

AD-A189 817

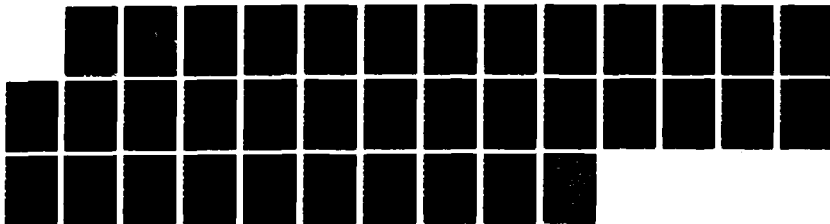
APPROXIMATIONS IN EXTREME VALUE THEORY(U) NORTH
CAROLINA UNIV AT CHAPEL HILL CENTER FOR STOCHASTIC
PROCESSES R L SMITH SEP 87 TR-205 AFOSR-TR-87-1848
F49620-85-C-0144

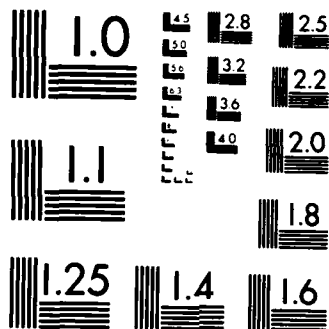
1/1

UNCLASSIFIED

F/G 12/3

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-A189 817

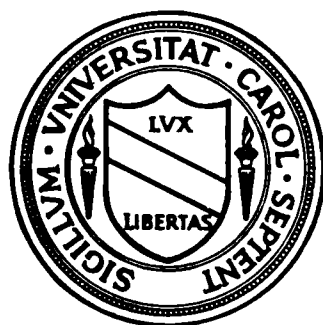
DTIC FILE COPY

2

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina

AFOSR-TR. 87-1848



DTIC
ELECTE
S JAN 07 1988 D
H

APPROXIMATIONS IN EXTREME VALUE THEORY

by

Richard L. Smith

Technical Report No. 205

September 1987

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution unlimited

87 12 29 111

2189 817

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 205			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR-87-1848		
6a. NAME OF PERFORMING ORGANIZATION University of North Carolina		6b. OFFICE SYMBOL (If applicable)		7a. NAME OF MONITORING ORGANIZATION AFOSR/NM	
6c. ADDRESS (City, State, and ZIP Code) Statistics Dept. 321-A Phillips Hall 039-A Chapel Hill, NC 27514				7b. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (If applicable) NM		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR No. F49620 85C 0144.	
8c. ADDRESS (City, State, and ZIP Code) AFOSR/NM Bldg 410 Bolling AFB DC 20332-6448		10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 61102F		PROJECT NO. 2304	TASK NO. A5
				WORK UNIT ACCESSION NO.	
11. TITLE (Include Security Classification) Approximations in extreme value theory					
12. PERSONAL AUTHOR(S) Smith, R.L.					
13a. TYPE OF REPORT Preprint		13b. TIME COVERED FROM 9/87 TO 8/88		14. DATE OF REPORT (Year, Month, Day) September 1987	
15. PAGE COUNT 52					
16. SUPPLEMENTARY NOTATION N/A					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB-GROUP	Key Words & Phrases: Extreme value theory, generalised extreme value distribution, generalised Pareto distribution, Hellinger distance, rates of convergence, regular variation with (over)		
19. ABSTRACT (Continue on reverse if necessary and identify by block number)					
Following a survey of rates of convergence in extreme value theory, a new class of approximations is developed and compared with existing approximations based on the extreme value distributions. Convergence in Hellinger distance is established, this distance measure being chosen because of its statistical applications. Numerical examples confirm the superiority of the new approximation. <i>Keywords:</i> <i>Remainder; total variation distance.</i>					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION Unclassified/unlimited		
22a. NAME OF RESPONSIBLE INDIVIDUAL Maj. Brian Woodruff			22b. TELEPHONE (Include Area Code) (202) 767-5027		22c. OFFICE SYMBOL NM

87 12 29 111

APPROXIMATIONS IN EXTREME VALUE THEORY

RICHARD L. SMITH*

University of Surrey, Guildford, U.K.

and

Center for Stochastic Processes

University of North Carolina, Chapel Hill, NC USA

Following a survey of rates of convergence in extreme value theory, a new class of approximations is developed and compared with existing approximations based on the extreme value distributions. Convergence in Hellinger distance is established, this distance measure being chosen because of its statistical applications. Numerical examples confirm the superiority of the new approximation.

Keywords: Extreme value theory, Generalised Extreme value distribution, Generalised Pareto distribution, Hellinger distance, Rates of convergence, Regular variation with remainder, Total variation distance.



Research supported by Air Force Office of Scientific Research Grant No. F49620 85C 0144.

*Permanent address: Department of Mathematics, University of Surrey, Guildford, GU2 5XH, England

Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

1. Rates of convergence in extreme value theory

Let F denote a probability distribution function and suppose there exist constants $a_n > 0$ and b_n , for $n \geq 1$, and a non-degenerate distribution function G such that

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = G(x). \quad (1.1)$$

Then G may be taken to be one of the "three types"

$$\Lambda(x) = \exp(-e^{-x}), \quad (1.2)$$

$$\phi_\alpha(x) = \begin{cases} 0 & x < 0, & (\alpha > 0) \\ \exp(-x^{-\alpha}), & x > 0 \end{cases} \quad (1.3)$$

$$\psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha), & x < 0, & (\alpha > 0) \\ 1 & x > 0. \end{cases} \quad (1.4)$$

Alternatively, G may be taken to be of "Generalized Extreme Value" form

$$G(x) = \exp\{-(1+\gamma x)_+^{-1/\gamma}\} \quad (1.5)$$

where $y_+ = \max(y, 0)$ and $-\infty < \gamma < \infty$: the case $\gamma = 0$ interpreted as the limit $\gamma \rightarrow 0$, which is (1.2). The range of the distribution in this case is the set

$$R_\gamma = \{x: 1+\gamma x > 0\},$$

i.e. $(-\gamma^{-1}, \infty)$ if $\gamma > 0$, $(-\infty, \gamma^{-1})$ if $\gamma < 0$, $(-\infty, \infty)$ if $\gamma = 0$. These results are well known and we refer to the books of Galambos (1978) and Leadbetter, Lindgren and Rootzén (1983) for details.

Interest in rates of convergence started with the very early paper of Fisher and Tippett (1928). They showed for normal extremes, that the appropriate limit is (1.2), but they argued that a "penultimate" approximation within the family (1.4) is better in practice. In the context of (1.5), this is equivalent to saying that the limiting value $\gamma=0$ is better replaced by a sequence of values γ_n , where $\gamma_n \uparrow 0$ as $n \rightarrow \infty$.

The modern theory of rates of convergence may be considered to have begun

with the works of Anderson (1971, 1976) and Galambos (1978, Section 2.10). They gave general formulae for computing pointwise rates of convergence. Since then, the theory has developed in three main directions.

The first direction has been towards the computation of explicit upper bounds for

$$\sup_x |F^n(a_n x + b_n) - G(x)| \quad (1.6)$$

when a_n, b_n are chosen appropriately. Hall and Wellner (1979) obtained the sharp upper bound $n^{-1}(2+n^{-1})e^{-2}$ when F is exponential, and Hall (1979) obtained the bound $3(\log n)^{-1}$ when F is normal, both with $G = \Lambda$. Davis (1982) combined the Hall-Wellner result with the probability integral transform to obtain a result for general F , but it requires rather detailed computations to apply it to any particular case. The best results in this direction have been obtained by Resnick (1986), who gave general results assuming essentially the von Mises conditions, introduced in Section 2. An interesting alternative approach, based on Zolotarev's method of ideal metrics, is given by Zolotarev and Rachev (1985), though this is currently confined to the Φ_α and Ψ_α limits.

The second direction of study stems from Anderson (1971), and is really more concerned with the structure of the remainder term than with explicit bounds. Smith (1982) derived uniform rates of convergence to Φ_α assuming a "slow variation with remainder" condition

$$\frac{-\log F(tx)}{-\log F(t)} = x^{-\alpha} \{1 + o(g(t))\}$$

for each fixed $x > 0$, where $g(t) \rightarrow 0$ at $t \rightarrow \infty$. A simple transformation allows this approach also to be applied to Ψ_α . Cohen (1982b) took rather a similar approach to the limit Λ , starting with the de Haan (1970) representation

$$-\log F(x) = c(x) \exp \left\{ -\int_X^x \frac{a(t)}{f(t)} dt \right\} \quad (x \geq X)$$

($c(x) \rightarrow c_1$, $a(x) \rightarrow 1$, f differentiable and $f'(x) \rightarrow 0$). As was pointed out by Anderson (1984), the alternative representation with $a(t) \equiv 1$, due to Balkema and de Haan (1972), allows some simplification of Cohen's results. In most

cases this approach leads to improved approximations for F^n . Rates of convergence of the penultimate approximation have also been established (Cohen 1982a,b, Gomes 1984), the normal case for instance being of $O\{(\log n)^{-2}\}$. The two directions for Φ_α have partly been brought together by Omev and Rachev (1987).

The third direction of study concerns the extension of the problem from statements about (1.1) or (1.6) to more general convergence criteria involving the joint distribution of several largest order statistics and convergence of densities instead of distribution functions. These considerations are especially relevant for statistical applications. Reiss (1981) obtained an asymptotic expansion for the distribution of the k largest order statistics from the uniform distribution, with rates of convergence (see also Kohne and Reiss, 1983) and Falk (1986) extended this to general distributions via the probability integral transform. This would appear to be a very powerful approach, though Falk's conditions are not easy to verify in particular cases. Weissman (1984) took a different point of view, asking how fast k could grow (as a function of n) for convergence to remain valid. Reiss (1984) pointed out the importance of Hellinger distance for statistical applications.

The present work is aimed at partly unifying these different approaches, both with a view to combining the results for the three domains of attraction, and incorporating the approach of Reiss and Falk within the general scheme. Convergence in Hellinger distance implies convergence in total variation distance, which in turn implies uniform convergence of distribution functions. Therefore it seems to us that Hellinger distance is the most appropriate distance measure to use. The usefulness of Hellinger distance in statistical applications is explained briefly in Section 3.

