Delta(E_n, G) \leq C n^{\beta(\gamma,a)}.$$

For $a = 0$ and $k = 1$, we deduce from (2.4.15), (2.4.17), and (2.4.20) the upper bound

$$\Delta(E_n, G) \leq C\big(n^{-\gamma} + n^{-1} + n^{-1/2}\big).$$

Thus, (2.4.21) is shown to be valid. ∎

REMARK. The explicit representation of the limit experiment, namely,

$$G = \big(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \{(S_m^{1/(1+a)} + t)_{m\in\mathbb{N}} : t \in \mathbb{R}\}\big)$$

was exhibited by *Janssen* (1989 b). Furthermore, G is a *stable* Poisson experiment with index of stability $1+a$. Stable experiments were introduced under the label "scale invariance" by *Müller* (1973) and were thoroughly investigated by *Strasser* (1985 b).

Since X^{-1} is Fréchet iff X is Weibull, one may conjecture that all results of this chapter carry over to location models of Fréchet type densities, where $S_m^{1/(1+a)}$ is replaced by $S_m^{-1/(1+a)}$. But this is no longer true, since the transformation X^{-1} does not lead to a location model. The Fréchet distribution has finite *Fisher* information for all shape parameters $a > -1$. Hence, we are in the usual LAN-case, i.e. the product experiment—rescaled with $\delta_n = n^{-1/2}$—converges weakly to a *Gaussian* shift. In this case, a fixed number of extremes asymptotically does not contain any information (see Section 4.2).

3 WEIBULL TYPE DENSITIES WITH COMPACT SUPPORT

The next four sections are concerned with Weibull type densities having a compact support. This chapter is organized similarly to Chapter 2.

3.1 Definition and Notations. Our starting point are Weibull type densities having a compact support $[0,b]$, $b>0$; that is for some x_0

$$\tilde{f}(x)=\begin{cases} x^a r_1(x), & \text{for } 0<x<x_0 \\ (b-x)^a r_2(b-x), & \text{for } b-x_o<x<b \\ 0, & \text{for } x\notin[0,b] \end{cases} \tag{3.1.1}$$

where $a>-1$ and r_i, $i=1,2$, are slowly varying functions at zero.

Throughout, it is assumed that the shape parameter a is known. Again, we consider the non-regular case $a\in(-1,1)$. For example, we get the density of the uniform distribution $\tilde{f}(x)=1_{(0,1)}(x)$ if $a=0$, $b=1$, and $r_1=r_2\equiv 1$ (see the comment on page 35).

Let $\tilde{P}_t$, $t\in\mathbb{R}$, be a location family defined via a density $\tilde{f}$ of type (3.1.1). Moreover, $X_{1:n},\ldots,X_{n:n}$ are the order statistics of a sample of size n, where X_i are i.i.d. random variables with common distribution P_t. In contrast to the densities of type (2.1.1), we have a further singularity at the right-hand side of the range of the distribution. So we have to include the upper extremes in our considerations. Inside $(x_0,b-x_0)$, we claim that the density f behaves well (see condition (3.3.6)(i)).

For convenience, we introduce the abbreviations

$$W_{n,k_1}=(X_{1:n},\ldots,X_{k_1:n}),\ Z_{n,k_2}=(X_{n-k_2+1:n},\ldots,X_{n:n}). \tag{3.1.2}$$

In Section 3.3, we prove the global sufficiency of the $k_1(n)$ lower and $k_2(n)$ upper extremes $(W_{n,k_1(n)},Z_{n,k_2(n)})$.

We introduce the following statistical experiments. Let

$$\tilde{E}_n=(\mathbb{R}^n,\mathcal{B}^n,\{\tilde{P}^n_{\delta_n t}:t\in\mathbb{R}\}) \tag{3.1.3}$$

where the sequence $(\delta_n)_n$ is explained in Section 3.3.

The second experiment is

$$\tilde{E}_{n,k_1,k_2}=(\mathbb{R}^{k_1+k_2},\mathcal{B}^{k_1+k_2},\{\tilde{V}_{n,k_1,k_2,t}:t\in\mathbb{R}\}) \tag{3.1.4}$$

where

$$\tilde{V}_{n,k_1,k_2,t}=\mathcal{L}(\delta_n^{-1}(W_{n,k_1},Z_{n,k_2})|\tilde{P}^n_{\delta_n t}). \tag{3.1.5}$$

Obviously

$$\tilde{V}_{n,k_1,k_2,t}=\mathcal{L}(\delta_n^{-1}(W_{n,k_1},Z_{n,k_2})+t|\tilde{P}^n_0).$$

Finally, we introduce the product experiments $\tilde{G}_{1,k_1}\otimes\tilde{G}_{2,k_2}$ and $\tilde{G}_1\otimes\tilde{G}_2$ which arise out of approximation to $\tilde{E}_{n,k_1,k_2}$ and $\tilde{E}_n$. Let $(Y_{i,l})_{l\in\mathbb{N}}$, $i=1,2$, be a sequence of independent random variables, where $Y_{1,l}$ and $Y_{2,l}$ are standard exponential and negative standard exponential, respectively. We will denote the $m-th$ partial sum by

$$S_{i,m}=\sum_{l=1}^{m}Y_{i,l}. \tag{3.1.6}$$

Define $(i = 1, 2)$

$$\tilde{Q}_{i,k_i,t} = \mathcal{L}\big((S_{i,m}^{1/(1+a)} + t)_{m \le k_i}\big) \tag{3.1.7}$$

and

$$\tilde{Q}_{i,t} = \mathcal{L}\big((S_{i,m}^{1/(1+a)} + t)_{m \in \mathbb{N}}\big). \tag{3.1.8}$$

Then

$$\tilde{G}_{i,k_i} = \big(\mathbb{R}^{k_i}, \mathcal{B}^{k_i}, \{\tilde{Q}_{i,k_i,t} : t \in \mathbb{R}\}\big) \tag{3.1.9}$$

and

$$\tilde{G}_i = \big(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}}, \{\tilde{Q}_{i,t} : t \in \mathbb{R}\}\big). \tag{3.1.10}$$

The comparison of the four statistical experiments is carried out in Section 3.4, according to the following diagram:

$$\begin{array}{ccccc} & \tilde{E}_n & \overset{(3.4.6)}{\longleftrightarrow} & \tilde{E}_{n,k_1,k_2} & \\ (3.4.25) & \updownarrow & & \updownarrow & (3.4.17) \\ & \tilde{G}_1 \otimes \tilde{G}_2 & \underset{(3.4.23)}{\longleftrightarrow} & \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2} & \end{array}$$

The decisive link between $\tilde{E}_{n,k_1,k_2}$ and $\tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}$ turns out to be the product experiment $\tilde{E}_{n,k_1}^{(1)} \otimes \tilde{E}_{n,k_2}^{(2)}$, where:

$$\tilde{E}_{n,k_i}^{(i)} = \big(\mathbb{R}^{k_i}, \mathcal{B}^{k_i}, \{\tilde{V}_{n,k_i,t}^{(i)} : t \in \mathbb{R}\}\big) \tag{3.1.11}$$

with

$$\begin{aligned} \tilde{V}_{n,k_1,t}^{(1)} &= \mathcal{L}\big(\delta_n^{-1} W_{n,k_1} | \tilde{P}_{\delta_n t}\big) \\ \tilde{V}_{n,k_2,t}^{(2)} &= \mathcal{L}\big(\delta_n^{-1} Z_{n,k_2} | \tilde{P}_{\delta_n t}\big). \end{aligned} \tag{3.1.12}$$

3.2 Upper Bound of the Deficiency between $\tilde{E}_n$ and $\tilde{E}_{n,k_1,k_2}$

To establish an upper bound of $\Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2})$ we choose the kernel

$$K_{x_1}^{(n,k_1,k_2)}(\cdot|w_{k_1}, z_{k_2}). \tag{3.2.1}$$

This is the conditional distribution of $(X_{1:n}, \dots, X_{n:n})$ given $W_{n,k_1} = w_{k_1} := (x_1, \dots, x_{k_1})$, $Z_{n,k_2} = z_{k_2} := (x_{n-k_2-1}, \dots, x_n)$ under the estimation $X_{1:n} = x_1$. Again, we know from *Reiss* (1989), Theorem 1.8.1, that

$$\begin{aligned} &K_t^{(n,k_1,k_2)}(\cdot|w_{k_1}, z_{k_2}) \\ &= \epsilon_{x_1} \times \cdots \times \epsilon_{x_{k_1}} \times \mathcal{L}(\tilde{Y}_{1:n-k_1-k_2}, \dots, \tilde{Y}_{n-k_1-k_2:n-k_1-k_2}) \times \epsilon_{x_{n-k_2+1}} \times \cdots \times \epsilon_{x_n} \end{aligned} \tag{3.2.2}$$

where $\tilde{Y}_i$, $i = 1, \dots, n - k_1 - k_2$, are i.i.d. random variables with common distribution $\tilde{P}_{t,x_{k_1},x_{n-k_2+1}}$ (the truncation of $\tilde{P}_0$ on the left at x_k and on the right at x_{n-k_2+1}).

We denote by $\tilde{F}$ the distribution function of $\tilde{P}_0$ and let $\tilde{f}_t(\cdot) = \tilde{f}(\cdot - t)$. If $\tilde{F}(x_{k_1} - t) \neq \tilde{F}(x_{n-k_2+1} - t)$ then the distribution $\tilde{P}_{t,x_{k_1},x_{n-k_2+1}}$ has the Lebesgue density

$$\tilde{f}_{t,x_{k_1},x_{n-k_2+1}} = \frac{\tilde{f}_t}{\tilde{F}(x_{n-k_2+1} - t) - \tilde{F}(x_{k_1} - t)} 1_{[x_{k_1},x_{n-k_2+1}]}. \tag{3.2.3}$$

We remark that

$$\mathcal{L}\big((X_{k_1}, X_{n-k_2+1})|\tilde{P}_t^n\big)\left(\{(x_{k_1}, x_{n-k_2+1}) : \tilde{F}(x_{k_1} - t) = \tilde{F}(x_{n-k_2+1} - t)\}\right) = 0 \tag{3.2.4}$$

and

$$\tilde{P}_{t,x_{k_1},x_{n-k_2+1}} \perp \tilde{P}_{0,x_{k_1},x_{n-k_2+1}} \text{ for } t > b. \tag{3.2.5}$$

The upper bound depends on the following auxiliary functions h, ψ, g_1, g_2, and g_3:

$$\begin{aligned} h(y) &= y^{a/2} 1_{(0,\infty)}(y) - (y-1)^{a/2} 1_{(1,\infty)}(y), \\ \psi(z) &= \int_z^\infty h^2(y), \; z > 1, \\ g_1(x) &= \int_1^{\epsilon/(2x)} \Big(\frac{(y^{\frac{a}{2}} r_1^{\frac{1}{2}}(xy) - (y-1)^{\frac{a}{2}} r_1^{\frac{1}{2}}(x(y-1))}{r_1^{1/2}(x)} - h(y)\Big)^2 dy, \\ g_2(x) &= \int_1^{1+\epsilon/(2x)} \Big(\frac{(y^{\frac{a}{2}} r_2^{\frac{1}{2}}(xy) - (y-1)^{\frac{a}{2}} r_2^{\frac{1}{2}}(x(y-1))}{r_2^{1/2}(x)} - h(y)\Big)^2 dy, \\ g_3(x) &= \big(x^{1+a} r_1(x)\big)^{-1} \int_{\epsilon/2}^{b-\epsilon/2} \big(\tilde{f}^{\frac{1}{2}}(y) - \tilde{f}^{\frac{1}{2}}(y-x)\big)^2 dy, \end{aligned} \tag{3.2.6}$$

for some $\epsilon \in (0, x_0/2)$ and $0 < x < \epsilon/2$.

Note that $h \in L_2(\lambda)$ and $\psi \equiv 0$ for $a = 0$.

3.2.1 Theorem. *Let $0 < \epsilon < x_0/2$. For $k_1,\ k_2 \in \{1,...,n\}$, such that $k_1 + k_2 \le n$, $1/n \le \tilde{F}(\epsilon/2)$, $k_1/n \le \tilde{F}(\epsilon) < \tilde{F}(b-\epsilon) \le (n-k_2)/n$, the following inequality holds:*

$$
\begin{aligned}
(3.2.7)\quad & \|\mathcal{L}((X_{1:n},\dots,X_{n:n})|\tilde{P}_t^n) - K_{x_1}^{(n,k_1,k_2)}\,\mathcal{L}((W_{n,k_1},Z_{n,k_2})|\tilde{P}_t^n)\| \\
& \le \big(\tilde{F}(b-3\epsilon/2)-\tilde{F}(\epsilon)\big)^{-1/2} \times \big\{(n-k_1-k_2)^{1/2}\,\tilde{I}_{1,n}\big(\tilde{I}_{2,n}+\tilde{I}_{3,n,k_1}+\tilde{I}_{4,n}\big) \\
& \qquad + (n-k_1-k_2)^{1/2}\,\tilde{I}_{5,n}\big(\tilde{I}_{6,n}+\tilde{I}_{7,n,k_2}\big)\big\} + R_{n,k_1,k_2}
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{I}_{1,n}^2 &= \int_0^{\epsilon/2} x_1^{a+1} r_1(x_1)\, d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1), \\
\tilde{I}_{2,n}^2 &= \int_0^{\epsilon/2} g_1(x_1)\, d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1), \\
\tilde{I}_{3,n,k_1}^2 &= \int_0^{\epsilon}\int_0^{\epsilon/2} \psi(x_{k_1}/x_1)\, d\mathcal{L}((X_{1:n},X_{k_1:n})|\tilde{P}_0^n)(x_1,x_{k_1}), \\
\tilde{I}_{4,n}^2 &= \int_0^{\epsilon/2} g_3(x_1)\, d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1), \\
\tilde{I}_{5,n}^2 &= \int_0^{\epsilon/2} x_1^{a+1} r_2(x_1)\, d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1), \\
\tilde{I}_{6,n}^2 &= \int_0^{\epsilon/2} g_2(x_1)\, d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1), \\
\tilde{I}_{7,n,k_1}^2 &= \int_0^{\epsilon}\int_0^{\epsilon/2} \psi(1+(b-x_{n-k_2+1})/x_1)\, d\mathcal{L}((X_{1:n},X_{n-k_2+1:n})|\tilde{P}_0^n)(x_1,x_{n-k_2+1})
\end{aligned}
$$

and

$$
\begin{aligned}
R_{n,k_1,k_2} &= exp\big(-n(\tilde{F}(\epsilon/2)-1/n)^2/3\big) \\
&+ exp\big(-n(\tilde{F}(\epsilon)-k_1/n)^2/3\big) \\
&+ exp\big(-n(1-\tilde{F}(b-\epsilon)-k_2/n)^2/3\big).
\end{aligned}
$$

PROOF. Denote the left-hand side of (3.2.7) by $\rho(n, k_1, k_2, t)$. Then

$$\begin{aligned}
&\rho(n, k_1, k_2, t) \\
&= \sup_{B\in\mathcal{B}^n} \Big| \int \big(K_t^{(n,k_1,k_2)}(B|(w_{k_1}, z_{k_2})) - K_{x_1}^{(n,k_1,k_2)}(B|(w_{k_1}, z_{k_2}))\big) \\
&\qquad\qquad d\mathcal{L}\big((W_{n,k_1}, Z_{n,k_2})|\tilde{P}_t^n\big)(w_{k_1}, z_{k_2})\Big| \\
&\le \int \sup_{B\in\mathcal{B}^n} \big|K_t^{(n,k_1,k_2)}(B|(w_{k_1}, z_{k_2})) - K_{x_1}^{(n,k_1,k_2)}(B|(w_{k_1}, z_{k_2}))\big| \\
&\qquad\qquad d\mathcal{L}\big((W_{n,k_1}, Z_{n,k_2})|\tilde{P}_t^n\big)(w_{k_1}, z_{k_2}) \\
&\le \iiint \big\|\tilde{P}^{n-k_1-k_2}_{t,x_{k_1},x_{n-k_2+1}} - \tilde{P}^{n-k_1-k_2}_{x_1,x_{k_1},x_{n-k_2+1}}\big\| \\
&\qquad\qquad d\mathcal{L}\big((X_{1:n}, X_{k_1:n}, X_{n-k_2+1:n})|\tilde{P}_t^n\big)(x_1, x_{k_1}, x_{n-k_2+1}) \\
&\le \int_{b-\epsilon+t}^{b+t} \int_t^{\epsilon+t} \int_t^{\frac{\epsilon}{2}+t} \big\|\tilde{P}^{n-k_1-k_2}_{t,x_{k_1},x_{n-k_2+1}} - \tilde{P}^{n-k_1-k_2}_{x_1,x_{k_1},x_{n-k_2+1}}\big\| \\
&\qquad\qquad d\mathcal{L}\big((X_{1:n}, X_{k_1:n}, X_{n-k_2+1:n})|\tilde{P}_t^n\big)(x_1, x_{k_1}, x_{n-k_2+1}) \\
&\qquad + \tilde{P}_t^n\{X_{1:n} > \frac{\epsilon}{2} + t\} + \tilde{P}_t^n\{X_{k_1:n} > \epsilon + t\} + \tilde{P}_t^n\{X_{n-k_2+1:n} < b - \epsilon + t\} \\
&\le \sqrt{2(n-k_1-k_2)} \int_{b-\epsilon+t}^{b+t} \int_t^{\epsilon+t} \int_t^{\frac{\epsilon}{2}+t} H\big(\tilde{P}_{t,x_{k_1},x_{n-k_2+1}}, \tilde{P}_{x_1,x_{k_1},x_{n-k_2+1}}\big) \\
&\qquad\qquad d\mathcal{L}\big((X_{1:n}, X_{k_1:n}, X_{n-k_2+1:n})|\tilde{P}_t^n\big)(x_1, x_{k_1}, x_{n-k_2+1}) \\
&\qquad + \tilde{P}_0^n\{X_{1:n} > \epsilon/2\} + \tilde{P}_0^n\{X_{k_1:n} > \epsilon\} + \tilde{P}_0^n\{X_{n-k_2+1:n} < b - \epsilon\}.
\end{aligned}$$

