A SURVEY OF NONREGULAR PROBLEMS

Richard L. Smith
Department of Mathematics
University of Surrey
Guildford, GU2 5XH, England.

Keywords

Bayesian inference, Brownian bridge, change-points, chi-square process, conditional inference, grouped likelihood, level crossings, likelihood ratio test, maximum likelihood, maximum product of spacings, mixture models, multivariate extremes, reliability, Weibull distribution.

1. TYPES OF NONREGULAR PROBLEM

Maximum likelihood estimates, as every student of Statistics learns, have a number of properties which make them desirable in large samples. They are consistent, asymptotically efficient and asymptotically normally distributed, with covariance matrix approximated by the inverse of either the observed or the Fisher information matrix. These results, along with the asymptotic normality of the score statistic and the asymptotic chi-square distribution of the log likelihood ratio statistic, may be considered the classical results of finite-parameter regular estimation theory.

Yet there are many "nonregular" problems for which these results fail to hold. These problems are not mere mathematical pathologies, constructed so as to impress upon the student the need for utmost rigour at all times, but include many examples derived from serious and important applications. Examples are the estimation of some distributions widely used in reliability and survival data analysis, mixtures of distributions, testing for a shift in the underlying parameters of a model (change-point problems), and testing for certain structures in regression or contingency table analysis. Indeed, nonregular problems, in the broadest sense, are so pervasive that they may be considered one of the major concerns of current work in theoretical statistics. As Weiss and Wolfowitz (1974, p.9) remarked, "the term 'regular' for the conditions under which the maximum likelihood estimator is efficient is more of a mathematical trick than anything which truly corresponds to the ordinary connotation of the word regular."

A review of this nature must necessarily be selective, and I confine myself entirely to finite-parameter problems, though even this excludes such classics as the Neyman-Scott problem (Cox and Hinkley 1974, p. 329) and Lindsay's beautiful work on mixture models with an unspecified number of components. Most of the examples concern independent, identically distributed observations, though there are some extensions to regression and simple time series models. The general theory of inference in stochastic processes, however, is much too broad to be covered here. I have also omitted any coverage of population size estimation (see for example Pickands and Raghavachari 1987, Raftery 1988) largely because I do not feel qualified to comment.

Having thus eliminated most of the problems my audience wanted to hear about, I consider three broad classes of nonregular problems. The first (Sections 2 and 3) concerns distributions such as the three-parameter Weibull, in which the range of the data is a function of unknown parameters. The second

(Section 4) concerns problems in which the parameters set is bounded, especially testing whether the parameters lie on the boundary of the parameter space. The third (Section 5) is all about testing problems in which some parameters are not identifiable under the null hypothesis. The best known example so this type are change-point and finite mixture problems, but there are many others and new ones are being discovered all the time. Finally, in Section 6 some examples of multivariate reliability and extreme-value distributions are given, the most complicated of which feature all three types of nonregularity in the same problem.

2. UNKNOWN ENDPOINT PROBLEMS

This section is concerned with problems in which one endpoint of the range is an unknown parameter. Such problems fail to satisfy Cramer's (1946) regularity conditions, though more general results such as Daniels (1961) (see also Williamson 1984) on asymptotic efficiency, and LeCam (1970) on asymptotic normality, show that the real issue has to do with discontinuities in the likelihood of its first two derivatives, rather than the endpoint problem as such. For instance, the three-parameter lognormal distribution is regular in the sense considered here, though there are still some practical difficulties with maximum likelihood estimation (Hill 1963, Griffiths 1980).

Following Smith (1985), consider the model with density

$$f\ (x\,;\theta,\phi) = (x-\theta)^{\alpha-1} g\ (x-\theta;\phi), \qquad x>\theta \qquad (2.1)$$

in which $\alpha>0$, θ is an unknown location parameter, g a known function satisfying $g\ (y\,;\phi) \to c\,(\phi)$ as $y \downarrow 0$ for some $c\,(\phi) > 0$, and ϕ a vector of unknown parameters. An example is the three-parameter Weibull model

$$f\ (x\,;\theta,\sigma,\alpha) = \alpha\sigma^{-\alpha}(x-\theta)^{\alpha-1}\exp[-\{(x-\theta)/\sigma\}^{\alpha}],\ x>\theta, \qquad (2.2)$$

for which $\phi=(\sigma,\alpha)$; $\sigma>0$, $\alpha>0$.

The following results apply to (2.2), and more generally to a wide class of distributions satisfying (2.1).

Consider first the case in which ϕ is known. It is readily checked that the Fisher information for θ is finite when $\alpha>2$, and then maximum likelihood estimation is regular. For $\alpha=2$, the maximum likelihood estimator is asymptotically both normal (Woodroofe 1972) and efficient (Weiss and Wolfowitz 1973), but the rate of convergence is $O\{(n\log n)^{-\frac{1}{2}}\}$ instead of $O(n^{-\frac{1}{2}})$. For $1<\alpha<2$ it has an extremely complicated limiting distribution described by Woodroofe (1974) but is not asymptotically efficient. For $0<\alpha\le1$ the likelihood is maximised by the sample minimum, which has the well-known Weibull limiting distribution (exact if the parent distribution is Weibull) but this has no optimality properties except when $\alpha=1$, when the sample minimum has an asymptotic sufficiency property (Weiss 1979). That the last result holds only when $\alpha=1$ is clear from the results of Janssen and Reiss (1988) on the asymptotic loss of information, incurred when restricting estimation to a finite number of lower order statistics. Asymptotic efficiency results are much harder to establish. Akahira (1975) established that the optimal rate of convergence is $O(n^{-\frac{1}{\alpha}})$ whenever $0<\alpha<2$, and Ibragimov and Has'minskii (1981) established very general results on the construction of asymptotically efficient estimators, but they are complicated and depend on the specified loss function. The situation is rather simpler if we restrict to $\alpha=1$, a case which covers the uniform and exponential distributions as well as most cases of truncated distributions. Some examples of optimal estimation in this case are the papers of Akahira (1982, 1988). Another line of attack has been generalisations of the Cramér-Rao inequality for the variance of unbiased estimators (Kiefer 1952, Polfeldt 1970, Vincze 1979, Akahira and Takeuchi 1987).

Smith (1985) generalised these results to problems of the form (2.1) in which ϕ as well as θ is unknown. In this case the key result is an orthogonality result (in a sense different from, but related to, the orthogonality property studied by Cox and Reid 1987) which essentially asserts that estimation of θ and ϕ, when the true α is less than 2, are asymptotically independent in the sense that neither one is affected by ignorance of the other. Thus the asymptotic distribution of the joint maximum likelihood estimates, when $1<\alpha<2$, was established as a product of Woodroofe's limiting distribution for $\hat{\theta}$ and the usual asymptotic normal distribution (constructed under the assumption θ known) for $\hat{\phi}$. Broader results along these lines, for instance using other than maximum likelihood estimates for θ, are possible.

