***matemática*:** Una publicación del ICM – ESPOL
2012, Vol. 10, No. 1

ASYMPTOTIC DISTRIBUTION THEORY FOR CONTAMINATION MODELS

Vera Francisco[1], Dickey David[2] & Lynch James[3]

Abstract. *In many situations one is interested in identifying observations that come from sources of variation other than the normal background or baseline source. A simple model for such situations is a two point mixture model where one component in the mixture corresponds to the baseline model and the second to the other sources (the contamination component). Here the goal is two-fold: (i) detect the overall presence of Contamination and (ii) identify observations that may be contaminated. A locally most powerful test is presented which gives some insights on how to accomplish this. Surprisingly, the test statistic can have an asymptotic distribution that is based on a stable law that is not the normal distribution. Examples and simulations are given to illustrate the approach.*

Keywords: multiple testing, anomaly detection, stable law, false discovery rate AMS 2000 Subject Classification: Primary 62F03, 62-07 Secondary 62E17, 62F30.

Resumen. *En muchas situaciones se tiene interés en identificar las observaciones que provienen de fuentes de variación distintas de la normal de base o de la fuente de referencia. Un modelo simple para tales situaciones es un modelo de mezcla de dos puntos, donde uno de los componentes en la mezcla corresponde al modelo de línea de base y la segunda a los de otras fuentes (el componente de la contaminación). Aquí el objetivo es doble: (i) detectar la presencia global de la Contaminación y (ii) identificar las observaciones que puedan estar contaminados. Una prueba localmente más poderosa se presenta la cual da algunas ideas sobre cómo lograr el objetivo. Sorprendentemente, la estadística de prueba puede tener una distribución asintótica que se basa en una ley estable que no es la distribución normal. Ejemplos y simulaciones se dan para ilustrar el enfoque.*

Key words: múltiples pruebas, detección de anomalías, ley estable, tasa de falso descubrimiento.

Recibido: Mayo, 2012
Aceptado: Julio, 2012

1. INTRODUCTION

The point of this paper is to present a contamination detection method based on the two point mixture model

$$f_p(x) \equiv \bar{p}f_0(x) + pf_1(x)\text{, where } \bar{p} = 1 - p\text{, and } 0 \le p \le 1 \quad \textbf{(1.1)}$$

and to investigate the asymptotic distribution of the maximum likelihood estimator (MLE) and the locally most powerful (LMP) test for the parameter *p*. The distribution, f_p, is the so called contaminated distribution model which is sometimes used to model outliers from the baseline model f_0. In this simple setting we shall see that, when $p = 0$, the MLE and the LMP test have asymptotic distributions that are non-standard. They exhibit the Chernoff phenomena (Chernoff, 1954) of being two point mixtures. These two points mixtures each have point masses at zero where the second component in the mixture is based on an α-stable law depending on the tail behavior of the likelihood ratio $f_1(X)/f_0(X)$ under f_0. In low contaminated situations $(p \approx 0)$, these asymptotics suggest using the LMP test to detect the presence of contamination. If the LMP test rejects $p = 0$, then we can use the empirical posterior $\frac{p^* f_1(x)}{f_{p^*}(x)}$, where p^* is the mle to investigate what observations may be contaminated (from f_1). Confidence bounds for this posterior can also be constructed using confidence intervals for p^*. The LMP for $p = 0$ suggests using the ratio $\frac{f_1(x)}{f_0(x)}$ to identify observations from f_1.

The asymptotics indicate that the determination of contamination when *p* is small can be problematic using classical frequentist approaches, especially if parameters need to be estimated. In addition, this has similar implications for multiple testing problems. E.g., in the analysis of microarrays, a mixture model f_0 is the model for the expression levels of the nonexpressed genes and f_1 for the differentially expressed genes. In particular, there can be a justification for the use of a central *t* distribution where the degrees of freedom is determined by the amount of replication in the

---

[1] *Vera Francisco, Ph. D., Profesor de la Escuela Superior Politécnica del Litoral (ESPOL). (e_mail: fvera@espol.edu.ec).*
[2] *Dickey David,Ph.D., University of South Carolina.*
[3] *Lynch James, Ph.D.,North Carolina State University.*

experiment or a central normal if the degrees of freedom is large. A similar justification can be used to use noncentral *t*'s or noncentral normals to model the differentially expressed genes. Here p is the proportion of differentially expressed genes.