For $x_1 \in (t, \frac{\epsilon}{2} + t)$, $x_{k_1} \in (t, \epsilon + t)$, and $x_{n-k_2+1} \in (b - \epsilon + t, b + t)$, we have

$$\tilde{F}(x_{n-k_2+1} - t) - \tilde{F}(x_{k_1} - t) \ge \tilde{F}(b - \epsilon) - \tilde{F}(\epsilon)$$
$$\tilde{F}(x_{n-k_2+1} - x_1) - \tilde{F}(x_{k_1} - x_1) \ge \tilde{F}(b - 3\epsilon/2) - \tilde{F}(\epsilon).$$

At this stage, we may apply the same arguments used in the proof of Theorem 2.2.1 which yield

(3.2.8)

$$\rho(n,k_1,k_2,t) \leq \left(\tilde{F}(b-3\epsilon/2)-\tilde{F}(\epsilon)\right)^{-1/2}(n-k_1-k_2)^{1/2}$$

$$\times \int_{b-\epsilon}^{b}\int_{0}^{\epsilon}\int_{0}^{\epsilon/2} \tilde{d}(x_1,x_{k_2},x_{n-k_2+1})\, d\mathcal{L}\big((X_{1:n},X_{k_1:n},X_{n-k_2+1:n})|\tilde{P}_0^n\big)(x_1,x_{k_1},x_{n-k_2+1})$$

$$+\tilde{P}_0^n\{X_{1:n}>\epsilon/2\}+\tilde{P}_0^n\{X_{k_1:n}>\epsilon\}+\tilde{P}_0^n\{X_{n-k_2+1:n}<b-\epsilon\}$$

where

$$\tilde{d}^2(x_1,x_{k_1},x_{n-k_2+1}) = \int_{x_{k_1}}^{x_{n-k_2+1}} \left(\tilde{f}^{1/2}(y)-\tilde{f}^{1/2}(y-x_1)\right)^2 dy.$$

Similar to (2.2.10), we deduce the following exponential bounds:

(3.2.9)

$$\tilde{P}_0^n\{X_{1:n}>\epsilon/2\} \leq exp\big(-n(\tilde{F}(\epsilon/2)-1/n)^2/3\big),\ for\ 1/n \leq \tilde{F}(\epsilon/2),$$

$$\tilde{P}_0^n\{X_{k_1:n}>\epsilon\} \leq exp\big(-n(\tilde{F}(\epsilon)-k_1/n)^2/3\big),\ for\ k_1/n \leq \tilde{F}(\epsilon),$$

$$\tilde{P}_0^n\{X_{n-k_2+1:n}<b-\epsilon\} \leq exp\big(-n(1-\tilde{F}(b-\epsilon)-k_2/n)^2/3\big),\ for\ k_2/n \leq 1-\tilde{F}(b-\epsilon).$$

Moreover, for $0<\epsilon<x_0/2$, $0<x_1<\epsilon/2$, and $b-\epsilon/2<x_{n-k_2+1}<b$

(3.2.10)

$$\tilde{d}^2(x_1,x_{k_1},x_{n-k_2+1})$$

$$\leq \int_{x_{k_1}}^{\epsilon/2} \left(y^{\frac{a}{2}}r_1^{\frac{1}{2}}(y)-(y-x_1)^{\frac{a}{2}}r_1^{\frac{1}{2}}(y-x_1)\right)^2 dy$$

$$+\int_{b-\epsilon/2}^{x_{n-k_2+1}} \left((b-y)^{\frac{a}{2}}r_2^{\frac{1}{2}}(b-y)-(b-y+x_1)^{\frac{a}{2}}r_2^{\frac{1}{2}}(b-y+x_1)\right)^2 dy$$

$$+\int_{\epsilon/2}^{b-\epsilon/2} \left(\tilde{f}^{1/2}(y)-\tilde{f}^{1/2}(y-x_1)\right)^2 dy$$

$$=: A_1(x_1,x_{k_1})+A_2(x_1,x_{n-k_2+1})+B(x_1).$$

Taking into account (3.2.6), we obtain similar results as we did in (2.2.11):

(3.2.11)

$$A_1(x_1,x_{k_1}) \leq x_1^{\frac{1+a}{2}}r_1^{\frac{1}{2}}(x_1)\big(g_1^{\frac{1}{2}}(x_1)+\psi^{\frac{1}{2}}(\frac{x_k}{x_1})\big)$$

$$A_2(x_1,x_{n-k_2+1}) \leq x_1^{\frac{1+a}{2}}r_2^{\frac{1}{2}}(x_1)\big(g_2^{\frac{1}{2}}(x_1)+\psi^{\frac{1}{2}}(1+\frac{b-x_{n-k_2+1}}{x_1})\big).$$

Applying the Cauchy-Schwarz inequality and combining (3.2.8) - (3.2.11), the proof is complete. ■

Denote the right-hand side of (3.2.7) by $D(n, k_1, k_2)$. Then

$$\Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) \leq D(n, k_1, k_2). \tag{3.2.12}$$

Notice that $D(n, k_1, k_2)$ is independent of t.

3.3 The Asymptotic Information Contained in the k_1 Lower and k_2 Upper Extremes

Denote by $\tilde{F}$ the distribution function of $\tilde{P}_0$. Since r_i is slowly varying, we already know that $\frac{1}{\tilde{F}^{-1}(1/n)} X_{1:n} \to S^1_{1,1}/(1+a)$ and $\frac{1}{\tilde{F}^{-1}(1-1/n)}(X_{n:n} - b) \to S^{1/(1+a)}_{2,1}$ weakly (or even in the strong sense) as $n \to \infty$. In general, $\tilde{F}^{-1}(1/n)$ is not the right normalizing sequence for the maximum, since the sequence $\left(\tilde{F}^{-1}(1/n)/\tilde{F}^{-1}(1-1/n)\right)_n$ does not necessarily converge (to some positive finite value). Therefore, we have to impose further assumptions. We claim that

$$r_1(x) \sim c r_2(x) \ (x \downarrow 0) \tag{3.3.1}$$

where $c \in (0, \infty)$.

Using the theory of regular variation, (3.3.1) implies

$$\tilde{F}^{-1}(1/n) \sim c^{-1} \tilde{F}^{-1}(1 - 1/n) \ (n \to \infty). \tag{3.3.2}$$

We define

$$\delta_n = \tilde{F}^{-1}(1/n). \tag{3.3.3}$$

Now, Slutsky's Theorem (see e.g. *Serfling* (1980), p. 19) states that $\delta_n^{-1}(X_{n:n} - b) \to cS^{1/(1+a)}_{2,1}$, $n \to \infty$.

Once again, we know from *Falk* (1985) and *Sweeting* (1985) that convergence of the extremes holds w.r.t. the variational distance, i.e.

$$\begin{aligned} &\|\mathcal{L}\big(\delta_n^{-1}(X_{1:n}, \ldots, X_{k_1:n} | \tilde{P}_0^n\big) - \mathcal{L}\big(S^{1/(1+a)}_{1,1}, \ldots, S^{1/(1+a)}_{1,k_1}\| \to 0 \\ &\|\mathcal{L}\big(\delta_n^{-1}(X_{n-k_2+1:n} - b, \ldots, X_{n:n} - b | \tilde{P}_0^n\big) - \mathcal{L}\big(cS^{1/(1+a)}_{2,k_2}, \ldots, cS^{1/(1+a)}_{2,1}\| \to 0 \end{aligned} \tag{3.3.4}$$

as $n \to \infty$.

In analogy to condition (2.3.4) and (2.3.4′), we assume that for some $\epsilon > 0$

$$r_i \text{ is bounded on } (0, \epsilon) \ \text{ and } \liminf_{x \downarrow 0} r_i(x) > 0, \tag{3.3.5}$$

or

$$r_i \text{ is decreasing on } (0, \epsilon) \tag{3.3.5′}$$

for i = 1,2.

In addition, we need conditions (cf. (3.27) in *Janssen* and *Reiss* (1988)) which ensure the convergence of $g_i(x)$ to zero as $x \downarrow 0$, i = 1,2.

First, we replace condition (2.3.5)(i) by

$$\int_{\epsilon}^{b-\epsilon} \left(\tilde{f}^{1/2}(x-t) - \tilde{f}^{1/2}(x)\right)^2 dx = o(t^{1+a} r(t)) \quad (t \downarrow 0) \tag{3.3.6(i)}$$

for each $\epsilon > 0$, where $r \in \{r_1, r_2\}$.

Under condition (3.3.1), we see that if (3.3.6) (i) holds for r_1 then it also holds for r_2 and vice versa. Condition (3.3.6)(i) says that the Hellinger distance is mainly determined by the local behaviour of the density $\tilde{f}$ at the singularities 0 and b.

Moreover, we assume that condition (2.3.5)(ii) is valid for r_1 and r_2. To be more precise, let

$$l_i(x)\tilde{r}_i(x) = l_i(x)exp\Big(\int_x^{x_0} \frac{b_i(u)}{u}\, du\Big)$$

be the *Karamata representation* of r_i, $i = 1, 2$. We assume that l_i is continuous on $[0, x_0]$ with $l_i(0) > 0$ and

$$\text{(3.3.6(ii))} \qquad \int_0^{x_0/2} \big(l_i^{1/2}(x+t) - l_i^{1/2}(x)\big)^2 x^a r_i(x)\, dx = o\big(t^{1+a} r(t)\big)$$

as $t \downarrow 0$.

Once again, Lemma 10.13 of *Janssen* and *Mason* (1989) implies that under condition (3.3.6)

$$\text{(3.3.7)} \qquad \lim_{x \downarrow 0} g_i(x) = 0$$

holds for i = 1,2.

Note that if different shape parameters occur in the representation (3.1.1), we are again in the situation of one singularity. Under the present assumptions, we see that only the singularity of higher order—that is the singularity with the smaller shape parameter—has a dominate influence.

Consider, for example, the generalized Pareto density $f(x) = (1 + a)x^a 1_{(0,1)}(x)$. We derive the representation

$$f(x) = \begin{cases} x^a r_1(x), & 0 < x < x_0 \\ (1-x)^0 r_2(1-x), & 1 - x_0 < x < 1 \end{cases}$$

for some appropriate x_0 with $r_1(x) = 1 + a$ and $r_2(x) = (1 + a)(1 - x)^a$. If $a \in (-1, 0)$ we have a pole at 0 and a jump at 1. In this case, the lower extremes are relevant. The *situation changes completely* for $a \in (0, 1)$. Only one singularity occurs, namely, a jump at the right endpoint. Hence, the maximum contains asymptotically all information. If $a = 0$ two jumps occur. In this case, it turns out that the statistic $(X_{1:n}, X_{n:n})$ is asymptotically sufficient.

Since $\tilde{P}_0$ has a compact support the condition

$$\text{(3.3.8)} \qquad \limsup_{n\to\infty} \int x^{1+a}\, d\mathcal{L}\big(\delta_n^{-1} X_{1:n} | \tilde{P}_0^n\big)(x) < \infty$$

trivially holds (cf. Lemma 2.3.1).

3.3.1 Theorem. *Let $\tilde{f}$ be a density of type (3.1.1) and assume that the conditions (3.3.5) ((3.3.5')) and (3.3.6) are fulfilled. There exists a constant $C > 0$ such that*

(3.3.9)
$$\lim_{n\to\infty} D(n,k_1,k_2) \le C\Big(\big(E\psi((\frac{S_{1,k_1}}{S_{1,1}})^{1/(1+a)})\big)^{1/2} + \big(E\psi(1+c(\frac{-S_{2,k_2}}{S_{1,1}})^{1/(1+a)})\big)^{1/2}\Big).$$

PROOF. As in the proof of Theorem 2.3.2, it is shown that

$$\limsup_{n\to\infty} (n-k_1-k_2)^{1/2}\tilde{I}_{j,n} < \infty \ for \ j \in \{1,5\}$$
$$\lim_{n\to\infty} \tilde{I}_{j,n} = 0 \ for \ j \in \{2,6\}$$
$$\lim_{n\to\infty} \tilde{I}^2_{3,n,k_1} \le E\psi((\frac{S_{1,k_1}}{S_{1,1}})^{1/(1+a)}).$$

The same arguments used for $\tilde{I}_{2,n}$ show

$$\lim_{n\to\infty} \tilde{I}^2_{4,n} = 0.$$

It remains to estimate the term $\tilde{I}_{7,n,k_2}$. We get

$$\tilde{I}^2_{7,n,k_2} \le E\psi(1+c(\frac{-S_{2,k_2}}{S_{1,1}})^{1/(1+a)}) + \|h\|_2^2 \, \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, b-X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}) \otimes \mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\|.$$

It is well known that the lower and upper extremes are asymptotically independent (see e.g. *Falk* and *Reiss* (1988)). This implies that

$$\|\mathcal{L}(\delta_n^{-1}(X_{1:n}, b-X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}) \otimes \mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\| \to 0 \quad (3.3.10)$$

as $n \to \infty$.

Summarizing the results above, the proof is complete. ∎

Now, we are able to show the global sufficiency of the $k_1(n)$ lower and $k_2(n)$ upper extremes.

3.3.2 Theorem (Global Sufficiency). *Suppose that the conditions of Theorem 3.3.1 are fullfilled.*

(i) *Let $a \in (-1,1)$, $a \ne 0$. There exists a constant $C > 0$ such that*

$$\Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) \le C\big(k_1^{(a-1)/(2(1+a))} + k_2^{(a-1)/(2(1+a))}\big) + o(n^0). \quad (3.3.11)$$

(ii) *Let $a = 0$. Then*

$$\Delta(\tilde{E}_n, \tilde{E}_{n,1,1}) = o(n^0). \quad (3.3.12)$$

PROOF. From the proof of Theorem 2.3.2, we already know that

$$E\psi\left(\left(\frac{S_{1,k_1}}{S_{1,1}}\right)^{1/(1+a)}\right) \leq Ck_1^{(a-1)/(1+a)}. \tag{3.3.13}$$

Thus in view of Theorem 3.3.1, it remains to be shown that

$$E\psi\left(1+c\left(\frac{-S_{2,k_2}}{S_{1,1}}\right)^{1/(1+a)}\right) \leq Ck_2^{(a-1)/(1+a)}. \tag{3.3.14}$$

Now an application of the inequality (2.3.16) yields

$$E\psi\left(1+c\left(\frac{-S_{2,k_2}}{S_{1,1}}\right)^{1/(1+a)}\right) \leq \frac{a^2c^{a-1}}{4(1+a)}ES_{1,1}^{(1-a)/(1+a)}E\psi\left(-S_{2,k_2}^{1/(1+a)}\right).$$

Since $-S_{2,k_2}$ is distributed according to a gamma distribution with parameter k_2 (thus having a density $x \to e^{-x}x^{k_2-1}/(k_2-1)!$, $x > 0$), it is not hard to verify (see Lemma 5.2 of *Janssen* and *Mason* (1989)) that

$$k_2^{-p} < E(-S_{2,k_2}^{-p}) \leq (k_2-1-k_0)^{-p} \tag{3.3.15}$$

for $k > p > 0$ where k_0 is defined by

$$k_0 = \begin{cases} p-1, & \text{for } p \in \mathbb{N} \\ [p], & \text{for } p \notin \mathbb{N} \end{cases}$$

and [] indicates the Gauss bracket.

Hence (3.3.14) is shown. ■

Remark. From Theorem (3.3.2) we deduce that

$$\lim_{n\to\infty} \Delta(\tilde{E}_n, \tilde{E}_{n,k_1(n),k_2(n)}) = 0 \tag{3.3.16}$$

whenever $min\{k_1(n), k_2(n)\} \to \infty$ as $n \to \infty$, i.e. the $k_1(n)$ smallest and $k_2(n)$ largest order statistics are ***asymptotically global sufficient.*** Moreover, in the case of jumps ($a = 0$) the statistic $(X_{1:n}, X_{n:n})$ is global sufficient. This generalizes the result of *Weiss* (1979). Finally, we remark that *Weiss* has also used the kernel $K_{X_{1:n}}^{(n,1,1)}$, though it is not explicit stated there.

3.4 Comparison of the Statistical Experiments $\tilde{E}_n$, $\tilde{E}_{n,k_1,k_2}$, $\tilde{G}_1 \otimes \tilde{G}_2$, and $\tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}$

We start with

3.4.1 Lemma. *Let $\tilde{f}$ be a density of type (3.1.1) and let $0 < \lambda < 1$. There exists a constant $C > 0$ such that*

(i) *for $a \in (-1,1)$, $a \neq 0$:*

(3.4.1)

$$\begin{aligned} D(n,k_1,k_2) &\le C\Big(k_1^{(a-1)/(2(1+a))} + k_2^{(a-1)/(2(1+a))} \\ &+ \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, X_{k_1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}, S_{1,k_1}^{1/(1+a)})\|^{1/2} \\ &+ \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, b - X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}) \otimes \mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\|^{1/2} \\ &+ \Big(\int_0^{\delta_n^{-1}\epsilon/2} g_1(\delta_n x_1)\, d\mathcal{L}(S_{1,1}^{1/(1+a)})(x_1)\Big)^{1/2} \\ &+ \Big(\int_0^{\delta_n^{-1}\epsilon/2} g_2(\delta_n x_1)\, d\mathcal{L}(S_{1,1}^{1/(1+a)})(x_1)\Big)^{1/2}\Big), \end{aligned}$$

(ii) *for $a = 0$:*

$$D(n,1,1) \le C\Big(\Big(\int_0^{\delta_n^{-1}\epsilon/2} g_1(\delta_n x_1)\, d\mathcal{L}(S_{1,1})(x_1)\Big)^{1/2} + \Big(\int_0^{\delta_n^{-1}\epsilon/2} g_2(\delta_n x_1)\, d\mathcal{L}(S_{1,1})(x_1)\Big)^{1/2}\Big), \tag{3.4.2}$$

for all $n \in \mathbb{N}$ and $max\{k_1, k_2\} \le \lambda n$.