An alternative approach, given in Section 5 of Smith (1985), is first to order the data $X_{n:1} \leq \ldots \leq X_{n:n}$, and then estimate ϕ by maximising

$$\sum_{i=2}^{n} \log f\,(X_{n:i}\;;X_{n,1}\,,\,\phi). \qquad (2.3)$$

This estimator of ϕ is consistent for all $\alpha>0$ and asymptotically efficient when $0<\alpha<2$. Proof of this proceeds via an asymptotic comparison of (2.3) with the ordinary log likelihood for ϕ when θ is known.

As a point of practical strategy, bearing in mind that α is usually unknown, my advice would usually be to start by treating the problem as a regular estimation problem, searching numerically for a local maximum of the log likelihood. If this fails, or leads to an estimate with $\hat{\alpha} < 2$, then it would be appropriate to switch to the more specialised techniques just described. Note, however, that there is another reason why maximum likelihood could fail, described in Section 3.5.

3. ALTERNATIVES TO MAXIMUM LIKELIHOOD

This section continues the discussion of Section 2, but is now focussed on alternative approaches to estimation.

3.1 Maximum product of spacings

Cheng and Amin (1983) defined a general method of estimation as follows. Let $X_1, \ldots, X_n$ be i.i.d. with distribution function $F\,(x,\Theta)$, where Θ is an unknown parameter vector. For given Θ, define $Y_0 = 0$, $Y_{n+1}=1$, $Y_i = F\,(X_{n:i}, \Theta)$ and thence

$$G = \left\{ \prod_{i=1}^{n+1} (Y_i - Y_{i-1}) \right\}^{1/(n+1)} .$$

Thus G is a normalised product of spacings of uniform order statistics, under Θ. The idea is to choose Θ to maximise G.

The method avoids some of the difficulties of maximum likelihood, since G is always bounded so there cannot be singularities. Cheng and Amin showed, for the kinds of problems discussed in Section 2, that the MPS estimate has all the usual asymptotic properties of maximum likelihood when $\alpha>2$ but also does well when $\alpha<2$: if $\Theta = (\theta,\phi)$ then ϕ is still efficiently estimated while the estimator of θ has properties which appear comparable with those of Section 2, though this last point has still not been completely settled. An advantage of the method is that it avoids the need to discriminate explicitly between the cases $\alpha>2$, $\alpha<2$ though this would still be necessary to form confidence intervals for θ. Ranneby (1984) independently discovered the MPS method, though without pointing out its usefulness for nonregular problems. Titterington (1985) gave it an interpretation as a grouped likelihood. This was developed further by Cheng and Iles (1987), who argued that both maximum likelihood and MPS estimators could be considered as approximations to a grouped likelihood estimator, but that only MPS is valid when $\alpha<2$. For more on grouped likelihood, see Section 3.4.

3.2 Bayesian estimation

There has been litte general discussion of Bayesian methods in the non- regular context. Dawid (1970) proved asymptotic normality of Bayes estimates when $\alpha>2$, but truly nonregular results do not seem to be available. Note, however, that Bayesian ideas underline the construction of many asymptotically efficient estimators, as in the work of Akahira and Ibragimov-Has'minskii cited in Section 2.

From the practical side Smith and Naylor (1987) presented a case study of the three-parameter Weibull distribution and concluded that Bayesian methods coped better than classical likelihood techniques with the highly unusual form of the log likelihood. In that case, however, the evidence was for $\alpha>>2$ and a different problem from the nonregularity arising when $\alpha<2$ (Section 3.5). The utility of Bayesian methods in the genuinely nonregular case has still to be explored.

3.3 Conditional and fiducial methods

For a pure location, or location-scale, model, the exact conditional method of Fisher (1934) is applicable. This method, which is operationally equivalent to Bayesian estimation with a suitable invariant prior, is not affected by the nonregularity of the problem. The extension to nuisance parameters, however, is not so easy. One could substitute a $n^{\frac{1}{2}}$ - consistent estimator for ϕ and then the asymptotic properties of estimation of θ would be unaffected, by the orthogonality results mentioned in Section 2, but Fisher's conditional distribution is no longer exact. Smith (1987) succeeded in applying the Cox-Reid (1987) technique to the three-parameter Weibull distribution, but with uncertain results. Ariyawansa and Templeton (1986), considering (2.2) and related models, proposed using Fisher's method to condition on $\tilde{\theta}$ and $\tilde{\sigma}$, where $\tilde{\theta}$ and $\tilde{\sigma}$ are equivarient estimators of θ and σ. Thus they obtained the joint conditional density of $(X_i - \tilde{\theta})/\tilde{\sigma}$ $(i=1,\dots,n-2)$ and maximised that to estimate α.

However, the motivation for this is unclear since $\tilde{\theta}$ and $\tilde{\sigma}$ are not ancillary for α. Another objection to their proposal is that, having estimated α by $\tilde{\alpha}$ say, they substituted $\tilde{\alpha}$ in Fisher's conditional density of $(\tilde{\theta},\tilde{\sigma})$ in order to obtain interval estimates for θ and σ. This ignores the estimation error in α. Overall, the application of conditional methods to nonregular problems is still a largely unexplored area.

3.4 Grouped Likelihood

Barnard (1967) and Giesbrecht and Kempthorne (1976) argued for a grouped likelihood approach, avoiding singularities. The motivation for this, very briefly, is that in the real world all data are rounded and so a grouped likelihood is really the correct likelihood function to use. Cheng and Iles (1987) took this as the starting point of their comparison of maximum likelihood and MPS methods.

A practical difficulty with grouped likelihood methods, when the grouping interval is not automatically specified by the data, is that the results can be sensitive to the choice of grouping interval. Nevertheless the idea can be recommended as a general technique for getting reasonable estimates in non-regular problems. Some recent examples include:

(a) Change-point problems. Matthews and Farewell (1985) used a grouped likelihood approach to estimate the change point in a hazard function.

(b) Transformations. Atkinson (1985) made a detailed study of Box-Cox transformations, but encountered difficulties with transformations of the icon

$$y \rightarrow \left\{(y+\mu)^{\lambda}-1\right\}/\lambda$$

with λ and μ both unknown; the difficulty arises from the constraint $y+\mu>0$. Atkinson, Pericchi and Smith (1989) gave a theoretical explanation for the difficulties of maximum likelihood estimation and proposed a grouped likelihood alternative.

(c) Time Series. A series of papers culminating with Lawrance and Lewis (1985) has proposed time series models with exponential and other non-normal marginal distributions. These models feature discontinuities in their conditional densities, and Raftery (1985) reviewed the implications of that from the point of view of non-regular estimation. Smith (1986) proposed a grouped likelihood approach in this case, but even the grouped likelihood is highly multimodal. In this case the nonregularity points to a difficulty with this class of models - do the discontinuities in the conditional densities correspond to physical features that one might reasonably expect to find in the data? If not, then the appropriateness of the model is very much in doubt.