Next section shows the data analytic model for detecting contamination, while Section 3 introduces the LMP test for *p*. Section 4 considers the asymptotic distribution of the MLE for p and the test statistic for the LMP, along with the tail behaviors of the terms in the LMP test statistic for the normal and exponential distributions.

2. POOLING AND MIXTURES

In many data analytic problems the observations $X_1,\ldots,X_n$ arise from pooling data from various sources of variation. In many cases, the pooling model has the following formulation for two sources of variation. In this formulation, a configuration *C* which is a subset of $\{1,2,\ldots,n\}$ indicates which observations come from one source and C^c from the other. For example, such a pooling model might occur in a binary network where the network is modeled by a Markov random field. In the spread of an infectious disease over the network, the nodes are partitioned into two groups, *C* and C^c, where C^c is the Collection of sites that have elevated levels of infections and *C* is the Collection of sites which are normal. In the normal case the number of infections is governed by f_0 while for the elevated level by f_1. Then,

$$p\left(\left(C,C^c\right),X_1,\ldots,X_n\right)$$
$$=K\exp\left\{E\left(\left(C,C^c\right)\right)\right\}\prod_{i\in C} f_0\left(X_i\right)\prod_{i\in C^c} f_1\left(X_i\right)$$

where $E\left(\left(C,C^c\right)\right)$ is related to the energy of the partition $\left(C,C^c\right)$ (Huang, 1963) and where we have suppressed parameters in $E\left(\left(C,C^c\right)\right)$ and the normalizing constant *K*. Here we have assumed the positivity condition that all partitions have positive probabilities. In general, the pooling model is given as follows.

- Generate a configuration *C* with probability *p*(*C*)
- Given *C*, for $i\in C, X_i$ are $\text{iid}\sim f_0$ and, for $i\in C^c, X_i$ are $\text{iid}\sim f_1$
  - *C* and C^c model a spatial or temporal (e.g., a change-point) pattern
  - You are "pooling" observations based on the configuration *C* where the configuration *C* is a hidden variable
  - The likelihood is then

$$\sum_C p(C)\prod_{i\in C} f_0\left(X_i\right)\prod_{i\in C^c} f_1\left(X_i\right)$$

Throughout we assume that all densities *f* are absolutely continuous with respect to a common measure *m* and absolutely continuous with respect to one another. The basic data analytic method is as follows:

- Envision that the data are the effects of pooling observations from f_0 and f_1 where f_0 is the background distribution and f_1 is the distribution of the contaminated observations.
- Treat the data as if it is from a mixture model and use a mixture model to estimate the mixing proportions for f_0 and f_1, that is, the proportions in *C* and C^c. Use the estimates to test the null hypothesis that one of the mixing proportions is equal to zero. If this hypothesis is rejected, see if the fitted mixture model can give insights into which observations came from f_0, that is, into the configuration *C*.

Formally, the basic data analytic model is the simple contaminated model

- $X_1,\ldots,X_n$ iid $\sim f_p=(1-p)f_0+pf_1$
  - f_0 is the density of the background mode.
  - f_1 models the contamination.
  - The likelihood is then.

$$\prod_{i=1}^{n}\left\{(1-p)f_0\left(X_i\right)+pf_1\left(X_i\right)\right\}$$
$$=\sum_{j=0}^{n}\sum_{C_j}(1-p)^j p^{n-j}\prod_{i\in C_j} f_0\left(X_i\right)\prod_{i\in C_j^c} f_1\left(X_i\right)$$

where C_j denotes a subset of size *j* from $\{1,\ldots,n\}$.