PROOF. For $max\{k_1,k_2\} \le \lambda n$ we have

$$\begin{aligned} exp(-n(\tilde{F}(\epsilon) - k_1/n)^2)/3) &\le exp(-k_1(\tilde{F}(\epsilon) - \lambda)^2/(3\lambda)) \\ exp(-n(1 - \tilde{F}(b-\epsilon) - k_2/n)^2)/3) &\le exp(-k_2(1 - \tilde{F}(b-\epsilon) - \lambda)^2/(3\lambda)). \end{aligned}$$

Moreover

$$\int_0^{\epsilon/2} g_i(x_1) d\mathcal{L}(X_{1:n}|\tilde{P}_0^n)(x_1)$$

$$\leq C\Big(\int_0^{\delta_n^{-1}\epsilon/2} g_i(\delta_n x_1) d\mathcal{L}(S_{1,1}^{1/(1+a)})(x_1) + \|\mathcal{L}(\delta_n^{-1}X_{1:n}|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)})\|\Big)$$

for i = 1,2.

From Section 3.3, we already know

$$\int_{(0,\epsilon/2)} \psi(\frac{x_{k_1}}{x_1})\, d\mathcal{L}((X_{1:n}, X_{k_1:n})|\tilde{P}_0^n)$$
$$\leq C\Big(k^{(a-1)/(1+a)} + \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, X_{k_1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}, S_{1,k_1}^{1/(1+a)})\|\Big),$$

and

$$\int_{(0,\epsilon/2)} \psi(1 + \frac{b - x_{n-k_2+1}}{x_1})\, d\mathcal{L}((X_{1:n}, X_{n-k_2+1:n})|\tilde{P}_0^n)$$
$$\leq C\Big(k^{(a-1)/(1+a)} + \|\mathcal{L}(\delta_n^{-1}(X_{1:n}, b - X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}) \otimes \mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\|\Big)$$

where $C > 0$ is a generic constant.

Recall that $\psi \equiv 0$ if $a = 0$. In view of (3.3.4) and (3.3.10), the assertion follows from Theorem 3.2.1. ∎

In order to establish rates of convergence, we have to impose further assumptions:

$$r_i(x) = c_i exp(\tilde{h}_i(x)),\ 0 < x < x_0 \tag{3.4.3}$$

where

$$|\tilde{h}_i(x)| \leq L_i x^{\gamma_i} \tag{3.4.4}$$

for some constant $L_i > 0$ and $\gamma_i > 0$, i = 1,2.

Obviously, condition (3.4.3) implies condition (3.3.5) and $c = c_1/c_2$, where c is the constant of (3.3.1).

Let $\tilde{\delta}_n$ be defined as in (2.4.7) (with c_1 instead of c). Note that (3.3.4), (3.3.10), (3.4.1), and (3.4.2) hold if δ_n is replaced by $\tilde{\delta}_n$. Again, we first treat the case of $a \neq 0$.

3.4.2 Theorem. *Assume that f is of type (3.1.1) for $a \in (-1,1)$, $a \neq 0$. In addition to (3.4.3) it is assumed that f is absolutely continuous on $(0, x_0)$ and $(b - x_0, b)$ with*

$$\int_0^{x_0} \Big(\frac{r_i'(x)}{r_i(x)}\Big)^2 x^a < \infty \tag{3.4.5}$$

$i = 1,2$.

Then for every $\lambda \in (0,1)$ there exists a constant $C > 0$ such that for all $n \in \mathbb{N}$, $k_1 + k_2 \leq n$, and $max\{k_1, k_2\} \leq \lambda n$ the following inequality holds:

(3.4.6)

$$\Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) \leq C\bigg(\Big(\frac{k_2}{n(n-k_2-1)}\Big)^{1/4} + \sum_{i=1}^{2} n^{max\{a-1,-2\gamma_i\}/(2(1+a))} + \sum_{i=1}^{2}\Big\{k_i^{(a-1)/(2(1+a))} + \Big(\frac{k_i}{n}\Big)^{\gamma_i/(2(1+a))} k_i^{1/4} + \Big(\frac{k_i}{n}\Big)^{1/2}\Big\}\bigg).$$

Before proving Theorem 3.4.2, we recall the following results:

From *Reiss* (1989), Corollary 5.5.5, we know that under condition (3.4.3)

$$\|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{1:n}, \ldots, X_{k_1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)}, \ldots, S_{1,k_1}^{1/(1+a)})\| \leq C\bigg(\Big(\frac{k_1}{n}\Big)^{\gamma_1/(1+a)} k_1^{1/2} + \frac{k_1}{n}\bigg) \tag{3.4.7}$$

and

$$\|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{n-k_2+1:n} - b, \ldots, X_{n:n} - b)|\tilde{P}_0^n) - \mathcal{L}(cS_{2,k_2}^{1/(1+a)}, \ldots, cS_{2,1}^{1/(1+a)})\| \leq C\bigg(\Big(\frac{k_2}{n}\Big)^{\gamma_2/(1+a)} k_2^{1/2} + \frac{k_2}{n}\bigg) \tag{3.4.8}$$

where $C > 0$ is a universal constant (cf. (2.4.13)).

Obviously, (3.4.7) and (3.4.8) imply

$$\Delta(\tilde{E}_{n,k_1}^{(1)}, \tilde{G}_{1,k_1}) \leq C\bigg(\Big(\frac{k_1}{n}\Big)^{\gamma_1/(1+a)} k_1^{1/2} + \frac{k_1}{n}\bigg), \tag{3.4.9}$$

$$\Delta(\tilde{E}_{n,k_2}^{(2)}, \tilde{G}_{2,k_2}) \leq C\bigg(\Big(\frac{k_2}{n}\Big)^{\gamma_2/(1+a)} k_2^{1/2} + \frac{k_2}{n}\bigg), \tag{3.4.10}$$

and

$$\Delta(\tilde{E}_{n,k_1}^{(1)} \otimes \tilde{E}_{n,k_2}^{(2)}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) \leq C\bigg(\sum_{i=1}^{2}\Big\{\Big(\frac{k_i}{n}\Big)^{\gamma_i/(1+a)} k_i^{1/2} + \frac{k_i}{n}\Big\}\bigg). \tag{3.4.11}$$

Falk and *Reiss* (1988 a) established the rate at which sample extremes become independent (see also the paper of *Falk* and *Kohne* (1986)). They have shown that for arbitrary random variables $\xi_1, \xi_2, \ldots$

$$(3.4.12)\quad \begin{aligned} &\|\mathcal{L}((\xi_{1:n},\ldots,\xi_{k_1:n},\xi_{n-k_2+1:n},\ldots,\xi_{n:n})|P) \\ &\quad - \mathcal{L}((\xi_{1:n},\ldots,\xi_{k_1:n})|P)\otimes\mathcal{L}((\xi_{n-k_2+1:n},\ldots,\xi_{n:n})|P)\| \\ &\qquad\qquad \le C\left(\left(\frac{k_1k_2}{n(n-k_1-k_2)}\right)^{1/2}\right). \end{aligned}$$

Note that (3.4.12) implies

$$(3.4.13)\qquad \Delta(\tilde{E}_{n,k_1,k_2}, \tilde{E}^{(1)}_{n,k_1}\otimes\tilde{E}^{(2)}_{n,k_2}) \le C\left(\left(\frac{k_1k_2}{n(n-k_1-k_2)}\right)^{1/2}\right).$$

PROOF OF THEOREM 3.4.2. Throughout the proof, C denotes a generic constant. Similar to (2.4.10), one shows

$$g_i(x) \le Cx^{min\{1-a,2\gamma_i\}}.$$

Hence,

$$\int_0^{\tilde{\delta}_n^{-1}\epsilon/2} g_i(\tilde{\delta}_n x_1)\, d\mathcal{L}(S_{1,1}^{1/(1+a)})(x_1) \le Cn^{max\{a-1,-2\gamma_i\}}$$

for i = 1,2.

Since

$$\begin{aligned} &\|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{1:n}, b-X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(S_{1,1}^{1/(1+a)})\otimes\mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\| \\ &\le \|\mathcal{L}(\tilde{\delta}_n^{-1}X_{1:n}|\tilde{P}_0^n) - \mathcal{L}(S_{1,k_1}^{1/(1+a)})\| \\ &+ \|\mathcal{L}(\tilde{\delta}_n^{-1}(b-X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(-cS_{2,k_2}^{1/(1+a)})\| \\ &+ \|\mathcal{L}(\tilde{\delta}_n^{-1}(X_{1:n}, b-X_{n-k_2+1:n})|\tilde{P}_0^n) - \mathcal{L}(\tilde{\delta}_n^{-1}X_{1:n}|\tilde{P}_0^n)\otimes\mathcal{L}(\tilde{\delta}_n^{-1}(b-X_{n-k_2+1:n})|\tilde{P}_0^n)\| \end{aligned}$$

the proof follows from (3.4.7), (3.4.8), (3.4.12), and Lemma 3.4.1.■

REMARK. To establish the inequality (3.4.6), we have used the asymptotic independence of extremes. The disadvantage of this is that we are in a trade off situation: The dependence decreases and the information increases as $k_1(n)$ and $k_2(n)$ tends to infinity. But it turns out that the bound occuring in (3.4.13) has no dominate influence on the rate of convergence of $\Delta(\tilde{E}_n, \tilde{G}_1\otimes\tilde{G}_2)$, as the proof of Theorem 3.4.7 will show.

Now, we consider the case $a = 0$.

3.4.3 Theorem. *Let f be a density of type (3.1.1) with $a = 0$. In addition to (3.4.3), assume that f is absolutely continuous on $(0, x_0)$ and $(b - x_0, 0)$ with*

$$(3.4.14)\qquad \int_0^{x_0} |f'(x)|^{\eta_1}/f(x)^{\eta_1-1}\,dx + \int_{b-x_0}^{b} |f'(x)|^{\eta_2}/f(x)^{\eta_2-1}\,dx < \infty$$

for $\eta_i \in (1,2]$, $i = 1,2$.

Then for every $\lambda \in (0,1)$, *there exists a constant* $C > 0$ *such that for all* $n \in \mathbb{N}$, $k_1 + k_2 \le n$, *and* $max\{k_1, k_2\} \le \lambda n$ *the following inequality holds:*

$$\Delta(\tilde{E}_n, \tilde{E}_{n,1,1}) \le C\Big(n^{(1-\eta_1)/2} + n^{(1-\eta_2)/2}\Big) \tag{3.4.15}$$

PROOF. Similar to the proof of Theorem (4.1) in *Janssen* and *Reiss* (1988) one shows that

$$g_i(x) = O(x^{(\eta_i - 1)/2}) \tag{3.4.16}$$

for i = 1,2.

Taking into account Lemma 3.4.1 (ii) the proof can be easily completed. ∎

3.4.4 Theorem. *Let f be a density of Type (3.1.1) for* $a \in (-1,1)$ *and let condition (3.4.3) be valid. There exists a constant* $C > 0$ *such that*

$$\begin{aligned} &\Delta(\tilde{E}_{n,k_1,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) \\ &\le C\Big(\big(\frac{k_1 k_2}{n(n-k_1-k_2)}\big)^{1/2} + \sum_{i=1}^{2}\{(\frac{k_i}{n})^{\gamma_i/(1+a)} k_i^{1/2} + \frac{k_i}{n}\}\Big). \end{aligned} \tag{3.4.17}$$

PROOF. Since

$$\begin{aligned} &\Delta(\tilde{E}_{n,k_1,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) \\ &\quad \le \Delta(\tilde{E}_{n,k_1,k_2}, \tilde{E}^{(1)}_{n,k_1} \otimes \tilde{E}^{(2)}_{n,k_2}) + \Delta(\tilde{E}^{(1)}_{n,k_1} \otimes \tilde{E}^{(2)}_{n,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}), \end{aligned}$$

the assertion is immediate from (3.4.13) and (3.4.11). ∎

The corresponding result to Theorem 2.4.1 is the following one:

3.4.5 Theorem (Strong Convergence). *Suppose that the conditions of Theorem 3.4.4 are fulfilled. In addition, assume that condition (3.3.6) holds. Then*

$$\Delta(\tilde{E}_n, \tilde{G}_1 \otimes \tilde{G}_2) \to 0 \text{ as } n \to \infty \tag{3.4.18}$$

and

$$\Delta(\tilde{E}_{n,k_1(n),k_2(n)}, \tilde{G}_1 \otimes \tilde{G}_2) \to 0 \text{ as } n \to \infty \tag{3.4.19}$$

whenever $min\{k_1(n), k_2(n)\} \to \infty$ *as* $n \to \infty$.

PROOF. We use the arguments of *Janssen* and *Reiss* (1988). First, we show that $(\tilde{E}_n)_n$ is a Cauchy sequence w.r.t. Δ.

Let $\epsilon > 0$. In view of Theorem 3.3.2, we can choose k_1, k_2, and $n \in \mathbb{N}$ such that

$$\Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) \le \epsilon/3$$

for all $n > n_0$.

Since

$$\Delta(\tilde{E}_{n,k_1,k_2}, \tilde{E}_{m,k_1,k_2}) \le \Delta(\tilde{E}_{n,k_1,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) + \Delta(\tilde{E}_{m,k_1,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2})$$

Theorem 3.4.4 implies that we can choose $n_1 \ge n_0$ such that for all $n, m \ge n_1$

$$\Delta(\tilde{E}_{n,k_1,k_2}, \tilde{E}_{m,k_1,k_2}) \le \epsilon/3.$$

Hence

$$\Delta(\tilde{E}_n, \tilde{E}_m) \le \epsilon$$

whenever $n, m \ge n_1$.

In view of the completeness of the distance Δ, the sequence $(\tilde{E}_n)_n$ converges to some experiment F. So it remains to show $F \sim \tilde{G}_1 \otimes \tilde{G}_2$.

For this, we choose $k_1(n)$, $k_2(n)$ with $min\{k_1(n), k_2(n)\} \to \infty$ as $n \to \infty$, such that

$$\Delta(\tilde{E}_{n,k_1(n),k_2(n)}, \tilde{G}_{1,k_1(n)} \otimes \tilde{G}_{2,k_2(n)}) \to 0 \tag{3.4.20}$$

as $n \to \infty$. This is possible in view of Theorem 3.4.4.

In addition, we know from Theorem 3.3.2 that

$$\Delta(\tilde{E}_n, \tilde{E}_{n,k_1(n),k_2(n)}) \to 0 \tag{3.4.21}$$

as $n \to \infty$.

Moreover, *Janssen (1989 b) showed that* $\tilde{G}_{i,k_i} \to \tilde{G}_i$ *weakly if* $k_i \to \infty$, $i = 1,2$. *Hence,*

$$\tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2} \to \tilde{G}_1 \otimes \tilde{G}_2 \tag{3.4.22}$$

weakly as $min\{k_1, k_2\} \to \infty$. Thus (3.4.20) – (3.4.22) shows $\Delta(F, \tilde{G}_1 \otimes \tilde{G}_2) = 0$ and the assertion *(3.4.18) is proved.*

The assertion (3.4.19) is obvious from

$$\Delta(\tilde{E}_{n,k_1(n),k_2(n)}, \tilde{G}_1 \otimes \tilde{G}_2) \le \Delta(\tilde{E}_n, \tilde{E}_{n,k_1(n),k_2(n)}) + \Delta(\tilde{E}_n, \tilde{G}_1 \otimes \tilde{G}_2). \blacksquare$$

We are now in the proper position to establish

3.4.6 Theorem.

(i) *Let* $a \in (-1, 1)$, $a \ne 0$. *There exists a constant* $C > 0$ *such that*

$$\Delta(\tilde{G}_1 \otimes \tilde{G}_2, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) \le C\left(k_1^{(a-1)/(2(1+a))} + k_2^{(a-1)/(2(1+a))}\right). \tag{3.4.23}$$

(ii) *For* $a = 0$

$$\Delta(\tilde{G}_1 \otimes \tilde{G}_2, \tilde{G}_{1,1} \otimes \tilde{G}_{2,1}) = 0. \tag{3.4.24}$$

PROOF. First, observe that

$$\Delta(\tilde{G}_1 \otimes \tilde{G}_2, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}) \leq \Delta(\tilde{G}_1 \otimes \tilde{G}_2, \tilde{E}_n) + \Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) + \Delta(\tilde{E}_{n,k_1,k_2}, \tilde{G}_{1,k_1} \otimes \tilde{G}_{2,k_2}).$$

Applying Theorems 3.4.5, 3.3.2, and 3.4.4 and letting n tend to infinity, the assertion follows. ∎

The last theorem of this section concerns the rate of convergence of the deficiency between $\tilde{E}_n$ and $\tilde{G}_1 \otimes \tilde{G}_2$.