The structure of the paper is as follows. Section 2 develops the approximations we use. The emphasis here is on having a single form of improved approximation valid for all three types. We also extend the notion of

penultimate approximation. In Section 3, proofs of convergence in Hellinger distance are given. These cover both the classical and threshold forms of extreme value approximation, and are expanded also to cover the joint distribution of k largest order statistics (for fixed k). Finally in Section 4 we give numerical examples of our new approximations, demonstrating that they really do make a considerable improvement on the classical extreme value approximations.

2. Development of the approximations

Suppose F has density $f(x) = dF(x)/dx$ defined on the range (x_*, x^*) where

$$x_* = \inf\{x: F(x) > 0\} \geq -\infty, \quad x^* = \sup\{x: F(x) < 1\} \leq \infty.$$

Then we may write

$$-\log F(x) = \exp \left\{ -\int_{x_*}^x \frac{dt}{\phi(t)} \right\}, \quad x_* < x < x^* \quad (2.1)$$

where

$$\phi(x) = \frac{-F(x) \log F(x)}{f(x)}. \quad (2.2)$$

Sometimes we use the alternative representation

$$1-F(x) = \exp \left\{ -\int_{x_*}^x \frac{dt}{\phi(t)} \right\}, \quad x_* < x < x^* \quad (2.3)$$

where

$$\phi(x) = \frac{1-F(x)}{f(x)}. \quad (2.4)$$

Whichever form is adopted, we shall assume ϕ is continuously differentiable and

$$\lim_{x \uparrow x^*} \phi'(x) = \gamma \quad (2.5)$$

for some real γ .

Equation (2.5) is one form of the well-known von Mises conditions which are sufficient though not necessary for the domain of attraction of an extreme value distribution (see de Haan (1976)). It makes no difference to the limit which of the two definitions of ϕ is adopted, and the limit is given by (1.5) with the same γ . The precise significance of (2.5) has been given by Pickands (1986): it is a necessary and sufficient condition for "twice-differentiable"

convergence, meaning that not only (1.1) holds but also convergence of the corresponding densities and derivatives of the densities. Convergence of densities alone has been studied also by Sweeting (1985), following de Haan and Resnick (1982). For our present purposes, convergence of densities is relevant but our main motivation for assuming (2.5) is mathematical tractability.

From (2.1) we have

$$\frac{-\log F(u + x\phi(u))}{-\log F(u)} = \exp \left\{ -\int_0^x \frac{\phi(u)}{\phi(u+s\phi(u))} ds \right\}, \quad (2.6)$$

$$\frac{\phi(u + s\phi(u))}{\phi(u)} = 1 + \int_0^s \phi'(u + w\phi(u)) dw. \quad (2.7)$$

By the mean value theorem, for each s

$$\frac{\phi(u + s\phi(u))}{\phi(u)} = 1 + s\phi'(y)$$

where y is between u and $u + s\phi(u)$. Consequently

$$\int_0^x \left\{ \frac{\phi(u)}{\phi(u + s\phi(u))} - \frac{1}{1 + s\phi'(y)} \right\} ds$$

is a continuous function of y , takes on both positive and negative values as y ranges from u to $u + x\phi(u)$ (unless ϕ' is constant), and so is zero for at least one y . Substituting in (2.6),

$$\frac{-\log F(u + x\phi(u))}{-\log F(u)} = \{1 + x\phi'(y)\}^{-1/\phi'(y)} \quad (2.8)$$

for some y between u and $u + x\phi(u)$. Now let us define, for each $n \geq 1$, b_n such that $-\log F(b_n) = n^{-1}$ (well-defined, since F is continuous) and let $a_n = \phi(b_n)$. Substituting $u = b_n$ in (2.8),

$$F^n(a_n x + b_n) = \exp[-\{1 + x\gamma_n(x)\}^{-1/\gamma_n(x)}] \quad (2.9)$$

where $\gamma_n(x) = \phi'(y)$, y being as in (2.8). If $a_n x + b_n$ is outside the range (x_*, x^*) then we interpret both sides of (2.9) to be 0 or 1 as appropriate.

Now suppose (2.5) holds, and let x be a fixed number in the range R_γ (recall (1.6)). It is easily verified from (2.5) that

$$\lim_{u \rightarrow \infty} \frac{\phi(u)}{u} = \gamma \quad (x^* = \infty); \quad \lim_{u \uparrow x^*} \frac{\phi(u)}{x^* - u} = -\gamma \quad (x^* < \infty) \quad (2.10)$$

and hence that $u + x\phi(u) \uparrow x^*$ uniformly over finite ranges of x as $u \uparrow x^*$.

Thus (2.9) tends to (1.5) as $n \rightarrow \infty$. This provides an independent proof of the sufficiency of (2.5) for (1.1), but in a form particularly well suited for the machinations to follow.

If we start with (2.3) in place of (2.1), then the argument is the same up to (2.8), which now reads

$$\frac{1 - F(u + x\phi(u))}{1 - F(u)} = \{1 + x\phi'(y)\}^{-1/\phi'(y)} \quad (2.11)$$

which in turn implies for $x > 0$ that

$$\lim_{u \uparrow x^*} \frac{1 - F(u + x\phi(u))}{1 - F(u)} = \begin{cases} (1 + x\gamma)^{-\frac{1}{\gamma}}, & (\gamma \geq 0 \text{ and } \gamma < 0, 0 < x < -\gamma^{-1}) \\ 0 & (\gamma < 0, x > -\gamma^{-1}) \end{cases} \quad (2.12)$$

This is the Generalised Pareto distribution introduced by Pickands (1975), which is particularly useful as a model of excesses over high thresholds. Some statistical applications are given by Smith (1984, 1987), Davison (1984), Joe (1987) and Hosking and Wallis (1987).

So far we have replaced $\phi'(y)$ by γ . In some sense, however, what we are doing is expanding the tail of F about u , so it may make more sense to approximate $\phi'(y)$ by $\phi'(u)$. This is especially true if $\gamma=0$ for then, by virtue of (2.10), $u + x\phi(u)$ is (for fixed x as $u \uparrow x^*$) much closer to u than to x^* . Thus we replace γ in (1.5) by $\gamma_n = \phi'(b_n)$, γ in (2.12) by $\gamma(u) = \phi'(u)$. The first of these is the penultimate approximation, precisely as it is defined by Gomes (1984) and equivalently to the definition of Cohen (1982b). Although Cohen and Gomes both prove that the penultimate approximation is better in general than the ultimate approximation (in the sense of giving a faster rate of convergence) they do not really give any motivation for considering it in the first place. The foregoing may provide some. Moreover, it also suggests that we could do the same thing when $\gamma \neq 0$, providing a penultimate approximation in this case also. Some of the evidence given later will suggest that this is an advantageous thing to do. So far as we are aware, this is the

first time that a penultimate approximation has been suggested when $\gamma \neq 0$.

If we want to go beyond this, the logical next step in view of (2.7) is to consider an expansion of $\phi'(u + w\phi(u))$ about $\phi'(u)$. At this point, however, we interrupt the proceedings to give some examples. These will serve both to illustrate what has been done so far, and to motivate the next step.

Example 1 Suppose $x^* = +\infty$ and

$$-\log F(x) = Cx^{-\alpha} \{1 + Dx^{-\beta} + O(x^{-\beta-\epsilon})\}, \quad x \rightarrow \infty. \quad (2.13)$$

where $C, \alpha, \beta, \epsilon$ are positive constants and D is real. This includes nearly all practical examples in the domain of attraction of (1.3), e.g. Pareto, Cauchy, t , F . We assume the relation (2.13) is twice differentiable, in the sense that we can differentiate term by term without affecting the order of the O -term. It follows that

$$\phi'(x) = \frac{1}{\alpha} + \frac{D\beta(\beta-1)}{\alpha^2} x^{-\beta} + O(x^{-\beta-\epsilon}). \quad (2.14)$$

Thus $\gamma = \alpha^{-1}$ and the rate of convergence in (1.1) is $O(\phi'(b_n) - \gamma) = O(b_n^{-\beta}) = O(n^{-\beta/\alpha})$ as in Smith (1982). However, in the case $\beta = 1$ the second term in (2.14) is 0 and so the rate of convergence is $o(n^{-1/\alpha})$. Smith (1982) showed the conventional approximation

$$F^n(b_n x) \rightarrow \phi_\alpha(x) \quad (F(b_n) = \exp(-n^{-1}))$$

achieves $O(n^{-\beta/\alpha})$ for all β and, though a way of reducing this to $o(n^{-1/\alpha})$ when $\beta = 1$ was proposed, the construction is artificial. Incidentally, the rate of $O(n^{-\beta/\alpha})$ is optimal (amongst all choices of a_n, b_n) when $\beta \neq 1$.

Continuing from (2.14), we have when $\beta \neq 1$

$$\begin{aligned} \phi'(u + x\phi(u)) - \phi'(u) &\sim \frac{D\beta(\beta-1)}{\alpha^2} u^{-\beta} \left[\left\{ 1 + x \frac{\phi(u)}{u} \right\}^{-\beta} - 1 \right] \\ &\sim \frac{D\beta(\beta-1)}{\alpha^2} u^{-\beta} \left[\{1 + x\phi'(u)\}^{-\beta} - 1 \right] \end{aligned} \quad (2.15)$$

using (2.5) and (2.10).

If we start with $1-F(x)$ in place of $-\log F(x)$ in (2.13), then the corresponding results hold for the threshold approximation (2.12).