For low contaminated models one approach is to calculate the mle, p^*, of *p*. Use p^* to test $H_0: p=0$ versus $H_1: p>0$. If H_0 is rejected see if the mixture model can give insights into the configuration C_j. For example, calculate the empirical Bayes posterior with prior $p\left(C_j\right)=\left(1-p^*\right)^j p^{*n-j}$. Then

$$p\left(C_j \middle| X_1,\ldots,X_n\right) \propto \left(1-p^*\right)^j p^{*n-j} \prod_{i\in C_j} f_0\left(X_i\right) \prod_{i\in C_j^c} f_1\left(X_i\right) \tag{2.1}$$

Another approach is the following two stage multiple testing type of method for $p \approx 0$. This suggests using the locally most powerful (LMP) test statistic (discussed in the next section) for testing $H_0 : p = 0$ versus $H_1 : p > 0$ as a screening test to detect if contamination is present. If the null hypothesis is rejected, then further diagnostic tools are used to try to identify which observations are contaminated.

One was given in (2.1) and some others are given below.

For a mixed distribution $f_p, A_0\left(X_i\right) = \left(1-p\right)\dfrac{f_0(X_i)}{f_p(X_i)}$, and $A_1\left(X_i\right) = 1 - A_0\left(X_i\right)$ are referred to as the assignment function (or membership function), of X_i to f_0 and f_1, respectively. The assignment function can be interpreted as the posterior probability that an observation came from one of the components of the mixture, and can be used to decide which observations are contaminated. Related to the assignment function is the contamination assignment set measure, $p_1\left(B\right) = p\dfrac{F_1\left(B\right)}{F_p\left(B\right)}$ where $F_i\left(B\right) = \int_B f_i\left(x\right) dm\left(x\right)$ for $i = 0,1,p$. The functions $A_0\left(X\right)$ and $p_0\left(B\right) = 1 - p_1\left(B\right)$ with $B = \left(-\infty,x\right]$ or $B = \left[x,\infty\right)$ are also referred to as the local false Discovery rate (FDR) and the FDR in multiple testing situations (Efron, 2007). Note that when the null hypothesis is rejected, $p_1\left(B\right)$ (with p replaced by its mle estimator) could be interpreted heuristically as an empirical Bayes posterior probability that an observation is contaminated given that it is in *B* and gives some indication of the proportion of contamination in *B* among the background. Also note that

$$p_1\left(\left[x,x+\varepsilon\right)\right) = p\frac{F_1\left(\left[x,x+\varepsilon\right)\right)/m\left(\left[x,x+\varepsilon\right)\right)}{F_p\left(\left[x,x+\varepsilon\right)\right)/m\left(\left[x,x+\varepsilon\right)\right)} \to A_1\left(X\right)$$
$$as\ \varepsilon \to 0$$

The LMP test (next section) suggests the use of $f_1/f_0\left(X_i\right)$ to detect the contaminated observations. A plot of this quantity should be centered around 1 when there is no contamination. To find a significant collection of spurious observations consider the following approach based on the LMP test statistic. Define $L_i = \left(f_1\left(X_i\right) - f_0\left(X_i\right)\right)/f_0\left(X_i\right)$. Let the order statistics be $L_{(1)} < L_{(2)} < \ldots < L_{(n)}$ and let $j\left(i\right)$ denote the inverse rank, i.e., $L_{(i)} = L_{j(i)}$. For mixture or scanning purposes, consider the sets

$$D_i = \left\{ j\left(n\right),\ldots, j\left(n-i+1\right)\right\} = \left\{k : L_{(n-i+1)} \le L_k\right\} \tag{2.2}$$

For mixtures with mle p^*, assign D_i to f_1 and D_i^c to f_0 where $i \approx np^*$. Look through the increasing sequence of sets D_i for a spatial pattern to emerge. Use (2.1) to determine which D_i^c is most probable.