3.4.7 Theorem. *Let $\tilde{f}$ be a density of type (3.1.1) with $a \in (-1, 1)$ and let (3.4.3), (3.4.5), and (3.4.14) be fulfilled. Then*

(3.4.25)
$$\Delta(\tilde{E}_n, \tilde{G}_1 \otimes \tilde{G}_2) = O(n^{\beta(\tilde{\gamma},a)})$$

where $\tilde{\gamma} = min\{\gamma_1, \gamma_2\}$ and

$$\beta(\tilde{\gamma}, a) = \begin{cases} \frac{a-1}{2(3+a)} & \text{for } \tilde{\gamma} \geq \frac{2(1+a)}{1-a},\ a \neq 0 \\ \frac{\tilde{\gamma}(a-1)}{(1+\tilde{\gamma})4(1+a)} & \text{for } 0 < \tilde{\gamma} < \frac{2(1+a)}{1-a},\ a \neq 0 \\ max\{-\tilde{\gamma}, -\frac{1}{2}\} & \text{for } a = 0. \end{cases}$$

PROOF. Combining (3.4.6), (3.4.17), and (3.4.23), we obtain

$$\begin{aligned} \Delta(\tilde{E}_n, \tilde{G}_1 \otimes \tilde{G}_2) &\leq \Delta(\tilde{E}_n, \tilde{E}_{n,k_1,k_2}) + \Delta(\tilde{E}_{n,k_1,k_2}, \tilde{G}_{k_1} \otimes \tilde{G}_{k_2}) + \Delta(\tilde{G}_{k_1} \otimes \tilde{G}_{k_2}, \tilde{G}_1 \otimes \tilde{G}_2) \\ &\leq C\Big(\big(\frac{k_1 k_2}{n(n-k_1-k_2)}\big)^{1/4} + 2n^{max\{a-1,-2\tilde{\gamma}\}/(2(1+a))} \\ &\quad + \sum_{i=1}^{2}\{k_i^{(a-1)/(2(1+a))} + \big(\frac{k_i}{n}\big)^{\tilde{\gamma}/(2(1+a))} k_i^{1/4} + \big(\frac{k_i}{n}\big)^{1/2}\}\Big). \end{aligned}$$

First, we examine the case $a \neq 0$.

As in the proof of Theorem 2.4.6, we choose

$$k_1(n) = k_2(n) = [n^{\tilde{\gamma}/(1+\tilde{\gamma})}].$$

Elementary calculations yield

$$\frac{k_1(n)k_2(n)}{n(n-k_1(n)-k_2(n))} = O\left(n^{-2/(1+\tilde{\gamma})}\right).$$

Thus, by repeating the arguments of the proof of Theorem 2.4.6, we obtain the assertion for $a \neq 0$.

The case $a = 0$ is immediate from (3.4.15) and (3.4.24). Note that the minimum $X_{1:n}$ and the maximum $X_{n:n}$ become independent with rate n^{-1} (see (3.4.12)). ■

4 GAUSSIAN SEQUENCES OF STATISTICAL EXPERIMENTS

In this chapter, we study order statistics in the case of Gaussian sequences.

4.1 The Almost Regular LAN Case of Weibull Type Densities

In this section, we discuss the borderline case $a = 1$. In the following let $X_1, \ldots, X_n$ be i.i.d. random variables with common density $f_t(\cdot) = f(\cdot - t)$, $t \in \mathbb{R}$, where

$$f(x) = xr(x)1_{(0,\infty)}(x). \tag{4.1.1}$$

Thereby, r is a normalized slowly varying function at zero; that is

$$r(x) = c\,exp\Big(\int_x^{\delta_0} b(u)\,du/u\Big),\ 0 < x < \delta_0, \tag{4.1.2}$$

for some δ_0 and some constant $c > 0$. Here, b is a function such that $b(u) \to 0$ as $u \downarrow 0$.

In addition, it is assumed that f is absolutely continuous on $(0, \infty)$ such that for all $\delta > 0$

$$\int_\delta^\infty \frac{(f'(x))^2}{f(x)}\,dx < \infty. \tag{4.1.3}$$

Then f has finite Fisher information iff

$$\int_0^{\delta_0} \frac{r(x)}{x}\,dx < \infty. \tag{4.1.4}$$

To verify (4.1.4) note that $f'(x) = (1 - cb(x))r(x)$ λ-a.e.

In the following, we restrict our attention to the case of infinite Fisher information. *Janssen* and *Mason* (1989) have recently shown that under condition (4.1.3), the sequence

$$U_{\delta_n}E^n = \big(\mathbb{R}^n, \mathcal{B}^n, \{P^n_{\delta_n t} : t \in \mathbb{R}\}\big) \tag{4.1.5}$$

is *local asymptotic normal*, where δ_n satisfies

$$\frac{H^2(P_0, P_{\delta_n t})}{(\delta_n t)^2 h(\delta_n t)} \to 1, \qquad h(t) := \frac{1}{8}\int_t^{\delta_0} \frac{r(x)}{x}\,dx.$$

If, in addition, $r(x)$ is convergent to some positive constant $c > 0$ as $x \downarrow 0$, then

$$\delta_n \sim (cnlog(n)/2)^{-1/2} \tag{4.1.6}$$

(see also *Ibragimov* and *Has'minskii* (1981), Chapter II, Section 5). The rate δ_n lies between the rate in the non-regular case $(n^{-1/(1+a)}, -1 < a < 1)$ and the rate in the regular case $(n^{-1/2}$, finite Fisher information). Densities of type (4.1.1) with $\lim_{x\downarrow 0} r(x) = c > 0$ are almost smooth densities in the sense of *Ibragimov* and *Has'minskii* (1981).

Moreover, *Janssen* and *Mason* (1989) showed that the sequence

$$Z_n(X_1,\ldots,X_n)=\delta_n\sum_{i=1}^{n}-f'(X_i)/f(X_i) \tag{4.1.7}$$

is central, i.e. $Z_n(X_1,\ldots,X_n)\to N(0,1)$ weakly under P_0^n. The proof of this result is non-trivial, since the underlying statistical experiment is non-differentiable.

With $k=k(n)$ tending to infinity sufficiently fast, we shall see that the $k(n)$ smallest order statistics are asymptotically sufficient. Recall, that in the non-regular case $k(n)$ is allowed to tend to infinity at any rate.

4.1.1 Theorem. *Let f be a density of type (4.1.1). We assume that $-f'/f$ is non-decreasing in a neighborhood of 0 and $\lim_{x\downarrow 0} r(x)=c\in(0,\infty)$. Let $k(n)$ be a sequence of positive integers, such that $1\le k(n)\le n+1-k(n)\le n$ and $k(n)\to\infty$ and $k(n)/n\to 0$ as $n\to\infty$.*

(i) *If $\lim_{n\to\infty} log(k(n))/log(n)=1$ then the $k(n)$ lower intermediate and extreme order statistics*

$$(X_{1:n},\ldots,X_{k(n):n})$$

are asymptotically sufficient uniformly on compact sets.

(ii) *If $\lim_{n\to\infty} log(k(n))/log(n)=0$ then the intermediate and central order statistics*

$$(X_{k(n)+1:n},\ldots,X_{n-k(n):n})$$

are asymptotically sufficient uniformly on compact sets.

PROOF. Ad (i): Let $F=F_0$ be the distribution function of P_0. First, we show

$$\delta_n\sum_{i=1}^{n}\Big(-\frac{f'(X_i)}{f(X_i)}1_{(0,\beta)}(X_i)-m_\beta\Big)\xrightarrow{P_0^n}N(0,1)$$

for some $\beta>0$, where $m_\beta=E_{P_0^n}\big(-\frac{f'(X_1)}{f(X_1)}1_{(0,\beta)}(X_1)\big)$.

Since

$$E_{P_0^n}\Big(\frac{f'(X_1)}{f(X_1)}\Big)=0$$

we get

$$E_{P_0^n}\Big(-\frac{f'(X_1)}{f(X_1)}1_{(\beta,\infty)}(X_1)\Big)=-m_\beta.$$

An application of the Tschebyschev inequality yields

$$\begin{aligned}
&P_0^n\Big\{\Big|\delta_n\sum_{i=1}^{n}\Big(-\frac{f'(X_i)}{f(X_i)}1_{(\beta,\infty)}(X_i)+m_\beta\Big)\Big|\ge\epsilon\Big\}\\
&\qquad\le\Big(\frac{\delta_n}{\epsilon}\Big)^2 Var_{P_0^n}\Big(\delta_n\sum_{i=1}^{n}\frac{f'(X_i)}{f(X_i)}1_{(\beta,\infty)}(X_i)\Big)\\
&\qquad=\Big(\frac{\delta_n}{\epsilon}\Big)^2 n\Big(\int_\beta^\infty\frac{\big(f'(x)\big)^2}{f(x)}\,dx-m_\beta^2\Big)\to 0
\end{aligned}$$

as $n\to\infty$. Recall that $\delta_n=o(n^{-1/2})$.

Denote by $\tilde{F}$ the distribution function of

$$\tilde{X}_1 := -\frac{f'(X_1)}{f(X_1)}1_{(0,\beta)}(X_1) = \frac{b(X_1)-1}{X_1}1_{(0,\beta)}(X_1).$$

For the remainder of the proof, we choose β such that $-f'/f$ is non-decreasing on $(0,\beta)$. Then

$$\tilde{X}_{i:n} = -\frac{f'(X_{i:n})}{f(X_{i:n})}1_{(0,\beta)}(X_{i:n}), \qquad i= 1, ..., n.$$

From the above considerations, we know that $\tilde{F}$ is in the domain of attraction of a stable law of index 2. Applying Theorem 1 of *S. Csörgő et al.* (1986) we can find sequences $A_n(k(n))$ and $C_n(k(n))$ such that

$$A_n(k(n))\left\{\sum_{i=k(n)+1}^{n-k(n)} \tilde{X}_{i:n} - C_n(k(n))\right\} \longrightarrow N(0,1).$$

Furthermore,

$$C_n(k(n)) = n\int_{k(n)/n}^{1-k(n)/n} \tilde{F}^{-1}(u)\,du$$

and

$$A_n(k(n)) = \frac{1}{n^{1/2}\sigma(k(n)/n)}$$

where

$$\sigma^2(s) = \int_s^{1-s}\int_s^{1-s} \big(min(u,v) - uv\big)\; d\tilde{F}^{-1}(u)d\tilde{F}^{-1}(v)$$

denotes the truncated variance function of $\tilde{F}$. Now, for some constants d_n, we get

$$\delta_n\left\{\sum_{i=1}^{k(n)} \tilde{X}_{i:n} - d_n\right\}$$

$$= \delta_n\sum_{i=1}^{n} \tilde{X}_{i:n} - \frac{\delta_n}{A_n(k(n))}A_n(k(n))\left\{\sum_{i=k(n)+1}^{n-k(n)} \tilde{X}_{i:n} - C_n(k(n))\right\} - \delta_n \sum_{i=n-k(n)+1}^{n} \tilde{X}_{i:n}.$$

It follows from the exponential bound for order statistics (*Reiss* (1989), Lemma 3.1.1) that

$$P_0^n\left\{\left|\delta_n \sum_{i=n-k(n)+1}^{n} \tilde{X}_{i:n}\right| \geq \epsilon\right\} \leq P_0^n\{X_{n-k(n)+1:n} < \beta\} \underset{n\to\infty}{\longrightarrow} 0.$$

Combining the above results, we conclude that

$$\delta_n\left\{\sum_{i=1}^{k(n)} \tilde{X}_{i:n} - d_n\right\}$$

is central if $\delta_n = o(A_n(k(n)))$.

A central sequence is known to be asymptotically sufficient. This holds uniform on compact sets if, in addition, the sequence $U_{\delta_n}E^n$ is equicontinuous (*Strasser* (1985 a), Corollary 81.5). But the equicontinuity follows from Lemma 2.1 of *Janssen* (1986), since the sequence $U_{\delta_n}E^n$ is translation invariant and the limit experiment is continuous.

To prove Theorem 4.1.1 (i), it remains to show that $\delta_n = o(A_n(k(n))$ is equivalent to $\lim_{n\to\infty} log(k(n))/log(n) = 1$. Since $\tilde{X}_i$ is in the domain of attraction of the normal law, we know from *M. Csörgő et al.* (1986) that the following two conditions are satisfied:

$$\sigma^2 \text{ is slowly varying at zero} \tag{4.1.8}$$

and

$$\lim_{u\downarrow 0} u\{(\tilde{F}^{-1}(u))^2 + (\tilde{F}^{-1}(1-u))^2\}/\sigma^2(u) = 0. \tag{4.1.9}$$

Moreover, they proved that whenever (4.1.8) and (4.1.9) hold, we have

$$\tau^2(s) \sim \sigma^2(s) \qquad (s \downarrow 0)$$

where

$$\tau^2(s) = \int_s^{1-s} (\tilde{F}^{-1}(u))^2\, du - \Big(\int_s^{1-s} \tilde{F}^{-1}(u)\, du\Big)^2. \tag{4.1.10}$$

Since $|b(x)| \to 0$ as $x \to 0$, for every $\epsilon > 0$ we can find $\beta = \beta(\epsilon)$ such that

$$-(1+\epsilon)\frac{1}{x} \le \frac{b(x)-1}{x} = -\frac{f'(x)}{f(x)} \le -(1-\epsilon)\frac{1}{x}$$

for $x \in (0,\beta)$. Denote by $\tilde{\tilde{F}}$ the distribution function of $\tilde{\tilde{X}}_1 = (-1/X_1)1_{(0,\beta)}(X_1)$. Then

$$\begin{aligned}
&(1-\epsilon)^2 \int_s^{1-s} \big(\tilde{\tilde{F}}^{-1}(u)\big)^2 du \le \int_s^{1-s} \big(\tilde{F}^{-1}(u)\big)^2 du \le (1+\epsilon)^2 \int_s^{1-s} \big(\tilde{\tilde{F}}^{-1}(u)\big)^2 du\\
&(1-\epsilon)^2\Big(\int_s^{1-s} \tilde{\tilde{F}}^{-1}(u)\, du\Big)^2 \le \Big(\int_s^{1-s} \tilde{F}^{-1}(u)\, du\Big)^2 \le (1+\epsilon)^2\Big(\int_s^{1-s} \tilde{\tilde{F}}^{-1}(u)\, du\Big)^2.
\end{aligned} \tag{4.1.11}$$

Elementary calculations yield

$$\tilde{\tilde{F}}(x) = \begin{cases} F(-1/x), & \text{for } x \le -1/\beta \\ F(\beta), & \text{for } x \in (-1/\beta, 0) \\ 1, & \text{for } x \ge 0 \end{cases}$$

and

$$\tilde{\tilde{F}}^{-1}(u) = \begin{cases} -\frac{1}{F^{-1}(u)}, & \text{for } u \in (0, F(\beta)] \\ 0, & \text{for } u \in (F(\beta), 1). \end{cases}$$

Using the quantile transformation, we get

$$\begin{aligned} \int_s^{1-s} \left(\tilde{\tilde{F}}^{-1}(u)\right)^2 du &= \int_s^{F(\beta)} \left(\tilde{\tilde{F}}^{-1}(u)\right)^2 du \\ &= \int_{\tilde{\tilde{F}}^{-1}(s)}^{\tilde{\tilde{F}}^{-1}(F(\beta))} u^2 \, d\tilde{\tilde{F}}(u) \\ &= \int_{-1/F^{-1}(s)}^{-1/\beta} u^2 \, d\tilde{\tilde{F}}(u) \\ &= \int_{1/\beta}^{1/F^{-1}(s)} u^2 \frac{1}{u^3} r(\frac{1}{u}) \, du \\ &= \int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} r(\frac{1}{u}) \, du. \end{aligned}$$

From

$$\int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} r(\frac{1}{u}) \, du \sim c \int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} \, du$$

(apply Lemma 1.2.1 (a) of *de Haan* (1970)) we conclude

$$\int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} r(\frac{1}{u}) \, du \sim c(-log(F^{-1}(s)) + log(\beta)).$$

From the theory of regular variation it is known that $F^{-1}(s) \sim s^{1/2} L(s)$, $s \downarrow 0$, for some slowly varying function L (*Bingham et al.* (1987), Theorem 1.5.12). Hence, we see that

$$\int_s^{1-s} \left(\tilde{\tilde{F}}^{-1}(u)\right)^2 du \sim -\frac{c}{2}(log(s) \qquad (4.1.12)$$

holds.

Next we show

$$(4.1.13) \qquad \Big(\int_s^{1-s} \tilde{\tilde{F}}^{-1}(u)\,du\Big)^2 = o(log(s)) \quad (s \downarrow 0).$$

Using the same arguments as above, we obtain

$$\Big(\int_s^{1-s} \tilde{\tilde{F}}^{-1}(u)\,du\Big)^2 = \Big(\int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u^2} r\big(\frac{1}{u}\big)\,du\Big)^2$$

$$\sim \Big(c \int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u^2}\,du\Big)^2$$

$$= c^2(F^{-1}(s) - \beta)^2 = o(log(s)).$$

Combining (4.1.10) – (4.1.13), we obtain

$$(1-\epsilon)^2 \le \liminf_{s\downarrow 0} \frac{\tau^2(s)}{-\frac{c}{2}log(s)} \le \limsup_{s\downarrow 0} \frac{\tau^2(s)}{-\frac{c}{2}log(s)} \le (1+\epsilon)^2.$$

From this we conclude that $\delta_n/A_n(k(n)) \to 0$ iff

$$\frac{n^{1/2}log^{1/2}(k(n)/n)}{n^{1/2}log^{1/2}(n)} = \Big(1 - \frac{log(k(n))}{log(n)}\Big)^{1/2} \to 0.$$

Ad (ii): From *S. Csörgő* and *Mason* (1986), we know that

$$A'_n(k(n))\Big\{\sum_{i=1}^{k(n)} \tilde{X}_{i:n} - C'_n(k(n))\Big\} \longrightarrow N(0,1),$$

where

$$C'_n(k(n)) = n\int_0^{k(n)/n} \tilde{F}^{-1}(u)\,du$$

and

$$A'_n(k(n)) = n^{-1/2}\Big(\int_{1/n}^{k(n)/n} (\tilde{F}^{-1}(u))^2\,du\Big)^{-1/2}.$$

Note that $\tilde{F}$ is regularly varying (at $-\infty$) with index -2. For some constants d'_n we get

$$\delta_n\Big\{\sum_{i=k(n)+1}^{n-k(n)} \tilde{X}_{i:n} - d'_n\Big\}$$

$$= \delta_n \sum_{i=1}^{n} \tilde{X}_{i:n} - \frac{\delta_n}{A'_n(k(n))} A'_n(k(n))\Big\{\sum_{i=1}^{k(n)} \tilde{X}_{i:n} - C'_n(k(n))\Big\} - \delta_n \sum_{i=n-k(n)+1}^{n} \tilde{X}_{i:n}.$$

Hence,

$$\delta_n \Big\{ \sum_{i=k(n)+1}^{n-k(n)} \tilde{X}_{i:n} - d'_n \Big\}$$

is central if $\delta_n / A'_n(k(n)) \to 0$.