3.5 The embedded model problem

The example studied by Smith and Naylor (1987) really illustrates a different kind of nonregularity in (2.2), close in spirit to the problems to be encountered in Section 4. If we write

$$\sigma = \psi a, \qquad \theta = \mu - \psi\alpha \qquad (3.1)$$

then as $\alpha \rightarrow \infty$ the density (2.2) tends to

$$\frac{1}{\psi}\exp\left\{\frac{x-\mu}{\psi} - \exp\left(\frac{x-\mu}{\psi}\right)\right\}, \qquad -\infty < x < \infty \qquad (3.2)$$

which is the Gumbel or Type I extreme value distribution for minima. Thus, if the data would be well modelled by (3.2), one would expect a degeneracy in the likelihood of (2.2) as $\alpha \rightarrow \infty$, or equivalently as $\theta \rightarrow -\infty$. A similar problem has long been known in the case of the three parameter lognormal distribution.

Recently Cheng and Iles (1989) have formalised this property by postulating the existence of an "embedded model" which arises as a limiting case of the original model. In the Weibull case (3.2) is the embedded model. In the lognormal case the embedded model turns out to be a normal distribution. Cheng and Iles proposed procedures for discriminating between the original and embedded models.

In the Weibull case the reparametrisation which combines (3.1) with $\gamma = -1/\alpha$ avoids the embedded model problem, since the resulting Generalised Extreme Value model is well-behaved near $\gamma = 0$. In this case Hosking (1984) reviewed tests for $\gamma = 0$ against alternatives $\gamma \neq 0$, but the nonregularity when $\gamma < -\frac{1}{2}$ of course remains.

3.6 Regression

The main model considered so far, in which observations are independent with common distributed satisfying (2.1) suggests many generalisations such as autoregressive models with (2.1) defining the innovations distribution, or regression models. The latter have been studied in Smith (1989). Suppose

$$y_i = \sum_{j=0}^{p} x_{ij}\,\beta_j + w_i, \quad 1 \le i \le n \tag{3.3}$$

where $\{x_{ij}\}$ are known covariates, $\{\beta_j\}$ unknown regression coefficients and $\{w_i\}$ independent random errors with density satisfying (2.1) with $\theta = 0$. We fix $x_{io} = 1$ so that β_0 is the constant in the regression; then there is no loss of generality in assuming

$$\sum_i x_{ij} = 0, \qquad j = 1, 2, \ldots, p. \tag{3.4}$$

n accordance with the preceding results when $\alpha > 2$ the problem is regular and I advocate maximum likelihood estimation. For $\alpha \le 2$, the problem is nonregular and the construction of asymptotically efficient estimators seems to be an even harder problem than in the situation of Section 2. To avoid these difficulties, Smith (1989) proposed the following solution generalising the estimators based on the sample minimum at the end of Section 2. Choose $\beta_0, \ldots, \beta_p$ to solve

$$\min \sum_i \Big(y_i - \sum_j x_{ij}\,\beta_j\Big) \text{ s.t. } y_i - \sum_j x_{ij}\,\beta_j \ge 0, \text{ each } i. \tag{3.5}$$

One motivation for (3.5) is that it is the maximum likelihood estimators of $\beta_0, \ldots, \beta_p$ when f is the exponential density; however, it is proposed for general f in the class (2.1). An argument based on the duality theorem of linear programming shows that, if we exclude a particular pathological case, the solution of (3.5) is such that $y_i = \sum_j x_{ij}\ \beta_j$ on a set of indexes $i \in \mathfrak{I}$, of cardinality $p+1$, such that the convex hull of the vectors

$$\{(x_{ij}, j = 1, \ldots, p), \quad i \in \mathfrak{I}\}$$

includes the zero vector. This assumes (3.4). For example if $p = 1$ the solution corresponds to a straight line joining exactly 2 points, whose x_{i1} values lie on opposite sides of 0, such that the straight line passes under all the other points. The pathological case just referred to arise when there is a point on the axis $x_{i1} = 0$, such that the pair (x_{i1}, y_i) lie below the convex hull formed by the other points. Then there are infinitely many straight lines through this point solving (3.5).

Smith (1989) uses his representation to find the exact and asymptotic distributions of the estimators, and also to study an estimator of ϕ, generalising (2.3), which is based on the residuals excluding points in $\mathfrak{I}$. Applications include extreme value problems with trends, a particular theme in hydrological and air pollution data analysis.

3.7 Summary

The MPS, Bayesian and conditional methods all have interesting features, but none appears at the present time to be a completely satisfactory alternative to likelihood procedures. This may change, however, as their properties become more fully developed. Other possibilities which have not been mentioned include the probability weighted moments methods of Hosking, Wallis and Wood (1985) and the idea of robust methods in nonregular problems, on which a start has been made by Boente and Fraiman (1988). The grouped likelihood idea is an all-purpose alternative. The embedded model problem has been mentioned mainly because it is a different kind of difficulty which arises with, amongst others, the three-parameter Weibull and lognormal distributions. Finally, Section 3.6 is a start on developing more general procedures based on nonregular families. There is clearly much scope for further developments of this nature.

4. PARAMETERS ON THE BOUNDARY OF THE PARAMETER SPACE

A quite different form of nonregularity arises when the true parameter lies on the boundary of the parameter space. Such problems have been considered by Chernoff (1954), Moran (1971), Chant (1974), Shapiro (1985) and Self and Liang (1987). We follow Self and Liang, whose paper reviewed all the earlier contributions.

It is assumed that the parameter space is a set $\Omega \subseteq \Re^p$, that the true parameter value is θ_0, and that in the neighbourhood of θ_0 the set Ω may be approximated by a cone, that is a set C such that if $x \in C$ then $\theta_0 + a(x - \theta_0) \in C$ for any $a > 0$. Self and Liang also assumed the classical Cramér conditions on the family of distributions - existence and positive definiteness of the Fisher information matrix $I(\theta)$, uniform boundedness of the third-order derivatives of the log likelihood by a function of finite expectation. Under these conditions they showed

(i) as sample size $n \to \infty$, there exists a sequence of points $\hat{\theta}_n \in \Omega$ which locally maximise the likelihood function, such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_o)$ is $0(1)$ in probability,

(ii) the limiting distribution of $n^{\frac{1}{2}}(\hat{\theta}_n - \theta_o)$ is the same as the distribution of the maximum likelihood estimators $\hat{\theta}$ based on a single realisation of a multivariate normal Z, with mean θ and covariance $I^{-1}(\theta_0)$, where the permitted range of Z is $C - \theta_0$ (i.e. C translated by θ_0).

Although the representation in (ii) makes clear the structure of the solution, there is still much work to be done to obtain explicit results. The simplest case is when $C - \theta_0$ is of the form $(0, \infty) \times \{(-\infty, \infty)\}^{p-1}$, so that only the first component of θ_0 lies on the boundary. Then the asymptotic distribution is a mixture of two multivariate normal components, one corresponding to $Z_1 > 0$ and the other to $Z_1 \leq 0$. Here Z_1 is the first component of Z. The case when two components of θ_0 lie on the boundary was also considered by Moran, Chant, and Self and Liang, but in this case Self and Liang represented it as a mixture of four components which are not multivariate normal. Cases with more than two components on the boundary lead to even more complicated mixtures.