Example 2 Suppose $x^* < \infty$ and

$$-\log F(x) = C(x^* - x)^\alpha \{1 + D(x^* - x)^\beta + O((x^* - x)^{\beta+\epsilon})\}, \quad x \uparrow x^* \quad (2.16)$$

($C, \alpha, \beta, \epsilon$ positive, D real) and that this relation is twice differentiable. If we replace $F(x)$ by $1 - F(x^* - x)$, this includes many distributions in the minimum domain of attraction of the Weibull distribution, with applications to reliability and elsewhere. In this case

$$\phi'(x) = -\frac{1}{\alpha} + \frac{D\beta(\beta+1)}{\alpha^2} (x^* - x)^\beta + O((x^* - x)^{\beta+\epsilon}) \quad (2.17)$$

so $\gamma = -\alpha^{-1}$ and

$$\begin{aligned} \phi'(u + x\phi(u)) - \phi'(u) &\sim \frac{D\beta(\beta+1)}{\alpha^2} \{(x^* - u - x\phi(u))^\beta - (x^* - u)^\beta\} \\ &\sim \frac{D\beta(\beta+1)}{\alpha^2} (x^* - u)^\beta \{(1 + x\phi'(u))^\beta - 1\}. \end{aligned} \quad (2.18)$$

In this case the rate of convergence in (1.1) is $O(n^{-\beta/\alpha})$ and there is no possibility of improving this by a different choice of a_n and b_n (Smith 1982). Again, if we start with $1 - F(x)$ in (2.16) then we get similar approximations for the threshold distribution.

In neither example so far have we emphasized the penultimate approximation, but numerical evidence of its efficacy will be given later.

Example 3 Let $\gamma = 0$. If we slightly strengthen the conditions for what Cohen (1982b) called Class N, then it is valid to make a Taylor expansion

$$\phi'(u + x\phi(u)) - \phi'(u) \sim x\phi(u)\phi''(u). \quad (2.19)$$

Examples include most well-known distributions in the domain of attraction of Λ , e.g. normal, log normal, gamma, Weibull, but not the exponential or logistic distributions for which ϕ' decreases exponentially fast. These are, in fact, the most important cases to which the theory we are going to develop does not apply, though since the reason is essentially that the convergence occurs too quickly, we would argue that this exclusion is not of importance for statistical applications.

It is not obvious how to combine (2.15), (2.18) and (2.19) into a single general formula. We shall, however, make a proposal. Define the family of functions on $0 < x < \infty$,

$$h_\rho(x) = \int_1^x u^{\rho-1} du = \begin{cases} \frac{x^\rho - 1}{\rho}, & \rho \neq 0 \\ \log x, & \rho = 0. \end{cases} \quad (2.20)$$

This function often arises as a remainder term in the theory of slow variation (Smith, 1982, Goldie and Smith 1987).

We assume that there exist real c and ρ and a non-negative function g , with $g(u) \rightarrow 0$ as $u \uparrow x^*$, such that

$$\lim_{u \uparrow x^*} \frac{\phi'(u) \{ \phi'(u + w\phi(u)) - \phi'(u) \}}{g(u) h_\rho(1 + w\phi'(u))} = c, \quad (2.21)$$

for each $w \in \mathbb{R}_+$. We further assume that $\phi'(x)$ is non-zero and of the same sign for all sufficiently large $x < x^*$, and that ρ is either 0 or of the opposite sign to ϕ' . Examples:

Example 1 $\rho = -\beta$, $g(u) = u^{-\beta}$, $c = -\frac{D\beta^2(\beta - 1)}{\alpha^3}$,

Example 2 $\rho = \beta$, $g(u) = (x^* - u)^\beta$, $c = -\frac{D\beta^2(\beta + 1)}{\alpha^3}$,

Example 3 ρ undetermined, $g(u) = \phi(u) |\phi''(u)|$, $c = \pm 1$.

Example 3 relies on $h_\rho(1 + w\phi'(u)) \sim w\phi'(u)$ as $\phi'(u) \rightarrow 0$. The fact that ρ is undetermined in this case is not important, since the results we derive are independent of ρ (in this case) up to the claimed order of approximation. Note that we also allow $c = 0$, so the $\beta = 1$ case of Example 1 is also included, though in this case a more logical approach would presumably be to take the next term in the expansion.

Substituting from (2.21) in (2.7) and then (2.6), setting $u = b_n$ where $-\log F(b_n) = n^{-1}$, $a_n = \phi(b_n)$, $\gamma_n = \phi'(b_n)$, $r_n = g(b_n)$, routine manipulations

lead to

$$F^n(a_n x + b_n) = \exp[-(1+x\gamma_n)_+^{-1/\gamma_n} \{1 + cr_n H_\rho(x, \gamma_n)\}] + o(r_n) \quad (2.22)$$

for each fixed x , where

$$H_\rho(x, \eta) = \begin{cases} \frac{h_\rho(1+x\eta) + \rho h_{-1}(1+x\eta) - (\rho+1)\log(1+x\eta)}{\rho(\rho+1)\eta^3}, & 1+x\eta > 0, \\ 0, & 1+x\eta \leq 0. \end{cases} \quad (2.23)$$

A rigorous derivation of (2.22) will be given in the next section. The cases

$\rho = 0$, $\rho = -1$ are defined by taking appropriate limits as

$$H_{-1}(x, \eta) = \frac{(1+x\eta)^{-1} \log(1+x\eta) + \log(1+x\eta) - 2(1-(1+x\eta)^{-1})}{\eta^3}$$

$$H_0(x, \eta) = \frac{\frac{1}{2} \log^2(1+x\eta) - \log(1+x\eta) + 1 - (1+x\eta)^{-1}}{\eta^3}$$

when $x > 0$. Note also that

$$\lim_{\eta \rightarrow 0} H_\rho(x, \eta) = \frac{x^3}{6}, \quad (2.24)$$

confirming that, in the case $\gamma_n \rightarrow 0$, $H_\rho(x, \gamma_n)$ in (2.22) may be replaced with $x^3/6$ (independent of ρ) without affecting the claimed rate of convergence.

For the threshold approximation (2.12), we should start with (2.3) instead of (2.1); the result then obtained is

$$\frac{1 - F(u + x\phi(u))}{1 - F(u)} = \{1 + x\phi'(u)\}_+^{-1/\phi'(u)} \{1 + cg(u)H_\rho(x, \phi'(u))\} + o(g(u)) \quad (2.25)$$

for each $x > 0$.

3. Hellinger convergence

Define

$$\left. \begin{aligned} F_n(x) &= F^n(a_n x + b_n), \\ G_n(x) &= \exp[-(1+x\gamma_n)_+^{-1/\gamma_n} \{1 + cr_n H_\rho(x, \gamma_n)\}]. \end{aligned} \right\} \quad (3.1)$$

In (2.22), we asserted that $|F_n(x) - G_n(x)| = o(r_n)$ for each fixed x . It is natural to ask whether this result holds uniformly over all x .

This is not the only sense, however, in which the closeness of F_n and G_n

could be measured. Another question is whether the densities $f_n = dF_n/dx$, $g_n = dG_n/dx$ converge uniformly at rate $o(r_n)$. If they do, then it follows from an easy extension of Scheffe's Lemma that

$$\sup_B \left| \int_B f_n(x) dx - \int_B g_n(x) dx \right| = o(r_n), \quad (3.2)$$

where the supremum is over all Borel sets B . This is the mode of convergence used by Falk (1986). Another measure studied by Reiss (1984) is Hellinger distance:

$$H(f_n, g_n) = [\int \{f_n^{1/2}(x) - g_n^{1/2}(x)\}^2 dx]^{1/2}. \quad (3.3)$$

If $H(f_n, g_n) = o(r_n)$ then (3.2) is immediate.

Equations (3.2) and (3.3) have direct statistical interpretation. For example, if B is the rejection region of some test calculated under the assumption that g_n is the correct distribution, then (3.2) says that the error in the computed probability of rejection is at most $o(r_n)$. The importance of Hellinger distance arises from the following inequality, pointed out by Reiss. Suppose we have N independent observations from each of f_n and g_n , and let $f_n^{(N)}$, $g_n^{(N)}$ denote the resulting joint densities. Then

$$H(f_n^{(N)}, g_n^{(N)}) \leq N^{\frac{1}{2}} H(f_n, g_n).$$

Suppose $H(f_n, g_n) = o(r_n)$ and $n \rightarrow \infty$, $N \rightarrow \infty$ such that Nr_n is bounded. Then

$$H(f_n^{(N)}, g_n^{(N)}) \rightarrow 0$$

so that the total variation distance between $f_n^{(N)}$ and $g_n^{(N)}$ is asymptotically negligible, i.e. statistical calculations carried out as if g_n was the correct density remain valid when sampling from f_n . This provides an alternative method of justifying statistical calculations based on extreme value approximations, avoiding the awkward moment-convergence technicalities of Goldie and Smith (1987), Smith (1987), Cohen (1987a, 1987b) and Joe (1987).

The main additional condition needed to prove Hellinger convergence is $\gamma > -\frac{1}{2}$. This condition is easily understood statistically, since when $\gamma \leq -\frac{1}{2}$ the problem is non-regular and standard maximum likelihood techniques fail.

Alternatives in these cases are proposed by Smith (1985, 1987).

For further information about Hellinger distance and total variation distance we refer to Ibragimov and Has'minskii (1981) or Section 4.2 of LeCam (1986). The following result is adapted from Theorem 7.6 of Ibragimov and Has'minskii:

Lemma 3.1 Let $f_o(x; \theta)$ denote a family of non-negative functions indexed by vector parameter $\theta \in \Theta$. Let $g_o(x; \theta) = f_o^{\frac{1}{2}}(x; \theta)$ with gradient vector ∇g with respect to θ . Suppose f_1, f_2 are two functions such that, for each x in a set B , there exist $\theta_i(x)$ ($i=1,2$) such that $f_i(x) = f(x; \theta_i(x))$. Suppose $\theta_i(x) \in \theta^* \subset \Theta$ for each $x \in B$, $i=1,2$. Then

$$\int_B \{f_1^{\frac{1}{2}}(x) - f_2^{\frac{1}{2}}(x)\}^2 dx \leq \sup_{x \in B} |\theta_1(x) - \theta_2(x)|^2 \int_B \sup_{\theta \in \theta^*} |\nabla g_o(x, \theta)|^2 dx. \quad (3.4)$$

Remark 3.2 This differs from Ibragimov and Has'minskii in that θ_1 and θ_2 depend on x ; i.e. f_i do not have to be members of the family $f_o(x; \theta)$ but only close to it. Finiteness of the integral in (3.4) is closely related to the boundedness (over θ^*) of the trace of the Fisher information matrix.

