

# ASYMPTOTIC SUFFICIENCY OF ORDER STATISTICS FOR ALMOST REGULAR WEIBULL TYPE DENSITIES

F. Marohn

*Received: Revised version: April 15, 1994*

## Abstract

Consider a location family which is defined via a Weibull type density having shape parameter  $a = 1$ . We treat the problem, which portion of the order statistics is asymptotically sufficient. It turns out that the intermediate order statistics are relevant.

## 1 Introduction and notation

Consider a location family defined via a Lebesgue-density  $f$  which is of Weibull type, that is,

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

where the shape parameter  $a > -1$  is known and  $r$  is slowly varying at zero assumed also to be known. The aim of the present paper is to clarify, which portion of the order statistics  $X_{1:n} \leq \dots \leq X_{n:n}$  of an iid sample  $X_1, \dots, X_n$  is asymptotically sufficient in the case  $a = 1$ . Thereby we restrict ourselves to the case of infinite Fisher information which covers all interesting distributions. For instance,

---

MSC 1991 subject classification: Primary 62G30

Key words and phrases: Weibull type density, intermediate order statistics, asymptotic sufficiency, local asymptotic normality

the Weibull density  $2x \exp(-x^2)$ ,  $x > 0$ , and the density of the generalized Pareto distribution  $2x1_{(0,1)}(x)$  have infinite Fisher information. Under a weak regularity condition (see (2) in the next section), Janssen and Mason [8] have shown that for the class of Weibull type densities with  $a = 1$  and infinite Fisher information the local asymptotic normality (LAN) expansion of the log-likelihood ratios of the corresponding densities holds. This result is crucial for our investigations. We remark that the case  $a = 1$  with infinite Fisher information was earlier investigated by Ibragimov and Has'minskii [5]. They established LAN under stronger regularity conditions (we repeat them after the proof of the Theorem), which imply that the slowly varying function  $r$  is a constant as  $x \downarrow 0$ .

Before we introduce some further notations and recall some facts concerning asymptotic sufficiency in the LAN case, we briefly discuss which portion of the order statistics is relevant in the two cases  $-1 < a < 1$  and  $a > 1$ . In the non-regular case  $-1 < a < 1$  the  $k(n)$  lower order statistics are sufficient whenever  $k(n)$  tends to infinity as  $n \rightarrow \infty$ , which was shown by Janssen and Reiss [9] (see also Marohn [10]). If  $a > 1$  and the Fisher information is finite, the situation is completely different since in this case the location family is  $L_2$ -differentiable and hence the LAN approximation holds: Then a fixed number  $k$  of extreme order statistics asymptotically does not contain any information on the unknown parameter. This is a consequence of a general result of Janssen and Marohn [7], where the role of extreme order statistics in general statistical models is studied. In the boundary case  $a = 1$  which has been left open, we will see that the following holds: If  $k_1(n)$  tends to infinity at a sufficiently slow rate and  $k_2(n)$  tends to infinity sufficiently fast then  $X_{k_1(n)+1:n}, \dots, X_{k_2(n):n}$  are asymptotically sufficient. Note that from the previous discussion of the LAN case we already know the existence of such a sequence  $k_1(n)$ .

Denote by  $E_P X$ ,  $Var_P X$  the expectation and the variance of a random element  $X$  under a probability measure  $P$ . The conditional expectation of  $Y$  given  $X$  w.r.t.  $P$  is denoted by  $E_P(Y|X)$ . Let  $E = (\Omega, \mathcal{A}, \{P_t : t \in T\})$  be a statistical experiment, that is,  $(\Omega, \mathcal{A})$  is a measurable space and  $\{P_t : t \in T\}$  is a family of probability measures indexed by some nonempty set  $T$ . A critical function or a test for  $E$  is a  $\mathcal{A}$ -measurable function  $\varphi : \Omega \rightarrow [0, 1]$ . By  $\mathcal{F}(\Omega, \mathcal{A})$  we denote the set of all critical functions for  $E$ . The  $n$ -fold product experiment of  $E$  is denoted by  $E^n = (\Omega^n, \mathcal{A}^n, \{P_t^n : t \in T\})$ .

