Likelihood-Based Inference in Some Continuous Exponential Families With Unknown Threshold Parameters

Thomas Michael Dubinin
William E. Wecker Associates, Inc.
505 San Marin Drive
Novato, CA 94945
dubinin@mail.wecker.com

Stephen Bruce Vardeman
Statistics and IMSE Departments
Iowa State University
Ames, IA 50011-1210
vardeman@iastate.edu

September 3, 2002

Abstract

We consider likelihood-based inference in some continuous exponential families with unknown threshold parameters. The introduction of threshold parameters necessitates modification of the standard asymptotic arguments and some possibly unexpected limiting distributions result.

1 Introduction

We consider likelihood-based inference in a context where there is an unknown threshold parameter and standard regularity conditions are well known to be violated. Our interest is in the effect of the lack of regularity on limiting distributional results.

The analysis in this paper was originally motivated by a problem of counting bacteria colonies in samples of tomato seed, where nonzero counts appeared to be approximately left-truncated lognormal *with unknown truncation parameter*. Figure 1 is a histogram for the $n = 198$ log-counts, 54 of which were "$-\infty$" corresponding to 0 counts or "non-detects." Figure 2 is a plot of 144 pairs

$$\left(ith \text{ smallest nonzero log-count}, \Phi^{-1}\left(\frac{i+21.5}{166}\right)\right)$$

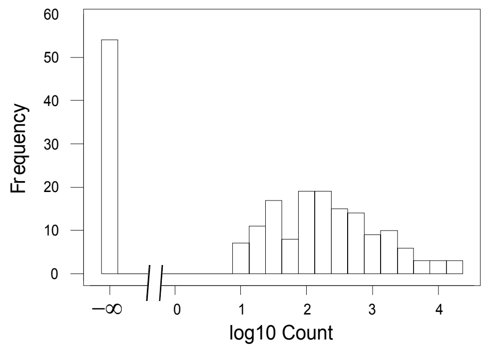

Figure 1: Histogram of 198 Bacteria log-Counts

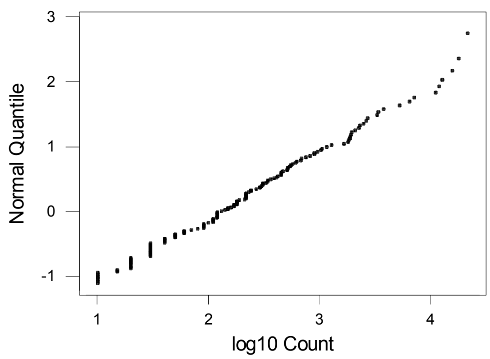

Figure 2: "Normal Plot" for 144 Finite log-Counts

which is a kind of "normal plot" adjusted for an estimate that a fraction of .1357 of an original normal distribution has been lost to left truncation.

For k real-valued functions $t_1(x), t_2(x), ..., t_k(x)$, a function $h(x) \geq 0$ and $\eta = (\eta_1, \eta_2, ..., \eta_k)' \in \mathcal{R}^k$ define

$$g_\eta(x) = h(x) \exp \sum_{j=1}^{k} \eta_j t_j(x)$$

We consider distributions absolutely continuous with respect to Lebesgue measure on $\mathcal{R}^1$derived from $g_\eta(x)$. To that end, let

$$\Omega = \left\{ \eta \in \mathcal{R}^k | \int_{-\infty}^{\infty} g_\eta(x) dx < \infty \right\}$$

and for $\eta \in \Omega$ and $d \in [-\infty, \infty)$, set

$$\mathcal{K}(\eta, d) = \left(\int_d^{\infty} g_\eta(x) dx \right)^{-1}$$

(Unless confusion is possible, we will write (η, d) in place of the more cumbersome $(\eta', d)'$.) Let F_η stand for the distribution function on $\mathcal{R}^1$ with density

$$f_\eta(x) = \mathcal{K}(\eta, -\infty) g_\eta(x)$$

and E_η and Var_η stand for expectation and variance of a function of $X \sim F_\eta$. For $d > -\infty$ let $\widetilde{F}_{\eta,d}$ stand for the distribution function on $\mathcal{R}^1$ with density

$$\begin{aligned} \widetilde{f}_{\eta,d}(x) &= \mathcal{K}(\eta, d) g_\eta(x) I\left[x \geq d\right] \\ &= \left(1 - F_\eta(d)\right)^{-1} f_\eta(x) I\left[x \geq d\right] \end{aligned}$$

and $\widetilde{\mathrm{E}}_{\eta,d}$ and $\widetilde{\mathrm{Var}}_{\eta,d}$ stand for expectation and variance of a function of $X \sim \widetilde{F}_{\eta,d}$. ($\widetilde{F}_{\eta,d}$ specifies the conditional distribution given $X \geq d$ of $X \sim F_\eta$.) On some occasions we will also want notation for "the other" conditional distribution of $X \sim F_\eta$. So for $d > -\infty$ let $\overline{\widetilde{F}}_{\eta,d}$ stand for the distribution function on $\mathcal{R}^1$ with density

$$\begin{aligned} \overline{\widetilde{f}}_{\eta,d}(x) &= \left(\mathcal{K}(\eta, -\infty)^{-1} - \mathcal{K}(\eta, d)^{-1}\right)^{-1} g_\eta(x) I\left[x < d\right] \\ &= F_\eta(d)^{-1} f_\eta(x) I\left[x < d\right] \end{aligned}$$

and $\overline{\widetilde{\mathrm{E}}}_{\eta,d}$ and $\overline{\widetilde{\mathrm{Var}}}_{\eta,d}$ stand for expectation and variance of a function of $X \sim \overline{\widetilde{F}}_{\eta,d}$. ($\overline{\widetilde{F}}_{\eta,d}$ specifies the conditional distribution given $X < d$ of $X \sim F_\eta$.)

This article treats three related one-sample inference problems. First we consider inference for $(\eta, d) \in \Omega \times \mathcal{R}^1$ based on

$$X_1, X_2, ..., X_n \sim \text{ iid } \widetilde{F}_{\eta,d}$$

This is inference with an unknown truncation point. Next, for

$$X_1, X_2, ..., X_n \sim \text{ iid } F_\eta$$

we treat inference for $(\eta, d) \in \Omega \times \mathcal{R}^1$ based not on the X_i, but rather on the n values

$$X_i^* = \begin{cases} -\infty & \text{if } X_i < d \\ X_i & \text{if } X_i \geq d \end{cases}$$

If we think of the value "$-\infty$" as a code for "nondetection" or "below some unknown minimum level of observation" this is inference under "type I" left censoring with an unknown censoring point. Finally, we consider inference for $(\eta, d) \in \Omega \times \mathcal{R}^1$and $p \in (0, 1)$ based on iid observations from a mixture of $\widetilde{F}_{\eta,d}$ and $\delta_{-\infty}$, a unit point mass at $-\infty$,

$$p\delta_{-\infty} + (1 - p)\, \widetilde{F}_{\eta,d}$$

(We are here letting the distribution function $\widetilde{F}_{\eta,d}$ stand for the probability distribution it specifies.) We pose this last problem because Figure 1 suggests that the count of non-detects in our motivating application may be too large to be explained as coming only from the left tail of an untruncated normal log-count distribution. The class of distributions in our second problem is a subclass of the third set of models defined by the parametric restriction $p = F_\eta(d)$.

None of these inference problems is regular in the classical sense of the likelihood satisfying smoothness conditions in the entire parameter vector. The lack of smoothness of $I[x \geq d]$ in d prevents application of standard results on likelihood-based inference. Our analysis shows that for some purposes, the fact that d is unknown may essentially be ignored in large samples, while for others it creates rather novel limiting distributional results.

There is a large literature on inference for truncated (and censored) families of distributions–represented for example by Cohen (1959) and Cohen (1991)– that almost exclusively treats the threshold as known. Smith (1985) considers likelihood-based inference for families that include translations of common life distributions. While our methods owe much to Smith's analysis, our results do not overlap his. Even the case where $g_\eta(x) = I[x \geq \zeta] \exp(-\eta x)$ and our truncation family is also a location-scale family (of two-parameter exponential distributions with threshold at least ζ) is not specifically treated by Smith, as he concentrates on families with "standard" (untranslated) densities that have limit 0 or $+\infty$ at 0.

While we will speak in terms of a lower threshold, upper threshold cases are covered by the obvious device of replacing X with $-X$. Further, though we won't provide the details, the results discussed here have analogs for cases where there are both upper *and* lower unknown thresholds.

2 Background Results

In this section we collect some background results and set the stage for our analysis of the problems introduced above.

2.1 Results for Exponential Families

It is well known that for fixed $d \in [-\infty, \infty)$, partial derivatives of the function $\ln \mathcal{K}(\eta, d)$ with respect to the entries of η are related to the moments of $\mathbf{t}(X) = (t_1(X), t_2(x), ..., t_k(X))'$ and to Fisher information matrices. For reference purposes, we record those relationships and also consider the differentiation of the mean of $\mathbf{t}(X)$ with respect to d.