Proof. We have

$$\begin{aligned} f_1^{\frac{1}{2}}(x) - f_2^{\frac{1}{2}}(x) &= g_o(x; \theta_1(x)) - g_o(x; \theta_2(x)) \\ &= \int_0^1 \{\theta_1(x) - \theta_2(x)\}^T \nabla g_o\{\theta_1(x) + t(\theta_2(x) - \theta_1(x))\} dt, \end{aligned}$$

so that

$$\{f_1^{\frac{1}{2}}(x) - f_2^{\frac{1}{2}}(x)\}^2 \leq |\theta_1(x) - \theta_2(x)|^2 \int_0^1 |\nabla g_o\{\theta_1(x) + t(\theta_2(x) - \theta_1(x))\}|^2 dt.$$

Now just integrate with respect to x .

We now come to our main result.

Theorem 3.3 Suppose ϕ , defined from (2.1), satisfies (2.5) with $\gamma > \frac{1}{2}$, and (2.21) with its associated conditions. Define b_n by $F(b_n) = \exp(-n^{-1})$, $a_n = \phi(b_n)$, $\gamma_n = \phi'(b_n)$, $r_n = g(b_n)$. Define F_n, G_n by (3.1) with associated density f_n, g_n . Suppose there exist, for each u , variables $s_1(u) > 0$, $s_2(u) < 0$ such that

$$\lim_{u \uparrow x} \frac{\{1 + \phi'(u)s_1(u)\}^{-1/\phi'(u)}}{g^2(u)} = 0, \quad (3.5)$$

$$\lim_{u \uparrow x} \frac{\exp[-\{1 + \phi'(u)s_2(u)\}^{-1/\phi'(u)}]}{g^2(u)} = 0, \quad (3.6)$$

$$\lim_{u \uparrow x} g(u) \max[\{1 + s\phi'(u)\}^\rho, \{1 + s\phi'(u)\}^{-1}, \log\{1 + s\phi'(u)\}] = 0 \quad (3.7)$$

uniformly on $s \in (s_2(u), s_1(u))$. Define $c(u, x)$ by

$$c(u, x) = \frac{\phi'(u)\{\phi'(u + x\phi(u)) - \phi'(u)\}}{g(u) h_\rho(1 + x\phi'(u))}$$

and suppose also that

$$\lim_{u \uparrow x} c(u, s) = c \quad (3.8)$$

uniformly on $s \in (s_2(u), s_1(u))$. Then $r_n H(f_n, g_n) \rightarrow 0$ as $n \rightarrow \infty$.

Remark 3.4 The simplest way to demonstrate (3.5)-(3.8) is to define s_1, s_2 by

$$\begin{aligned} \{1 + \phi'(u)s_1(u)\}^{-1/\phi'(u)} &= g^K(u), \\ \exp[-\{1 + \phi'(u)s_2(u)\}^{-1/\phi'(u)}] &= g^K(u) \end{aligned}$$

for some fixed $K > 2$, and then to show that (3.7), (3.8) hold for this choice of s_1, s_2 . For (3.7), considering first the upper limit $s \uparrow s_1$, we have

$$g(u) \{1 + \phi'(u)s_1(u)\}^\delta = g(u)^{1-K\delta\phi'(u)}$$

so we require $1-K\delta\phi'(u) \geq \delta_1 > 0$ as $u \uparrow x^*$. The only case that causes any difficulty is when $\gamma < 0$ and $\delta = -1$: then we do need $\gamma > \frac{1}{2}$. The limit as $s \rightarrow s_2$ is much easier since $\{1 + \phi'(u)s_2(u)\}^{-1/\phi'(u)}$ grows only logarithmically in $1/g(u)$. Thus (3.7) follows.

Now let us consider (3.8), breaking this up into cases $\gamma = 0$, $\gamma > 0$, $\gamma < 0$

0. For $\gamma = 0$, it suffices from (2.19) that

$$\frac{\phi''(u + s\phi(u))}{\phi''(u)} \rightarrow 1 \quad \text{uniformly on } |s| \leq K \log |\phi(u)\phi''(u)| \quad (3.9)$$

for some $K \geq 2$. This is similar to several conditions in Cohen (1982b), and is automatic if $\phi''(x)$ (in case $x^* = \infty$) or $\phi''(x^* - x^{-1})$ (in case $x^* < \infty$) is regularly varying. All of Cohen's "Class N" examples satisfy this.

For $\gamma > 0$, assuming (2.13) it follows that the relative error in (2.15) is $O(u^{-\epsilon})$ if $x > 0$, $O\{u^\beta(u + x\phi(u))^{-\beta-\epsilon}\}$ if $x < 0$. We must therefore show

$$u^\beta \{u + \phi(u)s_2(u)\}^{-\beta-\epsilon} \rightarrow 0. \quad (3.10)$$

But

$$\begin{aligned} u + \phi(u)s_2(u) &= \frac{\phi(u)}{\phi'(u)} \{1 + \phi'(u)s_2(u)\} + u \left\{1 - \frac{\phi(u)}{u\phi'(u)}\right\} \\ &= O\{u|\log g(u)|^{-\phi'(u)}\} + O(ug(u)) \end{aligned}$$

from which (3.10) follows.

For $\gamma < 0$, assuming (2.16), a very similar argument settles (3.8) as $s \rightarrow s_2$ but we have an additional complication as $s \rightarrow s_1$ because of the possibility $u + s_1(u)\phi(u) > x^*$. This is most easily settled by defining $\phi'(x)$ to be γ whenever $x > x^*$, $h_p(x)$ to be $-\rho^{-1}$ whenever $x < 0$ (assuming $\rho < 0$). Then it is easily seen that (3.8) holds.

Thus we would argue that (3.5)-(3.8) are reasonable assumptions which hold in most examples, after excluding certain cases which have been noted earlier.

Proof of Theorem 3.3 First we show

$$\int_{s_2(b_n)}^{s_1(b_n)} \left\{ f_n^{\frac{1}{2}}(n) - g_n^{\frac{1}{2}}(n) \right\}^2 dx = o(g^2(b_n)) \quad (3.11)$$

later extending the range of integration to $(-\infty, \infty)$.

We may write

$$\begin{aligned} f_n(x) &= na_n f(a_n x + b_n) F^{n-1}(a_n x + b_n) \\ &= \frac{\phi(b_n)}{-\log F(b_n)} \frac{-\log F(a_n x + b_n)}{\phi(a_n x + b_n)} \exp \left\{ - \frac{-\log F(a_n x + b_n)}{-\log F(b_n)} \right\} \end{aligned}$$

$$= \frac{\phi(b_n)}{\phi(b_n + x\phi(b_n))} \exp\left[-\int_0^x \frac{\phi(b_n)ds}{\phi(b_n + s\phi(b_n))}\right] - \exp\left\{-\int_0^x \frac{\phi(b_n)ds}{\phi(b_n + s\phi(b_n))}\right\}. \quad (3.12)$$

By (2.7),

$$\begin{aligned} \frac{\phi(u + s\phi(u))}{\phi(u)} &= 1 + s\phi'(u) + \int_0^s \frac{c(u,w)g(u)h_\rho(1 + w\phi'(u))}{\phi'(u)} dw \\ &= 1 + s\phi'(u) + \frac{c_1(u,s)g(u)}{\phi'(u)} \int_0^s h_\rho(1 + w\phi'(u)) dw \end{aligned}$$

where $c_1(u,s)$ is such that $c_1 \rightarrow c$ uniformly on $s \in (s_2(u), s_1(u))$. Evaluating the integral we have

$$\begin{aligned} \frac{\phi(u + s\phi(u))}{\phi(u)} &= \{1 + s\phi'(u)\} \left[1 + \frac{c_1(u,s)g(u)}{\phi'(u)} \right. \\ &\quad \left. \left\{ \frac{(1 + s\phi'(u))^\rho - (1 + s\phi'(u))^{-1}}{\phi'(u)\rho(\rho + 1)} - \frac{s(1 + s\phi'(u))^{-1}}{\rho} \right\} \right]. \end{aligned}$$

Now (3.7) shows that this is of form $\{1 + s\phi'(u)\}(1 + o(1))$ uniformly in s , so for the reciprocal we have

$$\begin{aligned} \frac{\phi(u)}{\phi(u + s\phi(u))} &= \{1 + s\phi'(u)\}^{-1} \left[1 - \frac{c_2(u,s)g(u)}{\phi'(u)} \right. \\ &\quad \left. \left\{ \frac{(1 + s\phi'(u))^\rho - (1 + s\phi'(u))^{-1}}{\phi'(u)\rho(\rho + 1)} - \frac{s(1 + s\phi'(u))^{-1}}{\rho} \right\} \right] \end{aligned}$$

where c_2 is another function such that $c_2(u,s) \rightarrow c$ uniformly on $(s_2(u), s_1(u))$.

This may also be written

$$\frac{\phi(u)}{\phi(u + s\phi(u))} = \{1 + s\phi'(u)\}^{-1} - c_2(u,s)g(u)H'_\rho(s, \phi'(u)) \quad (3.13)$$

where H'_ρ is the derivative with respect to the first component of H_ρ .

For later purposes, it is also convenient to write (3.13) in the form

$$\frac{\phi(u)}{\phi(u + s\phi(u))} = \{1 + s\phi'(u)\}^{-1} - \frac{c_3(u,s)g(u)H'_\rho(s, \phi'(u))}{1 + c_3(u,s)g(u)H_\rho(s, \phi'(u))} \quad (3.14)$$

where $c_3 \rightarrow c$ uniformly; this is equivalent to (3.13) because of (3.7).

Now take (3.13) and integrate:

$$\int_0^x \frac{\phi(u)ds}{\phi(u + s\phi(u))} = \frac{1}{\phi'(u)} \log\{1 + x\phi'(u)\} - g(u) \int_0^x c_2(u,s)H'_\rho(s, \phi'(u))ds$$

and hence

$$\exp\left\{-\int_0^x \frac{\phi(u)ds}{\phi(u+s\phi(u))}\right\} = \{1 + x\phi'(u)\}^{-1/\phi'(u)} \cdot \{1 + g(u)c_4(u,x)H_\rho(x,\phi'(u))\} \quad (3.15)$$

where $c_4(u,x)$ is yet another function satisfying $c_4 \rightarrow c$ uniformly on $(s_2(u), s_1(u))$.