A similar representation also applies to the asymptotic distribution of a likelihood ratio statistic. Consider the problem of testing $\theta \in \Omega_0$ against $\theta \in \Omega_1$, where Ω_0 is a subset of Ω and Ω_1 is its complement in Ω. Let C_0 and C_1 denotes cones approximating Ω_0 and Ω_1 respectively. Let λ_n denote the likelihood ratio statistic for testing Ω_0 against Ω_1. Then

(iii) the limiting distribution of $-2\log\lambda_n$, when θ_0 is the true parameter value, is the same as the distribution of the likelihood ratio test of $\theta \in C_0$ against $\theta \in C_1$ based on a single realisation of a multivariate normal Z with mean θ and covariance $I^{-1}(\theta_0)$.

In regular cases (no parameters on the boundary) it can easily be seen that this representation leads to the well-known asymptotic chi-squared distribution for the likelihood ratio statistic. In nonregular cases it leads to a variety of results depending on the numbers of test parameters and nuisance parameters on the boundary. Some of these lead to mixtures of chi-squared distributions, some to much less tractable solutions.

As previously mentioned, these results assume the Cramér conditions on the family of distribution, in particular, existence of the Fisher information matrix. In Section 6 we shall see some examples where this condition is not satisfied.

5. HYPOTHESIS TESTS INVOLVING PARAMETERS NOT IDENTIFIABLE

Suppose we have a model indexed by (vector) parameters (ξ,θ). We want to test $H_0 : \xi = \xi_0$ against $H_i : \xi \neq \xi_0$. The difficulty arises when some or all components of θ are not identifiable when H_0 is true. Examples arise in many contexts, such as mixture and change-point problems, model identification in time series analysis and econometrics, and categorical data analysis. For such problems the standard results, such as the asymptotic chi-squared distribution of the log likelihood ratio statistic, are genreally false, and the correct results depend very much on the precise problem being investigated.

Davies (1977, 1987) has presented a general approach to such problems. It is my opinion that Davies' approach deserves to be better known and understood, but that it nevertheless falls a long way short of being a complete solution to such problems. In numerous specific cases, better results are available, but overall this is an area still awaiting a general theory. An earlier review, with somewhat similar aims and conclusions to this, is that of Berman (1986).

5.1 Davies' Approach

Suppose ξ is p-dimensional and that, when H_0 is true, θ lies in a set $\Theta \subseteq \mathbb{R}^q$. Suppose $S_n(\theta)$ is a test statistic for H_0 valid for given $\theta \in \Theta$ and sample size n. For example, $S_n(\theta)$ could be twice the log likelihood ration. Suppose, as $n \to \infty$, we have

$$S_n \to S \qquad (5.1)$$

in the sense of weak convergence on Θ. Then it is possible to base a test of H_0 on $\max\{S_n(\theta) : \theta \in \Theta\}$. The limiting distribution is then calculated from the limiting process S.

Davies' results were confined to the case $q = 1$ and instances for which $S(\theta)$ is a continuous, one-parameter stochastic process with differentiable sample paths. For such processes, the celebrated Rice's formula stemming originally from Rice (1945) gives a good approximation to the tail distribution of the maximum of S. This formula states that, under suitable regularity conditions, the expected number of upcrossings of a curve $c(\theta)$ by $\{S(\theta), \theta \in (a, b)\}$ is given by

$$M(c) = \int_a^b f_{S(\theta)}(c(\theta))\, E\{(S''(\theta)) + \mid S(\theta) = c(\theta)\}\, d\theta, \qquad (5.2)$$

where $f_{S(\theta)}$ is the density of $S(\theta)$. Combined with the asymptotic Poisson character of the process of upcrossings, we have for a high curve $c(\theta)$,

$$P\left\{ \max_{a \le \theta \le b} (S(\theta) - c(\theta)) > 0 \right\} \approx P\{S(a) > c(a)\} + M(c). \qquad (5.3)$$

Sometimes we want to work with $|S|$ instead of S and then a similar expression for the lower tail must be added to (5.3).

Davies (1977) considered the case $p = 1$, when S may generally be taken to be a continuous-parameter Gaussian process with zero mean and constant variance (in the likelihood ratio case, via a signed square root transformation). Let $\rho(\theta_1, \theta_2)$ denote the covariance function of S and let

$\rho_{11}(\theta) = [\partial^2\rho / \partial\theta^2{}_1]_{\theta_1=\theta_2=\theta}$, assumed to exist.

Then by Davies (1977), equation (3.6), adapting Section 13.2 of Cramér and Leadbetter (1967), we have

$$M(c) = \frac{1}{2\pi}\int_a^b \exp\left\{-\frac{c^2(\theta)}{2}\right\}\{-\rho_{11}(\theta)\}^{\frac{1}{2}}\,\psi\left[\frac{c'(\theta)}{\{-\rho_{11}(\theta)\}^{\frac{1}{2}}}\right]d\theta \quad (5.4)$$

where $\psi(x) = \exp(-x^2/2) - x\int_x^\infty \exp(-t^2/2)dt$.. Davies (1987) has suggested some quick sample-based approximations to (5.4).

Davies (1987) extended this to p>1. In this case S may be taken to be a chi-squared process, which is a sum of squares of independent Gaussian processes. Here he gave a formula evaluating (5.2) in the case of components with different covariance functions, thus generalising the identically-distributed summands case which had been developed in great detail by Sharpe (1978), Lindgren (1980) and Aronowich and Adler (1985, 1986). Detailed discussion of Rice's formula and its generalisations is due to Belyaev (1968) and Marcus (1977). For broader reviews of extreme values in continuous-parameter processes see Leadbetter, Lindgren and Rootzén (1983) and Leadbetter and Rootzén (1988). For a different but non-rigorous approach, see Aldous (1989).

In cases where these formulae apply, they give simple and reasonably accurate approximations to the tail probabilities of the test statistic.

5.2 Deficiencies in Davies' approach

(a) If q>1 then the limiting process is a random field, whose extremes are much less well understood than those of a one-parameter process. Early papers in this area include Belyaev (1967) and Qualls and Watanabe (1973). More recent references include Sigmund (1988), whose technique seems powerful but restricted in the class of fields to which it applies, and Adler and Samorodnitsky (1987), whose Theorem 2.1 provides a general bound for crossing probabilities of a Gaussian random field. With this result we can get an approximation in the case $p=1$, $q>1$. When p and q are both >1 we need a theory of chi-squared random fields, for which I am not aware of any results at the present time.

(b) For certain problems, especially those connected with change-points, the limiting process S has non-differentiable sample paths. For extreme values in such cases see Chapter 12 of Leadbetter, Lindgren and Rootzén (1983); however, the detailed results are confined to stationery Gaussian processes. Alternatively, Leadbetter and Rootzén (1988) review diffusion processes. The best known examples of this type all concern change-points, for which alternative methods have been developed (Section 5.3). Some other examples where the Davies approach fails were given by Berman (1986).