Let  $(E_n) = (\Omega_n, \mathcal{A}_n, \{P_{n,h} : h \in H\})$  be a sequence of statistical experiments, where  $H$  is an Euclidean space. Assume that  $(E_n)_n$  is asymptotically normal with central sequence  $(Z_n)_n$ . Then  $(Z_n)_n$  is known to be *asymptotically sufficient* for  $(E_n)_n$  in the sense that

$$\lim_{n \rightarrow \infty} \left| \int_{\Omega_n} \varphi_n dP_{n,h} - \int_{\Omega_n} E_{P_{n,0}}(\varphi_n | Z_n) dP_{n,h} \right| = 0$$

for every  $h \in H$  and every sequence of critical functions  $\varphi_n \in \mathcal{F}(\Omega_n, \mathcal{A}_n)$ ,  $n \in \mathbb{N}$ , which is a consequence of the exponential approximation of  $(E_n)_n$ , see Strasser [12], Theorem 81.4. Moreover, if  $(X_n)_n$  is a sequence such that the  $\sigma$ -field generated by  $X_n$  contains the  $\sigma$ -field generated by  $Z_n$ ,  $n \in \mathbb{N}$ , then  $(X_n)_n$  is also asymptotically

sufficient (see the proof of Theorem 81.4 in Strasser [12]). Assume, in addition, that  $(E_n)_n$  is equicontinuous, that means, for every  $h \in H$  and every  $\varepsilon > 0$  there exists a  $\delta(\varepsilon, h)$  such that  $\|h_1 - h\| < \delta(\varepsilon, h)$  implies  $d_1(P_{n,h}, P_{n,h_1}) < \varepsilon, h_1 \in H$ , where  $d_1$  denotes the variational distance. Then  $(Z_n)_n$  is asymptotically sufficient for  $(E_n)_n$  uniformly on compact sets i.e.,

$$\lim_{n \rightarrow \infty} \sup_{t \in K} \left| \int_{\Omega_n} \varphi_n dP_{n,h} - \int_{\Omega_n} E_{P_{n,0}}(\varphi_n | Z_n) dP_{n,h} \right| = 0$$

for every compact set  $K \subset H$  and every sequence of critical functions  $\varphi_n \in \mathcal{F}(\Omega_n, \mathcal{A}_n)$  (see Strasser [12], Corollary 81.5).

## 2 The Main Result

Denote by  $X_{1:n} \leq \dots \leq X_{n:n}$  the order statistics of the canonical projections  $X_i : \mathbb{R}^n \rightarrow \mathbb{R}$  on the  $i$ -th coordinate and denote by  $P_t$  the location family with density  $f_t(\cdot) = f(\cdot - t)$ , where

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

We require  $r$  to be slowly varying at zero. By Karamata's Theorem (see Bingham et al. [1], Theorem 1.3.1)  $r$  admits in the neighborhood of 0 the representation

$$r(x) = c \exp \left( \int_x^{\delta_0} \frac{b(u)}{u} du \right), \quad 0 < x < \delta_0,$$

for some  $\delta_0 > 0$  and some constant  $c > 0$ . Here,  $b$  is a function such that  $b(u) \rightarrow 0$  as  $u \downarrow 0$ . In addition it is assumed that  $f$  is absolutely continuous on  $(0, \infty)$  such that for all  $\delta > 0$

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

In the following, we restrict ourselves to the case of infinite Fisher information  $I := \int (f'(x))^2 / f(x) dx = \infty$ . Note that  $I < \infty$  if