From the first part of the proof, we already know that

$$\begin{aligned} A'_n(k(n)) &\sim n^{-1/2} \Big(\int_{1/n}^{k(n)/n} (\tilde{\bar{F}}^{-1}(u))^2 \, du \Big)^{-1/2} \\ &\sim n^{-1/2} \big(\frac{c}{2} (-log(n/k(n)) + log(n)) \big)^{-1/2} \\ &= \big(\frac{c}{2} n log(k(n)) \big)^{-1/2}. \end{aligned}$$

Thus

$$\frac{\delta_n}{A'_n(k(n))} \sim \Big(\frac{log(k(n))}{log(n)} \Big)^{1/2}.$$

The proof is complete. ■

An example for a sequence $k(n)$ which satisfies the condition of Theorem 4.1.1 (i) is $k(n) = n^{1-\alpha_n}$ with $\alpha_n \to 0$ and $\alpha_n log(n) \to \infty$.

The simple reason that a fixed number of extremes does not contain any information is that they have the wrong rate of convergence: The rate of the normalizing constants of extremes (which ensure a non-degenerate limit law) does not coincide with the rate of local alternatives of the underlying experiment. Recall, that in the non-regular case these rates are equal. To highlight this point consider the following example:

Let f be the Weibull density; that is $r(x) = 2exp(-x^2)$. In this case we have $\delta_n = (nlog(n))^{-1/2}$. For notational simplicity let again $W_{n,k_1} = (X_{1:n}, \ldots, X_{k_1:n})$ and $Z_{n,k_2} = (X_{n-k_2+1:n}, \ldots, X_{n:n})$. Then

$$(4.1.14) \qquad (\mathbb{R}^{k_1+k_2}, \mathcal{B}^{k_1+k_2}, \{\mathcal{L}((W_{n,k_1}, Z_{n,k_2}) | P^n_{\delta_n t}) : t \in \mathbb{R}\}) \longrightarrow E_0$$

where $E_0 = (\mathbb{R}, \mathcal{B}, \{\epsilon_0\})$ denotes the totally uninformative experiment.

To verify (4.1.14), we have to show that the log-likelihood process of the binary experiment $(\mathcal{L}((W_{n,k_1}, Z_{n,k_2}) | P^n_0), \mathcal{L}((W_{n,k_1}, Z_{n,k_2}) | P^n_{\delta_n t}))$ with base 0 converges weakly (on $[-\infty, \infty]$) to ϵ_0 as $n \to \infty$. In order to show this denote by F the Weibull distribution function and let $m = n - k_1 - k_2$.

First, since $F^{-1}(1/n) \sim n^{-1/2}$ we obtain $n^{1/2}X_{i:n} \to F_i$, where F_i is the distribution function $F_i(x) = F(x)\sum_{l=1}^{i-1}(-log(F(x))^l/l!$, $x > 0$. Moreover, note that the upper extremes belong to the domain of attraction of the Gumbel distribution. We know that we can choose the normalizing constants $b_n = F^{-1}(1-1/n)$ and $a_n = (nf(b_n))^{-1}$ (see e.g. *David* (1981), p. 262 f.). We obtain $b_n = log^{1/2}(n)$ and $a_n = (2log^{1/2}(n))^{-1}$.

At this point, straightforward calculations yield:

$$\frac{d\mathcal{L}((W_{n,k_1}, Z_{n,k_2})|P^n_{\delta_n t})}{d\mathcal{L}((W_{n,k_1}, Z_{n,k_2})|P^n_0)}$$

$$= \prod_{i=1}^{k_1} \frac{f(X_{i:n}-\delta_n t)}{f(X_{i:n})} \prod_{j=1}^{k_2} \frac{f(X_{n-j+1:n}-\delta_n t)}{f(X_{n-j+1:n})} \times \Big(\frac{(F(X_{n-k_2+1:n}-\delta_n t) - F(X_{k_1:n}-\delta_n t)}{F(X_{n-k_2+1:n}) - F(X_{k_1:n})}\Big)^m$$

$$= \prod_{i=1}^{k_1} \frac{X_{i:n}-\delta_n t}{X_{i:n}} \frac{exp(-(X_{i:n}-\delta_n t)^2)}{exp(-X_{i:n}^2)} \prod_{j=1}^{k_2} \frac{X_{n-j+1:n}-\delta_n t}{X_{n-j+1:n}} \frac{exp(-(X_{n-j+1:n}-\delta_n t)^2)}{exp(-X_{n-j+1:n}^2)}$$

$$\times \left(\frac{exp(-(X_{k_1:n}-\delta_n t)^2) - exp(-(X_{n-k_2+1:n}-\delta_n t)^2)}{exp(-X_{k_1:n}^2) - exp(-X_{n-k_2+1:n}^2)}\right)^m$$

$$= \prod_{i=1}^{k_1} \frac{X_{i:n}-\delta_n t}{X_{i:n}} exp(2\delta_n t X_{i:n} - \delta_n^2 t^2) \prod_{j=1}^{k_2} \frac{X_{n-j+1:n}-\delta_n t}{X_{n-j+1:n}} exp(2\delta_n t X_{n-j+1:n} - \delta_n^2 t^2)$$

$$\times \left(exp(-(X_{k_1:n}-\delta_n t)^2 + X_{k_1:n}^2)\right)^m$$

$$\times \frac{\left(1 - \frac{1}{n} exp(log(n) + (X_{k_1:n}-\delta_n t)^2 - (X_{n-k_2+1:n}-\delta_n t)^2)\right)^m}{\left(1 - \frac{1}{n} exp(log(n) + X_{k_1:n}^2 - X_{n-k_2+1:n}^2)\right)^m}$$

$$= \prod_{i=1}^{k_1} \frac{X_{i:n}-\delta_n t}{X_{i:n}} exp(2\delta_n t X_{i:n} - \delta_n^2 t^2) \prod_{j=1}^{k_2} \frac{X_{n-j+1:n}-\delta_n t}{X_{n-j+1:n}} exp(2\delta_n t X_{n-j+1:n} - \delta_n^2 t^2)$$

$$\times exp(2m\delta_n t X_{k_1:n} - m(\delta_n t)^2)$$

$$\times \frac{\left(1 - \frac{1}{n} exp(-a_n^{-1}(X_{n-k_2+1:n} - b_n)a_n(X_{n-k_2+1:n} + b_n) + o_{P_0^n}(n^0))\right)^m}{\left(1 - \frac{1}{n} exp(-a_n^{-1}(X_{n-k_2+1:n} - b_n)a_n(X_{n-k_2+1:n} + b_n))\right)^m}.$$

We have

$$\frac{X_{i:n}-\delta_n t}{X_{i:n}} = 1 - \frac{n^{1/2}\delta_n t}{n^{1/2}X_{i:n}} \to 1,$$

$$\frac{X_{n-j+1:n}-\delta_n t}{X_{n-j+1:n}} = 1 - \frac{a_n^{-1}\delta_n t}{a_n^{-1}(X_{n-j+1:n} - b_n) + a_n^{-1}b_n} \to 1,$$

$$\delta_n X_{n-j+1:n} = (\delta_n a_n)a_n^{-1}(X_{n-j+1:n} - b_n) + \delta_n b_n \to 0$$

and

$$a_n(X_{n-k_2+1} + b_n) \longrightarrow 1$$

under P_0^n as $n \to \infty$.

Combining these results, we see that

$$\frac{d\,\mathcal{L}((W_{n,k_1}, Z_{n,k_2})|P^n_{\delta_n t})}{d\,\mathcal{L}((W_{n,k_1}, Z_{n,k_2})|P^n_0)} \to 1$$

holds.

REMARK. For $a > 1$ the Weibull distribution has finite *Fisher information*. We treat this case in a more general context in the next section.

4.2 The Regular LAN Case

In this section, we show that under the regular LAN condition (L_2-differentiability) a fixed number of extremes does not contain any information. Though this result is intuitively clear it is not stated in literature. Nevertheless, a mathematically rigorous proof is given.

First, let us recall some facts concerning ***differentiable curves***. Suppose $(\Omega, \mathcal{A}, \nu)$ is a σ-finite measure space and let $\mathcal{P}$ be the set of all probability measures on $(\Omega, \mathcal{A})$ which satisfy $P \ll \nu$. The set $\mathcal{P}$ is identified with a subset of $L_2(\Omega, \mathcal{A}, \nu)$ by $P \longrightarrow \left(\frac{dP}{d\nu}\right)^{1/2}$, $P \in \mathcal{P}$.

Let $\epsilon > 0$. A curve $t \to P_t$ from $(-\epsilon, \epsilon)$ to $\mathcal{P}$ is *differentiable* at $t = 0$ (or at P_0) if the map

$$t \longrightarrow \left(\frac{dP_t}{d\nu}\right)^{1/2}, \qquad t \in (-\epsilon, \epsilon)$$

is differentiable at $t = 0$. If $t \to P_t$ is differentiable at $t = 0$ then the derivative is of the form $t \to \frac{t}{2}h$, where $h \in L_2(\Omega, \mathcal{A}, \nu)$.

Moreover, h is a derivative if and only if

$$h = g\sqrt{\frac{dP}{d\nu}}$$

for some $g \in L_2(P_0)$ and

$$\int g \, dP_0 = 0$$

(e.g. *Strasser* (1985 a), Theorem 75.2).

The element g is called a *tangent vector* at P_0. The differentiability of a path implies

$$P_t(N_t) = o(t^2) \tag{4.2.1}$$

(*Strasser* (1985 a), Lemma 75.7), where N_t denotes the singularity part of the Lebesgue decomposition of P_t w.r.t. P_0.

Moreover, we derive the expansion

$$\left(\frac{dP_t}{dP_0}\right)^{1/2} = 1 + \frac{t}{2}g + tr_t \tag{4.2.2}$$

with

$$\int r_t^2 \, dP_0 = o(t^o). \tag{4.2.3}$$

A path which admits an expansion (4.2.2) is called *Hellinger differentiable*. The Hellinger differentiability is closely related to various other differentiable concepts (*DDC*-differentiable, weak differentiable, see the book of *Pfanzagl* and *Wefelmeyer* (1985)).

Now suppose that $t \to P_t$ is differentiable at P_0 with tangent vector g. Then the log-likelihood process admits the fundamental expansion (*Strasser* (1985 a), Theorem 75.8)

$$\text{(4.2.4)} \qquad log \frac{dP^n_{n^{-1/2}t}}{dP^n_0}(X_1, \ldots, X_n) = \frac{t}{\sqrt{n}} \sum_{i=1}^{n} g(X_i) - \frac{t^2}{2} \|g\|^2_{P_0} + o_{P_0^n}(n^o),$$

i.e. the rescaled experiment

$$E_n = \left(\mathbb{R}^n, \mathcal{B}^n, \{P^n_{n^{-1/2}t} : t \in \mathbb{R}\}\right)$$

converges weakly to the Gaussian shift

$$\left(\mathbb{R}, \mathcal{B}, \{\mathcal{N}(t\|g\|^2_{L_2(P_0)}, \|g\|^2_{L_2(P_0)}) : t \in \mathbb{R}\}\right).$$

The following theorem states that the extreme order statistics asymptotically contain no information. Denote by $E_0 = (\mathbb{R}, \mathcal{B}, \{\epsilon_0\})$ the totally uninformative experiment. Let F_t be the distribution function of P_t, $W_{n,k} = (X_{1:n}, \ldots, X_{k:n})$, and $Z_{n,k} = (X_{n-k+1:n}, \ldots, X_{n:n})$.

4.2.1 Theorem. *Let X_i, $1 \le i \le n$, be i.i.d. random variables with common distribution P_t, $t \in \mathbb{R}$. We assume that P_t has a Lebesgue density f_t and that $t \to P_t$ is differentiable at $t = 0$ with tanget vector g. Then*

$$\text{(4.2.5)} \qquad \begin{aligned} W_{n,k}E_n &= \left(\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(W_{n,k}|P^n_{n^{-1/2}t}) : t \in \mathbb{R}\}\right) \longrightarrow E_0 \\ Z_{n,k}E_n &= \left(\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(Z_{n,k}|P^n_{n^{-1/2}t}) : t \in \mathbb{R}\}\right) \longrightarrow E_0 \end{aligned}$$

for $n \to \infty$.

Proof. Since *Gaussian* shifts are homogeneous, the sequence $(\mathbb{R}^n, \mathcal{B}^n, \{P^n_{n^{-1/2}t} : t \in \mathbb{R}\})$ is contiguous. This implies the contiguity of $W_{n,k}E_n$ and $Z_{n,k}E_n$. Hence, (4.2.5) is proved, if

$$log \frac{d\mathcal{L}(W_{n,k}|P^n_{n^{-1/2}t})}{d\mathcal{L}(W_{n,k}|P^n_0)} \xrightarrow{\mathcal{L}(W_{n,k}|P^n_0)} 0$$

$$log \frac{d\mathcal{L}(Z_{n,k}|P^n_{n^{-1/2}t})}{d\mathcal{L}(Z_{n,k}|P^n_0)} \xrightarrow{\mathcal{L}(Z_{n,k}|P^n_0)} 0$$

which is equivalent to

$$\text{(4.2.6)} \qquad \begin{aligned} &\sum_{i=1}^{k} log \frac{dP_{n^{-1/2}t}}{dP_0}(X_{i:n}) \; + \; (n-k) log \frac{1 - F_{n^{-1/2}t}(X_{k:n})}{1 - F_0(X_{k:n})} \xrightarrow{P_0^n} 0 \\ &\sum_{i=1}^{k} log \frac{dP_{n^{-1/2}t}}{dP_0}(X_{n-i+1:n}) \; + \; (n-k) log \frac{F_{n^{-1/2}t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})} \xrightarrow{P_0^n} 0. \end{aligned}$$

First, we know that

$$g_{ni}(X_1, \ldots, X_n) = 2\left(\sqrt{\frac{dP_{n^{-1/2}t}}{dP_0}(X_i)} - 1\right)$$

is a *Lindeberg array* in the sense of *Strasser* (1985 a); that is

$$\lim_{n\to\infty} n \int_{\{|g_{ni}|>\epsilon\}} g_{ni}^2 \, dP_0^n = 0.$$

Hence, as $n \to \infty$,

$$(4.2.7) \qquad \begin{aligned} g_{n,1:n} &:= \min_{1\leq i\leq n} g_{ni}(X_1,\dots,X_n) \xrightarrow{P_0^n} 0 \\ g_{n,n:n} &:= \max_{1\leq i\leq n} g_{ni}(X_1,\dots,X_n) \xrightarrow{P_0^n} 0. \end{aligned}$$

Since

$$log\frac{dP_{n^{-1/2}t}}{dP_0}(X_i) = 2log(\frac{1}{2}g_{ni}+1),$$

we get

$$2log(\frac{1}{2}g_{n,1:n}+1) \leq log\frac{dP_{n^{-1/2}t}}{dP_0}(X_i) \leq 2log(\frac{1}{2}g_{n,n:n}+1).$$

Taking account of (4.2.7)

$$log\frac{dP_{n^{-1/2}t}}{dP_0}(X_{j:n}) \xrightarrow{P_0^n} 0, \; j = 1,...,n.$$

Next, we treat the term $(n-k)log\big(F_{n^{-1/2}t}(X_{n-k+1:n})/F_0(X_{n-k+1:n})\big)$. The Taylor expansion of $x \to log(1-x)$, $|x|<1$, yields

$$(n-1)log\frac{F_{n^{-1/2}t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})} = -(n-1)\sum_{j\geq 1}\frac{1}{j}\Big(\frac{F_0(X_{n-k+1:n})-F_{n^{-1/2}t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})}\Big)^j.$$

Moreover,

$$\begin{aligned} &n\big(F_0(X_{n-k+1:n}) - F_{n^{-1/2}t}(X_{n-k+1:n})\big) \\ &= n\int_{X_{n-k+1:n}}^{\omega(P_0)} (f_{n^{-1/2}t} - f_0)\, d\lambda \\ &= n\int_{X_{n-k+1:n}}^{\omega(P_0)} \big(\frac{dP_{n^{-1/2}t}}{dP_0} - 1\big)\, dP_0 + nP_{n^{-1/2}t}\big((X_{n-k+1:n},\omega(P_0))\cap N_{n^{-1/2}t}\big). \end{aligned}$$

From (4.2.1), we already know that

$$\lim_{n\to\infty} nP_{n^{-1/2}t}\big((X_{n-k+1:n},\omega(P_0))\cap N_{n^{-1/2}t}\big) = 0.$$

Paying attention to (4.2.2), we obtain

$$n\int_{X_{n-k+1:n}}^{\omega(P_0)}\Big(\sqrt{\frac{dP_{n^{-1/2}t}}{P_0}}-1\Big)\Big(\sqrt{\frac{dP_{n^{-1/2}t}}{P_0}}+1\Big)\,dP_0$$
$$= n\int_{X_{n-k+1:n}}^{\omega(P_0)}\Big(\frac{n^{-1/2}t}{2}g+n^{-1/2}tr_{n^{-1/2}t}\Big)\Big(2+\frac{n^{-1/2}t}{2}g+n^{-1/2}tr_{n^{-1/2}t}\Big)\,dP_0$$
$$= n^{1/2}t\int_{X_{n-k+1:n}}^{\omega(P_0)} g\,dP_0+2n^{1/2}t\int_{X_{n-k+1:n}}^{\omega(P_0)} r_{n^{-1/2}t}\,dP_0+o_{P_0^n}(n^0).$$

An application of the Cauchy-Schwarz inequality yields

$$\Big|n^{1/2}\int_{X_{n-k+1:n}}^{\omega(P_0)} g\,dP_0\Big|\le n^{1/2}\Big(\int_{X_{n-k+1:n}}^{\omega(P_0)} dP_0\Big)^{1/2}\Big(\int_{X_{n-k+1:n}}^{\omega(P_0)} g^2\,dP_0\Big)^{1/2}$$
$$=\big(n(1-F_0(X_{n-k+1:n}))\big)^{1/2}\Big(\int_{X_{n-k+1:n}}^{\omega(P_0)} g^2\,dP_0\Big)^{1/2}.$$

We show that

$$\int_{X_{n-k+1:n}}^{\omega(P_0)} g^2\,dP_0 \xrightarrow{P_0^n} 0. \tag{4.2.8}$$

Let $\epsilon>0$. Since $g\in L_2(P_0)$, we can find a real number $a<\omega(P_0)$ such that

$$\int_a^{\omega(P_0)} g^2\,dP_0<\epsilon/2.$$

Define

$$A_{n,k,a}=\{\underline{x}\in\mathbb{R}^n: X_{n-k+1:n}(\underline{x})\le a\}.$$

We derive

$$P_0^n\Big(\Big\{\underline{x}\in\mathbb{R}^n:\int_{X_{n-k+1:n}(\underline{x})}^{\omega(P_0)} g^2\,dP_0>\epsilon\Big\}\Big)$$
$$=P_0^n\Big(\Big\{\underline{x}\in\mathbb{R}^n:\int_{X_{n-k+1:n}(\underline{x})}^{\omega(P_0)} g^2\,dP_0>\epsilon\Big\}\cap A_{n,k,a}\Big)$$
$$+P_0^n\Big(\Big\{\underline{x}\in\mathbb{R}^n:\int_{X_{n-k+1:n}(\underline{x})}^{\omega(P_0)} g^2\,dP_0>\epsilon\Big\}\cap A_{n,k,a}^c\Big)$$
$$\le P_0^n(A_{n,k,a})+P_0^n(\emptyset)$$
$$=P_0^n(\{X_{n-k+1:n}\le a\})\underset{n\to\infty}{\longrightarrow}0.$$

Hence, (4.2.8) is shown.