Denote the mean vector and covariance matrix of $\mathbf{t}(X)$ under $\widetilde{F}_{\eta,d}$ by respectively

$$\mu(\eta, d) = \widetilde{\mathrm{E}}_{\eta,d}\mathbf{t}(X) \text{ and } \mathbf{\Sigma}(\eta, d) = \widetilde{\mathrm{Var}}_{\eta,d}\mathbf{t}(X)$$

Then with this notation, there is

Proposition 1 *(Moments of* $\mathbf{t}(X)$ *and differentiation of* $-\ln \mathcal{K}(\eta, d)$ *with respect to* η*)*

1. $\mu(\eta, d) = \left(\frac{\partial}{\partial \eta_1}\left(-\ln \mathcal{K}(\eta, d)\right), \frac{\partial}{\partial \eta_2}\left(-\ln \mathcal{K}(\eta, d)\right), ..., \frac{\partial}{\partial \eta_k}\left(-\ln \mathcal{K}(\eta, d)\right)\right)'$
2. $\boldsymbol{\Sigma}(\eta, d) = \left(\frac{\partial^2}{\partial \eta_l \partial \eta_m}\left(-\ln \mathcal{K}(\eta, d)\right)\right)_{l=1,...,k;\ m=1,...,k}$

It is immediate from the proposition that

$$\boldsymbol{\Sigma}(\eta, d) = \frac{\partial}{\partial \eta}\mu(\eta, d) \tag{1}$$

In addition, for fixed d the matrix $\boldsymbol{\Sigma}(\eta, d)$ is the Fisher information matrix for η based on a single observation $X \sim \widetilde{F}_{\eta,d}$. And since for fixed $d > -\infty$ the distributions $\overline{\overline{\widetilde{F}}}_{\eta,d}$ form a regular exponential family, exactly as in (1)

$$\frac{\partial}{\partial \eta}\overline{\overline{\widetilde{\mathrm{E}}}}_{\eta,d}\mathbf{t}(X) = \overline{\overline{\widetilde{\mathrm{Var}}}}_{\eta,d}\mathbf{t}(X) \tag{2}$$

It will be useful to differentiate $\mu(\eta, d)$ not only with respect to the entries of η, but also with respect to d. Writing out

$$\mu_j(\eta, d) = \int_d^\infty t_j(x)\mathcal{K}(\eta, d)g_\eta(x)dx$$

and applying the Leibnitz rule, one can establish

Proposition 2 *(Differentiation of the mean function with respect to d)*

$$\frac{\partial}{\partial d}\mu(\eta, d) = \widetilde{f}_{\eta,d}(d)\left[\mu(\eta, d) - \mathbf{t}(d)\right]$$

Finally, for the case of censoring, it is useful to note what is (for a fixed d) the Fisher information matrix for η based on a single observation $X^* = -\infty \cdot I\left[X < d\right] + X \cdot I\left[X \geq d\right]$ where $X \sim F_\eta$. This is, by definition,

$$\begin{aligned}\mathbf{I}^*\left(\eta, d\right) = {} & F_\eta(d)\left(-\frac{\partial^2}{\partial \eta_l \partial \eta_m}\left(\ln F_\eta(d)\right)\right)_{l=1,...,k;\ m=1,...,k} \\ & + \left(\int_d^\infty \left(-\frac{\partial^2}{\partial \eta_l \partial \eta_m}\ln f_\eta(x)\right) f_\eta(x)dx\right)_{l=1,...,k;\ m=1,...,k}\end{aligned}$$

and it has several equivalent representations, two of which are given next.

Proposition 3 *(Fisher information under censoring)*

$$\begin{aligned}\mathbf{I}^*\left(\eta, d\right) &= \boldsymbol{\Sigma}(\eta, -\infty) - F_\eta(d)\overline{\overline{\widetilde{Var}}}_{\eta,d}\mathbf{t}(X) \\ &= (1 - F_\eta(d))\boldsymbol{\Sigma}(\eta, d) \\ &\quad + F_\eta(d)(1 - F_\eta(d))\left(\mu(\eta, d) - \overline{\overline{\widetilde{E}}}_{\eta,d}\mathbf{t}(X)\right)\left(\mu(\eta, d) - \overline{\overline{\widetilde{E}}}_{\eta,d}\mathbf{t}(X)\right)'\end{aligned}$$

Hollander, Proschan and Sconing (1990) have considered Kullback-Leibler divergence measures under censoring, and the Fisher information could be obtained from second partials of their measures.

2.2 Limiting Results for Sample Minima and Means

In the problems we will consider, a natural likelihood-based estimator of d proves to be a sample minimum. On the other hand, the natural likelihood-based estimators of the elements of η (and of p) are derived from averages. Therefore, a technology appropriate for studying the asymptotics of likelihood-based inference in these families must deal simultaneously with extremes and means. Chow and Teugels (1978) (used earlier by Smith (1985)) provide exactly the right type of probability tools to support our study. They show that under appropriate conditions on $Y_1, Y_2, ...$ iid from a continuous distribution on $\mathcal{R}^1$ and on a function q, the sample mean of the first n values $q(Y_i)$ is asymptotically independent of an extreme order statistic for the first n values Y_i. This gives hope that the nonregularity in the inference problems we consider will be of a kind that is nevertheless amenable to asymptotic analysis.

3 Inference in the Truncated Exponential Family

Consider then inference for $(\eta, d) \in \Omega \times \mathcal{R}^1$ based on $X_1, X_2, ..., X_n$ that are iid $\widetilde{F}_{\eta,d}$. The loglikelihood function is

$$L(\eta, d) = \ln\left(\prod_{i=1}^{n} \widetilde{f}_{\eta,d}(X_i)\right)$$

When $\min\{X_1, X_2, ..., X_n\} < d$, $L(\eta, d) = -\infty$. With

$$\bar{t}_j = \frac{1}{n}\sum_{i=1}^{n} t_j(X_i)$$

when $\min\{X_1, X_2, ..., X_n\} \geq d$ the loglikelihood is

$$L(\eta, d) = n\sum_{j=1}^{k} \eta_j \bar{t}_j + \sum_{i=1}^{n} \ln h(X_i) + n \ln \mathcal{K}(\eta, d)$$

Since for every η the normalizing constant $\mathcal{K}(\eta, d)$ is nondecreasing in d,

$$\widehat{d} = \min\{X_1, X_2, ..., X_n\}$$

is an obvious estimator of d. (A referee has noted that bias correction could improve the small sample properties of this estimator. While that is true, it is tangential to our main story and we will simply consider $\widehat{d}$.) Plugging $\widehat{d}$ into the loglikelihood, we maximize $L(\eta, \widehat{d})$ in order to find an estimator for η. Recalling Proposition 1, upon differentiating $L(\eta, \widehat{d})$ with respect to each of its arguments and setting those partial derivatives equal to zero, the resulting set of estimating equations has the form

$$\bar{\mathbf{t}} = (\bar{t}_1, \bar{t}_2, ..., \bar{t}_k)' = \mu(\eta, \widehat{d}) \tag{3}$$

That is, one sets d at the minimum observed value and then seeks a parameter vector η that makes the theoretical and empirical means of the $t_j(X)$'s the same. Let $\widehat{\eta}$ denote a solution of equation (3). The balance of this section concerns the limiting behavior of $(\widehat{\eta}, \widehat{d})$, and some implications of that behavior.

3.1 The Limiting Distribution of $(\widehat{\eta}, \widehat{d})$

The fundamental ingredients of $(\widehat{\eta}, \widehat{d})$ are the minimum observation, $\widehat{d}$, and the vector of sample means $\overline{\mathbf{t}}$. It is possible to use the Cramér-Wold device and the Chow and Teugels arguments to establish the following.

Theorem 4 *Suppose that $X_1, X_2, ..., X_n$ are iid $\widetilde{F}_{\eta,d}$. If $g_\eta(x)$ is positive and right continuous at d, then $\left(\sqrt{n}\left(\overline{\mathbf{t}} - \mu(\eta, d)\right)', n(\widehat{d} - d)\right)'$ has a limiting distribution. This joint distribution is one of independence where the first marginal is k-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}(\eta, d)$ and the second is exponential with mean $1/\widetilde{f}_{\eta,d}(d)$.*

Proof. Full details of the proof can be found in Dubinin (2000). ■

So we know the joint limiting behavior of $\overline{\mathbf{t}}$ and $\widehat{d}$. The smoothness of $\mu(\eta, d)$ (that derives from properties of regular exponential families) can then be used to deduce the joint limiting behavior of $\widehat{\eta}$ and $\widehat{d}$. Key to a proof of the next theorem is the $\sqrt{n}$ equivalence of $\widehat{\eta}$ and an MLE of η based on full knowledge of d. We note that Smith (1985) proved an equivalence stronger than this in his context. (We note also that since we may appeal to "fixed d" estimation results for regular exponential families in the truncation problem, the proof of Theorem 5 doesn't use the particular form of the limiting $\overline{\mathbf{t}}$ marginal given in Theorem 4. However, that marginal result is important for our subsequent discussion of the censoring problem.)