3. THE LMP TEST

In this section we discuss the LMP and the MLE of the simple contaminated model (1.1). To obtain the LMP test we need the following. Let $\varphi\left(f\left(X_1\right),\ldots,f\left(X_n\right)\right) = \varphi\left(f\right) = \log \prod_i f\left(X_i\right)$ denote the log likelihood of a set of observations from a common distribution *f* and let

$$\Phi_{p_0}\left(f_1;f_0\right) \equiv \lim_{p\to p_0} \frac{\varphi\left(f_p\right) - \varphi\left(f_{p_0}\right)}{p - p_0}$$
$$= \frac{\partial}{\partial p} \log \prod_{i=1}^{n} f_p\left(X_i\right)\Bigg|_{p=p_0}$$
$$= \sum_{i=1}^{n} \frac{f_p(X_i) - f_0(X_i)}{f_{p0}\left(X_i\right)}$$

From the generalized Neyman-Pearson lemma (cf., Ferguson, 1967, Sections 5.1 and 5.5), it is easy to show that the LMP test for testing $H_0 : p = p_0$ versus $H_1 : p > p_0$ is based on $\Phi_{p_0}\left(f_1;f_0\right)$ (see Ferguson, 1967, equation 5.78).

The LMP test statistic is related to the gradient plot introduced by Lindsay (1983a) in the study of mixed distribution models of which (1.1) is a special case. He uses the gradient plot to determine when the one point mixture mle (i.e., *p* = 0) is the global mixture mle. When it isn't, this suggests that some contamination is present. However, as shown in the next section, when the sample size is large and *p* = 0, the MLE p^* will be greater than 0 with probability 0.5. The function $\Phi_p\left(f_1;f_0\right)$ plays a prominent role in the análisis of data from mixtures models where it is the directional derivative $D(\theta;Q) = \Phi\left(f_\theta; f_Q\right) = \sum_{i=1}^{n}\left\{\frac{f\theta(X_i)}{f_Q(X_i)} - 1\right\}$ defined below. Here the mixture is over a family of densities $\left\{f_\theta : \theta \in \Theta\right\}$. Let M denote the set of probability measures on Θ. For $Q \in M$ denote the mixed distribution over the family with mixing distribution Q by

$$f_Q = \int f_\theta dQ(\theta)$$

For $X_1, \ldots, X_n$ being iid from f_Q, the likelihood and log likelihood are given by

$$L(Q) = \prod_i f_Q\left(X_i\right) \text{ and } \varphi\left(f_Q\right) = \log \prod_i f_Q(X_i)$$

where $f_Q = \left(f_Q\left(X_1\right), \ldots, f_Q\left(X_n\right)\right)$. The directional derivative of φ at f_{Q_0} towards f_{Q_1} is

$$\Phi\left(f_{Q_1}; f_{Q_0}\right) = \lim_{\varepsilon \to 0}\left(\varphi\left((1-\varepsilon) f_{Q_0} + \varepsilon f_{Q_1}\right) - \varphi\left(f_{Q_0}\right)\right) \Big/ \varepsilon$$

$$= \sum_{i=1}^{n} \frac{f_{Q_1}\left(X_i\right) - f_{Q_0}\left(X_i\right)}{f_{Q_0}\left(X_i\right)} = \sum_{i=1}^{n}\left\{\frac{f_{Q_1}\left(X_i\right)}{f_{Q_0}\left(X_i\right)} - 1\right\}$$

$$= \int D(\theta, Q) dQ_1(\theta)$$

The directional derivative D is used to identify when a k-point MLE, Q_k^*, for $L(Q)$ is the global mle Q^* (a k-point mle maximizes the likelihood function restricted to mixtures with k components). The basic idea is that $D(\theta;Q) = 0$ at the support points of the k- point MLE Q^* and $D(\theta;Q) \le 0$ if and only if Q^* is the global MLE (Lindsay, 1983a,b).

4. ASYMPTOTIC CONSIDERATIONS

In this section, we determine the asymptotic distributions of the MLE p^* of p and the LMP test statistic for testing $H_0 : p = 0$. When testing $H_0 : p = p_0$ and p_0 is in the interior of the parameter space, i.e., $0 < p_0 < 1$, the usual asymptotics go through, since they are based on sums of bounded random variables (see Proposition 4.1). Therefore, we focus only in the case when testing $H_0 : p = 0$. Section 4.1 considers the case when the true value of the parameter $p = 0$. Since $p = 0$ is on the boundary, this leads to asymptotics under nonstandard conditions. In particular, the asymptotic distribution of the MLE p^* is a mixed distribution, where one of the components is degenerate at 0, and the other is either half normal when the Fisher information $I_0 = E_0\left(\left[\left(f_1 - f_0\right)/f_0\right]^2\right) < \infty$ or is a stable law when $I_0 = \infty$.