Moreover,

$$P(\{n(U_{n-k+1:n}-1)\leq x\})\longrightarrow exp(x)\sum_{j=0}^{k-1}\frac{(-x)^j}{j!},\ x<0,$$

where $U_{i:n}$ denotes the i-th order statistic of n independent $(0,1)$-uniformly distributed random variables. Thus,

$$n^{1/2}\int_{X_{n-k+1:n}}^{\omega(P_0)} g\,dP_0 \xrightarrow{P_0^n} 0.$$

The same arguments show that

$$n^{1/2}\int_{X_{n-k+1:n}}^{\omega(P_0)} r_{n^{-1/2}t}\,dP_0 \xrightarrow{P_0^n} 0.$$

Combining the above results, we see that

$$(n-k)log\frac{F_{n^{-1/2}t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})}\xrightarrow{P_0^n} 0$$

holds. In a similar way, we deduce

$$(n-k)log\frac{1-F_{n^{-1/2}t}(X_{k:n})}{1-F_0(X_{k:n})}\xrightarrow{P_0^n} 0. \blacksquare$$

4.3 A Characterization Theorem of Gaussian Sequences via Extremes

In this section, we show that under monotone likelihood ratios and certain regularity conditions the following holds: A limit experiment is non-*Gaussian* if and only if a fixed number of extreme order statistics asymptotically contains information.

Consider a homogeneous experiment $E = \big(\mathbb{R}, \mathcal{B}, \{P_t : t \in \mathbb{R}\}\big)$ with *monotone likelihood ratios*; that is

$$\frac{dP_{t_2}}{dP_{t_1}} = h_{t_1,t_2} \circ S$$

where $h_{t_1,t_2} : \mathbb{R} \longrightarrow [0,\infty]$ is an increasing function for $t_1 < t_2$ and S is any statistic.

W.l.g. we assume that $S \equiv$ *identity*. Otherwise, consider the experiment $S_* E = \big(\mathbb{R}, \mathcal{B}, \{\mathcal{L}(S|P_t) : t \in \mathbb{R}\}\big)$ which is equivalent to E since S is sufficient. Note that $\mathcal{L}(S|P_t)$ has monotone likelihood ratios in the identity.

In addition, we assume that the family $\{P_t : t \in \mathbb{R}\}$ is L_1-*differentiable* at P_0 with derivative $g \in L_1(P_0)$; that is, the likelihood ratio admits an expansion

$$\frac{dP_t}{dP_0} = 1 + tg + tr_t, \tag{4.3.1}$$

where $r_t \in L_1(P_0)$ with $\|r_t\|_{L_1(P_0)} \to 0$ as $t \to 0$. The concept of L_1-differentiability is important in connection with local tests, see the book by *Witting* (1985).

4.3.1 Theorem. *Suppose that $E = \big(\mathbb{R}, \mathcal{B}, \{P_t : t \in \mathbb{R}\}\big)$ is a homogeneous experiment with monotone likelihood ratios (in the identity), P_t is continuous, and that $t \to P_t$ is L_1-differentiable at P_0. Assume that*

$$U_{\delta_n} E^n = E_n = \big(\mathbb{R}^n, \mathcal{B}^n, \{P^n_{\delta_n t} : t \in \mathbb{R}\}\big) \longrightarrow F$$

where $U_{\delta_n} E^1$ is infinitesimal and F is homogeneous and not totally uninformative. In addition it is assumed that for some $p \geq 1$ the normalizing sequence $(\delta_n)_{n \in \mathbb{N}}$ satisfies $\delta_n = O(n^{-1/p})$, g is p-integrable in a neighborhood of $\alpha(P_0)$ and $\omega(P_0)$, and that the remainder term satisfies $\|r_t\|_{L_p(P_0)} \to 0$ as $t \to 0$.

Then the following assertions are equivalent:

(i) F is non-Gaussian.

(ii) A fixed number of extreme order statistics asymptotically contains information; that is, the sequences of statistical experiments

$$W_{n,k} E_n = \big(\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(W_{n,k}|P^n_{\delta_n t}) : t \in \mathbb{R}\}\big)$$
$$Z_{n,k} E_n = \big(\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(Z_{n,k}|P^n_{\delta_n t}) : t \in \mathbb{R}\}\big)$$

have accumulation points, which are unequal the totally uninformative experiment.

In testing theory, Theorem 4.3.1 has the following meaning: If $(P^n_{\delta_n t_0}, P^n_{\delta_n t_1}) \to F$ and F is non-Gaussian, then we can find a test sequence based on a fixed number of extremes which separates $\{t_0\}$ and $\{t_1\}$.

REMARK. The assumptions of this theorem necessarily imply that $\liminf_{n\to\infty} n\delta_n \geq K$ for some positive constant K. Otherwise, if $n\delta_n \to 0$, the L_1-differentiability implies

$$\|P_0^n - P_{\delta_n t}^n\| \leq n\|P_0 - P_{\delta_n t}\| = n\delta_n t\|g + r_{\delta_n t}\|_{L_1(P_0)} \to 0$$

which contradicts our assumption that F is not totally uninformative.

Before giving the proof, we remark that this result is motivated by a well-known result in the theory of sums of independent random variables. Let $(\xi_{nk})_{1\leq k\leq k_n}$ be a triangular array of rowwise independent real valued random variables (over some probability space $(\Omega, \mathcal{A}, P)$). The variables ξ_{nk} are said to be *infinitesimal* if

$$\max_{1\leq k\leq k_n} P\{|\xi_{nk}| \geq \epsilon\} \longrightarrow 0 \tag{4.3.2}$$

as $n \to \infty$.

Suppose that

$$\sum_{k=1}^{k_n} \xi_{nk}$$

converges to some (non-degenerate) limit law Q. Then Q is Gaussian iff

$$P\{\sup_{1\leq k\leq k_n} |\xi_{nk}| \geq \epsilon\} \longrightarrow 0, \tag{4.3.3}$$

see *Gnedenko* and *Kolmogorov* (1968), p. 126, 127. Condition (4.3.3) is, in turn, equivalent to

$$\begin{aligned} \min_{1\leq k\leq k_n} \xi_{nk} &\longrightarrow 0 \\ \max_{1\leq k\leq k_n} \xi_{nk} &\longrightarrow 0. \end{aligned} \tag{4.3.4}$$

In this statistical context, we consider the array

$$(\xi_i)_{1\leq i\leq n} = \left(log\frac{dP_{\delta_n t}}{dP_0}(X_i)\right)_{1\leq i\leq n}.$$

Concerning the equivalence of (4.3.3) and (4.3.4), the assumption of monotone likelihood ratios becomes plausible since $(t > 0)$

$$\xi_{i:n} = log\frac{dP_{\delta_n t}}{dP_0}(X_{i:n}).$$

PROOF OF THEOREM 4.3.1. We show:

$$F \text{ is } \textit{Gaussian} \iff W_{n,k}E_n \longrightarrow E_0 \text{ and } Z_{n,k}E_n \longrightarrow E_0$$

where $E_0 = (\mathbb{R}, \mathcal{B}, \{\epsilon_0\})$ denotes the totally uninformative experiment.

We only consider the log-likelihood process with base 0. This is sufficient in view of the contiguity of the sequence $U_{\delta_n} E^n$.

The log-likelihood processes of $W_{n,k}E_n$ and $Z_{n,k}E_n$ are

$$(4.3.5)\qquad \begin{aligned} log\frac{d\mathcal{L}(W_{n,k}|P^n_{\delta_n t})}{d\mathcal{L}(W_{n,k}|P^n_0)} &= \sum_{i=1}^{k} log\frac{dP_{\delta_n t}}{dP_0}(X_{i:n}) + (n-k)log\frac{1-F_{\delta_n t}(X_{k:n})}{1-F_0(X_{k:n})}\\ log\frac{d\mathcal{L}(Z_{n,k}|P^n_{\delta_n t})}{d\mathcal{L}(Z_{n,k}|P^n_0)} &= \sum_{i=1}^{k} log\frac{dP_{\delta_n t}}{dP_0}(X_{n-i+1:n}) + (n-k)log\frac{F_{\delta_n t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})}. \end{aligned}$$

To verify (4.3.5) (note, that we did not assume $P_t \ll \lambda$) we make use of the fact that under P^n_t the statistics $W_{n,k}$ and $Z_{n,k}$ have the P^k_t-density

$$\begin{aligned} \frac{d\mathcal{L}(W_{n,k}|P^n_t)}{dP^k_t}(x_1,\dots,x_k) &= \frac{n!}{(n-k)!}(1-F_t(x_k))^{n-k}\\ \frac{d\mathcal{L}(Z_{n,k}|P^n_t)}{dP^k_t}(x_{n-k+1},\dots,x_n) &= \frac{n!}{(n-k)!}(F_t(x_{n-k+1}))^{n-k}. \end{aligned}$$

if $x_1 < x_2 < \cdots < x_k$, $x_{n-k+1} < \cdots < x_n$, $= 0$ otherwise (see *Reiss* (1989), p. 33). Since $P_t \ll P_0$, we have $\mathcal{L}(W_{n,k}|P^n_t) \ll \mathcal{L}(W_{n,k}|P^n_0)$. For $B \in \mathcal{B}^k$ we get ($\underline{x} = x_1,\dots,x_k$)

$$\begin{aligned} &P^n_t(\{W_{n,k} \in B\})\\ &= \mathcal{L}(W_{n,k}|P^n_t)(B)\\ &= \int_B \frac{d\mathcal{L}(W_{n,k}|P^n_t)}{d\mathcal{L}(W_{n,k}|P^n_0)}(\underline{x})\, d\mathcal{L}(W_{n,k}|P^n_0)(\underline{x})\\ &= \int_B \frac{dP^k_t}{dP^k_0}(\underline{x})\frac{d\mathcal{L}(W_{n,k}|P^n_t)}{dP^k_t}(\underline{x})\Big(\frac{d\mathcal{L}(W_{n,k}|P^n_0)}{dP^k_0}(\underline{x})\Big)^{-1} d\mathcal{L}(W_{n,k}|P^n_0)(\underline{x})\\ &= \int_B \frac{dP^k_t}{dP^k_0}(x_1,\dots,x_k)\Big(\frac{1-F_t(x_k)}{1-F_0(x_k)}\Big)^{n-k} d\mathcal{L}(W_{n,k}|P^n_0)(x_1,\dots,x_k)\\ &= \int_{W_{n,k}\in B} \frac{dP^k_t}{dP^k_0}(X_{1:n},\dots,X_{k:n})\Big(\frac{1-F_t(X_{k:n})}{1-F_0(X_{k:n})}\Big)^{n-k} dP^n_0. \end{aligned}$$

Analogue for $Z_{n,k}$.

First, we prove that under the present regularity conditions

$$(4.3.6)\qquad (n-k)log\frac{F_{\delta_n t}(X_{n-k+1:n})}{F_0(X_{n-k+1:n})} \xrightarrow{P^n_0} 0.$$

holds. As in the proof of Theorem 4.2.1, we have to show

$$n\big(F_{\delta_n t}(X_{n-k+1:n}) - F_0(X_{n-k+1:n})\big) \xrightarrow{P^n_0} 0$$

as $n \to \infty$.

The L_1-differentiability implies

$$\begin{aligned}\big|n\big(F_{\delta_n t}(X_{n-k+1:n}) - F_0(X_{n-k+1:n})\big)\big| &= \Big|n\int_{X_{n-k+1:n}}^{\omega(P_0)} \big(\frac{dP_{\delta_n t}}{dP_0} - 1\big)\, dP_0\Big| \\ &\le n\delta_n t \int_{X_{n-k+1:n}}^{\omega(P_0)} |g|\, dP_0 + n\delta_n t \int_{X_{n-k+1:n}}^{\omega(P_0)} |r_{\delta_n t}|\, dP_0.\end{aligned}$$

Note that by assumption P_t, $t \in \mathbb{R}$, are pairwise equivalent. For $p = 1$ assertion (4.3.6) is obviously valid.

Let $p > 1$. Choose a sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n \uparrow \omega(P_0)$ and $P_0^n(\{X_{n-k+1:n} \le a_n\}) \to 0$ as $n \to \infty$. For sufficiently large n we have $g1_{(a_n,\omega(P_0))} \in L_p(P_0)$. Define

$$A_{n,k,a} = \{X_{n-k+1:n} \le a\}.$$

Then for $\epsilon > 0$

$$\begin{aligned}&P_0^n\Big(\Big\{n\delta_n \int_{X_{n-k+1:n}}^{\omega(P_0)} |g|\, dP_0 > \epsilon\Big\}\Big) \\ &\le P_0^n(A_{n,k,a_n}) + P_0^n\Big(\Big\{n\delta_n \int_{X_{n-k+1:n}}^{\omega(P_0)} |g|\, dP_0 > \epsilon\Big\} \cap A_{n,k,a_n}^c\Big) \\ &\le P_0^n(\{X_{n-k+1:n} \le a_n\}) \\ &\quad + P_0^n\Big(\Big\{n^{1/p}\delta_n\big(n(1 - F_0(X_{n-k+1:n}))\big)^{(p-1)/p}\Big(\int_{X_{n-k+1:n}}^{\omega(P_0)} |g|^p\, dP_0\Big)^{\frac{1}{p}} > \epsilon\Big\} \cap A_{n,k,a_n}^c\Big) \\ &\le P_0^n(\{X_{n-k+1:n} \le a_n\}) \\ &\quad + P_0^n\Big(\Big\{n^{1/p}\delta_n\big(n(1 - F_0(X_{n-k+1:n}))\big)^{(p-1)/p}\Big(\int_{a_n}^{\omega(P_0)} |g|^p\, dP_0\Big)^{\frac{1}{p}} > \epsilon\Big\}\Big).\end{aligned}$$

Since $\|g1_{(a_n,\omega(P_0))}\|_{L_p(P_0)} \to 0$ as $n \to \infty$, we see that

$$n\delta_n \int_{X_{n-k+1:n}}^{\omega(P_0)} |g|\, dP_0 \xrightarrow{P_0^n} 0$$

holds.

Moreover,

$$n\delta_n \int_{X_{n:n}}^{\omega(P_0)} |r_{\delta_n t}|\, dP_0 \xrightarrow{P_0^n} 0.$$

This follows form

$$\begin{aligned}&n\delta_n \int_{X_{n-k+1:n}}^{\omega(P_0)} |r_{\delta_n t}|\, dP_0 \\ &\qquad\le n^{1/p}\delta_n\big(n(1 - F(X_{n-k+1:n}))\big)^{(p-1)/p}\Big(\int_{X_{n-k+1:n}}^{\omega(P_0)} |r_{\delta_n t}|^p\, dP_0\Big)^{1/p}\end{aligned}$$

and

$$\int_{X_{n-k+1:n}}^{\omega(P_0)} |r_{\delta_n t}|^p \, dP_0 \xrightarrow{P_0^n} 0.$$

Hence, (4.3.6) is proved.