Theorem 5 *Suppose that $X_1, X_2, ..., X_n$ are iid $\widetilde{F}_{\eta,d}$, $g_\eta(x)$ is positive and right continuous at d, and $\mathbf{\Sigma}(\eta, d)$ is nonsingular. If $\widehat{\eta}$ is consistent for η, then $\left(\sqrt{n}\left(\widehat{\eta} - \eta\right)', n(\widehat{d} - d)\right)'$ has a limiting joint distribution. This is one of independence where the first marginal is k-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}(\eta, d)^{-1}$ and the second is exponential with mean $1/\widetilde{f}_{\eta,d}(d)$.*

Proof. An outline of the proof is in the appendix. ■

As far as inference for η alone is concerned, the fact that $\widehat{d}$ converges to d at rate n^{-1} allows one to simply treat d as it if were known to be $\widehat{d}$ and suffer no additional adverse consequences. A consistent root of the "likelihood" equation (3) has the same asymptotic distribution as a maximum likelihood estimator of η based on full knowledge of d. And, for example, one can make Wald type confidence sets for η (or a sub-vector of η) using the approximate distribution and the consistent estimator of the covariance matrix for that distribution, $\mathbf{\Sigma}(\widehat{\eta}, \widehat{d})^{-1}$. However, as the next sub-section shows, Theorem 5 does have some interesting (and perhaps even unexpected) implications where the object of inference involves the whole parameter vector (η, d).

3.2 Likelihood Ratio Testing

Consider first testing the hypotheses

$$\mathrm{H_0} : (\eta, d) = (\eta_0, d_0) \quad \text{versus } \mathrm{H_a} : \text{not } \mathrm{H_0}$$

under the assumptions of Theorem 5 at (η_0, d_0). A likelihood ratio test statistic for these hypotheses is

$$\Lambda(\eta_0, d_0) = \frac{\sup_{(\eta,d)} \prod_{i=1}^{n} \widetilde{f}_{\eta,d}(X_i)}{\prod_{i=1}^{n} \widetilde{f}_{\eta_0,d_0}(X_i)}$$

Classical analyses under regularity conditions produce χ^2 limiting null distributions for twice the natural log of this kind of statistic, where the degrees of freedom correspond to the dimension of the parameter vector. But it is not *a priori* obvious whether *any* standard distribution will describe the limiting null behavior of $2 \ln \Lambda(\eta_0, d_0)$ in the present context.

Consider a second order Taylor expansion of

$$\ln \prod_{i=1}^{n} \widetilde{f}_{\eta_0,d_0}(X_i)$$

at the point $(\widehat{\eta}, \widehat{d})$. All second and higher order derivatives of $\ln \widetilde{f}_{\eta,d}(x)$ with respect to the entries of η are constant in x, as are first and higher order derivatives with respect to d. Further, for any j, the sum $\sum_{i=1}^{n} \frac{\partial}{\partial \eta_j} \ln \widetilde{f}_{\eta,d}(X_i)\Big|_{\widehat{\eta},\widehat{d}}$ is 0 by design. So we may write

$$\begin{aligned}\sum_{i=1}^{n} \ln \widetilde{f}_{\eta_0,d_0}(X_i) &= \sum_{i=1}^{n} \ln \widetilde{f}_{\widehat{\eta},\widehat{d}}(X_i) + n(d_0 - \widehat{d}) \left.\frac{\partial}{\partial d} \ln \widetilde{f}_{\eta,d}\right|_{\widehat{\eta},\widehat{d}} \\ &+ \left(\frac{n}{2}\right) \sum_{j=1}^{k} (\eta_{0j} - \widehat{\eta}_j)^2 \left.\frac{\partial^2}{\partial \eta_j^2} \ln \widetilde{f}_{\eta,d}\right|_{\widehat{\eta},\widehat{d}} \\ &+ n \sum_{l \neq m} (\eta_{0l} - \widehat{\eta}_l)(\eta_{0m} - \widehat{\eta}_m) \left.\frac{\partial^2}{\partial \eta_l \partial \eta_m} \ln \widetilde{f}_{\eta,d}\right|_{\widehat{\eta},\widehat{d}} \qquad (4) \\ &+ n \sum_{j=1}^{k} (\eta_{0j} - \widehat{\eta}_j)(d_0 - \widehat{d}) \left.\frac{\partial^2}{\partial \eta_j \partial d} \ln \widetilde{f}_{\eta,d}\right|_{\widehat{\eta},\widehat{d}} \\ &+ \left(\frac{n}{2}\right) (d_0 - \widehat{d})^2 \left.\frac{\partial^2}{\partial d^2} \ln \widetilde{f}_{\eta,d}\right|_{\widehat{\eta},\widehat{d}} + R_n(\widehat{\eta}, \widehat{d})\end{aligned}$$

where the remainder term $R_n(\widehat{\eta}, \widehat{d})$ involves sums of products of third order terms in $(d_0 - \widehat{d})$ and the $(\eta_{0j} - \widehat{\eta}_j)$ and third order partials evaluated at $(\eta^{\#}, d^{\#}) = \alpha(\eta_0, d_0) + (1 - \alpha)(\widehat{\eta}, \widehat{d})$ for some $\alpha \in (0, 1)$. (With (η_0, d_0) probability 1, $d_0 < \widehat{d}$ and this expansion makes sense. The loglikelihood is smooth for $d < \widehat{d}$ and its discontinuity along the $d = \widehat{d}$ hyperplane is irrelevant to the validity of this expansion as long as partials with respect to d are interpreted as left partials.)

The negligibility of the last three summands in expansion (4) (including $R_n(\widehat{\eta}, \widehat{d})$) follows immediately from the consistency of $(\widehat{\eta}, \widehat{d})$ and $\left(\eta^{\#}, d^{\#}\right)$ for (η_0, d_0) under the null hypothesis, the fact that under the null hypothesis $n(d_0 - \widehat{d})$ and each $n(\eta_{0j} - \widehat{\eta}_j)^2$ converge in distribution and the continuity of the derivatives in the parameters. So we may analyze the limiting null behavior of

$$\lambda(\eta_0, d_0) = 2\left(L(\widehat{\eta}, \widehat{d}) - L(\eta_0, d_0)\right) = -2\left(\sum_{i=1}^{n} \ln \widetilde{f}_{\eta_0, d_0}(X_i) - \sum_{i=1}^{n} \ln \widetilde{f}_{\widehat{\eta}, \widehat{d}}(X_i)\right)$$

(which will often be the limiting null behavior of $2 \ln \Lambda(\eta_0, d_0)$) by considering -2 times the 2nd, 3rd and 4th terms on the right of expansion (4).

First,

$$-2n(d_0 - \widehat{d})\, \frac{\partial}{\partial d} \ln \widetilde{f}_{\eta, d}\Big|_{\widehat{\eta}, \widehat{d}} = 2\left(\frac{\partial}{\partial d} \ln \widetilde{f}_{\eta, d}\Big|_{\widehat{\eta}, \widehat{d}}\right) n(\widehat{d} - d_0)$$

The consistency of $(\widehat{\eta}, \widehat{d})$ for (η_0, d_0) under the null hypothesis, the continuity of the partial derivative in the parameter vector, and the limiting distribution for $n(\widehat{d} - d_0)$ promised by Theorem 5 show that under the null hypothesis this is asymptotically $2\left(\frac{\partial}{\partial d} \ln \widetilde{f}_{\eta, d}\Big|_{\eta_0, d_0}\right)$ times an exponential variable with mean $1/\widetilde{f}_{\eta_0, d_0}(d_0)$. But

$$\frac{\partial}{\partial d} \ln \widetilde{f}_{\eta, d} = \frac{\partial}{\partial d} \ln \mathcal{K}(\eta, d) = -\frac{\partial}{\partial d} \ln \int_d^{\infty} g_{\eta}(x) dx = \widetilde{f}_{\eta, d}(d)$$

so that the null limit of $-2n(d_0 - \widehat{d})\, \frac{\partial}{\partial d} \ln \widetilde{f}_{\eta, d}\Big|_{\widehat{\eta}, \widehat{d}}$ is exponential with mean 2, that is χ^2_2.

Then

$$-2\left(\left(\frac{n}{2}\right) \sum_{j=1}^{k} (\eta_{0j} - \widehat{\eta}_j)^2\, \frac{\partial^2}{\partial \eta_j^2} \ln \widetilde{f}_{\eta, d}\Big|_{\widehat{\eta}, \widehat{d}} + n \sum_{l \neq m} (\eta_{0l} - \widehat{\eta}_l)(\eta_{0m} - \widehat{\eta}_m)\, \frac{\partial^2}{\partial \eta_l \partial \eta_m} \ln \widetilde{f}_{\eta, d}\Big|_{\widehat{\eta}, \widehat{d}}\right)$$

is the quadratic form

$$\sqrt{n}\,(\eta_0 - \widehat{\eta})' \left(\frac{\partial^2}{\partial \eta_l \partial \eta_m} \left(-\ln \mathcal{K}(\eta, d)\right)\Big|_{\widehat{\eta}, \widehat{d}}\right)_{l=1,\ldots,k;\; m=1,\ldots,k} \left(\sqrt{n}\,(\eta_0 - \widehat{\eta})\right)$$

Theorem 5 guarantees that under the null hypothesis $\sqrt{n}\,(\eta_0 - \widehat{\eta})$ is asymptotically k-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}(\eta_0, d_0)^{-1}$ while the null consistency of $(\widehat{\eta}, \widehat{d})$ for (η_0, d_0), the continuity of derivatives of $-\ln \mathcal{K}(\eta, d)$

and Proposition 1 imply that the $k \times k$ matrix above converges in (η_0, d_0) probability to $\mathbf{\Sigma}(\eta_0, d_0)$. The quadratic form is then asymptotically χ^2_k under the null hypothesis, and by Theorem 5 asymptotically independent of the linear term in d_0.