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

which follows from the equation  $f'(x) = (1 - b(x))r(x)$   $\lambda$ -a.e. In the case  $I = \infty$ , Janssen and Mason [8], Theorem 14.4, showed that under condition (2) the sequence of statistical experiments

$$E_n = (\mathbb{R}^n, \mathcal{B}^n, \{P_{\delta_n t}^n : t \in \mathbb{R}\})$$

is locally asymptotic normal with central sequence

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

where the rescaling rate  $\delta_n$  can be chosen according to

$$(3) \quad d_2^2(P_0, P_{\delta_n}) = \frac{1}{8n}$$

and  $d_2$  denotes the Hellinger distance. (Formula (14.68) of Theorem 14.4 contains a misprint; the limit value in (14.68) must be  $1/8$  instead of  $1$ ). The proof is based on the fact that LAN of  $(E_n)_n$  is equivalent to the condition

$$(4) \quad \frac{d_2^2(P_0, P_t)}{t^2 h(t)} \xrightarrow{t \downarrow 0} 1, \text{ where } h(t) := \frac{1}{8} \int_t^{\delta_0} \frac{r(x)}{x} dx,$$

(see Janssen [6], Theorem 2.3). Note that  $h$  is slowly varying at  $0$  and that  $h(t) \rightarrow \infty$  as  $t \downarrow 0$ . In the following we will assume that the condition  $r(x) \rightarrow c > 0$  as  $x \downarrow 0$  holds which implies that the Fisher information is infinite. Replacing  $t$  by  $\delta_n$  in (4), condition (3) implies

$$\delta_n \sim (cn \log n/2)^{-1/2}$$

(cf. Ibragimov and Has'minskii [5], Section 5 of Chapter II). The rate  $\delta_n$  lies between the rate  $n^{-1/(1+a)}$  in the non-regular case ( $-1 < a < 1$ ), and the rate  $n^{-1/2}$  in the regular case ( $a > 1$ , finite Fisher information).

Now we are ready to state our result.

**Theorem.** *Let  $f$  be a density of type (1) and let condition (2) be satisfied. Assume that  $r(x) \rightarrow c > 0$  as  $x \downarrow 0$ . Let  $k_1(n)$  and  $k_2(n)$  be sequences of positive integers such that  $k_1(n) < k_2(n)$ ,  $k_1(n) \rightarrow \infty$ ,  $k_2(n)/n \rightarrow 0$  as  $n \rightarrow \infty$ . If  $\limsup_{n \rightarrow \infty} \delta_n k_1(n) < \infty$  and  $\lim_{n \rightarrow \infty} \log k_2(n)/\log n = 1$  then the sequence of intermediate order statistics  $X_{k_1(n)+1:n}, \dots, X_{k_2(n):n}$  is asymptotically sufficient for  $(E_n)_n$  uniformly on compact sets.*

Examples for sequences  $k_i(n)$ ,  $i = 1, 2$ , which satisfy the condition of the Theorem, are  $k_1(n) = [1/\delta_n]$ ,  $k_2(n) = [n\alpha_n]$  with  $\alpha_n \rightarrow 0$  and  $\log \alpha_n/\log n \rightarrow 0$  (take for instance  $\alpha_n = 1/\log n$ ), and  $k_2(n) = [n^{1-\alpha_n}]$  with  $\alpha_n \rightarrow 0$  and  $\alpha_n \log n \rightarrow \infty$  (take for instance  $\alpha_n = 1/\log \log n$ ), where  $[ \ ]$  denotes the integer part.

**PROOF OF THE THEOREM:** The aim of the proof is to show the existence of a central sequence which is based on  $X_{k_1(n)+1:n}, \dots, X_{k_2(n):n}$ , where  $k_1(n)$  and  $k_2(n)$  satisfy the condition of the theorem. This is done by making use of a result of Csörgö et al. [3]. Throughout the proof  $Z$  denotes a standard normal random variable,  $F := F_0$  the distribution function of  $P_0$ , and  $F^{-1}(q) = \inf\{t \in \mathbb{R} : F(t) \geq q\}$  the generalized inverse of  $F$ .