(c) The integral in (5.2) or (5.4) may not be finite without an artificial restriction on the interval (a,b). In this case an alternative scaling may be needed to obtain a limiting result.

A still-unresolved example of this difficulty was given by Hartigan (1985) and concerns the simplest kind of mixture problem, for which

$$f(x;\xi,\theta) = (2\pi)^{-\frac{1}{2}}\left\{(1-\xi)e^{-x^2/2} + \xi\, e^{-(x-\theta)^2/2}\right\} \quad (5.5)$$

and the null hypothesis is $\xi=0$. For fixed $\theta\neq 0$, Taylor expansion about $\xi=0$ shows that the likelihood ratio test may be based on

$$S_n(\theta) = \Sigma Z_i(\theta)/\{\Sigma Z_i^2(\theta)\}^{\frac{1}{2}}$$

where $Z_i(\theta) = \exp(X_i\ \theta - \frac{1}{2}\theta^2)$ -1, X_i being the i'th datum. As $n \to \infty$, so S_n converges to a Gaussian process with covariance function

$$\rho(\theta_1, \theta_2) = \{\exp(\theta_1, \theta_2)-1\}\{\exp(\theta_1^2)-1\}^{-\frac{1}{2}}\{\exp(\theta_2^2)-1\}^{-\frac{1}{2}}. \qquad (5.6)$$

For this it can be checked that (5.4) is valid provided (a,b) is a finite interval bounded away from zero, but as $\theta \to 0$ or $\pm\infty$ the integral blows up. This shows that the likelihood ratio test statistic tends to ∞ under the null hypothesis. Hartigan conjectured that the rate of growth is 0(log log n) but neither this not any limiting distribution have been proved.

There is a large literature on mixture problems, including the review paper of Redner and Walker (1984) and the books of Titterington, Smith and Makov (1985) and MacLachlan and Basford (1988). Nevertheless, this example shows that there are some quite fundamental problems still unresolved.

(d) Of course, one way to resolve all these difficulties is to use simulation or a bootstrap-type technique, but this raises another question : why use a likelihod-ratio procedure in the first place? The regular asymptotic optimality results do not go through and, in such cases where Bayesian approaches have been developed, they appear to have been successful, though direct comparisons are lacking in most cases.

5.3 Alternative methods for particular problems

In the absence of a fully satisfactory general theory, most of the individual problems in this field have developed a substantial literature of their own. In this section I review some of these development.

(a) Change-point problems. These problems have been intensively studied and have spawned a vast literature which can only very briefly be covered here. The simplest problem concerns an independent sequence $X_1, ..., X_n$, where $X_i \sim N(\mu_i, 1)$, and the hypothesis to be tested are

$$H_0 : \mu_1 = \ldots = \mu_n$$

$$H_1 : \text{For some } m < n,\ \mu_1 = \ldots = \mu_m \neq \mu_{m+1} = \ldots = \mu_n$$

Then the square root of the log likelihood ratio statistic (James et al. 1987) is

$$\max_{1 \le m < n} \left[\left| mS_n/n - S_m \right| / \{m(1-m/n)\}^{\frac{1}{2}} \right] \qquad (5.7)$$

James et al. described approximations to the null distribution of (5.7) based on methods of sequential analysis. Worsley (1986) has extended the method to a general exponential family and proposed a direct computational method for the size of the test. Siegmund (1988) reviewed these and many related developments which go back to Chernoff and Zacks (1964) and Hinkley (1970).

In these cases the direct approach just outlined seems superior to the Davies approach based on a limiting stochastic process. There are related problems, however, relating to change-points in continuous-parameter processes such as Poisson process (Kendall and Kendall 1980, Akman and Raftery 1986) and hazard functions (Matthews, Farewell and Pyke 1985). For these problems the Davies approach seems to be needed, and reduces to calculating probabilities of the form

$$P\left\{ \max_{t_0 \le t \le t_1} \frac{W_0(t)}{\{t(1-t)\}^{\frac{1}{2}}} \ge c \right\} \tag{5.8}$$

where $0 < t_0 < t_1 < 1$ and W_0 is a Brownian bridge on (0,1). For this process equation (5.4) is inapplicable but direct calculations are due to Mandl (1962), Dirkse (1975), Kielson and Ross (1975) and De Long (1981); see also equation (26) of James et al. (1987). A complication here is the restriction to $t_0 > 0$, $t_1 < 1$ meaning that the change-point has to be bounded away from the endpoints - an instance of problem (c) mentioned in the previous subsection. However, it appears that it is only with the direct approach of Sigmund and Worsley that this difficulty is avoidable.

Alternatives to likelihood ratio and score-test procedures have also been proposed. For example Cobb (1978) and Hinkley and Schechtman (1987) proposed conditional approaches, while Smith (1975) and Raftery and Akman (1986) took a Bayesian line. Comparisons by James et al (1987) suggest that the Bayesian approach has superior power when the change-point is near the middle of the time interval, likelihood ratio procedures doing better only when the change-point is near the ends.

This brief review has concentated on testing for the existence of a change-point, which is only part of the problem of inference. There are many other papers on estimating change-points. Overall, it appears that the approach of Siegmund and Worsley is the most powerful for handling simple change-point problems but there are many more complicated problems for which the Davies approach seems to be required. In these cases the key formulae are (5.8) and the various approximate methods for its evaluation.

(b) Certain structural problems in categorical data analysis are of the type being considered here. For example Haberman (1981) considered a two-way table with cell probabilities of form

$$p_{ij} = \exp(\alpha + \beta_i + \gamma_j + \psi\mu_i \upsilon_j). \tag{5.9}$$

Under $H_0 : \psi = 0$ the parameters μ_i, υ_j are not identifiable so we have a problem of the type being studied. Haberman showed that the likelihood ratio statistic is equivalent to a test based on the canonical correlation, and hence that its limiting distribution is that of the largest root of Wishart matrix. There is no obvious connection between this and the Davies approach.

More examples of models and tests of this form are contained in the paper and discussion of Anderson (1984).

(c) Time series and econometrics are full of identifiability problems; see for example Sargan (1983) for some of the nonregular asymptotics that can ensue. As an example of the Davies approach in this context, consider the problem of testing whether the autoregressive and moving-average components of an ARMA model have a common root. I consider here the simplest case, that of an ARMA (1,1) model with known mean

$$X_n - aX_{n-1} = \varepsilon_n - b\,\varepsilon_{n-1} \tag{5.10}$$

with $|a| < 1$, $|b| < 1$ and $\{\varepsilon_n\}$ a sequence of independent $N(0,1)$ variables. Under $H_0 : a = b$, (5.10) reduces to $X_n = \varepsilon_n$ and we have a white noise sequence.