So, under the null hypotheses $\lambda(\eta_0, d_0)$ has the limiting distribution of a sum of independent χ^2 random variables and

$$\lambda(\eta_0, d_0) \xrightarrow{\mathcal{L}} \chi^2_{k+2}$$

This is a novel and possibly surprising result. The "classical" χ^2 nature of the limiting null distribution is preserved in this partially nonregular context, *but the appropriate degrees of freedom are neither the dimension of the parameter vector ($k+1$), nor the dimension of* η (which is what one might possibly expect taking too seriously the notion that the estimation of d is of little asymptotic consequence in this problem). The fact that the linear term in d_0 in the Taylor expansion for $\ln \prod_{i=1}^n \widetilde{f}_{\eta_0, d_0}(X_i)$ is not negligible (and asymptotically χ^2_2) produces the novel limiting distribution.

We remarked in the Introduction that the techniques of this paper can be extended to families with both a lower and an upper threshold. In those families, the limiting null distribution associated with a likelihood ratio statistic for a point null hypothesis for the full parameter vector will be χ^2_{k+4} (following from a distributional result that shows the sample minimum, the sample maximum and an appropriately defined $\widehat{\eta}$ to be asymptotically independent).

Further, this kind of analysis extends to likelihood ratio tests of other possibly more interesting hypotheses. For k' smooth functions $\theta_1, \theta_2, ..., \theta_{k'}$ each mapping $\Omega \times \mathcal{R}^1$ to $\mathcal{R}^1$, consider a null hypothesis of the form $\mathrm{H}_0 : \theta_i(\eta, d) = 0$ for $i = 1, 2, ..., k'$. For example, a point null hypothesis concerning the mean of $X \sim \widetilde{F}_{\eta,d}$ is of this form with $k' = 1$. Or in a truncated normal family, a point null hypothesis about the pair (μ, σ)–the mean and standard deviation of the untruncated distribution–is of this form with $k' = 2$. A likelihood ratio test statistic for this null hypothesis is

$$\Lambda_\theta = \frac{\sup_{(\eta,d)} \prod_{i=1}^n \widetilde{f}_{\eta,d}(X_i)}{\sup_{(\eta,d) \text{ with } \theta(\eta,d)=\mathbf{0}} \prod_{i=1}^n \widetilde{f}_{\eta,d}(X_i)}$$

and analysis of the null behavior of Λ_θ employs an expansion of the form (4) where (η_0, d_0) is replaced by $(\widehat{\eta}_\theta, \widehat{d}_\theta)$, the vector of maximum likelihood estimates *under the* k' *constraints* $\theta_i(\eta, d) = 0$.

At a parameter vector (η_0, d_0) satisfying the constraints, these are at least locally linear and the nature of the asymptotics depends on the character of the set of gradient vectors $\nabla\theta_i$ at (η_0, d_0). Consider the situation where the k' gradient vectors $\nabla\theta_i(\eta_0, d_0)$ are linearly independent. There are two subcases, according to whether or not there is some i for which the first k entries of $\nabla\theta_i(\eta_0, d_0)$ are all 0, that is whether or not (at least locally) the constraints completely specify $d = d_0$. Where d is (locally) completely specified, the analysis proceeds almost exactly as in the earlier point null hypothesis case. The linear term in $\widehat{d}_\theta$ in the expansion of $\ln \prod_{i=1}^n \widetilde{f}_{\widehat{\eta}_\theta,\widehat{d}_\theta}(X_i)$ makes a χ_2^2 contribution to the final limit. The quadratic form in $(\widehat{\eta}_\theta - \widehat{\eta})$ is asymptotically a quadratic form in the difference between a k-variate normal vector and its projection on a $(k'-1)$-dimensional space containing its mean vector, and is asymptotically $\chi^2_{k'-1}$. So a $\chi^2_{k'+1}$ null limit results. Where d is not (locally) completely specified, one can argue based on the null consistency of both $(\widehat{\eta},\widehat{d})$ and $(\widehat{\eta}_\theta,\widehat{d}_\theta)$ and the fact that while first partials of $n^{-1}\ln \prod_{i=1}^n \widetilde{f}_{\eta,d}(X_i)$ with respect to the η_j's at $(\widehat{\eta},\widehat{d})$ are 0 the first partial with respect to d tends to something positive, that with probability tending to 1, $\widehat{d}_\theta = \widehat{d}$ exactly. This means that asymptotically only the quadratic form in $(\widehat{\eta}_\theta - \widehat{\eta})$ is important, and $(\widehat{\eta}_\theta - \widehat{\eta})$ is the difference between a k-variate normal vector and its projection on a k'-dimensional space containing its mean vector. So a $\chi^2_{k'}$ limit results.

One can go on to consider likelihood ratio testing of (sets of) "one-sided" hypotheses of the form $\mathrm{H}_0 : \theta_i(\eta,d) \le 0$ for $i = 1,2,...,k'$ after the fashion of Chernoff (1954) or Ferguson (1996, page 150). Here cutoff points are based on the limiting distribution(s) of the test statistic for parameter vectors on the boundary of the null region. Unless $k' = 1$ several fundamentally different analysis must be considered, depending upon how many (and which) of the of $\theta_i(\eta,d)$ are strictly negative, i.e. depending upon the local character of the boundary at the parameter vector under consideration. (Asymptotics at "corners" are different than at points on "edges," which are different from those at points "in the middle of a face," etc.)

As the simplest example of what happens with one-sided likelihood ratio testing, consider $k' = 1$ cases. For (η_0, d_0) with $\theta(\eta_0,d_0) = 0$, suppose first that $\theta(\eta,d) = 0$ (at least locally) specifies $d = d_0$. Where $\theta(\eta,d) \le 0$ locally amounts to $d \le d_0$, the (η_0,d_0) limiting distribution of twice the log of the likelihood ratio statistic will be χ_2^2. Where $\theta(\eta,d) \le 0$ locally amounts to $d \ge d_0$ twice the log of the likelihood ratio statistic is 0 when $\widehat{d} \ge d_0$ and so the (η_0,d_0) limiting distribution is a point mass at 0. Where $\theta(\eta,d) = 0$ does not locally specify $d = d_0$ the (η_0,d_0) limiting distribution of twice the log of the likelihood ratio statistic will be an equal-parts mixture of a unit point mass at 0 and a χ_1^2 distribution, as argued in Chernoff (1954). For typical one-sided testing problems like ones concerning the mean of $X \sim \widetilde{F}_{\eta,d}$, this last limit will hold for all parameter points on the boundary of the null region.

3.3 Confidence Set Estimation

Theorem 5 immediately enables Wald type confidence set estimation for η.

$$\left\{\eta |\; n\,(\widehat{\eta}-\eta)'\,\mathbf{\Sigma}(\widehat{\eta},\widehat{d})\,(\widehat{\eta}-\eta) < \chi^2_{k,\gamma}\right\}$$

functions as a large sample γ-level confidence set for η (and the obvious modifications can be made to produce confidence sets for sub-vectors). Further, Theorem 5 implies that $n(\widehat{d}-d)\widetilde{f}_{\widehat{\eta},\widehat{d}}(\widehat{d}) \xrightarrow{\mathcal{L}} \text{Exp}(1)$ so that for $s>0$

$$\left(\widehat{d}-\frac{s}{n\widetilde{f}_{\widehat{\eta},\widehat{d}}(\widehat{d})},\widehat{d}\right)$$

is a large sample $(1-\exp(-s))$-level confidence interval for d.

But it also possible to invert the likelihood ratio tests of the previous subsection to do set estimation for one or more functions of (η,d). For example, the set

$$\left\{(\eta,d)|\ \lambda(\eta,d)<\chi^2_{k+2,\gamma}\right\}$$

is an approximate γ-level confidence set for the entire parameter vector (η,d). And with $\theta_i(\eta,d)=\rho_i(\eta,d)-c_i$ for constants c_i, inversion of likelihood ratio tests for $\text{H}_0:\theta_i(\eta,d)=0$ for $i=1,2,...,k'$ as $\mathbf{c}$ ranges over $\mathcal{R}^{k'}$ produces a joint confidence set for values of the k' parametric functions $\rho_i(\eta,d)$.