Section 4.2 considers the distribution of the LMP test statistic for testing $H_0 : p = 0$ when the true value of the parameter $0 < p < 1$. The results therein can be used for power calculations. Section 4.3 gives the distributional properties of the ratio of two densities for the cases used in the examples and simulations. Though out this section, let $X_1, \ldots, X_n$ be iid with density $f_p(x) = (1-p) f_0(x) + p f_1(x)$ where all the random variables are assumed to be defined on the same probability space. Also let $Z_i = f_1(X_i)/f_0(X_i)$ and $L_i = Z_i - 1 = \frac{f_1(X_i) - f_0(X_i)}{f_0(X_i)}$. The LMP test statistic from Section 3 corresponding to the null hypothesis $H_0 : p = 0$ is denoted by $T_n = \sum_{i=1}^{n} L_i$. Let $I_0 = E_0\left(L_i^2\right)$ and $W_i = E_0\left(\left|L_i\right|^3\right), i = 0,1$ where E_0 denotes expectation under H_0. Note that I_0 is the Fisher information under H_0. Also, throughout this section, G_α represents the cumulative distribution function of a stable law with parameter $\alpha \in (0,2]$, i.e., its characteristic function is (A.1). Define $\overline{G}_\alpha = 1 - G_\alpha$.

The next proposition is used in some parts of this section and is the basis for the claim that when p_0 is in the interior of the parameter space the terms in the LMP test statistic are all bounded.

PROPOSITION 4.1.

$$\frac{f_1(x) - f_0(x)}{(1-p) f_0(x) + p f_1(x)}$$

is bounded for $0 < p < 1$ (hence all its moments are finite).

Proof. Notice that

$$\frac{f_1(x)-f_0(x)}{(1-p)f_0(x)+pf_1(x)} = \frac{1}{p}\left(\frac{f_1(x)}{\frac{1-p}{p}f_0(x)+f_1(x)}\right)$$

$$-\frac{1}{1-p}\left(\frac{f_0(x)}{f_0(x)+\frac{p}{1-p}f_1(x)}\right)$$

Therefore,

$$\left|\frac{f_1(x)-f_0(x)}{(1-p)f_0(x)+pf_1(x)}\right| \le \frac{1}{p}+\frac{1}{1-p}$$

4.1 FIRST CASE: $p = 0$

The next few lemmas show the distribution of the MLE p^* when $p = 0$ under different conditions.

Lemma 4.2. Under $H_0 : p = 0, p^*$ converges to 0 almost surely.

Proof. Let

$$l(p) \equiv \varphi(f_p)$$

$$l'(p) = \frac{\partial}{\partial p}\varphi(f_p) = \frac{\partial}{\partial p}\log\prod_{i=1}^{n} f_p(X_i) = \sum_{i=1}^{n}\frac{f_1(X_i)-f_0(X_i)}{f_p(X_i)}$$

and note that

$$l''(p) = -\sum_{i=1}^{n}\frac{\left[f_1(X_i)-f_0(X_i)\right]^2}{f_p(X_i)^2} \le 0$$

So $l(p)$ is concave and attains its maximum, p^*, either at 0 or 1 or on (0; 1). Let $U_n(p) = l'(p)$ where $U_n = U_n(0)$ and note that $U_n(p)$ is the sum of n iid random variables with

$$E_0\left(\frac{f_1(X_i)-f_0(X_i)}{f_p(X_i)}\right) = 0 \text{ when } p = 0 \text{ and}$$