First, a careful study of the proof of Lemma 14.3 in Janssen and Mason [8] shows that the assertion of this lemma remains true if the function  $x \rightarrow f'(x)/f(x)$  is replaced by  $x \rightarrow 1/x$  (the assumption  $r(x) \rightarrow_{x \downarrow 0} c$  is not needed here). Hence, the

arguments used in the proof of Theorem 14.4 in Janssen and Mason [8] show that the central sequence  $(Z_n)_n$  can be replaced by the central sequence

$$\tilde{Z}_n(X_1, \dots, X_n) := -\delta_n \sum_{i=1}^n ((1/X_i) - E_{P_0^n}(1/X_i)).$$

Note that  $E_{P_0^n}(1/X_1) = \int r(x) dx < \infty$ . Next we show that a truncated version of  $(\tilde{Z}_n)_n$  is central, precisely,

$$(5) \quad \tilde{Z}_n - \delta_n \sum_{i=1}^n \left( -\frac{1}{X_i} 1_{(0,\beta)}(X_i) - m_\beta \right) = o_{P_0^n}(1)$$

for some  $\beta > 0$ , where  $m_\beta := E_{P_0^n}((-1/X_1)1_{(0,\beta)}(X_1))$ . An application of the Tschebyschev inequality yields

$$\begin{aligned} & P_0^n \left\{ \left| \delta_n \sum_{i=1}^n \left( -\frac{1}{X_i} 1_{(\beta,\infty)}(X_i) + E_{P_0^n} \left( \frac{1}{X_1} 1_{(\beta,\infty)}(X_1) \right) \right) \right| > \varepsilon \right\} \\ & \leq \left( \frac{\delta_n}{\varepsilon} \right)^2 n \text{Var}_{P_0^n} \left( \frac{1}{X_i} 1_{(\beta,\infty)}(X_i) \right) \\ & = \left( \frac{\delta_n}{\varepsilon} \right)^2 n \left( \int_\beta^\infty \frac{r(x)}{x} dx - \left( \int_\beta^\infty r(x) dx \right)^2 \right) \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . Recall that  $\delta_n = o(n^{-1/2})$ . Hence (5) is shown.

Define

$$Y_i := -\frac{1}{X_i} 1_{(0,\beta)}(X_i), \quad 1 \leq i \leq n.$$

Then we have

$$Y_{i:n} = -\frac{1}{X_{i:n}} 1_{(0,\beta)}(X_{i:n}), \quad i = 1, \dots, n.$$

In the following we denote by  $G = G_\beta$  the distribution function of  $Y_1$ . From (5) we know that  $G$  is in the domain of attraction of the normal law. Hence, applying Theorem 1 of Csörgö et al. [3] we can find sequences  $A_n(k(n))$  and  $C_n(k(n))$  with  $k(n) \rightarrow \infty$  and  $k(n)/n \rightarrow 0$  as  $n \rightarrow \infty$  such that

$$(6) \quad A_n(k(n)) \left\{ \sum_{i=k(n)+1}^{n-k(n)} Y_{i:n} - C_n(k(n)) \right\} \rightarrow Z$$

under  $P_0^n$ . Moreover, the normalizing constants are explicitly given by

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

and

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

where  $(0 < s < 1/2)$

$$\sigma^2(s) := \int_s^{1-s} \int_s^{1-s} (\min(u, v) - uv) dG^{-1}(u) dG^{-1}(v)$$

denotes the truncated variance function of  $G$ . We choose  $k(n) := k_2(n)$ , where  $k(n)$  is the sequence defined in (6). Now, for  $D_n := nm_\beta - C_n(k_2(n))$  we obtain

$$\begin{aligned} & \delta_n \left\{ \sum_{i=k_1(n)+1}^{k_2(n)} Y_{i:n} - D_n \right\} \\ &= \delta_n \left( \sum_{i=1}^n (Y_{i:n} - m_\beta) \right) - \delta_n \sum_{i=1}^{k_1(n)} Y_{i:n} - \delta_n \sum_{i=n-k_2(n)+1}^n Y_{i:n} \\ & \quad - \frac{\delta_n}{A_n(k_2(n))} A_n(k_2(n)) \left\{ \sum_{i=k_2(n)+1}^{n-k_2(n)} Y_{i:n} - C_n(k_2(n)) \right\}. \end{aligned}$$