4 Inference in the Censored Exponential Family

Consider next inference for $(\eta,d)\in\Omega\times\mathcal{R}^1$ based on $X_1^*,X_2^*,...,X_n^*$ that are iid from a distribution on $[-\infty,\infty)$ of the form

$$F_\eta(d)\delta_{-\infty}+(1-F_\eta(d))\widetilde{F}_{\eta,d}$$

Let $n_{-\infty}$ denote the count of observations X_i^* that are "$-\infty$" or "nondetects" and abbreviate $n-n_{-\infty}$ as $n_{\mathcal{R}}$. Unless $d\le\min\{X_i^*|\ X_i^*>-\infty\}$the loglikelihood is $-\infty$. For d no more than the minimum "detect," and

$$\overline{t}_j^*=\frac{1}{n_{\mathcal{R}}}\sum_{i\text{ s.t. }X_i^*>-\infty}t_j(X_i^*)$$

the loglikelihood function for this second problem is

$$L^*(\eta,d)=n_{-\infty}\ln F_\eta(d)+n_{\mathcal{R}}\sum_{j=1}^{k}\eta_j\overline{t}_j^*+\sum_{i\text{ s.t. }X_i^*>-\infty}\ln h(X_i^*)+n_{\mathcal{R}}\ln\mathcal{K}(\eta,-\infty)$$

and we consider inference based on this.

For any η, $L^*(\eta,d)$ is nondecreasing in $d\le\min\{X_i^*|\ X_i^*>-\infty\}$. So an obvious likelihood-based estimator of d is

$$\widehat{d^*}=\min\{X_i^*|\ X_i^*>-\infty\}$$

and we maximize $L^*(\eta,\widehat{d^*})$ in order to find an estimator for η.

Now

$$\frac{\partial}{\partial \eta_j} L^*(\eta, d) = n_{-\infty} \cdot \frac{\frac{\partial}{\partial \eta_j} F_\eta(d)}{F_\eta(d)} + n_{\mathcal{R}} \bar{t}_j^* - n_{\mathcal{R}} \mu_j(\eta, -\infty)$$

An implication of part 1 of Proposition 1 is that for $d \geq -\infty$

$$\frac{\partial}{\partial \eta_j} \mathcal{K}(\eta, d) = -\mathcal{K}(\eta, d) \widetilde{\mathrm{E}}_{\eta,d} t_j(X) = -\mathcal{K}(\eta, d) \mu_j(\eta, d)$$

and upon using this and the fact that

$$F_\eta(d) = 1 - \frac{\mathcal{K}(\eta, -\infty)}{\mathcal{K}(\eta, d)}$$

it is easy to establish that

$$\frac{\partial}{\partial \eta_j} F_\eta(d) = (1 - F_\eta(d)) (\mu_j(\eta, -\infty) - \mu_j(\eta, d))$$

Combining this with a small amount of algebra

$$\frac{\partial}{\partial \eta_j} L^*(\eta, d) = n_{-\infty} \overline{\widetilde{\mathrm{E}}}_{\eta,d} t_j(X) + n_{\mathcal{R}} \bar{t}_j^* - n \mu_j(\eta, -\infty) \tag{5}$$

Finally setting each $\frac{\partial}{\partial \eta_j} L^*(\eta, \widehat{d^*}) = 0$, the estimating equations for η are

$$\frac{n_{-\infty}}{n} \overline{\widetilde{\mathrm{E}}}_{\eta,\widehat{d^*}} \mathbf{t}(X) + \frac{n_{\mathcal{R}}}{n} \overline{\mathbf{t}}^* = \mu(\eta, -\infty) \tag{6}$$

where in the obvious way $\overline{\mathbf{t}}^* = \left(\bar{t}_1^*, \bar{t}_2^*, ..., \bar{t}_k^*\right)'$. One sets d at the minimum observed "detect" value and then seeks η that makes a natural (censored-data) empirical approximation of $\mathrm{E}_\eta \mathbf{t}(X)$ equal to that (without-censoring) mean vector. Let $\widehat{\eta}^*$ denote a solution of equation (6). The balance of this section concerns the limiting behavior of $(\widehat{\eta}^*, \widehat{d^*})$ and some of its implications.

4.1 The Limiting Distribution of $(\widehat{\eta}^*, \widehat{d^*})$

The ingredients of $(\widehat{\eta}^*, \widehat{d^*})$ are the minimum "detect" $\widehat{d^*}$, the vector of "detect" sample means $\overline{\mathbf{t}}^*$ and the count of "nondetects" $n_{-\infty}$. We may use what is known from Section 3 about the behavior of $\widehat{d}$ and $\overline{\mathbf{t}}$ as the basis for an analysis of the behavior of $\widehat{d^*}$, $\overline{\mathbf{t}}^*$ and $n_{-\infty}$ if we first establish an appropriate lemma.

Lemma 6 *Suppose random vectors $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, ...$ converge in distribution to $\mathbf{Y}$ and that nonnegative integer-valued random variables $B_1, B_2, B_3, ...$ converge to ∞ with probability 1. Suppose further that $Z_1, Z_2, Z_3, ...$ are one-to-one transformations of the B_n converging in distribution to Z. If $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, ...$ are such that conditional on $B_n = b$, $\mathbf{W}_n$ has the same distribution as $\mathbf{Y}_b$,*

$$(Z_n, \mathbf{W}_n) \xrightarrow{\mathcal{L}} (Z, \mathbf{Y})$$

where Z and $\mathbf{Y}$ are independent.

Proof. Let $\psi_n(\mathbf{s})$ be the characteristic function of $\mathbf{Y}_n$ and $\psi(\mathbf{s})$ be the characteristic function of $\mathbf{Y}$, and consider the characteristic function of $(Z_n, \mathbf{W}_n)$, $\Psi_n(u, \mathbf{s})$. For any fixed $(u, \mathbf{s})$

$$\begin{aligned}\Psi_n(u, \mathbf{s}) &= \mathrm{EE}\left[\exp\left(iuZ_n + i\mathbf{s}'\mathbf{W}_n\right)|B_n\right] \\ &= \mathrm{E}\exp\left(iuZ_n\right)\psi_{B_n}(\mathbf{s}) \\ &= \mathrm{E}\exp\left(iuZ_n\right)\psi(\mathbf{s}) + \mathrm{E}\exp\left(iuZ_n\right)\left[\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})\right]\end{aligned}$$

Now,

$$\begin{aligned}\left|\mathrm{E}\exp\left(iuZ_n\right)\left[\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})\right]\right| &\leq \mathrm{E}\left|\exp\left(iuZ_n\right)\left[\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})\right]\right| \\ &\leq \mathrm{E}\left|\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})\right|\end{aligned}$$

But the almost sure convergence of B_n to ∞ implies that $|\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})| \to 0$ almost surely. Then since $|\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})| \leq 2$ the dominated convergence theorem implies that $\mathrm{E}|\psi_{B_n}(\mathbf{s}) - \psi(\mathbf{s})| \to 0$. Finally, since by hypothesis $\mathrm{E}\exp\left(iuZ_n\right) \to \mathrm{E}\exp\left(iuZ\right)$ the result follows. ■

Theorem 7 *Suppose $X_1^*, X_2^*, ..., X_n^*$ are iid $F_\eta(d)\delta_{-\infty} + (1 - F_\eta(d))\widetilde{F}_{\eta,d}$. If $g_\eta(x)$ is positive and right continuous at d and $F_\eta(d) \in (0, 1)$,*

$$\left(\sqrt{n_{\mathcal{R}}}\left(\overline{\mathbf{t}}^* - \mu(\eta, d)\right)', \frac{n_{-\infty} - nF_\eta(d)}{\sqrt{nF_\eta(d)\left(1 - F_\eta(d)\right)}}, n_{\mathcal{R}}\left(\widehat{d}^* - d\right)\right)'$$

has a limiting distribution. This is one of independence where the first marginal is k-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}(\eta, d)$, the second is standard normal, and the third is exponential with mean $1/\widetilde{f}_{\eta,d}(d)$.

Proof. Conditional on $X_i^* > -\infty$, X_i^* is from the distribution $\widetilde{F}_{\eta,d}$. Therefore, conditioned on $n_{\mathcal{R}}$

$$\left(\sqrt{n_{\mathcal{R}}}\left(\overline{\mathbf{t}}^* - \mu(\eta, d)\right)', n_{\mathcal{R}}\left(\widehat{d}^* - d\right)\right)'$$

has the same distribution as $\left(\sqrt{n}\left(\overline{\mathbf{t}} - \mu(\eta, d)\right)', n(\widehat{d} - d)\right)'$ considered in Theorem 4 *for a sample of size $n = n_{\mathcal{R}}$*. By Theorem 4 these distributions converge to the product of a k-variate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}(\eta, d)$ and an exponential with mean $1/\widetilde{f}_{\eta,d}(d)$.

Now $n_{\mathcal{R}} \to \infty$ almost surely, and since $n_{-\infty} = n - n_{\mathcal{R}}$,

$$Z_n = \frac{n_{-\infty} - nF_\eta(d)}{\sqrt{nF_\eta(d)\left(1 - F_\eta(d)\right)}}$$

is a one-to-one transformation of $n_{\mathcal{R}}$. Z_n converges in distribution to standard normal, so applying Lemma 6 with $B_n = n_{\mathcal{R}}$, the theorem follows. ■

The analysis of the asymptotic behavior of $(\widehat{\eta}^*, \widehat{d}^*)$ follows from smoothness properties and Theorem 7, much as the asymptotic behavior of $(\widehat{\eta}, \widehat{d})$ in the truncation problem follows from smoothness properties and Theorem 4. The details of proof are more tedious, but in the end one has the following.