Define a new parametric family by

$$f_0(x; \theta_1, \theta_2) = \exp[-(1+x\eta)^{-1/\eta} \{1 + \theta_1 H_\rho(x, \eta)\}].$$

$$(1+x\eta)^{-1/\eta} \{1 + \theta_1 H_\rho(x, \eta)\} [(1+x\eta)^{-1} - \frac{\theta_2 H'_\rho(x, \eta)}{1 + \theta_2 H_\rho(x, \eta)}]$$

where the parameters η , which we shall identify with γ_n , and ρ are not shown explicitly as parameters of f_0 .

By (3.12), (3.14) and (3.15), we have

$$f_n(x) = f_0(x; c_4(b_n, x)r_n, c_3(b_n, x)r_n).$$

But directly from (3.1) we have

$$g_n(x) = f_0(x; cr_n, cr_n).$$

We have therefore set everything up to apply Lemma 3.1; we let $g_0 = f_0^{\frac{1}{2}}$, $B = (s_2(b_n), s_1(b_n))$ and take θ^* to be some small interval around (cr_n, cr_n) . The only thing to show is that the integral in (3.4) is finite.

Consider first what happens as $x \rightarrow s_1(b_n)$. Note that

$$\left[\frac{\partial g_0}{\partial \theta_1} \right]^2 = \frac{1}{f_0} \left[\frac{\partial f_0}{\partial \theta_1} \right]^2, \quad i=1,2.$$

As $x \rightarrow s_1$, we have $f_0 \asymp (1+x\eta)^{-1/\eta-1}$. Here we use $\asymp$ to denote "is the same order of magnitude as" and always keep in mind (3.7). Consider first $\eta < 0$. We have

$$\begin{aligned} \frac{\partial f_0}{\partial \theta_1} &= \exp[-(1+x\eta)^{-1/\eta} \{1 + \theta_1 H_\rho(x, \eta)\}] [(1+x\eta)^{-1/\eta} H_\rho(x, \eta) \\ &\quad - (1+x\eta)^{-2/\eta} H_\rho(x, \eta) \{1 + \theta_1 H_\rho(x, \eta)\}] [(1+x\eta)^{-1} - \frac{\theta_2 H'_\rho(x, \eta)}{1 + \theta_2 H_\rho(x, \eta)}] \end{aligned}$$

the dominant term in which is $\asymp (1 + x\eta)^{-1/\eta-2}$. Hence

$$\left[\frac{\partial g_0}{\partial \theta_1} \right]^2 \asymp (1 + x\eta)^{-1/\eta-3}$$

which is integrable as $1 + x\eta \rightarrow 0$ because $\eta > -\frac{1}{2}$. A very similar calculation

shows that $\left[\frac{\partial g_0}{\partial \theta_2} \right]^2$ is of the same order of magnitude, as $1 + x\eta \rightarrow 0$.

Now suppose $\eta > 0$. In this case $s_1 \rightarrow \infty$ and we may assume $\rho \leq 0$. Hence

$$\frac{\partial f_0}{\partial \theta_1} \asymp (1 + x\eta)^{-1/\eta-1} \log(1 + x\eta).$$

$$\left[\frac{\partial g_0}{\partial \theta_1} \right]^2 \asymp (1 + x\eta)^{-1/\eta-1} \log^2(1 + x\eta).$$

Similarly we have

$$\left[\frac{\partial g_0}{\partial \theta_2} \right]^2 \asymp (1 + x\eta)^{-1/\eta-1}.$$

So in this case the required integrals are finite for each $\eta > 0$ and even uniformly as $\eta \rightarrow 0$.

Similar calculations may be made as $x \rightarrow s_2$, but in this case there is no problem because everything is decaying exponentially. Hence we conclude that the integral in (3.4) is indeed bounded, so we deduce (3.11).

To complete the proof, it will suffice from

$$\int_{s_1}^{\infty} \{f_n^{\frac{1}{2}}(x) - g_n^{\frac{1}{2}}(x)\}^2 dx = \int_{s_1}^{\infty} f_n(x) dx - 2 \int_{s_1}^{\infty} f_n^{\frac{1}{2}}(x) g_n^{\frac{1}{2}}(x) dx + \int_{s_1}^{\infty} g_n(x) dx$$

to show that $1 - F_n(s_1(b_n)) = o(r_n^2)$, $1 - G_n(s_1(b_n)) = o(r_n^2)$, and similarly that $F_n(s_2(b_n))$ and $G_n(s_2(b_n))$ are each $o(r_n^2)$. In the case of G_n , these results follow directly from (3.5) and (3.6), also using (3.7) to show that the r_n term in the definition of G_n may be ignored for the purpose of this comparison. In the case of F_n , note that (3.15) is an expression for $-\log F_n(a_n x + b_n)$; using (3.5), (3.6) and (3.7) again, the result follows. With this the proof of the theorem is complete.

A similar result is obtained for threshold convergence. We state the following without proof:

Theorem 3.5 Suppose ϕ , defined from (2.3), satisfies (2.5) with $\gamma > \frac{1}{2}$ and

(2.21). Define F_u, G_u by

$$\left. \begin{aligned} F_u(x) &= \frac{F(u + x\phi(u)) - F(u)}{1 - F(u)} \\ G_u(x) &= 1 - \{1 + x\phi'(u)\}_+^{-1/\phi'(u)} [1 + cg(u)H_\rho(x, \phi'(u))] \end{aligned} \right\} \quad (3.16)$$

($x > 0$), with associated densities f_u, g_u . Defining $s_2(u)$ to be 0, suppose $s_1(u)$ exists such that (3.5), (3.7) and (3.8) are satisfied. Then $g(u)H(f_u, g_u) \rightarrow 0$ as $u \uparrow x^*$.

For the k largest order statistics (k fixed, $n \rightarrow \infty$) it seems impossible to avoid an additional error term of $O(n^{-1})$ (cf. Falk 1986). This does not matter, of course, if $nr_n \rightarrow \infty$, which is usually the case in practice. Also, in this case, it does not matter whether we start with (2.1) or (2.3) as our definition of ϕ .

Theorem 3.6 Suppose the assumptions of Theorem 3.3 are satisfied, with ϕ defined from either (2.1) or (2.3). Let $Y_{1:n} \leq \dots \leq Y_{n:n}$ denote the order statistics of a sample from F , and let $X_i^{(n)} = (Y_{n-i+1:n} - b_n)/a_n$ for $i=1, 2, \dots, k$, where k is a fixed positive integer. Let $f_n(x_1, \dots, x_k)$ denote the joint density of $X_1^{(n)}, \dots, X_k^{(n)}$. Define

$$\begin{aligned} \psi_n(x; \theta) &= (1 + x\gamma_n)^{-1/\gamma_n} \{1 + \theta H_\rho(x, \gamma_n)\}, \\ g_n(x_1, \dots, x_k) &= \prod_{i=1}^k \{-\psi_n'(x_i; cr_n)\} \exp\{-\psi_n(x_k; cr_n)\}. \end{aligned}$$

defined when $x_1 \geq \dots \geq x_k$, $1 + x_i \gamma_n > 0$ for each i . Then

$$H(f_n, g_n) = o(r_n) + O(n^{-1}).$$

Proof. Assume ϕ has been defined from (2.1). We have

$$\begin{aligned} f_n(x_1, \dots, x_k) &= \frac{n!}{(n-k)!} \prod_{i=1}^k \{a_n f(a_n x_i + b_n)\} F^{n-k}(a_n x_k + b_n), \\ &\quad x_1 \geq \dots \geq x_k. \end{aligned}$$

Let us first replace this by

$$f_n^*(x_1, \dots, x_k) = \prod_{i=1}^k \left\{ \frac{n a_n f(a_n x_i + b_n)}{F(a_n x_i + b_n)} \right\} F^n(a_n x_k + b_n).$$

It is easy to see, by writing the likelihood ratio f_n^*/f_n in terms of uniform order statistics, that $H(f_n, f_n^*)$ is $O(n^{-1})$. Define ψ_n as above.

$$\xi_n(x; \theta) = - \frac{\psi'_n(x; \theta)}{\psi_n(x; \theta)} = (1 + x\gamma_n)^{-1} - \frac{\theta H'_\rho(x, \gamma_n)}{1 + \theta H_\rho(x, \gamma_n)}.$$

Now, using (3.14) and (3.15),

$$\begin{aligned} f_n^*(x_1, \dots, x_k) &= \\ &= \prod_{i=1}^k \left\{ \frac{\phi(b_n)}{\phi(b_n + x_i \phi(b_n))} \frac{\log F(b_n + x_i \phi(b_n))}{\log F(b_n)} \right\} \exp \left\{ - \frac{\log F(b_n + x_k \phi(b_n))}{\log F(b_n)} \right\} \\ &= \prod_{i=1}^k \{ \xi_n(x_i; c_3(b_n, x_i)r_n) \psi_n(x_i; c_4(b_n, x_i)r_n) \} \exp \{ - \psi_n(x_k; c_4(b_n, x_k)r_n) \}. \end{aligned}$$

We also have

$$g_n(x_1, \dots, x_k) = \prod_{i=1}^k \{ \xi_n(x_i; cr_n) \psi_n(x_i; cr_n) \} \exp \{ - \psi_n(x_k; cr_n) \}.$$

The proof is now similar to that of Theorem 3.3, in that we define a $2k$ -parameter family

$$f_o(x_1, \dots, x_k; \theta^{(1)}, \theta^{(2)}) = \prod_{i=1}^k \{ \xi_n(x_i; \theta_i^{(2)}) \psi_n(x_i; \theta_i^{(1)}) \} \exp \{ - \psi_n(x_k; \theta_k^{(1)}) \}$$

with parameters $\theta^{(j)} = (\theta_1^{(j)}, \dots, \theta_k^{(j)})$, $j = 1, 2$. We apply Lemma 3.1 with

$B = (s_2(b_n), s_1(b_n))^k$. The proof that the integral in (3.4) is bounded is similar to the corresponding proof in Theorem 3.3, and the extension from B to $\mathbb{R}^k$ is also similar. With this the proof of Theorem 3.6 is complete.