The exponential bound for order statistics of iid uniformly distributed random variables on  $(0,1)$  (see Reiss [11], Lemma 3.1.1) and the quantile transformation yield

$$P_0^n \left\{ \left| \delta_n \sum_{i=n-k_2(n)+1}^n Y_{i:n} \right| > \varepsilon \right\} \leq P_0^n \{ X_{n-k_2(n)+1:n} < \beta \} \rightarrow_{n \rightarrow \infty} 0.$$

Similar arguments show

$$P_0^n \left\{ \left| \delta_n \sum_{i=1}^{k_1(n)} Y_{i:n} \right| > \varepsilon \right\} \leq P_0^n \{ \delta_n k_1(n) |Y_{1:n}| > \varepsilon \} \rightarrow_{n \rightarrow \infty} 0$$

whenever  $\limsup_{n \rightarrow \infty} \delta_n k_1(n) < \infty$ . Combining the above results, we conclude that  $\delta_n (\sum_{i=k_1(n)+1}^{k_2(n)} Y_{i:n} - D_n)$  is central, that means,

$$\delta_n \left\{ \sum_{i=k_1(n)+1}^{k_2(n)} Y_{i:n} - D_n \right\} \rightarrow Z$$

under  $P_0^n$ , if  $\delta_n = o(A_n(k_2(n)))$ .

Moreover, the sequence  $(E_n)_n$  is equicontinuous. This follows from Lemma 2.1 of Janssen [6], since the sequence  $(E_n)_n$  is translation invariant and the limit experiment is continuous. Therefore, to prove the Theorem it remains to show that  $\delta_n = o(A_n(k_2(n)))$  is equivalent to  $\lim_{n \rightarrow \infty} \log k_2(n) / \log n = 1$ .

We show  $\tau(s) \sim \sigma^2(s)$  as  $s \downarrow 0$ , where

$$\tau(s) := \int_s^{1-s} (G^{-1}(u))^2 du.$$

We have explicitly

$$G(y) = \begin{cases} F(-1/y) & , \text{ for } y < -1/\beta \\ F(\beta) & , \text{ for } y \in [-1/\beta, 0) \\ 1 & , \text{ for } y \geq 0 \end{cases}$$

and

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

Using the quantile transformation and the transformation theorem of integrals we obtain for small  $s$

$$\begin{aligned} (7) \quad \int_s^{1-s} (G^{-1}(u))^2 du &= \int_s^{F(\beta)} (G^{-1}(u))^2 du \\ &= \int_{G^{-1}(s)}^{G^{-1}(F(\beta))} u^2 dG(u) \\ &= \int_{-1/F^{-1}(s)}^{-1/\beta} u^2 dG(u) \\ &= - \int_{-1/F^{-1}(s)}^{-1/\beta} u^2 \frac{1}{u^3} r\left(-\frac{1}{u}\right) du \\ &= \int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} r\left(\frac{1}{u}\right) du. \end{aligned}$$

Since the integral (7) diverges as  $s \downarrow 0$  we have  $\lim_{s \downarrow 0} \tau(s) = \infty$ . This, together with the fact that  $G$  is in the domain of attraction of the normal law, implies  $\tau(s) \sim \sigma^2(s)$ , see the proof of Proposition A.2 given in the Appendix of Csörgő et al. [2]. An application of the first part of Lemma 1.2.1 of de Haan [4] shows

$$\begin{aligned} \int_s^{1-s} (G^{-1}(u))^2 du &\sim c \int_{1/\beta}^{1/F^{-1}(s)} \frac{1}{u} du \\ &\sim c(\log(\beta) - \log(F^{-1}(s))). \end{aligned}$$

Combining the above results and taking into account  $F^{-1}(s) = s^{1/2}L(s)$  for some slowly varying function  $L$  (see Bingham et al. [1], Theorem 1.5.12), we obtain  $\tau(s) \sim -\frac{\varepsilon}{2} \log s$ . From this we conclude that  $\delta_n/A_n(k_2(n)) \rightarrow 0$  iff

$$(8) \quad -\frac{n^{1/2} \log^{1/2}(k_2(n)/n)}{n^{1/2} \log^{1/2}(n)} = \left(1 - \frac{\log(k_2(n))}{\log(n)}\right)^{1/2} \rightarrow 0.$$