Theorem 8 *Suppose* $X_1^*, X_2^*, ..., X_n^*$ *are iid* $F_\eta(d)\delta_{-\infty} + (1 - F_\eta(d))\widetilde{F}_{\eta,d}$, $g_\eta(x)$ *is positive and right continuous at* d, $F_\eta(d) \in (0,1)$, $\mathbf{I}^*(\eta, d)$ *is nonsingular and* $\widehat{\eta}^*$ *is consistent for* η*. Then* $\sqrt{n}\,(\widehat{\eta}^* - \eta)$ *and* $n(\widehat{d}^* - d)$ *have limiting distributions. These are respectively k-variate normal with mean* $\mathbf{0}$ *and covariance matrix* $\mathbf{I}^*(\eta, d)^{-1}$ *and exponential with mean* $1/f_\eta(d)$*.*

If in addition, there exists $\widetilde{\eta}^*$ *that solves*

$$\frac{n_{-\infty}}{n}\overline{\overline{E}}_{\eta,d}\mathbf{t}(X) + \frac{n_{\mathcal{R}}}{n}\overline{\mathbf{t}}^* = \mu(\eta, -\infty)$$

and is consistent for η*, then* $\left(\sqrt{n}\,(\widehat{\eta}^* - \eta)', n(\widehat{d}^* - d)\right)'$ *has a limiting joint distribution of independence between* $\sqrt{n}\,(\widehat{\eta}^* - \eta)$ *and* $n(\widehat{d}^* - d)$*.*

Proof. An outline of the proof is in the appendix. ■

Theorem 8 says that the asymptotics of maximum likelihood estimation under censoring with unknown threshold are analogous to those under truncation. As far as inference for η alone is concerned, the fact that $\widehat{d}^*$ converges to d at rate n^{-1} allows one to treat d as it if were known to be $\widehat{d}^*$ and suffer no additional adverse consequences. A consistent root of the "likelihood" equation (6) has the same asymptotic distribution as a maximum likelihood estimator of η based on full knowledge of d. And Theorem 8 has implications for testing and interval estimation analogous to what we found in Sections 3.2 and 3.3 under truncation.

4.2 Likelihood Ratio Testing

As in the truncation case, consider first testing

$$\mathrm{H}_0 : (\eta, d) = (\eta_0, d_0) \;\text{ versus } \mathrm{H}_\mathrm{a} : \text{not } \mathrm{H}_0$$

under the assumptions of Theorem 8 at (η_0, d_0). Our interest is in the asymptotic null behavior of

$$\lambda^*(\eta_0, d_0) = 2\left(L^*(\widehat{\eta}^*, \widehat{d}^*) - L^*(\eta_0, d_0)\right)$$

and we proceed by considering the nature of a Taylor expansion for $L^*(\eta_0, d_0)$ at the point $(\widehat{\eta}^*, \widehat{d}^*)$.

The first order terms in the elements of η_0 are 0 by design. The first order term in d_0 is

$$n_{-\infty}\left(d_0 - \widehat{d}^*\right)\left(\frac{f_{\widehat{\eta}^*}(\widehat{d}^*)}{F_{\widehat{\eta}^*}(\widehat{d}^*)}\right) = -\left(\frac{n_{-\infty}}{n_{\mathcal{R}}}\right)\left(\frac{1 - F_{\widehat{\eta}^*}(\widehat{d}^*)}{F_{\widehat{\eta}^*}(\widehat{d}^*)}\right)\widetilde{f}_{\widehat{\eta}^*,\widehat{d}^*}(\widehat{d}^*)\cdot n_{\mathcal{R}}\left(\widehat{d}^* - d_0\right)$$

which in light of Theorem 7 is asymptotically -1 times an Exp(1) random variable under H_0. So, just as in the truncation case, the contribution of the linear term in d_0 to the null limiting behavior of $\lambda^*(\eta_0, d_0)$ is χ_2^2.

Expressions (1), (2) and (5) provide a simple way of representing the matrix of second partials of $L^*(\eta, d)$ with respect to the entries of η. That is,

$$\frac{\partial^2}{\partial \eta^2} L^*(\eta, d) = -\left(n\mathbf{\Sigma}(\eta, -\infty) - n_{-\infty} \widetilde{\overline{\text{Var}}}_{\eta,d} \mathbf{t}(X) \right)$$

and it is then obvious in light of Proposition 3 that under the null hypothesis

$$\frac{1}{n} \frac{\partial^2}{\partial \eta^2} L^*(\eta, d) \bigg|_{\widehat{\eta}^*, \widehat{d}^*} \to -\mathbf{I}^* \left(\eta_0, d_0 \right)$$

So the contribution to $\lambda^*(\eta_0, d_0)$ of the second order terms in the elements of η_0 in the Taylor expansion for $L^*(\eta_0, d_0)$ is

$$-2 \left(\frac{1}{2} \left(\sqrt{n} \left(\eta_0 - \widehat{\eta}^* \right) \right)' \left(\frac{1}{n} \frac{\partial^2}{\partial \eta^2} L^*(\eta, d) \bigg|_{\widehat{\eta}^*, \widehat{d}^*} \right) \sqrt{n} \left(\eta_0 - \widehat{\eta}^* \right) \right)$$

which in view of Theorem 8 is asymptotically χ^2_k and independent of the linear term in d_0 under the null hypothesis.

These together provide the anticipated χ^2_{k+2} null limit for $\lambda^*(\eta_0, d_0)$, provided the "mixed" second order terms and the third order remainder terms are negligible. And this follows easily from the orders of convergence of $(\widehat{\eta}^*, \widehat{d}^*)$ to (η_0, d_0) guaranteed by Theorem 8 and the continuity of partials of $\ln F_\eta(d)$ and $\ln \mathcal{K}(\eta, -\infty)$ in the parameter vector.

It is clear that the asymptotics of likelihood ratio testing for $\text{H}_0 : \theta_i(\eta, d) = 0$ for $i = 1, 2, ..., k'$ (and for $\text{H}_0 : \theta_i(\eta, d) \leq 0$ for $i = 1, 2, ..., k'$) are exactly as for the truncation problem. And finally, just as for the truncation problem, the analysis can easily be extended to a doubly censored case where there is both an unknown lower censoring point and an unknown upper censoring point. Under a null hypothesis of equality constraints, one "counts each fully constrained threshold parameter twice" in identifying appropriate χ^2 null limits.

4.3 Confidence Set Estimation

The implications for confidence set estimation under censoring of the limiting distributional results in Theorem 8 and in Section 4.2 are exactly analogous to those of Theorem 5 and Section 3.2 under truncation. Both

$$\left\{ \eta | \ -n \left(\widehat{\eta}^* - \eta \right)' \mathbf{I}^* (\widehat{\eta}^*, \widehat{d}^*) \left(\widehat{\eta}^* - \eta \right) < \chi^2_{k,\gamma} \right\}$$

and

$$\left\{ \eta | \ \left(\widehat{\eta}^* - \eta \right)' \frac{\partial^2}{\partial \eta^2} L^*(\eta, d) \bigg|_{\widehat{\eta}^*, \widehat{d}^*} \left(\widehat{\eta}^* - \eta \right) < \chi^2_{k,\gamma} \right\}$$

function as Wald type large sample γ-level confidence sets for η (and the obvious modifications can be made to produce confidence sets for sub-vectors). For

$s > 0$,

$$\left(\widehat{d}^* - \frac{s}{n_{-\infty}\left(\frac{f_{\widehat{\eta}^*}(\widehat{d}^*)}{F_{\widehat{\eta}^*}(\widehat{d}^*)}\right)}, \widehat{d}^*\right) \quad \text{and} \quad \left(\widehat{d}^* - \frac{s}{n f_{\widehat{\eta}^*}(\widehat{d}^*)}, \widehat{d}^*\right)$$

are both large sample $(1 - \exp(-s))$-level confidence intervals for d. And, as in the truncation, problem it also possible to invert the likelihood ratio tests of the previous section to do set estimation. The set

$$\left\{(\eta, d) | \ \lambda^*(\eta, d) < \chi^2_{k+2,\gamma}\right\} \tag{7}$$

is a large sample γ-level confidence set for the entire parameter vector (η, d) and inversion of likelihood ratio tests for (sets of) parametric functions produces useful confidence procedures.

5 Inference for Mixtures of a Point Mass at $-\infty$ and a Member of the Truncated Exponential Family

Consider now the last of the three problems posed in the Introduction, namely inference based on iid observations from

$$p\delta_{-\infty} + (1 - p)\,\widetilde{F}_{\eta,d}$$

We continue use of the notation of the previous sections and note again that this model 1) generalizes the censoring model of Section 4 by dropping the requirement that $p = F_\eta(d)$ and 2) is motivated by our application to bacteria counts. In this section we observe that what has gone before makes obvious the nature of limiting results for inference in this model and apply some of these to the motivating data set.