Consider first the case where b is known. By re-writing (5.10) in the form

$$\varepsilon_n = X_n + \sum_{r=1}^{\infty} (b^r - ab^{r-1}) X_{n-r}$$

the "Conditional Least Squares" (Box and Jenkins 1976) estimator of a based on $\{X_n, 1 \le n \le N\}$ is obtained by minimising

$$\sum_{n=1}^{N} \{X_n - \sum_{r=1}^{n-1} (b^r - ab^{r-1}) X_{n-r}\}^2. \tag{5.11}$$

Here (5.11) is approximately twice the log likelihood. By minimising with respect to a and comparing with the case $a = b$ we obtain the signed square root of the likelihood ratio statistic in the form

$$\left(\sum_{n=1}^{N} \sum_{r=1}^{n-1} X_n X_{n-r} b^{r-1}\right) / \left\{\sum_{n=1}^{N} \sum_{r=1}^{n-1} b^{2r-2} X^2_{n-r}\right\}^{\frac{1}{2}}$$

and the law of large numbers applied to the denominator shows that this in turn is asymptotically equivalent to

$$S_n(b) = \left(\frac{1-b^2}{N}\right)^{\frac{1}{2}} \sum_{n=1}^{N} \sum_{r=1}^{n-1} X_n X_{n-r} b^{r-1}.$$

An easy application of the Martingale CLT shows that $S_n(b)$ is asymptotically $N(0,1)$ for fixed b, and that the limiting correlation of $S_n(b_1), S_n(b_2)$ is given by

$$\rho(b_1, b_2) = (1-b_1^2)^{\frac{1}{2}} (1-b_2^2)^{\frac{1}{2}} (1-b_1 b_2)^{-1}.$$

It is then easily checked that ρ satisfies the condtions required for (5.4) with

$$-\rho_{11}(b) = (1-b^2)^{-2}.$$

Consequently Davies' approach is applicable provided we restrict b to a closed subinterval of (-1, 1).

To sum up, Davies' approach certainly has the potential to be applied to a far wider range of problems than it has in the literature so far. However, a number of alternative solutions exist for particular problems and this demonstrates that it is not a universally applicable method.

6. MULTIVARIATE EXTREME VALUE AND RELIABILITY DISTRIBUTIONS

In this section I briefly review some recent work which encompasses all three types of nonregularity considered in this paper.

Multivariate extreme value distributions (Resnick 1987, Smith, Tawn and Yuen 1989) generalise the extreme-values stability property to higher dimensions: if $\underset{\sim}{X}_1, ..., \underset{\sim}{X}_n$ are independent p - vectors from a multivariate extreme value distribution and M_n denotes the vector of componentwise maxima, then a suitable location-scale transformation, applied separately to each component, reduces the distribution of $\underset{\sim}{M}_n$ to that of $\underset{\sim}{X}_1$. Such distributions are useful, for example, in studying the joint distribution of floods at several sites, or of lifetimes of interrelated components in a reliability context.

The bivariate case has been studied in detail by Tawn (1988). Following a representation of Pickands (1981), it is possible to transform the margins to unit exponentiality and maxima into minima, when we get a representation

$$P\{X > x, Y > y\} = \exp\{-(x+y)A(y/(x+y)\}$$

for some convex function A on [0,1] satisfying $A(0) = A(1) = 1$. Tawn considered several parametric models for A, the most successful being Gumbel's logistic model

$$A(w) = \{(1-w)^r + w^r\}^{\frac{1}{r}}, \quad r \geq 1 \qquad (6.1)$$

and an asymmetric extension

$$A(w) = \{\theta^r(1-w)^r + \phi^r w^r\}^{\frac{1}{r}} + (\theta-\phi)w + 1-\theta. \quad (0 \leq \theta, \phi \leq 1, r \geq 1) \qquad (6.2)$$

Some non-regular problems asociated with these are:

(a) In (6.1), the Fisher information for r blows up as $r \rightarrow 1$; hence the problem of testing independence ($r = 1$) is not only one-sided (cf. Section 4 above) but also not solvable by standard central-limit theory. In this case Tawn applied the theory of stable distributions to show that the results of Section 4 essentially still hold, but with a non-standard rate of convergence.

(b) If we apply (6.1) in a context with the marginal distributions are initially unknown, but follow a generalised extreme value distribution then the nonregularity just mentioned must be combined with the kind studied in Section 2. In this case Tawn was able to prove an orthogonality result analogous to that described in Section 2.

(c) Model (6.2) introduces additional complications: independence may correspond to any of θ=0, ϕ=0 or $r = 1$ so we have a degeneracy problem of the type considered in Section 5. This has still not been resolved.

In spite of the compliations these ideas introduce, the main message of Tawn's paper is that likelihood methods are applicable to these families, in contrast to the more <u>ad hoc</u> methods of Tiago de Oliveira (1984) and others. Extensions to higher dimensions have been considered by Tawn (1989).

Crowder (1989a) considered the family

$$P\{T_1 > t_1, \ldots, T_p > t_p\} = \exp\{\kappa^\nu - (\kappa + \sum_1^p \xi_j t_j^{\phi_j})^\nu\} \qquad (6.3)$$

for the joint distribution of survival times $T_1, \ldots, T_p$; here $\kappa \geq 0$, $0 < \nu \leq 1$ and ξ_j, ϕ_j positive. The case $\kappa = 0$ reduces to a well-known multivariate Weibull family, the multivariate generalisation of (6.1), but the case $\kappa > 0$ is new. Crowder proposed its use as a distribution for repeated measures analysis.

Crowder (1989b) has studied some of the nonregular aspects of this. Testing $H_0 : \kappa = 0$ is within the framework of Section 4, but in this case the score statistic turns out to be infinite if $\nu < 1$. Crowder proposed a modified version of the score-statistic and established its asymptotic distribution. When $0 < \nu < 1$ this turns out to be stable with index ν. Crowder also considers the test of $H_0 : \nu = 0$, in which case κ is not identifiable and we again have a problem of the type studied in Section 5. In this case Crowder showed that Davies' approach is applicable though it is not easy to evaluate (5.4) directly.

7. CONCLUSIONS

The wide range of examples considered in this paper shows that nonregular problems are not mere curiosities but arise in many practical contexts. The discussion of this paper has focussed on first-order asymptotic results since in most cases these are the only ones available. Such studies as hae been performed have suggested that such results may often given rather poor approximations in practice, so that one should consider supplementing the theoretical results by direct calculation or simulation. However, there are at least two reasons why I do not think that theoretical treatment of these problems should be repalced entirely by simulation. The first is that in many of these problems the asymptotic results is a fairly simple one, even though non-standard, and can therefore very usefully be applied in the preliminary stages of an analysis, when precise computation of significance levels is not too important. Simulation can be used later, if desired, when a specific model has been identified. The second reason is that simulation tells us nothing about which estimator or test to use. In many cases the nonregular theory points to something other than mechanical application of maximum likelihood estimation.

ACKNOWLEDGEMENTS

This paper has grown from a talk I gave at the RSS-sponsored Workshop on Asymptotic Statistics at Edinburgh 1986, funded by the SERC. I thank numerous participants at the Workshop for their comments but especially my Surrey colleagues Martin Crowder, Trevor Sweeting and Jonathan Tawn. Thanks also to Nancy Reid for the invitation to present the paper here.