The proof is complete.  $\square$

**Remark.** The preceding proof shows in particular that also in the case  $a = 1$  a fixed number of extreme order statistics  $X_{1:n}, \dots, X_{k:n}$  ( $k$  fixed) asymptotically does not contain any information on the unknown (location) parameter i.e., is not asymptotically sufficient. We mention that throughout the proof  $k_1(n)$  and  $k_2(n)$  can be replaced by 1 and  $k$ , respectively (in particular (6) remains true, see Remark 2 in Csörgő et al. [3]). But (8) is violated, since  $\delta_n \sim A_n(k)$ .

Under the stronger regularity assumption that  $f$  admits in the right neighborhood of 0 the representation

$$f(x) = xr(x) = cx + \psi(x), \quad x > 0,$$

where  $\psi(x) := (r(x) - c)x$  is twice continuously differentiable with  $\psi(0) = \psi'(0) = 0$ , Ibragimov and Has'minskii [5], Chapter 2, Section 5, showed LAN with central sequence

$$\zeta_n = \delta_n \sum_{i=1}^n \left( \frac{1}{X_i} 1_{A_{n,i}} - E_{P_0^n} \left( \frac{1}{X_i} 1_{A_{n,i}} \right) \right),$$

where  $\delta_n = (c \log n/2)^{-1/2}$  and

$$A_{n,i} = \left\{ 1/(n^{1/2} \log^{1/4} n) < X_i \leq 1/\log n \right\}.$$

Note that Theorem 1 of Csörgő et al. [3], which was an essential tool in our proof, is not applicable to  $(\zeta_n)_n$  since the distribution of  $(1/X_i)1_{A_{n,i}}$  depends on  $n$ .

**Acknowledgement.** The author is indebted to M. Falk and A. Janssen for fruitful discussions.

## References

- [1] Bingham, N.H., Goldie, C.M. and Teugels, J.L.: Regular Variation. Cambridge University Press, Cambridge, 1987
- [2] Csörgő, M., Csörgő, S., Horváth, L. and Mason, D.M.: Normal and stable convergence of integral functions of the empirical distribution function. Ann. Probab. 14 (1986), 86–118
- [3] Csörgő, S., Horváth, L. and Mason D.M.: What portion of a sample makes a partial sum asymptotically stable or normal? Probab. Theory Rel. Fields 72 (1986), 1–16
- [4] Haan, L. de: On Regular Variation and its Application to the Weak Convergence of Sample Extremes. Math. Centre Tracts 32, Amsterdam, 1975
- [5] Ibragimov, I.A. and Has'minskii, R.Z.: Statistical Estimation. Application of Mathematics, Vol. 16. Springer, New York, 1981



- [6] Janssen, A.: Limit of translation invariant experiments. *J. Multivar. Anal.* 20 (1986), 129–142
- [7] Janssen, A. and Marohn, F.: On statistical information of order statistics, local extreme value alternatives, and Poisson point processes. *J. Multivar. Anal.* 48 (1994), 1–30
- [8] Janssen, A. and Mason, D. M.: *Non-Standard Rank Tests. Lecture Notes in Statistics 65.* Springer, Berlin-Heidelberg, 1990
- [9] Janssen, A. and Reiss, R.-D.: Comparison of location models of Weibull type samples and extreme value processes. *Probab. Theory Rel. Fields* 78 (1988), 273–292
- [10] Marohn, F.: Global sufficiency of extreme order statistics in location models of Weibull type. *Probab. Theory Rel. Fields* 88 (1991), 261–268
- [11] Reiss, R.-D.: *Approximate Distributions of Order Statistics (with Applications to Nonparametric Statistics).* Springer Series in Statistics, New York, 1989
- [12] Strasser, H.: *Mathematical Theory of Statistics. De Gruyter Studies in Math. 7,* Berlin-New York, 1985

Frank Marohn  
Mathematisch-Geographische Fakultät  
Katholische Universität Eichstätt  
D-85071 Eichstätt, Germany

FRANK.MAROHN@KU-EICHSTAETT.DE