5.1 Generalities

For d no more than the minimum "detect," the loglikelihood in this problem is

$$L^{**}(\eta, d, p) = n_{-\infty} \ln p + n_{\mathcal{R}} \ln(1-p) + n_{\mathcal{R}} \sum_{j=1}^{k} \eta_j \overline{t}^*_j + \sum_{i \text{ s.t. } X^*_i > -\infty} \ln h(X^*_i) + n_{\mathcal{R}} \ln \mathcal{K}(\eta, d)$$

and it is obvious that inference "separates" cleanly into the smaller problems of inference for p and for (η, d). Natural likelihood-based estimators are $\widehat{d}^*$, $\widehat{p} = n_{-\infty}/n$ and $\widehat{\eta}^{**}$ a solution to the set of equations

$$\overline{\mathbf{t}}^* = \mu(\eta, \widehat{d}^*) \tag{8}$$

The argument of Theorem 7 easily establishes that provided $g_\eta(x)$ is positive and right continuous at d and $p \in (0,1)$

$$\left(\sqrt{n_{\mathcal{R}}}\left(\overline{\mathbf{t}}^{*}-\mu(\eta,d)\right)', \frac{n\widehat{p}-np}{\sqrt{np\,(1-p))}}, n_{\mathcal{R}}\left(\widehat{d}^{*}-d\right)\right)'$$

has the same limiting distribution promised in that result. Then essentially the same argument made for Theorem 5 will show that if $\widehat{\eta}^{**}$ is consistent for η, the vector $\left(\sqrt{n}\left(\widehat{\eta}^{**}-\eta\right)', \sqrt{n}(\widehat{p}-p), n(\widehat{d}-d)\right)'$ has a limiting joint distribution of independence where the first marginal is k-variate normal with mean $\mathbf{0}$ and covariance matrix $(1-p)^{-1}\mathbf{\Sigma}(\eta,d)^{-1}$, the second is normal with mean 0 and variance $p\,(1-p))$ and the third is exponential with mean $((1-p)\widetilde{f}_{\eta,d}(d))^{-1}$. Further, with the exception that there are now $k+1$ "regular" parameters to consider instead of just k, all that has been said in Sections 3 and 4 about the implications of this kind of limiting distributional result for testing and set estimation carries over verbatim to this setting.

5.2 Application to the Bacteria Counts

Consider some applications of the foregoing theory to the (base 10) log-counts represented in Figures 1 and 2. These have $n = 198, n_{\mathcal{R}} = 144$ and $\widehat{d}^{*} = 1.00000$. The nonzero log-counts have sample mean 2.32889 and sample standard deviation .799915. We'll apply the normal version of our framework (where $T_1(x) = x^2$ and $T_2(x) = x$ and in terms of the mean and standard deviation of the untruncated distribution $\eta_1 = -1/2\sigma^2$ and $\eta_2 = \mu/\sigma^2$).

Maximum likelihood estimates in the mixture model are $\widehat{d}^{*} = 1.00000, \widehat{p} = 54/198 = .2727, \widehat{\mu} = 2.08306$ and $\widehat{\sigma} = .980867$. The maximum of $L^{**}(\eta,d,p)$ for these data is then -276.796. On the other hand, maximum likelihood estimates in the censoring model are $\widehat{d}^{*} = 1.00000, \widehat{\mu} = 1.77491$ and $\widehat{\sigma} = 1.17115$, and the maximum of $L^{*}(\eta,d)$ for these data is -278.526. Since with $\theta(\eta,d,p) = p - F_\eta(d) = p - \Phi(\frac{d-\mu}{\sigma})$ the hypothesis $\mathrm{H}_0 : \theta(\eta,d) = 0$ doesn't constrain d, our analysis guarantees a χ^2_1 null limit for comparing the censoring and mixture models. Since $2(-276.796-(-278.526)) = 3.46$ a p-value of .063 is indicated and the bacteria counts present strong but not conclusive evidence that more was at work in the application than simple left censoring. Since the comparison between the censoring and mixture models is not absolutely clear-cut, we'll provide illustrative numerical results under both models.

As examples where the "extra degree of freedom" is not relevant, consider estimating the median bacteria log-count (including $-\infty$'s). In the censoring model this is simply μ (provided $d < \mu$) and in mixture model this is

$$\rho(\mu,\sigma,d,p) = \mu + \sigma\Phi^{-1}\left(\frac{(.5-p)}{(1-p)}(1-\Phi(\frac{d-\mu}{\sigma})) + \Phi(\frac{d-\mu}{\sigma})\right)$$

(provided $p < .5$). Applying the χ^2_1 limit and inverting likelihood ratio tests of $\mathrm{H}_0 : \mu - c = 0$ and $\mathrm{H}_0 : \rho(\mu,\sigma,d,p) - c = 0$, approximate 95% confidence

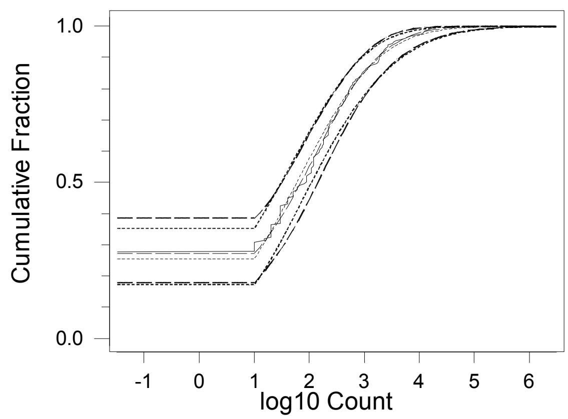

Figure 3: Empirical CDF, Two Estimated Parametric CDFs and Confidence Bands for the CDF Under the Censoring and Mixture Models

limits are $(1.597, 1.942)$ in the censoring model and $(1.649, 2.033)$ in the mixture model.

And finally, consider an application of the confidence sets like (7) for the whole parameter vector. For real x, the c.d.f. of log-counts under censoring is

$$C(x|\mu,\sigma,d) = \begin{cases} \Phi(\frac{d-\mu}{\sigma}) & \text{if } x \le d \\ \Phi(\frac{x-\mu}{\sigma}) & \text{if } x > d \end{cases}$$

while for the mixture model it is

$$M(x|\mu,\sigma,d,p) = \begin{cases} p & \text{if } x \le d \\ p + (1-p)\frac{\Phi(\frac{x-\mu}{\sigma}) - \Phi(\frac{d-\mu}{\sigma})}{1-\Phi(\frac{d-\mu}{\sigma})} & \text{if } x > d \end{cases}$$

Confidence bands for these functions of x can be made by for each x finding minimum and maximum values of C and M over all parameter vectors in a confidence set for that vector. We used our χ^2_4 and χ^2_5 limits and produced the two sets of simultaneous confidence bounds for the log-count c.d.f. shown in Figure 3. (Short dashes are for the censoring model and long ones are for the mixture model.) This application shows clearly the relevance of the "extra degree of freedom" result in real data analysis.

6 Conclusion

We have considered the impact of the existence of an unknown threshold parameter on the asymptotics of likelihood-based inference in otherwise regular continuous exponential families. Some limited experience with simulations in the normal family (recorded in Dubinin (2000)) suggests that the effects we predict (both to be present and to be absent) are evident in samples of practical size. We are thus convinced that results of this paper, beyond being a satisfying mathematical extension of what is well known about likelihood-based inference in regular problems, have implications for practice.

We note finally that a referee has suggested that the results of this article can be extended beyond exponential families to any parametric family where regular asymptotics hold. We strongly suspect that this person is correct, but presently do not see through the details clearly enough to restate our results in this more general context.

7 References

Chernoff, H. (1954). "On the distribution of the likelihood ratio," *Annals of Mathematical Statistics*, **25**, pp. 573-578.

Chow, T.L and Teugels, J.L. (1978). "The sum and maximum of i.i.d. random variables," *Proceedings of the 2nd Prague Symposium on Asymptotic Statistics*, pp. 394-403.

Cohen, C.A. (1959). "Simplified estimators for the normal distribution when samples are singly censored or truncated," *Technometrics*, **1**, pp.217-237.

Cohen, C.A. (1991). *Truncated and censored samples: Theory and Applications*, Marcel Dekker Inc., New York.

Dubinin, T.M. (2000). "Likelihood-based inference in some partially nonregular exponential families," unpublished Iowa State University Ph.D. dissertation.

Ferguson, T.S. (1996). *A Course in Large Sample Theory*, Chapman & Hall, London & New York.

Hollander, M., Proschan, F. and Sconing, J. (1990). "Information, censoring and dependence," in *Topics in Statistical Dependence*, Block, Sampson and Savits ed., Institute of Mathematical Statistics, Hayward, CA, pp. 257-268.

Schervish, M.J. (1995). *Theory of Statistics*, Springer-Verlag Inc., New York.

Smith, R.L. (1985). "Maximum likelihood estimation in a class of nonregular cases," *Biometrika*, **72**, pp. 67-90.