$$< 0 \text{ for } p > 0 \qquad \textbf{(4.1)}$$

When $U_n \le 0, U_n(p) \le U_n$ since $l(p)$ is concave. Thus, $l(p)$ attains it maximum at 0 on $\{U_n \le 0\}$. When $U_n > 0, l(p)$ attains its maximum on (0,1]. Since $U_n(p)$ has mean less than 0 for $0 < p < 1$, $U_n(p)/n$ converges almost surely to a negative number (because of Proposition 4.1). When $U_n(p) < 0$ and $U_n > 0, 0 < p^* < p$ with $U_n(p^*) = 0$ since $U_n(0) > 0$ and $U_n(p) < 0$. Thus, $\lim p^* < p$ almost surely on the set where $U_n(p)/n$ converges to its mean. Since $p > 0$ is arbitrary, this with the previous paragraph implies that p^* converges to zero almost surely.

Lemma 4.3. If $I_0 < \infty$ and $W_i < \infty, i = 0,1$, then, under $H_0, \sqrt{n}p^*$ converges in distribution to X where $X = 0$ with probability .5 and $= \left|N\left(0, I_0^{-1}\right)\right|$ with probability .5.

Proof. If $p^* \in (0,1)$, then $l'(p^*) = 0$ and

$$l'(0) = l'(0) - l'(p^*) = -l''(0)(p^*) - \frac{l'''(p')(p^*)^2}{2} \qquad (4.2)$$

where p' is between 0 and p^* and

$$l'''(p) = 2\sum_{i=1}^{n}\frac{\left[f_1(X_i)-f_0(X_i)\right]^3}{f_p(X_i)^3}$$

Note that since the derivative of $l'''(p)$ is nonpositive, $l'''(p)$ is nonincreasing and $l'''(1) \le l'''(p) \le l'''(0)$. Thus, since $W_i < \infty$ for $i = 0,1$, the sequence $l'''(p)/n, n = 1,2,...$ is bounded almost surely. It follows from (4.2) that when $U_n > 0$ and $U_n(1) < 0, p^* \in (0,1)$ and

$$\frac{l'(0)}{\sqrt{n}} = \frac{-l''(0)}{n}\sqrt{n}p^* - \frac{l'''(p')\sqrt{n}(p^*)^2}{2n}$$

$$= \frac{-l''(0)}{n}\left(\sqrt{n}p^*\right)(1+R_n) \qquad (4.3)$$

In (4.3) R_n goes to zero almost surely since p^* converges to zero almost surely, the sequence $l'''(p)/n, n = 1,2,...$ is bounded almost surely and $-l''(0)/n$ converges almost surely to I_0. When $U_n \le 0, p^* = 0$. Since $U_n/\sqrt{n}$ is asymptotically $N(0, I_0)$ and $U_n(1)/n$ converges almost surely to a negative number by (4.2), $P(U_n \le 0)$ and $P(U_n > 0 \text{ and } U_n(p) < 0)$ both converge to 1/2. The second part of this lemma follows from this and from (4.3) since $-l''(0)/n$ converges almost surely to $E_0\left(\left[f_1 - f_0/f_0\right]^2\right) = I_0$ and $-l'(0)/\sqrt{n}$ converges in distribution to $N(0, I_0)$.

For the next lemma, $\{a_n\}$ is a sequence of real numbers and

$$V_n(p) = \frac{1}{a_n^2}\sum \frac{\left(f_1(X_i) - f_0(X_i)\right)^2}{f_0(X_i)\, f_p(X_i)}$$

Lemma 4.4. If Z_1 Satisfies (A.2) for some $1 < \alpha \le 2$ (i.e., is in the domain of attractation of an α-stable law) and a_n satisfies (A.4), then, under H_0, $a_n p^* V_n(p^*)$ converges in distribution to X where $X = 0$ with probability $G_\alpha(0)$ and $P(X > d) = \bar{G}_\alpha(d)$ for $d > 0$.