4. Examples

Three examples will be used to illustrate the foregoing theory. These are normal maxima, lognormal minima and minima from a Gamma distribution with index $\alpha > 1$. The last two are treated by reflecting about the origin so as to use the

theory for maxima. All three examples have $nr_n \rightarrow \infty$, and this allows us to make two small changes in the procedures without affecting the claimed rates of convergence. These are to define ϕ from (2.3) instead of (2.1), and to define b_n by $F(b_n) = 1 - n^{-1}$ instead of $\exp(-n^{-1})$. Since we are involving the normal distribution, we use ϕ to denote the standard normal distribution function but keep ϕ in the sense in which it has been used throughout the paper. The normal density will be written $\phi'(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.

With ϕ and b_n as just defined, we may write

$$a_n = \phi(b_n) = \{nf(b_n)\}^{-1}. \quad (4.1)$$

$$\gamma_n = \phi'(b_n) = - \frac{\phi(b_n)f'(b_n)}{f(b_n)} - 1. \quad (4.2)$$

We define $\epsilon_n = cg(b_n)$, which is taken to be $\phi(b_n)\phi''(b_n)$ when F is in the domain of attraction of Λ . In this case, further application of (4.2) gives

$$\epsilon_n = \gamma_n^2 + \gamma_n - a_n^2 \left(\frac{f'(x)}{f(x)} \right)' \Big|_{x=b_n}. \quad (4.3)$$

Experience has shown that it is important to use the exact constants; even minor variations on the foregoing scheme upset the comparisons to follow.

Normal distribution Take $F = \phi$, $f(x) = \phi'(x)$ and so

$$\frac{f'(x)}{f(x)} = -x, \quad \left[\frac{f'(x)}{f(x)} \right]' = -1.$$

We define b_n by $\phi(b_n) = 1 - n^{-1}$; application of (4.1)-(4.3) yields

$$a_n = n^{-1}(2\pi)^{1/2} \exp(b_n^2/2),$$

$$\gamma_n = a_n b_n^{-1},$$

$$\epsilon_n = \gamma_n^2 + \gamma_n + a_n^2.$$

The expansion $\{1 - \phi(x)\}/\phi'(x) = x^{-1} - x^{-3} + 3x^{-5} - \dots$ shows that $\phi(x) \sim x^{-1}$, $\phi'(x) \sim -x^{-2}$, $\phi(x)\phi''(x) \sim x^{-4}$. Since $b_n = O\{(\log n)^{1/2}\}$ we have that the rates of convergence of the ultimate and penultimate approximations are $O\{(\log n)^{-1}\}$, $O\{(\log n)^{-2}\}$.

Lognormal distribution Take $F(x) = \Phi(-\sigma^{-1} \log|x|)$ for $x < 0$, where $\sigma > 0$. In this case

$$f(x) = \sigma^{-1} |x|^{-1} (2\pi)^{-1/2} \exp\left\{-\frac{(\log|x|)^2}{2\sigma^2}\right\}.$$

$$\frac{f'(x)}{f(x)} = |x|^{-1} \{1 + \sigma^{-2} \log|x|\}.$$

$$\left[\frac{f'(x)}{f(x)}\right]' = |x|^{-2} \{1 - \sigma^{-2} + \sigma^{-2} \log|x|\}.$$

Hence with B_n satisfying $\Phi(B_n) = 1 - n^{-1}$, we have

$$b_n = -\exp(-\sigma B_n).$$

$$a_n = n^{-1} \sigma |b_n| (2\pi)^{1/2} \exp(B_n^2/2).$$

$$\gamma_n = -a_n |b_n|^{-1} \{1 - \sigma^{-1} B_n\} - 1.$$

$$\epsilon_n = \gamma_n^2 + \gamma_n - a_n^2 |b_n|^{-2} \{1 - \sigma^{-2} + \sigma^{-2} \log|b_n|\}.$$

Since $\phi(x) \sim \sigma^2 |x| (-\log|x|)^{-1}$, $\phi'(x) \sim \sigma^2 (\log|x|)^{-1}$, $\phi(x)\phi''(x) \sim \sigma^4 (\log|x|)^{-3}$ the rates of convergence are $O(B_n^{-1}) = O((\log n)^{-1/2})$ for the ultimate approximation, $O(B_n^{-3}) = O((\log n)^{-3/2})$ for the penultimate approximation.

Gamma distribution Take $f(x) = |x|^{\alpha-1} e^x / \Gamma(\alpha)$ for $x < 0$, where $\alpha > 1$. In this case

$$\frac{f'(x)}{f(x)} = 1 - \frac{\alpha-1}{|x|}.$$

So with b_n satisfying $1 - F(b_n) = n^{-1}$, we have

$$a_n = n^{-1} |b_n|^{-\alpha+1} \exp(|b_n|) \Gamma(\alpha).$$

$$\gamma_n = a_n \{|b_n|^{-1}(\alpha-1) - 1\} - 1.$$

$$\rho = 1, \epsilon_n = 2|b_n| \alpha^{-2} (\alpha+1)^{-1}.$$

The values for ρ and ϵ_n follow from the expansion $1 - F(x) = \{\alpha \Gamma(\alpha)\}^{-1} |x|^\alpha \{1 - \alpha(\alpha+1)^{-1} |x| + \dots\}$ of the form (2.16) with $\beta = 1$, $D = -\alpha(\alpha+1)^{-1}$. Then we take $\rho = \beta$, $cg(u) = -|u|^\beta D \beta^2 (\beta+1) \alpha^{-3}$ as in Section 2.

Figure 1 shows the exact density for normal maxima with $n=100$, together with our three approximations, i.e. (1.5) with $\gamma = 0$, (1.5) with $\gamma = \gamma_n$ and (2.22). All three approximations are close to the true density, but the first approximation is perceptibly the worst of the three, and the third approximation the best. Table 1 gives more details of the exact and three

approximate distributions, including mean, variance, skewness and kurtosis of each, and three measures of discrepancy between the approximations and exact densities: the uniform or Kolmogorov-Smirnov distance (1.6), the Hellinger distance (3.3) and total variation distance which, in the same notation as (3.3), is calculated by

$$V(f_n, g_n) = \int \{f_n(x) - g_n(x)\}_+ dx \quad (4.4)$$

The calculations confirm our overall claim about the ranking of the three approximations. Also shown are the corresponding calculations for the threshold distribution, i.e. (2.12) with $\gamma=0$ (exponential distribution), (2.12) with $\gamma = \gamma_n$, (2.25). It is noticeable that the first approximation is very poor when assessed by skewness and kurtosis, but much better when assessed by the other criteria. This is mainly responsible for the adverse comments made by Fisher and Tippett (1928), who took skewness and kurtosis as their main criterion of fit. It also warns of the danger in using moments for statistical comparison.

Figure 2 and Table 2 show corresponding calculations for the lognormal distribution with $\sigma = 1$, $n = 250$. We took a larger sample size here because of the poorer overall fit. The most striking thing here is that the first approximation is very much worse than the other two. Note also $\gamma_n = -0.4422$ - a long way from its limiting value $\gamma = 0$.

Figure 3 and Table 3 are for the Gamma distribution with $\alpha=5$, $n=100$. For the first approximation in this case we took

$$F_n(x) \approx \exp\{-(x/b_n)^\alpha\}, \quad x < 0$$

equivalent to the classical two-parameter Weibull approximation usually assumed in this situation. Figure 3 shows strikingly how poor it is. The other two approximations are indistinguishable from the true density, except in one tail.

Finally, in Table 4 we give calculations for the normal distribution at sample sizes $n=10^m$, $m=1, \dots, 5$. The decrease in distance from approximate to exact agrees very well with the theoretical rates of decay, of $O\{(\log n)^{-1}\}$.

$O\{(\log n)^{-2}\}$ and $O\{(\log n)^{-3}\}$, for the three approximations.

Acknowledgement

This paper was written during a visit to the Center for Stochastic Processes, University of North Carolina, Chapel Hill. I would like to thank Ross Leadbetter for the invitation, and the staff of the Center for their hospitality. I also thank Jonathan Cohen, Rolf Reiss and Ishay Weissman for sending me copies of unpublished papers.

REFERENCES

- Anderson, C.W. (1971), Contributions to the Asymptotic Theory of Extreme Values. Ph.D. Thesis, University of London.
- Anderson, C.W. (1976), Extreme value theory and its approximations. Proc. Symp. Reliability Technology. U.K. Atomic Energy Authority, Bradford.
- Anderson, C.W. (1984), Large deviations of extremes. In Statistical Extremes and Applications, J. Tiago de Oliveira (ed), Reidel, Dordrecht, 325-340.
- Balkema, A.A. and Haan, L. de (1972), On R. Von Mises' condition for the domain of attraction of $\exp(-e^{-x})$. Ann. Math. Statist. 43, 1352-1354.
- Cohen, J.P. (1982a), The penultimate form of approximation to normal extremes. Adv. Appl. Prob. 14, 324-339.
- Cohen, J.P. (1982b), Convergence rates for the ultimate and penultimate approximations in extreme value theory. Adv. Appl. Prob. 14, 833-854.
- Cohen, J.P. (1987a), Large sample theory for fitting an approximate Gumbel model to maxima. Sankhya A 48, 372-392.
- Cohen, J.P. (1987b), Fitting extreme value distributions to maxima. Preprint, University of Kentucky.
- Davis, R.A. (1982), The rate of convergence in distribution of maxima. Stat. Neerlandica 36, 31-35.
- Davison, A.C. (1984), Modelling excesses over high thresholds, with an application. In Statistical Extremes and Applications, J. Tiago de Oliveira (ed.), Reidel, Dordrecht, 461-482.
- Falk, M. (1986), Rates of uniform convergence of extreme order statistics. Ann. Inst. Statist. Math. 38, 245-262.

Fisher, R.A. and Tippett, L.H.C. (1928), Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proc. Camb. Phil. Soc.* 24, 180-190.

Galambos, J. (1978), *The Asymptotic Theory of Extreme Order Statistics*. Wiley, New York.

Goldie, C.M. and Smith, R.L. (1987), Slow variation with remainder: Theory and Application. *Quart. J. Math. Oxford* (2) 38, 45-71.