8 Appendix

Here we provide outlines of proofs for Theorems 5 and 8.

Proof of Theorem 5. Theorem 7.57 of Schervish (1995) implies that with (η, d) probability approaching 1, there is (a "known d MLE") $\widetilde{\eta}$ that solves $\mu(\eta, d) = \overline{\mathbf{t}}$ for which $\sqrt{n}(\widetilde{\eta} - \eta)$ is asymptotically $MVN_k(\mathbf{0}, \mathbf{\Sigma}(\eta, d)^{-1})$. The asymptotic independence of $\overline{\mathbf{t}}$ and $\widehat{d}$ implies that of $\widetilde{\eta}$ and $\widehat{d}$. The convergence of $\sqrt{n}(\widehat{\eta} - \widetilde{\eta})$ to 0 in probability then suffices to establish the asymptotic independence of $\widehat{\eta}$ and $\widehat{d}$.

Write

$$\mu(\widehat{\eta}, d) - \mu(\widetilde{\eta}, d) = \left(\left. \frac{\partial \mu_l(\eta, d)}{\partial \eta_m} \right|_{\eta^{\#l}} \right)_{l=1,\ldots,k;\ m=1,\ldots,k} (\widehat{\eta} - \widetilde{\eta})$$

where each $\eta^{\#l}$ is on the line segment between $\widehat{\eta}$ and $\widetilde{\eta}$. So with probability approaching 1

$$(\widehat{\eta} - \widetilde{\eta}) = \left(\left. \frac{\partial \mu_l(\eta, d)}{\partial \eta_m} \right|_{\eta^{\#l}} \right)^{-1}_{l=1,\ldots,k;\ m=1,\ldots,k} (\mu(\widehat{\eta}, d) - \mu(\widetilde{\eta}, d))$$

Then, since $\mu(\widetilde{\eta}, d) = \overline{\mathbf{t}} = \mu(\widehat{\eta}, \widehat{d})$, linearizing the $\mu_j(\widehat{\eta}, \widehat{d})$ at $\mu_j(\widehat{\eta}, d)$

$$(\widehat{\eta} - \widetilde{\eta}) = \left(\left. \frac{\partial \mu_l(\eta, d)}{\partial \eta_m} \right|_{\eta^{\#l}} \right)^{-1}_{l=1,\ldots,k;\ m=1,\ldots,k} \left(-(\widehat{d} - d) \begin{bmatrix} \frac{\partial}{\partial d} \mu_1(\widehat{\eta}, d)\big|_{d^1} \\ \frac{\partial}{\partial d} \mu_2(\widehat{\eta}, d)\big|_{d^2} \\ \vdots \\ \frac{\partial}{\partial d} \mu_k(\widehat{\eta}, d)\big|_{d^k} \end{bmatrix} \right)$$

where each $d^j \in (d, \widehat{d})$ and the convergence in distribution of $n(\widehat{d} - d)$ implies $\sqrt{n}(\widehat{\eta} - \widetilde{\eta}) \to 0$. ■

Proof of Theorem 8. For $\eta \in \Omega, p \in [0,1], \mathbf{s} \in \mathcal{R}^k$ and $d \in \mathcal{R}$, let

$$\mathbf{\Gamma}(\eta, p, \mathbf{s}, d) = p\overline{\overline{\mathrm{E}}}_{\eta,d}\mathbf{t}(X) + (1-p)\mathbf{s} - \mu(\eta, -\infty)$$

Equation (6) is

$$\mathbf{\Gamma}\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, \widehat{d}^*\right) = \mathbf{0}$$

Linearize the $\mathbf{\Gamma}_j\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right)$ at $\left(\widehat{\eta}^*, \widehat{d}^*\right)$ and write

$$\mathbf{\Gamma}\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right) = \left(\left. \frac{\partial \mathbf{\Gamma}_l(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)}{\partial \eta_m} \right|_{(\eta^l, d^l)} \right)_{l=1,\ldots,k;\ m=1,\ldots,k} (\eta - \widehat{\eta}^*)$$
$$+ \left(d - \widehat{d}^*\right) \begin{bmatrix} \frac{\partial}{\partial d} \mathbf{\Gamma}_1(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\big|_{(\eta^1, d^1)} \\ \frac{\partial}{\partial d} \mathbf{\Gamma}_2(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\big|_{(\eta^2, d^2)} \\ \vdots \\ \frac{\partial}{\partial d} \mathbf{\Gamma}_k(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\big|_{(\eta^k, d^k)} \end{bmatrix} \qquad (9)$$

where each $\left(\eta^j, d^j\right)$ is on the line segment between $\left(\widehat{\eta}^*, \widehat{d}^*\right)$ and (η, d). Use this to conclude that

$$\left(\left.\frac{\partial \mathbf{\Gamma}_l(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)}{\partial \eta_m}\right|_{(\eta^l, d^l)}\right)_{l=1,\ldots,k;\ m=1,\ldots,k} \sqrt{n}(\widehat{\eta}^* - \eta) \tag{10}$$

and $-\sqrt{n}\mathbf{\Gamma}\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right)$ have the same limiting behavior.

Write

$$\begin{aligned}-\sqrt{n}\mathbf{\Gamma}\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right) &= \sqrt{n}\left(1 - \frac{n_{-\infty}}{n}\right)\left(\overline{\mathbf{t}}^* - \mu(\eta, d)\right) \\ &+ \sqrt{n}\left(\frac{n_{-\infty}}{n} \cdot \overline{\overline{\mathrm{E}}}_{\eta,d}\mathbf{t}(X) + \left(1 - \frac{n_{-\infty}}{n}\right)\mu(\eta, d) - \mu(\eta, -\infty)\right)\end{aligned}$$

and using Theorem 7 and the 2nd representation in Proposition 3, conclude this has a $MVN_k\left(\mathbf{0}, \mathbf{I}^*(\eta, d)\right)$ limit. Then argue using relationship (2) and the 1st representation in Proposition 3 that the matrix of partials in (10) converges to $-\mathbf{I}^*(\eta, d)$. The marginal limit for $\sqrt{n}(\widehat{\eta}^* - \eta)$ follows.

The advertised limiting marginal for $n(\widehat{d}^* - d)$ is immediate from the limiting distribution of $n_{\mathcal{R}}\left(\widehat{d}^* - d\right)$ promised in Theorem 7, the fact that $\frac{n_{\mathcal{R}}}{n} \to F_\eta(d)$ and the relationship between $f_\eta(x)$ and $\widetilde{f}_{\eta,d}(x)$.

Regarding the asymptotic independence of $\widehat{\eta}^*$ and $\widehat{d}^*$, Theorem 7 implies the asymptotic independence of the entries of the vector $(\sqrt{n}\left(\overline{\mathbf{t}}^* - \mu(\eta, d)\right)', (n_{-\infty} - nF_\eta(d))/\sqrt{nF_\eta(d)\left(1 - F_\eta(d)\right)}, n\left(\widehat{d}^* - d\right))'$. As hypothesized, let $\widetilde{\eta}^*$ be a solution of $\mathbf{\Gamma}\left(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right) = \mathbf{0}$ that is consistent for η. The asymptotic independence of $\overline{\mathbf{t}}^*$ and $\widehat{d}^*$ implies that of $\widetilde{\eta}^*$ and $\widehat{d}^*$. It then suffices to show that $\sqrt{n}\left(\widehat{\eta}^* - \widetilde{\eta}^*\right) \to 0$.

Expand $\mathbf{0} = \mathbf{\Gamma}\left(\widetilde{\eta}^*, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d\right)$ about the point $\left(\widehat{\eta}^*, \widehat{d}^*\right)$as in (9). (Replace (η, d) with $(\widetilde{\eta}^*, d)$ in (9) where each $\left(\eta^{\#j}, d^{\#j}\right)$ is on the line segment between $\left(\widehat{\eta}^*, \widehat{d}^*\right)$ and $(\widetilde{\eta}^*, d)$.) Then with probability approaching 1

$$(\widehat{\eta}^* - \widetilde{\eta}^*) = \left(d - \widehat{d}^*\right)\left(\left.\frac{\partial \mathbf{\Gamma}_l(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)}{\partial \eta_m}\right|_{(\eta^{\#l}, d^{\#l})}\right)^{-1}_{\substack{l=1,\ldots,k;\\ m=1,\ldots,k}} \begin{bmatrix} \left.\frac{\partial}{\partial d}\mathbf{\Gamma}_1(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\right|_{(\eta^{\#1}, d^{\#1})} \\ \left.\frac{\partial}{\partial d}\mathbf{\Gamma}_2(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\right|_{(\eta^{\#2}, d^{\#2})} \\ \vdots \\ \left.\frac{\partial}{\partial d}\mathbf{\Gamma}_k(\eta, \frac{n_{-\infty}}{n}, \overline{\mathbf{t}}^*, d)\right|_{(\eta^{\#k}, d^{\#k})} \end{bmatrix}$$

Since both the matrix and the vector on the right converge and $n\left(d - \widehat{d}^*\right)$ has a limiting distribution, $\sqrt{n}(\widehat{\eta}^* - \widetilde{\eta}^*) \to 0$. ■