Proof. For $p^* \in (0,1), l'(p^*) = 0$ and

$$\begin{aligned} l'(0) &= l'(0) - l'(p^*) \\ &= \sum_{i=1}^{n}\left(f_1(X_i) - f_0(X_i)\right)\left(\frac{1}{f_0(X_i)} - \frac{1}{f_{p^*}(X_i)}\right) \\ &= p^* \sum \frac{\left(f_1(X_i) - f_0(X_i)\right)^2}{f_0(X_i)\, f_{p^*}(X_i)} \\ &= p^* a_n^2 V_n(p^*) \end{aligned} \tag{4.4}$$

Note that when $p^* \in (0,1)$,

$$\frac{l'(0)}{a_n} = a_n p^* V_n(p^*)$$

and $l'(0)/a_n$ converges in distribution to G_α by Lemma A.1 since $E_0(Z_1) = 1$. Since $p^* = 0$ when $l'(0) < 0$, the results follows.

From the proof of Lemma 4.4, by setting $a_n = \sqrt{n}$ we can get the asymptotic distribution of p^* without the third moment assumption given in Lemma 4.3. The next corollary states this result.

Corollary 4.5. If $I_0 < \infty$, then, under H_0, $\alpha = 2$ and $p^* \sqrt{n} V_n(p^*)$ converges in distribution to X where $X = 0$ with probability $.5 = \left|N\left(0, I_0^{-1}\right)\right|$ and with probability .5. Moreover, $V_n(0)$ converges almost surely to I_0.

Remark 4.6. When $1 < \alpha < 2, V_n(0)$ converges in distribution to a stable law with parameter $\alpha/2$ (by Corollary A.2). So, one Could replace $V_n(p^*)$ with $V_n(0)$ in (4.4), except that one could not justify this replacement without putting some condition on $l'''(p)$ The next few lemmas show the distribution of the LMP test statistic T_n for various cases.

Lemma 4.7. If $I_0 < 1$, then, under $H_0, T_n/\sqrt{n}$ converges in distribution to $N(0, I_0)$.

Proof. The proof follows by a direct application of the central limit theorem.

Lemma 4.8. If Z_1 and a_n satisfy conditions (A.2) and (A.4), respectively, for some $1 < \alpha \le 2$, then, under $H_0, T_n/a_n$ converges in distribution to G_α (a stable law with parameter α).

Proof. The proof follows by a direct application of Lemma A.1

If f_1 has an unknown parameter and $p = 0$, an identifiability issue surfaces that makes it impossible to estimate that parameter. In this case, if the parameter is estimated from the data and used to calculate T_n, it is not clear to what limit distribution T_n is converging. The simulations shown later illustrate this point.

4.2 Case 2: $p > 0$

The asymptotic distribution of T_n given in Lemmas 4.7 and 4.8 is for $p = 0$. The next two lemmas give the asymptotic distribution of T_n when $p > 0$. For this, assume that $X_1 \sim f_p = (1-p) f_0 + p f_1$ and let $W_0' = E_0\left(L_1^3\right)$.

Lemma 4.9. If $I_0 < \infty$ and $W_0 < \infty$, then $\left(T_n - n p I_0\right)/\sqrt{n}$ converges in distribution to $N\left(0, I_0 + pW_0' - p^2 I_0^2\right)$.

Proof. It is easy to prove that $E_p(L_1) = pI_0$ and $E_p\left(L_1^2\right) = I_0 + pW_0'$. The result then follows from a direct application of the central limit theorem.

Lemma 4.10. Suppose Z_1 and a_n satisfy conditions (A.2) and (A.4), respectively, for some $0 < \alpha \le 2$. If $1 < \alpha \le 2$, then $I_0 < \infty$ and $\left(T_n - npI_0\right)/a_n$ converges in distribution to G_α, while if $0 < \alpha < 1$, T_n/a_n converges in distribution to G_α. If $\alpha = 1$, then $\left(T_n - \mu_n\right)/a_n$ converges in distribution to a stable law with parameter 1, where μ_n is defined as in (A.3).

Proof. The proof is a direct application of the results in Appendix A.

4.3 DISTRIBUTIONAL PROPERTIES OF DENSITY RATIOS

In this section, we consider the properties of $Z = f_1(X)/f_0(X)$ for some frequently used distributions. These properties are required to use the lemmas in Sections 4.1 and 4.2. Section 4.3.1 considers the case when both, f_0 and f_1 are exponential distributions, and Section 4.3.2 considers the case of the normal distribution.