Gomes, M.I. (1984), Penultimate limiting forms in extreme value theory. *Ann. Inst. Statist. Math.* 36, 71-85.

Haan, L. de (1970), On Regular Variation and its Application to the Weak Convergence of Sample Extremes. *Mathematical Centre Tracts No. 32*, Amsterdam.

Haan, L. de and Resnick, S. (1982), Local limit theorems for sample extremes. *Ann. Probab.* 10, 396-413.

Hall, P. (1979), On the rate of convergence of normal extremes. *J. Appl. Prob.* 16, 433-439.

Hall, P. (1980), Estimating probabilities for normal extremes. *Adv. Appl. Prob.* 12, 491-500.

Hall, W.J. and Wellner, J.A. (1979), The rate of convergence in law of the maximum of an exponential sample. *Statist. Neerlandica* 33, 151-154.

Hosking, J.R.M., and Wallis, J.R. (1987), Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29, 339-349.

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

Joe, H. (1987), Estimation of quantiles of the maximum of N observations. *Biometrika*, 74, 347-354.

Kohne, W. and Reiss, R.-D. (1983), A note on uniform approximation to extreme order statistics. *Ann. Inst. Statist. Math.* 35, 343-345.

Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983), *Extremes and Related Properties of Random Sequences and Series*. Springer Verlag, New York.

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

Omey, E. and Rachev, S.T. (1987), On the rate of convergence in extreme value theory. To appear.

Pickands, J. (1986), The continuous and differentiable domains of attraction of the extreme-value distributions. *Ann. Probab.* 14, 996-1004.

Reiss, R.-D. (1981), Uniform approximation to distribution of extreme order statistics. *Ad. Appl. Prob.* 13, 533-547.

Reiss, R.-D. (1984), Statistical inference using appropriate extreme value models. Preprint 124, University of Siegen.

- Resnick, S. (1986), Uniform rates of convergence to extreme value distributions. In J. Srivastava (ed.), Probability and Statistics; Essays in Honor of Franklin A. Graybill. North Holland, Amsterdam.
- Smith, R.L. (1982), Uniform rates of convergence in extreme value theory. Adv. Appl. Prob. 14, 600-622.
- Smith, R.L. (1984), Threshold methods for sample extremes. In Statistical Extremes and Applications, J. Tiago de Oliveira (ed.), 621-638, Reidel, Dordrecht.
- Smith, R.L. (1985), Maximum likelihood estimation in a class of non-regular cases. Biometrika 72, 67-92.
- Smith, R.L. (1987), Estimating tails of probability distributions. Ann. Statist., September 1987.
- Sweeting, T.J. (1985), On domains of uniform local attraction in extreme value theory. Ann. Probab. 13, 196-205.
- Weissman, I. (1984), On the asymptotic distribution of an unbounded number of sample extremes. Preprint, Technion, Haifa.
- Zolotarev, V.M. and Rachev, S.T. (1985), Rate of convergence in limit theorems for the max-scheme. In, Lecture Notes in Mathematics #1155, Springer, Berlin, 415-442.

TABLE 1

STANDARD NORMAL DISTRIBUTION, $n = 100$.

LOCATION CONST = 2.3263; SCALE CONST = .3752; GAMMA = -.1271
RHO = 1.0; EPSILON = .0298

I. DISTRIBUTION OF SAMPLE MAXIMA

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	2.518	2.553	2.511	2.517
VARIANCE:	.184	.232	.175	.187
SKEWNESS:	.429	1.298	.276	.422
KURTOSIS:	3.765	5.399	3.309	3.709
KOLM-SMIR DIST:		.0272	.0052	.0027
TOTAL VARIATION DIST:		.0390	.0065	.0041
HELLINGER DIST:		.083	.026	.011

II. THRESHOLD DISTRIBUTION

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	.344	.380	.338	.345
VARIANCE:	.097	.141	.088	.098
SKEWNESS:	2.529	4.001	2.004	2.512
KURTOSIS:	6.302	9.001	5.246	6.129
KOLM-SMIR DIST:		.0311	.0055	.0013
TOTAL VARIATION DIST:		.0311	.0055	.0014
HELLINGER DIST:		.073	.026	.009

TABLE 2

LOG NORMAL DISTRIBUTION, SIGMA = 1, n = 250.

LOCATION CONST = -.0705; SCALE CONST = .0238; GAMMA = -.4422
 RHO = 1.0; EPSILON = .0557

I. DISTRIBUTION OF SAMPLE MAXIMA

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	-.0635	-.0563	-.0639	-.0636
VARIANCE:	.00053	.00093	.00050	.00053
SKEWNESS:	.174	1.298	.226	.208
KURTOSIS:	3.088	5.400	2.993	3.121
KOLM-SMIR DIST:		.1117	.0137	.0053
TOTAL VARIATION DIST:		.1450	.0151	.0062
HELLINGER DIST:		.325	.085	.070

II. THRESHOLD DISTRIBUTION

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	.0179	.0248	.0175	.0179
VARIANCE:	.00016	.00057	.00014	.00016
SKEWNESS:	.63	4.01	.44	.55
KURTOSIS:	2.96	9.01	2.59	2.76
KOLM-SMIR DIST:		.1262	.0143	.0051
TOTAL VARIATION DIST:		.1262	.0143	.0051
HELLINGER DIST:		.303	.084	.072

TABLE 3

GAMMA DISTRIBUTION, ALPHA = 5, n = 100.

LOCATION CONST = -1.2791; SCALE CONST = .3222; GAMMA = -.3147
RHO = 1.0; EPSILON = .0381

I. DISTRIBUTION OF SAMPLE MAXIMA

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	-1.165	-1.173	-1.170	-1.167
VARIANCE:	.104	.072	.100	.105
SKEWNESS:	.0009	.0646	.0127	.0039
KURTOSIS:	2.834	2.880	2.715	2.808
KOLM-SMIR DIST:		.0539	.0078	.0031
TOTAL VARIATION DIST:		.1027	.0117	.0066
HELLINGER DIST:		.155	.060	.039

II. THRESHOLD DISTRIBUTION

	EXACT	1ST APPR	2ND APPR	3RD APPR
MEAN:	.2505	.2142	.2461	.2504
VARIANCE:	.040	.032	.037	.040
SKEWNESS:	1.055	1.400	.810	.980
KURTOSIS:	3.687	4.200	3.210	3.484
KOLM-SMIR DIST:		.0758	.0081	.0032
TOTAL VARIATION DIST:		.0785	.0109	.0040
HELLINGER DIST:		.100	.059	.037

TABLE 4

STANDARD NORMAL DISTRIBUTION

DISTRIBUTION OF MAXIMA

FIT OF THREE APPROXIMATIONS FOR VARIOUS SAMPLE SIZES

	1ST APPR	2ND APPR	3RD APPR
SAMPLE SIZE 10			
KOLM-SMIR DIST:	.052	.026	.028
TOTAL VARIATION DIST:	.0789	.0350	.0351
HELLINGER DIST:	.1404	.1024	.0876
SAMPLE SIZE 100			
KOLM-SMIR DIST:	.0272	.0052	.0027
TOTAL VARIATION DIST:	.0390	.0065	.0040
HELLINGER DIST:	.083	.026	.011
SAMPLE SIZE 1,000			
KOLM-SMIR DIST:	.01825	.00236	.00039
TOTAL VARIATION DIST:	.02595	.00273	.00067
HELLINGER DIST:	.0576	.0122	.0026
SAMPLE SIZE 10,000			
KOLM-SMIR DIST:	.01368	.00133	.00017
TOTAL VARIATION DIST:	.01930	.00157	.00020
HELLINGER DIST:	.0433	.0070	.0011
SAMPLE SIZE 100,000			
KOLM-SMIR DIST:	.01092	.00085	.00009
TOTAL VARIATION DIST:	.01531	.00101	.00010
HELLINGER DIST:	.0345	.0045	.0006