REFERENCES

Adler, R.J. and Samorodnitsky, G. (1987), Tail behaviour for the suprema of Gaussian processes with applications to empirical processes. Ann. Probab. 15, 1339-1351.

Akahira, M. (1975), Asymptotic theory for estimation of location in non-regular cases. I: Order of convergence of consistent estimatores. II: Bounds of asymptotic distributions of consistent estimators. Rep. Statist. Appl. Res. Union Jap. Sci. Eng. 22, 8-26 and 99-115.

Akahira, M. (1982), Asymptotic optimality in non-regular cases. Ann. Inst. Stat. Math. 34, 69-82.

Akahira, M. (1988), Second order optimality of estimators for a density with finite cusps. Ann. Inst. Stat. Math. 40, 311-328.

Akahira, M. and Takeuchi, K. (1987), Locally minimum variance unbiased estimators in a discontinuous density function. Metrika 34, 1-15.

Akman, V. and Raftery, A.E. (1986), Asymptotic inference for a change-point Poisson process. Ann. Statist. 14, 1583-1590.

Aldous, D.J. (1989), Probability Approximations via the Poisson Clumping Heuristic. Applied Mathematical Sciences series no. 77, Springer Verlag, New York.

Anderson, J.A. (1984), Regression and ordered categorical variables (with discussion). J.R. Statist. Soc. B 46, 1-30.

Ariyawansa, K.A. and Templeton, J.G.C. (1986), Structural inference for parameters of a power function distribution. Stat. Hefte 27, 117-139.

Aronowich, M. and Adler, R.J. (1985), Behavior of chi-squared processes at extrema. Adv. Appl. Probab. 17, 280-297.

Aronovich, M. and Adler R.J. (1986), Extrema and level crossings of chi-squared processes. Adv. Appl. Prob. 18, 901-920.

Atkinson, A.C. (1985), Plots, Transformations and Regression. OUP, Oxford.

Atkinson, A.C., Pericchi, L.R. and Smith, R.L. (1989), Grouped likelihood for the shifted power transformation. Submitted.

Barnard, G.A. (1965), The use of the likelihood function in statistical practice. Proc. 5th Berk. Symp. I, 27.

Belayev, Yu. K. (1967), Bursts and shines of random fields. Soviet. Math. Dokl. 8, 1107-1109.

Belyaev, Yu. K. (1968), On the number of exits across the boundary of a region by a vector stochastic process. Theor. Probab. Appl. 13, 320-324.

Berman, M. (1986), Some unusual examples of extrema associated with hypothesis tests when nuisance parameters are present only under the alternative. Proc. Pacific Stat. Congress, ed. I.S. Francis et al. Elsevier Science Publishers B.V., North Holland.

Boente, G. and Fraiman, R. (1988), On the asymptotic behaviour of general maximum likelihood estimates for the nonregular case under nonstandard conditions". Biometrika 75, 35-44.

Box, G.E.P. and Jenkins, G.M. (1976), Time Series Analysis, Forecasting and Control (Second Edition). Holden Day, San Francisco.

Chant, D. (1974), On asymptotic tests of composite hypotheses in nonstandard conditions. Biometrika 61, 291-298.

Cheng, R.C.H. and Amin, N.A.K. (1983), Estimating parameters in continuous univariate distributions with shifted origin. J. R. Statist. Soc. B 45, 394-403.

Cheng, R.C.H. and Iles, T.C. (1987), Corrected maximum likelihood in non-regular problems. J.R. Statist. Soc. B. 49, 95-101.

Cheng, R.C.H. and Iles, T.C. (1989), Embedded models in three parameter distributions and their estimation. J.R. Statist. Soc. B., to appear.

Chernoff, H. (1954), On the distribution of the likelihood ratio. Ann. Math. Stat. 25, 573-578.

Chernoff, H. and Zacks, S. (1964), Estimating the current mean of a normal distribution which is subject to changes in time. Ann. Math. Statist. 35, 999-1028.

Cobb, G.W. (1978), The problem of the Nile: conditional solution to a change point problem. Biometrika 62, 243-251.

Cox, D.R. and Hinkley, D.V. (1974), Theoretical Statistics. Chapman and Hall, London.

Cox, D.R. and Reid, N. (1987), Parameter orthogonality and conditional inference. J.R. Statist. Soc. B 49, 1-39.

Cramér, H. (1946), Mathematical Methods of Statistics. Princeton University Press.

Cramér, H. and Leadbetter, M.R. (1967). Stationary and Related Stochastic Processes. Wiley, New York.

Crowder, M. (1989a), A multivariate distribution with Weibull connections. J.R. Statist. Soc. B 51, 93-107.

Crowder, M. (1989b), On some nonregular tests for a Weibull repeated measures model. Preprint, University of Surrey.

Daniels, H.E. (1961), The asymptotic efficiency of a maximum likelihood estimator. Proc. Fourth Berk. Symp. I, 151-163.

Davies, R.G. (1977), Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 64, 247-254.

Davies, R.G. (1987), Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33-43.

Dawid, A.P. (1970), On the limiting normality of posterior distributions. Proc. Camb. Phil Soc. 67, 625-633.

De Long, D.M. (1981), Crossing probabilities for a square root boundary by a Bessel process. Comm. Statist. A10, 2197-2213.

Dirkse, J.P. (1975), An absorption probability for the Ornstein-Uhlenbeck process. J. Appl. Prob. 12, 595-599.

Fisher, R.A. (1934), Two new properties of mathematical likelihood. Proc. R. Soc. A 144, 285-307.

Giesbrecht, F. and Kempthorne, O. (1976), Maximum likelihood estimation in the three-parameter lognormal distribution. J. R. Statist. Soc. B 38, 257-264.

Griffiths, D.A. (1980), Interval estimation for the three-parameter lognormal distribution via the likelihood function. Appl. Statist. 29 (1980), 58-68.

Haberman, S.J. (1981), Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction. Ann. Statist. 9, 1178-1186.

Hartigan, J.A. (1985), A failure of likelihood asymptotics for the mixture model. Proceedings of the Berkeley Symposium in honor of Jerzy Neyman and Jack Kiefer, Vol. II, L. LeCam and R.A. Olshen (eds.), Wadsworth, New York, pp. 807-810.

Hill, B.M. (1963), The three-parameter lognormal distribution and Bayesian analysis of a point-source epidemic. J. Amer. Statist. Assoc. 58, 72-84.

Hinkley, D.V. (1970), Inference about the change-point in a sequence of random variables. Biometrika 57, 1-17.

Hinkley, D.V. and Schechtman, E. (1987), Conditional bootstrap methods in the mean shift model. Biometrika 74, 85-94.

Hosking, J.R.M. (1984), Testing whether the shape parameter is zero in the generalized extreme-value distribution. Biometrika 71, 367-374.

Hosking, J.R.M., Wallis, J.R. and Wood, E.F. (1985), Estimation of the generalised extreme-value distribution by the method of probability-weighted moments. Technometrics 27, 251-261.