Similar to (4.3.6) one shows

$$(n-k)log\frac{1-F_{\delta_n t}(X_{k:n})}{1-F_0(X_{k:n})} \xrightarrow{P_0^n} 0. \tag{4.3.7}$$

Now, suppose that F is *Gaussian.* (For the proof of this direction, the assumption of monotone likelihood ratios is not needed.) From Theorem (6.3) of *Milbrodt* and *Strasser* (1985), we know that

$$\lim_{n\to\infty} nP_0^n\big(\{|log\frac{dP_{\delta_n t}}{dP_0}(X_1)| > \epsilon\}\big) = 0.$$

This together with

$$\begin{aligned} P_0^n\big(\{|log\frac{dP_{\delta_n t}}{dP_0}(X_{j:n})| > \epsilon\}\big) &\le P_0^n\big(\{\max_{1\le i\le n} |log\frac{dP_{\delta_n t}}{dP_0}(X_i)| > \epsilon\}\big) \\ &= P_0^n\big(\cup_{1\le i\le n}\{|log\frac{dP_{\delta_n t}}{dP_0}(X_i)| > \epsilon\}\big) \\ &\le nP_0^n\big(\{|log\frac{dP_{\delta_n t}}{dP_0}(X_1)| > \epsilon\}\big) \end{aligned}$$

implies

$$log\frac{dP_{\delta_n t}}{dP_0}(X_{j:n}) \xrightarrow{P_0^n} 0, \; j = 1, ..., n. \tag{4.3.8}$$

This proves one half of the assertion, if we combine (4.3.5) – (4.3.8).

To prove the converse, note that for $k > 1$

$$W_{n,k}E_n \longrightarrow E_0 \text{ implies } W_{n,1}E_n = X_{1:n}E_n \longrightarrow E_0$$

(There are several ways to see this: $W_{n,k}E_n \to E_0$ is equivalent to $\|\mathcal{L}(W_{n,k}|P_{\delta_n s}^n) - \mathcal{L}(W_{n,k}|P_{\delta_n t}^n)\| \to 0$ as $n \to \infty$, $s, t \in \mathbb{R}$. Now, the assertion follows from $\|\mathcal{L}(X_{1:n}|P_{\delta_n s}^n) - \mathcal{L}(X_{1:n}|P_{\delta_n t}^n)\| \le \|\mathcal{L}(W_{n,k}|P_{\delta_n s}^n) - \mathcal{L}(W_{n,k}|P_{\delta_n t}^n)\|$. Another way is to argue with the *error function* or the *Mellin transform* (see *Strasser* (1985 a), Chapter 3). Observe thereby, that $W_{n,k}E_n$ is more informative than $X_{1:n}E_n$.)
and

$$Z_{n,k}E_n \longrightarrow E_0 \text{ implies } Z_{n,1}E_n = X_{n:n}E_n \longrightarrow E_0.$$

Taking into account (4.3.5), (4.3.6), and (4.3.7) for $k = 1$, we obtain

$$log\frac{dP_{\delta_n t}}{dP_0}(X_{1:n}) \xrightarrow{P_0^n} 0$$

$$log\frac{dP_{\delta_n t}}{dP_0}(X_{n:n}) \xrightarrow{P_0^n} 0.$$

Since $\{P_t : t \in \mathbb{R}\}$ has monotone likelihood ratios, we deduce that

$$(4.3.9) \qquad P_0^n\{\max_{1\le i\le n} |log\frac{dP_{\delta_n t}}{dP_0}(X_i)| \geq \epsilon\} \longrightarrow 0.$$

Notice that for $t < 0$, $dP_{\delta_n t}/dP_0$ is decreasing. By assumption, $\sum_{i=1}^n log\frac{dP_{\delta_n t}}{dP_0}(X_i)$ converges to some non-degenerate law. Combining these results, we obtain

$$(4.3.10) \qquad \sum_{i=1}^n log\frac{dP_{\delta_n t}}{dP_0}(X_i) \longrightarrow \mathcal{N}(\mu(t), \sigma^2(t))$$

for some $\mu(t) \in \mathbb{R}$ and $\sigma^2(t) > 0$. The homogeneity of F implies the contiguity of $U_{\delta_n}E^n$. This together with (4.3.10) implies the weak convergence of the binary experiment $(P_0^n, P_{\delta_n t}^n)$ to a binary Gaussian experiment, say, (Q_0, Q_t), $t \in \mathbb{R}$. To prove that F is Gaussian, we have to show that for finite subsets $J \in \mathbb{R}$, $0 \notin J$,

$$\mu = \mathcal{L}((log\frac{dQ_j}{dQ_0})_{j\in J}|Q_0)$$

is a normal distribution. This is shown as follows (cf. the proof of Theorem 3.1 in *Janssen* (1989 b)): First, F is infinitely divisible, since F is the limit of an infinitesimal array (*Milbrodt* and *Strasser* (1985), Theorem 5.11). Hence, μ is an infinitely divisible distribution on $\mathbb{R}^J$. From (4.3.10), we know that $\mathcal{L}(p_j|\mu)$ is a normal distribution for each projection $p_j : \mathbb{R}^J \to \mathbb{R}$ on the j-th coordinate, $j \in J$. Let $(\mu_t)_{t>0}$ be the continuous convolution semigroup generated by $\mu(= \mu_1)$. Then μ is Gaussian iff the Levy-measure of the convolution semigroup vanishes. But this is equivalent to $\frac{1}{t}\mu_t(\mathbb{R}^J - (-\epsilon, \epsilon)^J) \to 0$ for $t \to 0$ and each $\epsilon > 0$. Using the sub-additivity of measures, we obtain

$$\frac{1}{t}\mu_t(\mathbb{R}^J \setminus (-\epsilon, \epsilon)^J) \leq \frac{1}{t}\sum_{j\in J}\mathcal{L}(p_j|\mu_t)(\mathbb{R} \setminus (-\epsilon, \epsilon)) \longrightarrow 0.$$

Hence, μ is normal. ∎

REMARK 2. For $p = 1$, the assumption of homogeneity of E can be dropped; for $p > 1$, this assumption can be dropped whenever

$$nP_{\delta_n t}\{(X_{n:n}, \omega(P_0)) \cap N_{\delta_n t}\} \to 0$$
$$nP_{\delta_n t}\{(\alpha(P_0), X_{1:n}) \cap N_{\delta_n t}\} \to 0$$

as $n \to \infty$. Note that the L_1-differentiability implies $P_t(N_t) = o(t)$, where N_t is the singularity part of P_t w.r.t. P_0.

5 POINT PROCESSES

The last chapter is devoted to point processes.

5.1 The Concept of Point Processes

First, we recall the definition of a point process and fix some further notations.

Suppose S is locally compact with countable basis. We denote the corresponding Borel σ-field with the symbol $\mathcal{B}$. Designate by $M(S,\mathcal{B})$ the set of all *point measures* defined on S. Recall that $\mu \in M(S,\mathcal{B})$, if there exists a denumerable set of points $x_i \in S$, $i \in I$, such that

$$\mu = \sum_{i \in I} \epsilon_{x_i}$$

and

$$\mu(K) < \infty$$

for every compact set K. The set $M(S,\mathcal{B})$ is endowed with the σ-field $\mathcal{M}(S,\mathcal{B})$, which is by definition the smallest σ-field such that the "projections" $\mu \longrightarrow \mu(B)$, $B \in \mathcal{B}$, 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 e.g. *Kallenberg* (1986)).

A *point process* (over some probability space $(\Omega, \mathcal{A}, P)$) is a measurable map

$$N : (\Omega, \mathcal{A}, P) \longrightarrow (M(S,\mathcal{B}), \mathcal{M}(S,\mathcal{B})).$$

In other words, a point process is a random variable with values in the space of point measures.

For example,

$$N_n(\cdot) = \sum_{i=1}^{n} \epsilon_{X_i}(\cdot \cap D)$$

defines a point process, where $D \subseteq S$ and X_i are random variables; N_n is called *truncated empirical point process* or, if $S = D$, *empirical point process*.

The most important class of point processes are the Poisson processes: Given a Radon measure ν on $(S,\mathcal{B})$, a point process N is called *Poisson process* with *mean measure* (also called *intensity*) ν, if N satisfies

(i) $\forall$ B $\in \mathcal{B}$ $\forall$ k $\in \mathbb{N}_0$:

$$P\{N(B) = k\} = \begin{cases} \frac{(\nu(B))^k}{k!} e^{-\nu(B)}, & \text{if } \nu(B) < \infty \\ 0, & \text{if } \nu(B) = \infty \end{cases}$$

(ii) $\forall$ n $\in \mathbb{N}$ $\forall$ $B_i \in \mathcal{B}$, $1 \le i \le n$, mutally disjoint: $N(B_i)$, $1 \le i \le n$, are independent.

Poisson processes occur as (weak) limit processes in many cases. They play the same superior role as the normal distribution for sums of random variables or as the extreme value distributions for extremes.

Point processes in connection with extreme value theory can be found in the book by *Resnick* (1987).

An interesting characterization was given in *Falk* and *Reiss* (1988 b): It was shown that the convergence of appropriately truncated empirical point processes to certain Poisson processes (extreme value processes), measured w.r.t. the variational distance, holds if, and only if, the underlying distribution belongs to the strong domain of attraction of an extreme value distribution.

An important reference for the statistical inference within Poisson models is the book by *Karr* (1986).

5.2 Statistical Experiments of Point Processes

The aim of this section is to show, that the results of the second chapter can be reformulated in terms of point processes. In the following, we consider the measurable space $(S, \mathcal{B}) = (\mathbb{R}, \mathcal{B})$. We shall see, that in the i.i.d. case the original experiment and the corresponding point process experiment are equivalent. A simple consequence of this fact is that the loss of information due to a reduction of order statistics in the original experiment is the same as in the corresponding point process experiment.

Let us introduce the sub-space M_n of all point measures with total mass n for $n \in \mathbb{N}$:

$$M_n = M_n(\mathbb{R}, \mathcal{B}) = \left\{ \mu_n \in M(\mathbb{R}, \mathcal{B}) : \mu_n = \sum_{i=1}^{n} \epsilon_{x_i}, \, x_i \in \mathbb{R} \right\}.$$

It is obvious that $M_n \in \mathcal{M}(\mathbb{R}, \mathcal{B})$. The space M_n is endowed with the trace σ-field

$$\mathcal{M}_n = \mathcal{M}_n(\mathbb{R}, \mathcal{B}) = \mathcal{M}(\mathbb{R}, \mathcal{B}) \cap M_n(\mathbb{R}, \mathcal{B}).$$

Now we consider the surjective map

$$\begin{aligned} \tilde{N}_n : \quad \mathbb{R}^n &\longrightarrow M_n \\ (x_1, \ldots, x_n) &\longrightarrow \sum_{i=1}^{n} \epsilon_{x_i}. \end{aligned}$$

It is clear that $\tilde{N}_n$ is $\mathcal{B}^n_{sym}, \mathcal{M}_n$-measurable, where $\mathcal{B}^n_{sym}$ denotes the σ-field of measurable sets which are invariant under permutation:

$$\mathcal{B}^n_{sym} = \{ B \in \mathcal{B} : \pi(B) = B \quad \forall \pi \in \mathcal{S}_n \}.$$

Here, we denote by $\mathcal{S}_n$ the permutation group of order n.

But more can be said about $\tilde{N}_n$.

5.2.1 Proposition. *Let $T \neq \emptyset$ be an arbitrary set. Then $\tilde{N}_n$ is a sufficient statistic for $E_n = (\mathbb{R}^n, \mathcal{B}^n, \{Q^n_t : t \in T\})$.*

PROOF. Let $\mathbb{R}^n_{\leq} = \{(x_1, ..., x_n) \in \mathbb{R}^n : x_1 \leq x_2 \leq \cdots \leq x_n\}$ be the cone of ordered values of the Euclidian space $\mathbb{R}^n$. The space $\mathbb{R}^n_{\leq}$ is equipped with the trace σ-field $\mathcal{B}^n_{\leq} = \mathcal{B}^n \cap \mathbb{R}^n_{\leq}$. We consider the $\mathcal{B}^n_{\leq}, \mathcal{M}_n$-measurable map

$$\begin{aligned} \tau_n : \quad \mathbb{R}^n_{\leq} &\longrightarrow M_n \\ (x_{1:n}, ..., x_{n:n}) &\longrightarrow \sum_{i=1}^{n} \epsilon_{x_i} \end{aligned}$$

Then $\tilde{N}_n = \tau_n \circ (x_{1:n}, \ldots, x_{n:n})$. Since the order statistic is sufficient and τ_n is bijective, the proposition is proved if

$$\tau_n^{-1}(\mathcal{M}_n) = \mathcal{B}^n_{\leq}. \tag{5.2.1}$$

First

$$\tau_n^{-1}(\mathcal{M}_n)$$

is a σ-field. Since

$$\mathcal{G} = \{B_1 \times \cdots \times B_n : B_i \in \mathcal{B}, i = 1, \ldots, n\}$$

generates $\mathcal{B}^n$ we find that $\mathcal{B}^n_\leq$ is generated by

$$\begin{aligned}\mathcal{G}_\leq &= \mathcal{G} \cap \mathbb{R}^n_\leq \\ &= \{B_1 \times \cdots \times B_n \cap \mathbb{R}^n_\leq : B_i \in \mathcal{B}, i = 1, \ldots, n\}.\end{aligned}$$

We show that $\mathcal{G}_\leq \subset \tau_n^{-1}(\mathcal{M}_n)$ which implies (5.2.1). The set

$$\begin{aligned}&\Big\{\sum_{i=1}^n \epsilon_{x_i} \in M_n : (x_1, \ldots, x_n) \in B_1 \times \cdots \times B_n\Big\} \\ &\quad = \Big\{\sum_{i=1}^n \epsilon_{x_i} \in M_n : \sum_{i=1}^n \epsilon_{x_i}(B_j) \geq 1, j = 1, ..., n, \sum_{i=1}^n \epsilon_{x_i}\big(\bigcup_{j=1}^n B_j\big) = n\Big\} \in \mathcal{M}_n\end{aligned}$$

satisfies

$$\tau_n^{-1}(M) = B_1 \times \cdots \times B_n \cap \mathbb{R}^n_\leq. \blacksquare$$

5.2.2 Corollary. *Let E_n as in Proposition 5.2.1 and let*

$$E_n^{(N)} = (M_n, \mathcal{M}_n, \{\mathcal{L}(\tilde{N}_n | Q_t^n) : t \in T\})$$

be the corresponding point process experiment. Then E_n and $E_n^{(N)}$ are equivalent:

(5.2.2) $$\Delta\big(E_n, E_n^{(N)}\big) = 0.$$

We draw our attention to sparse order statistics $Z_{n,r_1,\ldots,r_k} = (X_{r_1:n}, \ldots, X_{r_k:n})$ with $1 \leq r_1 < r_2 \cdots < r_k \leq n$, $1 \leq k \leq n$. Define

$$N_{n,r_1,\ldots,r_k} = \sum_{i=1}^k \epsilon_{X_{r_i:n}}.$$

Since

$$N_{n,r_1,\ldots,r_k}(\cdot) = \sum_{i=1}^n \epsilon_{X_i}\big(\cdot \cap [X_{r_1:n}, X_{r_k:n}]\big)$$

in distribution if Q is continuous , this point process is a truncated empirical point process, whereby the truncation is random.

We define the statistical experiment

$$E_{n,r_1,\dots,r_n} = \left(\mathbb{R}^k, \mathcal{B}^k, \{\mathcal{L}(Z_{n,r_1,\dots,r_k}|Q_t^n) : t \in T\}\right)$$

and the corresponding point process experiment

$$E^{(N)}_{n,r_1,\dots,r_k} = \left(M_k, \mathcal{M}_k, \{\mathcal{L}(N_{n,r_1,\dots,r_k}|Q_t^n) : t \in T\}\right).$$

5.2.3 Corollary. *Adapting the notations of above, we get*

$$\Delta\left(E^{(N)}_n, E^{(N)}_{n,r_1,\dots,r_k}\right) = \Delta\left(E_n, E_{n,r_1,\dots,r_k}\right). \tag{5.2.3}$$

PROOF. We already know that

$$\Delta\left(E_{n,r_1,\dots,r_k}, E^{(N)}_{n,r_1,\dots,r_k}\right) = 0.$$

Applying the triangular inequality, the assertion follows from Corollary 5.2.2. ▪

To treat point measures with infinite mass as well, we consider the space $(S,\mathcal{B}) = (\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+)$ with $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$. Let

$$M_\infty^+ = \left\{\mu \in M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) : \mu = \sum_{i\in\mathbb{N}} \epsilon_{x_i}\right\}$$

Moreover, let $\mathcal{M}_\infty^+$ be the restriction of $\mathcal{M}(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+)$ to M_∞^+. The space M_∞^+ is Polish w.r.t. the vague topology, because it is a "$\mathcal{G}_\delta$–set", i.e. M_∞^+ is a countable intersection of open sets. We have

$$\begin{aligned} M_\infty^+ &= M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) \setminus \bigcup_{n\in\mathbb{N}} \{\mu \in M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) : \mu(\mathbb{R}_+) \leq n\} \\ &= \bigcap_{n\in\mathbb{N}} \Big(M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) \cap \{\mu \in M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) : \mu(\mathbb{R}_+) > n\}\Big). \end{aligned}$$

Note that the set $\{\mu \in M(\mathbb{R}_+, \mathcal{B}\cap\mathbb{R}_+) : \mu(\mathbb{R}_+) \leq n\}$ is vaguely closed.

Since $\mu(K) = \sum_{i\in\mathbb{N}} \epsilon_{x_i}(K) < \infty$ for compact sets K (by definition) the sequence $(x_i)_{i\in\mathbb{N}}$ has no (finite) accumulation points.

We have

$$M_\infty^+ = \left\{\sum_{i\in\mathbb{N}} \epsilon_{x_i} : (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_{+,\leq}\right\}$$

where

$$\mathbb{R}^{\mathbb{N}}_{+,\leq} = \left\{(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}_+ : x_i \leq x_{i+1},\ i \in \mathbb{N},\ \liminf_{i\to\infty} x_i \to \infty\right\}$$

is the set of non-negative, increasing, and unbounded sequences.