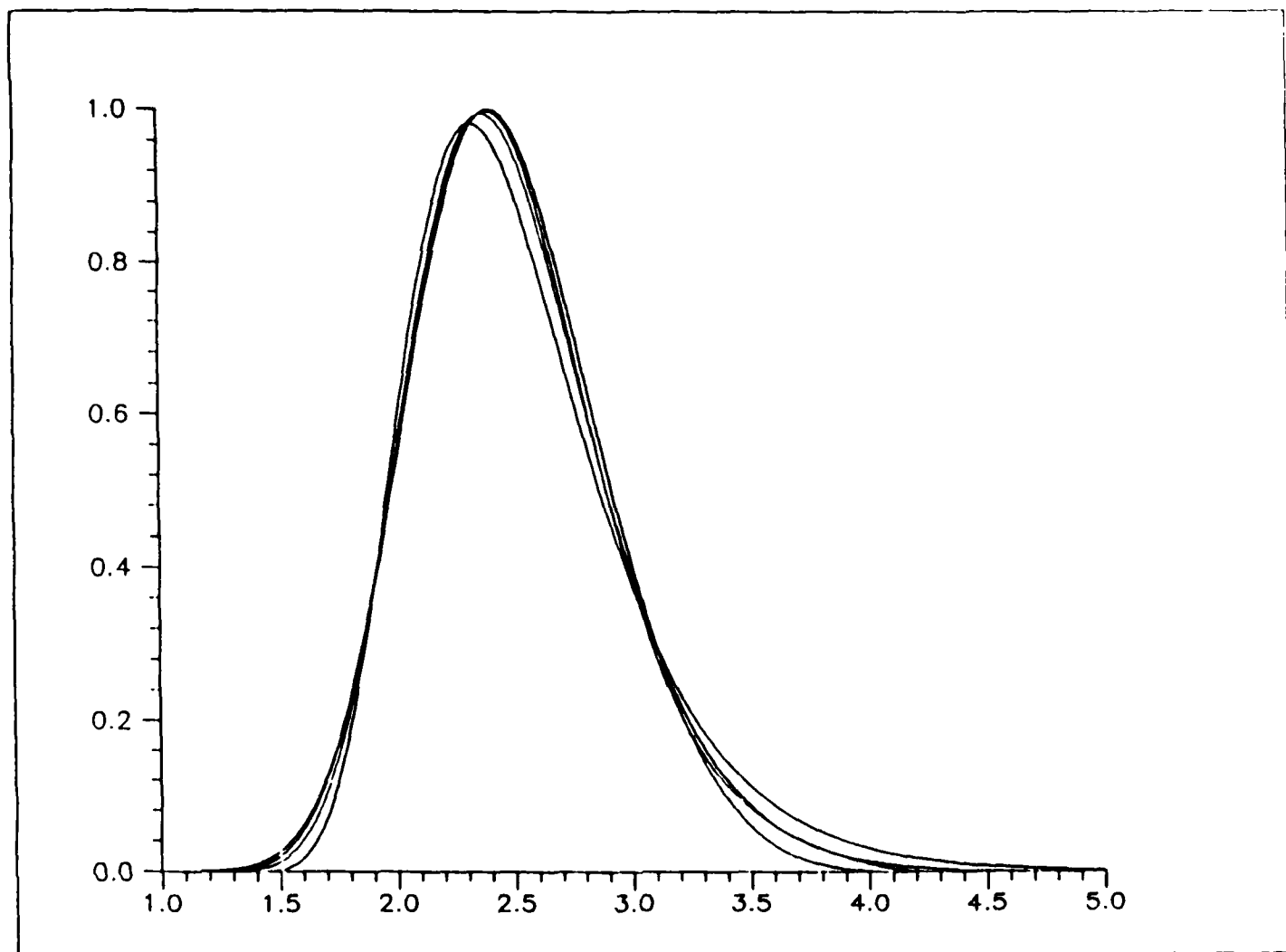


FIGURE 1: DENSITY OF SAMPLE MAXIMA FOR STANDARD NORMAL DISTRIBUTION, $n=100$.

Top to bottom at $x=2.6$:
 2nd approx.
 3rd approx.
 Exact
 1st approx.

Top to bottom at $x=3.3$:
 1st approx.
 Exact
 3rd approx.
 2nd approx.

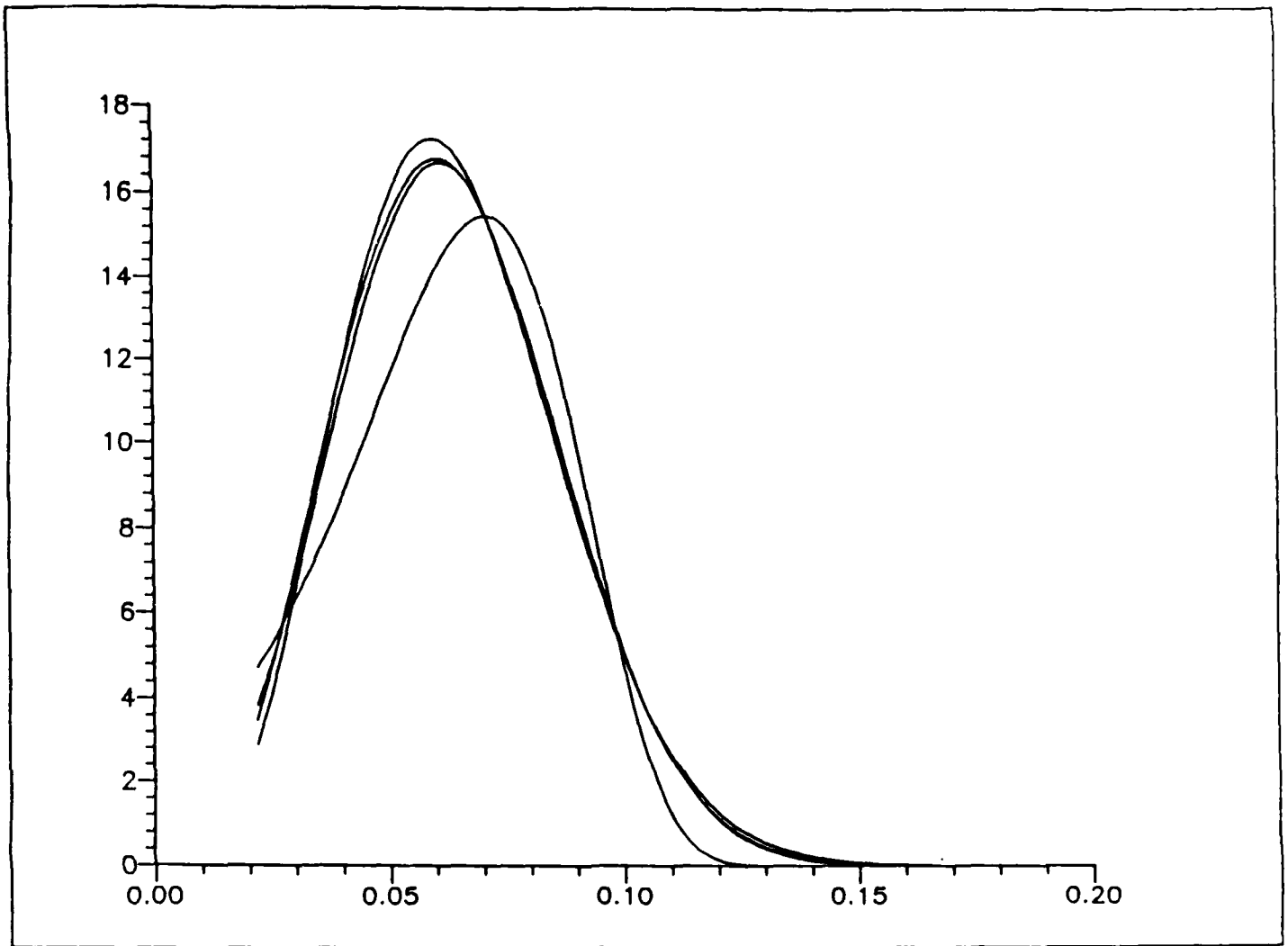


FIGURE 2: DENSITY OF SAMPLE MINIMA FOR LOGNORMAL DISTRIBUTION, $\sigma = 1$, $n = 250$.

Top to bottom at $x=0.5$:

- Exact
- 2nd approx.
- 3rd approx.
- 1st approx.

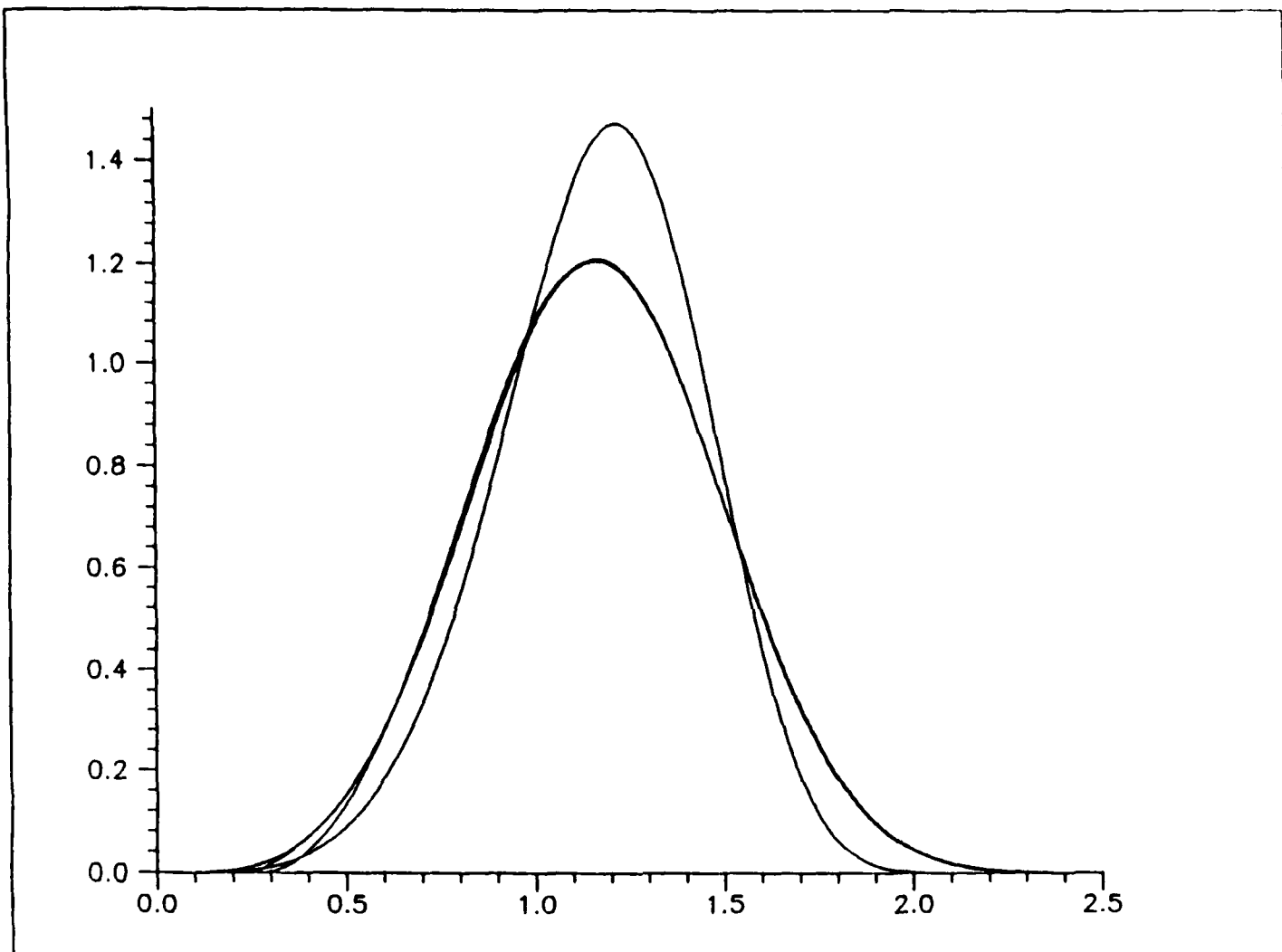


FIGURE 3: DENSITY OF SAMPLE MINIMA FOR GAMMA DISTRIBUTION, $\alpha = 5$, $n = 100$.
The 1st approximation is the curve visibly removed from the others; the
exact density and the 2nd and 3rd approximations are virtually indistinguishable.

END

DATE

3-88

DTIC