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

James, B., James, K.L. and Siegmund, D. (1987), Tests for a change point. Biometrika 74, 71-83.

Janssen, A. and Reiss, R.-D. (1988), Comparison of location models of Weibull type samples and extreme value processes. Probab. Th. Rel. Fields 78, 273-292.

Kendall, D.G. and Kendall, W.S. (1980), Alignments in two-dimensional random sets of points. Adv. Appl. Prob. 12, 380-424.

Kiefer, J. (1952), On minimum variance in non-regular estimation. Ann. Math. Statist. 23, 627-629.

Kielson, J. and Ross, H.F. (1975), Passage time distributions of Gaussian Markov (Ornstein-Uhlenbeck) processes. In, Selected Tables in Mathematical Statistics 3, Institute of Mathematical Statistics. Providemce: American Mathematical Society, pp. 233-328.

Lawrance, A.J. and Lewis, P.A.W. (1985), Modelling and residual analysis of nonlinear autoregressive time series in exponential variables. J. Roy. Statist. Soc. B.

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

Leadbetter, M.R. and Rootzén, H. (1988), Extremal theory for stochastic processes. Ann. Probab. 16, 431-478.

LeCam, L. (1970), On the assumptions used to prove asymptotic normality of maximum likelihood estimators. Ann. Math. Statist. 41, 802-828.

Lindgren, G. (1980), Extreme values and crossings for the chi-squared process and other functions of multidimensional Gaussian processes, with reliabaility applications. Adv. Appl. Prob. 12, 746-774.

Mandl, P. (1962), On the distribution of the time which the Uhlenbeck process requires to exceed a boundary (in Czech) Apl. Mat. 7, 141-148.

Marcus, M.B. (1977), Level crossings of a stochastic process with absolutely continuous sample paths. Ann. Probab. 5, 52-71.

Matthews, D.E., Farewell, V. and Pyke, R. (1985), Asymptotic score statistic processes and tests for constant hazard against a change-point alternative. Ann. Statist. 13, 583-591.

McLachlan, G.J. and Basford, K.E. (1988), Mixture Models: Inference and Applications to Clustering. Dekker, New York.

Moran, P.A.P. (1971), Maximum likelihood estimation in non-standard problems. Proc. Camb. Phil. Soc. 70, 441-450.

Pickands, J. (1981), Multivariate extreme value distributions. Proceedings of the 43th Session of the I.S.I. (Buenos Aires) 859-878.

Pickands, J. and Raghavachari, M. (1987), Exact and asymptotic inference for the size of a population. Biometrika 74, 355.

Polfeldt, T. (1970), The order of the minimum variance in a non-regular case. Ann. Math. Statist. 41, 667-672.

Qualls, C. and Watanabe, H. (1973), Asymptotic properties of Gaussian random fields. Trans. Amer. Math. Soc. 177, 155-171.

Raftery, A.E. (1985), Discussion of Lawrance and Lewis (1985). J. R. Statist. Soc. B 47, 185-187.

Raftery, A.E. (1988), Inference for the binomial N parameter: a hierachical Bayes approach. Biometrika 75, 223-228.

Raftery, A.E., and Akman, V.E. (1986), Bayesian analysis of a Poisson process with a change-point. Biometrika 73, 85-89.

Ranneby, B. (1984), The maximum spacing method, an estimation method related to the maximum likelihood method. Scand. J. Statist. 11, 93-112.

Redner, R.A. and Walker, H.F. (1984), Mixture densities, maximum likelihood and the EM algorithm. SIAM Review 26, 195-239.

Resnick, S. (1987), Extreme Values, Point Processes and Regular Variation. Springer Verlag, New York.

Rice, S.O. (1944,1945) Mathematical analysis of random noise. Bell Syst. Tech. J. 23, 282-332 and 24, 46-156.

Sargan, J.D. (1983), Identification and lack of identification. Econometrica 51, 799-820.

Self, S.G. and Liang, K.-Y. (1987), Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Amer. Statist. Assoc. 82, 605-610.

Shapiro, A. (1985), Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. Biometrika 72, 133-144.

Sharpe, K. (1978), Some properties of the crossings process generated by a stationary chi-squared process. Adv. Appl. Prob. 10, 373-391.

Siegmund, D. (1988a), Confidence sets in change-point problems. Int. Statist. Rev. 56, 31-48.

Siegmund, D. (1988b), Approximate tail probabilities for the maxima of some random fields. Ann. Probab. 16, 487-501.

Smith, A.F.M. (1975), A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika 62, 407-416.

Smith, R.L. (1985), Maximum likelihood estimation in a class of non-regular cases. Biometrika 72, 67-92.

Smith, R.L. (1986), Maximum likelihood estimation for the NEAR(2) model. J. R. Statist. Soc. B 48, 251-257.

Smith, R.L. (1987), Discussion of Cox and Reid (1987). J.R. Statist. Soc. B, 49, 21-22.

Smith, R.L. (1989), Nonregular regression. Submitted.

Smith, R.L. and Naylor, J.C. (1987), A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. Appl. Statist. 36, 358-369.

Smith, R.L., Tawn, J.A. and Yuen, H.-K. (1989), Statistics of multivariate extremes. Submitted.

Tawn, J.A. (1988), Bivariate extreme value theory - models and estimation. Biometrika 75, 397-415

Tawn, J.A. (1989), Modelling multivariate extreme value distributions. Submitted.

Tiago de Oliveira, J. (1984), Bivariate models for extremes. In, Statistical Extremes and Applications, J. Tiago de Oliveira (ed.), 131-153, Reidel, Dordrecht.

Titterington, D.M. (1985), Comment on "Estimating parameters in continuous univariate distributions". J.R. Statist. Soc. B 47, 115-116.

Titterington, D.M., Smith, A.F.M. and Makov, U.E. (1985), The Statistical Analysis of Finite Mixture Distributions. Wiley, Chichester.

Vincze, I. (1979), On the Cramér-Fréchet-Rao inequality in the non-regular case. In, Contributions to Statistics, the Jaroslav Hajek Memorial Volume, Academia, Prague, 253-262.

Weiss, L. (1979), Asymptotic sufficiency in a class of non-regular cases. Selecta Statistica Canadiana 5, 143-150.

Weiss, L. and Wolfowitz, J. (1973), Maximum likelihood estimation of a translation parameter of a truncated distribution. Ann. Statist. 1, 944-947.

Weiss, L. and Wolfowitz, J. (1974), Maximum Probability Estimators and Related Topics. Lecture Notes in Mathematics 424, Springer, New York.

Williamson, J.A. (1984), A note on the proof by H.E. Daniels of the asymptotic efficiency of a maximum likelihood estimator. Biometrika 71, 651-653.

Woodroofe, M. (1972), Maximum likelihood estimation of a translation parameter of a truncated distribution. Ann. Math. Stat. 43, 113-122.

Woodroofe, M. (1974), Maximum likelihood estimation of translation parameter of truncated distribution II. Ann. Statist. 2, 474-488.

Worsley, K.J. (1986), Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika 73, 91-104.