Denote by $\mathcal{B}^{\mathbb{N}}_{+,\leq}$ the restriction of $\mathcal{B}^{\mathbb{N}}$ to $\mathbb{R}^{\mathbb{N}}_{+,\leq}$. Recall, that $\mathbb{R}^{\mathbb{N}}_{+}$ is Polish as countable product of Polish spaces. Since $\mathbb{R}^{\mathbb{N}}_{+,\leq}$ is a closed subset of $\mathbb{R}^{\mathbb{N}}_{+}$, we know that $\mathbb{R}^{\mathbb{N}}_{+,\leq}$ is also Polish.

The following lemma has an application in the next section.

5.2.4 Lemma. *The bijective map*

$$\begin{aligned} \tau_\infty : \quad \mathbb{R}^{\mathbb{N}}_{+,\leq} &\longrightarrow M^+_\infty \\ (x_i)_{i\in\mathbb{N}} &\longrightarrow \sum_{i\in\mathbb{N}} \epsilon_{x_i} \end{aligned}$$

is $\mathcal{B}^{\mathbb{N}}_{+,\leq}, \mathcal{M}^+_\infty$-measurable and

$$\tau_\infty\left(\mathcal{B}^{\mathbb{N}}_{+,\leq}\right) = \mathcal{M}^+_\infty. \tag{5.2.4}$$

Proof. To verify the measurability, it is sufficient to show that for $B \in \mathcal{B} \cap \mathbb{R}_+$ and $k \in \mathbb{N}_0$

$$\tau_\infty^{-1}\left(M_k(B)\right) \in \mathcal{B}^{\mathbb{N}}_{+,\leq}$$

where

$$M_k(B) := \left\{\mu \in M^+_\infty : \mu(B) = k\right\}.$$

For $k = 0$ we obtain

$$\begin{aligned} M_0(B) &= \left\{\mu \in M^+_\infty : \mu(B) = 0\right\} \\ &= \left\{\sum_{i\in\mathbb{N}} \epsilon_{x_i} \in M^+_\infty : x_i \notin B \ \forall i \in \mathbb{N}\right\}. \end{aligned}$$

Hence

$$\tau_\infty^{-1}(M_0(B)) = (B^c)^{\mathbb{N}} \cap \mathbb{R}^{\mathbb{N}}_{+,\leq}.$$

For $k \in \mathbb{N}$ we obtain

$$\begin{aligned} M_k(B) &= \left\{\mu \in M^+_\infty : \mu(B) = k\right\} \\ &= \Bigg\{\sum_{i\in\mathbb{N}} \epsilon_{x_i} \in M^+_\infty : \exists\, x_{i_1}, \ldots, x_{i_k} \text{ with } x_{i_j} \in B, \\ &\qquad\qquad 1 \leq j \leq k,\ x_j \notin B, j \in \mathbb{N} \setminus \{x_{i_1}, \ldots, x_{i_k}\}\Bigg\}. \end{aligned}$$

Hence

$$\tau_\infty^{-1}(M_k(B)) = \bigcup_{n>k}\left(\left(\bigcup_{\pi\in\mathcal{S}_n} \pi(B^k \times (B^c)^{n-k}) \times (B^c)^{\mathbb{N}\setminus\{1,\ldots,n\}}\right) \bigcap \mathbb{R}^{\mathbb{N}}_{+,\leq}\right) \in \mathcal{B}^{\mathbb{N}}_{+,\leq}.$$

The measurability is shown.

Now, we proof the validity of the equality (5.2.4). The inclusion $\mathcal{M}_{\infty}^{+} \subseteq \tau_{\infty}(\mathcal{B}_{+,\leq}^{\mathbb{N}})$ is immediate from the properties of the map τ_{∞}. Since the underlying spaces are Polish, the converse inclusion $\tau_{\infty}(\mathcal{B}_{+,\leq}^{\mathbb{N}}) \subseteq \mathcal{M}_{\infty}^{+}$ follows directly from the famous theorem of Kuratowski (see, for instance, *Jacobs* (1978), p. 420). ■

5.3 Empirical Point Processes of Weibull Type Samples

In this section, we combine the results of Chapter 2 and the previous section of this chapter. We assume that X_i, $i \in \mathbb{N}$, are i.i.d. random variables with density of Weibull type (2.1.1). Let S_j, $j \in \mathbb{N}$, be defined as in (2.1.5). We consider the point processes

$$\begin{aligned} N_n &= \sum_{i=1}^{n} \epsilon_{\delta_n^{-1} X_i} \\ N_{n,k} &= \sum_{i=1}^{k} \epsilon_{\delta_n^{-1} X_{i:n}} \\ N_t^* &= \sum_{j \in \mathbb{N}} \epsilon_{S_j^{1/(1+a)}+t} \\ N_{k,t}^* &= \sum_{j=1}^{k} \epsilon_{S_j^{1/(1+a)}+t}. \end{aligned}$$

Note, that N_t^* is a Poisson process with mean value function $x \longrightarrow (x-t)^{1+a}$, $x > t$ (see *Resnick* (1987), Corollary 4.19).

We now define the corresponding point process experiments of $E_n, E_{n,k}, G$ and G_k:

$$\begin{aligned} E_n^{(N)} &= \big(M_n, \mathcal{M}_n, \{\mathcal{L}(N_n | P_{\delta_n t}^n) : t \in \mathbb{R}\}\big) \\ E_{n,k}^{(N)} &= \big(M_k, \mathcal{M}_k, \{\mathcal{L}(N_{n,k} | P_{\delta_n t}^n) : t \in \mathbb{R}\}\big) \\ G^{(N)} &= \big(M_\infty, \mathcal{M}_\infty, \{\mathcal{L}(N_t^*) | P_0) : t \in \mathbb{R}\}\big) \\ G_k^{(N)} &= \big(M_k, \mathcal{M}_k, \{\mathcal{L}(N_{k,t}^*) | P_0) : t \in \mathbb{R}\}\big). \end{aligned}$$

From the results of Section 5.2, we already know that

$$\Delta(E_n^{(N)}, E_{n,k}^{(N)}) = \Delta(E_n, E_{n,k})$$

and

$$\Delta(E_{n,k}^{(N)}, G_k^{(N)}) = \Delta(E_{n,k}, G_k)$$

The last equality follows from $\Delta(G_k^{(N)}, G_k) = 0$.

Moreover, we have $\Delta(G, G^{(N)}) = 0$. This follows from Lemma 5.2.4 which states the sufficiency of τ_∞. Thus

$$\Delta(G_k^{(N)}, G^{(N)}) = \Delta(G_k, G).$$

Consequently, all results established in Chapter 2 can be carried over; in particular, we have (see Theorem 2.4.1)

$$\Delta(E_n^{(N)}, G^{(N)}) = o(n^0)$$

and

$$\Delta(E_{n,k(n)}^{(N)}, G^{(N)}) = o(n^0)$$

whenever $k(n)$ tends to infinity as n tends to infinity.

Taking into account Theorem 2.4.5, we see that

$$\Delta(G_{k(n)}^{(N)}, G^{(N)}) \to 0$$

holds for $k(n) \to \infty$ as $n \to \infty$. Observe, that

$$\sup_{M \in \mathcal{M}(\mathbb{R}_+, \mathcal{B} \cap \mathbb{R}_+)} |\mathcal{L}(N_{k,t}^\star | P_0)(M) - \mathcal{L}(N_t^\star | P_0)(M)| = \mathcal{L}(N_t^\star | P_0)(M_k^c) = 1.$$

REMARK 1. In the sense of Definition (12.2) in *Milbrodt* (1985), $G^{(N)}$ is a *standard Poisson experiment* with intensities

$$B \longrightarrow E_{P_0} N_t^\star(B), \quad t \in \mathbb{R}.$$

REMARK 2. Let $K_t^{(n,k)}(\cdot|\underline{x})$, $\underline{x} = (x_1, .., x_k)$, Y_i as in (2.2.2) and $\tilde{N}_n$ as in Section 5.2. The conditional distribution of $\sum_{i=1}^n \epsilon_{X_i}$ given $\sum_{i=1}^k \epsilon_{X_{i:n}} = \sum_{i=1}^k \epsilon_{x_i}$ is equal to

$$\mathcal{L}(\tilde{N}_n | K_t^{(n,k)}(\cdot|\underline{x})) = (\sum_{i=1}^{k} \epsilon_{x_i} \star \mathcal{L}(\sum_{i=1}^{n-k} \epsilon_{Y_i}))(\cdot).$$

This becomes obvious from the fact, that $\sigma(N_{n,k,t}) = \sigma(X_{1:n}, \ldots, X_{k:n})$.

References

N.H. Bingham, C.M. Goldie and J.L. Teugels (1987), "Regular Variation," Cambridge University Press, Cambridge.

D. Blackwell (1951), *Comparison of experiments*, Proc. 2nd Berkeley Symp. Math. Statist. Probab., 93–102.

M. Csörgő, S. Csörgő, L. Horváth, and D.M. Mason (1986), *Normal and stable convergence of integral functions of the empirical distribution function*, Ann. Probab. **14**, 86–118.

S. Csörgő, L. Horváth, and D.M. Mason (1986), *What portion of a sample makes a partial sum asymptotically stable or normal?*, Probab. Theory Rel. Fields **72**, 1–16.

S. Csörgő and D.M. Mason (1986), *The asymptotic distribution of sums of extreme values from a regulary varying distribution*, Ann. Probab. **14**, 974–983.

H.A. David (1981), "Order Statistics," 2nd edition,Wiley, New York.

M. Falk (1985), *Rates of uniform convergence of extreme order statistics*, Habilitationsschrift, University of Siegen.

M. Falk and W. Kohne (1986), *On the rate at which sample extremes become independent*, Ann. Statist. **14**, 1339–1346.

M. Falk and R.-D. Reiss (1988 a), *Independence of order statistics*, Ann. Statist. **16**, 854–862.

M. Falk and R.-D. Reiss (1988 b), *Poisson approximation of empirical processes*, Preprint 218, University of Siegen.

B.V. Gnedenko and A.N. Kolmogorov (1968), "Limit Distributions for Sums of Independent Random Variables," Addison-Wesley, Reading, Massachusetts.

Haan, L. de (1976), "On Regular Variation, and its Application to the Weak Convergence of Sample Extremes," Math. Centre Tracts, Amsterdam.

J. Helgeland (1982), *Additional observations and statistical information in the case of 1-parameter exponential distributions*, Z. Wahrsch. verw. Gebiete **59**, 77–100.

H. Heyer (1982), "Theory of Statistical Experiments," Springer Series in Statistics, Springer, New York.

I.A. Ibragimov and R.Z. Has'minskii (1981), "Statistical Estimation," Application of Mathematics, Vol. 16, Springer, Berlin-Heidelberg-New York-Tokyo.

K. Jacobs (1978), "Measure and Integral," Academic Press, London-New York.

A. Janssen (1986), *Limit of translation invariant experiments*, J. Multivar. Anal. **20**, 129–142.

A. Janssen (1989 a), *The role of extreme order statistics for exponential families*, in: J. Hüsler and R.-D. Reiss. Extreme Value Theory, Lecture Notes in Statistics 51, Springer, New York.

A. Janssen (1989 b), *Statistical experiments with nonregular densities*, Appendix to the monograph: A. Janssen and D.M. Mason: Non-standard Rank Tests. Forthcoming monograph.

A. Janssen and D.M. Mason (1989), "Non-standard Rank Tests," Forthcoming monograph. To appear in: Lecture Notes in Statistics. Springer.

A. Janssen and R.-D. Reiss (1988), *Comparison of location models of Weibull type samples and extreme value processes*, Probab. Th. Rel. Fields **78**, 273–292.

O. Kallenberg (1986), "Random Measures," 4th edition, Academic Press, Berlin-London.

A.F. Karr (1986), "Point Processes an their Statistical Inference," Marcel Dekker, New York.

J.F. Lawless (1982), "Statistical Models and Methods for Lifetime Data," Wiley, New York.

L. LeCam (1956), *On the asymptotic theory of estimation and testing hypotheses*, Proc. 3rd Berkeley Symp. Math. Statist. Probab. **1**, 129–156.

L. LeCam (1964), *Sufficiency and approximate sufficiency*, Ann. Math. Statist. **35**, 1419–1455.

L. LeCam (1986), "Asymptotic Methods in Statistical Decison Theory," Springer Series in Statistics, Springer, New York.

E.L. Lehmann (1989), *Comparing location experiments*, Ann. Statist. **16**, 521–533.

B. Lindqvist (1984), *On the memory of an absorbing Markov chain*, Rev. Roumaine Math. Pures Appl., Tome XXIX, N^o 2, 171–182.

H. Luschgy (1987), *Comparison of shift experiments on a Banach space*, in: M.L. Puri et al. (eds.), Mathematical Statistics and Probability Theory, Vol. A, 217-230, Reidel, Dordrecht.

E. Mammen (1983), *Die statistische Information zusätzlicher Beobachtungen. Inaugural-Dissertation*, Heidelberg.

E. Mammen (1986), *The statistical information contained in additional observations*, Ann. Statist. **14**, 665–678.

H. Milbrodt (1985), *Representation of Poisson experiments*, in: A. Janssen, H. Milbrodt, and H. Strasser (1985), Infinitely Divisible Statistical Experiments, Lecture Notes in Statistics 27, Springer, Berlin-Heidelberg, 106–123.

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

D.W. Müller (1973), *Contribution to the discussion of LeCam (1973): Sur les contraintes imposées par les passages à la limite usuels en statistique*, Bull. I.S.I, Proc. 39th Session, Vienna, XLV, Book 4.

Th. Polfeldt (1970), *Asymptotic results in non-regular estimation*, Scand. Act. Journ., 1–2 Supplement, 1–78.

J. Pfanzagl and W. Wefelmeyer (1985), "Asymptotic Expansions for General Statistical Models," Lecture Notes in Statistics 31, Springer, Berlin-Heidelberg.

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

S.I. Resnick (1987), "Extreme Values, Regular Variation, and Point Processes," Applied Probability, Vol. 4, Springer, New York.

R.J. Serfling (1980), "Approximation Theorems of Mathematical Statistics," Wiley, New York.

H. Strasser (1985 a), "Mathematical Theory of Statistics," De Gruyter Studies in Math. 7, De Gruyter, Berlin-New York.

H. Strasser (1985 b), *Stability and translation invariance of statistical experiments*, Probab. Math. Statist. **5**, 1–20.

T.J. Sweeting (1985), *On domains of uniform local attraction in extreme value theory*, Ann. Probab. **13**, 196-205.

E.N. Torgersen (1970), *Comparison of experiments when the parameter space is finite*, Z. Wahrsch. v. Geb. **16**, 219–249.

E.N. Torgersen (1972), *Comparison of translation experiments*, Ann. Math. Statist. **43**, 1383–1399.

L. Weiss (1979), *The asymptotic sufficiency in a class of nonregular cases*, Sel. Statist. Canad. **V**, 141–150.

H. Witting (1985), "Mathematische Statistik I," Teubner, Stuttgart.

LIST OF SYMBOLS

$\exists$	existence
$\forall$	for all
$\mathbb{N}$	positve integers
$\mathbb{N}_0$	$= \mathbb{N} \cup \{0\}$
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	$\{x \in \mathbb{R} : x \geq 0\}$
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{R}^{\mathbb{N}}$	$\{(x_n)_{n \in \mathbb{N}} : x_n \in \mathbb{R}\}$
$\mathbb{R}^n_{\leq}$	$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1 \leq \cdots \leq x_n\}$
1_A	indicator function of a set A
A^c	complement of a set A
$\setminus$	set difference
$\sim$	asymptotically equivalent
o, O	Landau symbols
f'	derivative of f
$X_{i:n}$	i-th order statistic of $X_1, \ldots, X_n$
$W_{n,k}$	$= (X_{1:n}, \ldots, X_{k:n})$
$Z_{n,k}$	$= (X_{n-k+1:n}, \ldots, X_{n:n})$
$\sigma(X)$	σ-field generated by X
$\alpha(P)$	"left endpoint of a probability measure P"
$\omega(P)$	"right endpoint of a probability measure P"
$\mathcal{L}(X\|P)$	distribution of X under P
KP	distribution of a Markov kernel K under P
F^{-1}	quantile function of a distribution function F
$E_P X$	expectation of X *w.r.t.* P
$Var_P X$	variance of X *w.r.t.* P
$E_P(Y\|X)$	conditional expectation of Y given X *w.r.t.* P
$N(\mu, \sigma^2)$	normal distribution with parameters μ and σ^2
ϵ_x	dirac measure in x
λ	Lebesgue measure
dy	integrating *w.r.t.* λ
$\frac{dQ}{dP}$	likelihood ratio of Q *w.r.t.* P
$\mu \ll \nu$	μ absolutely continuous *w.r.t.* ν
$\mu \perp \nu$	$\mu(A) = 0$ and $\nu(A^c) = 0$ for some A
$L_p(\mu)$	space of the p-fold integrable functions *w.r.t.* μ
$\\| \\|_{L_p(\mu)}$	norm of $L_p(\mu)$
$\\| \\|$	variational distance

H	Hellinger distance
$\stackrel{d}{=}$	equality in distribution
$\stackrel{P}{\longrightarrow}$	convergence in P-probability
$o_{P_n}(n^0)$	$\lim_{n\to\infty} P_n\{\lvert X_n\rvert > \epsilon\} = 0$ for some sequence $(X$
$E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$	statistical experiment for a parameter set T
$\mathcal{E}(T)$	collection of all experiments for T
$\mathcal{F}(\Omega, \mathcal{A})$	set of critical functions defined on $(\Omega, \mathcal{A})$
(P, Q)	binary experiment
$P \star Q$	convolution of P and Q
$P \otimes Q$	product measure of P and Q
$E \otimes F$	product experiment of E and F
E^n	n-fold product experiment of E
$\delta(E, F)$	deficiency of E *w.r.t.* F
$\Delta(E, F)$	deficiency between E and F
Δ_α	restriction to the parameter set α
$E \sim F$	E equivalent to F
$E_n \to E$	weak convergence of E_n to E