4.3.1 EXPONENTIAL DISTRIBUTION

Suppose $f_\theta(x) = \frac{1}{\theta} e^{-x/\theta} 1_{[x>0]}$ and let $f_j = f_{\theta_j}$ for $j = 0,1$. If $X \sim f_\theta$ then

$$P_\theta(Z > z) = \begin{cases} 1_{[z\le 0]} + \left(1 - (z/\beta)^{-\alpha}\right) 1_{[0<z<\beta]} \\ \quad \text{when } \theta_1 < \theta_0 \\ 1_{[z\le\beta]} + (z/\beta)^{-\alpha} 1_{[z>\beta]} \\ \quad \text{when } \theta_1 > \theta_0 \end{cases}$$

Where $\beta = \frac{\theta_0}{\theta_1}$ and $\alpha = \frac{\theta_0\theta_1}{\theta(\theta_1 - \theta_0)}$.

For $\theta_1 > \theta_0, z^\alpha P_\theta(Z > z) \to c = \beta^\alpha$ as $z \to \infty$, which corresponds to the first row in Tables 5 and 6. In particular, if $0 < \alpha < 2$, condition (A.2) is satisfied and therefore Z is in the domain of attraction of an α-stable law. If $\alpha \ge 2$ then condition (A.6) is satisfied and Z would be in the domain of attraction of a normal distribution. The appropriate normalizing constant is $a_n = \left(\beta^\alpha s_\alpha n\right)^{1/\alpha}$ for $0 < \alpha < 2$, and $a_n = \sqrt{0.5\beta^2 n \log n}$ for $\alpha = 2$, and $a_n = \sqrt{0.5n \operatorname{var}_\theta(Z)}$ for $\alpha > 2$. The mean of the density ratio Z is

$$E_\theta(Z) = \begin{cases} \dfrac{\alpha\beta}{\alpha - 1} & \text{when } \theta_1 < \theta_0 \text{ or } (\theta_1 > \theta_0 \text{ and } \alpha > 1) \\ \infty & \text{when } \theta_1 > \theta_0 \text{ and } \alpha \le 1 \end{cases}$$

and the variance is

$$\operatorname{var}_\theta(Z) = \begin{cases} \dfrac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)} & \text{when } \theta_1 < \theta_0 \text{ or} \\ & (\theta_1 > \theta_0 \text{ and } \alpha > 2) \\ \infty & \text{when } \theta_1 > \theta_0 \text{ and } 1 < \alpha \le 2 \\ \text{undefined} & \text{when } \theta_1 > \theta_0 \text{ and } \alpha \le 1 \end{cases}$$

In particular, when $\theta = \theta_0, \alpha = \theta_1/(\theta_1 - \theta_0)$ and the mean becomes $E_{\theta_0}(Z) = 1$ and the variance, which is actually I_0, becomes

$$I_0 = \operatorname{var}_{\theta_0}(Z) = \begin{cases} \dfrac{(\theta_1 - \theta_0)^2}{\theta_1(2\theta_0 - \theta_1)} & \text{when } \theta_1 < 2\theta_0 \\ \infty & \text{when } \theta_1 \ge 2\theta_0 \end{cases}$$

For $\theta = \theta_1$ the mean becomes $E_{\theta_1}(Z) = I_0 + 1$, which is infinite when $\theta_1 \ge 2\theta_0$, and the variance becomes

$$\operatorname{var}_{\theta_1}(Z) = \begin{cases} \dfrac{\theta_0^3(\theta_1 - \theta_0)^2}{2\theta_1^2(2\theta_0 - \theta_1)^2(1.5\theta_0 - \theta_1)} & \text{when } \theta_1 < 1.5\theta_0 \\ \infty & \text{when } 1.5\theta_0 \le \theta_1 < 2\theta_0 \\ \text{undefined} & \text{when } \theta_1 \ge 2\theta_0 \end{cases}$$

4.3.2 NORMAL DISTRIBUTION

Let $\phi(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ denote the density of a standard normal distribution. Suppose now that $f_{\mu,\sigma^2}(x) = \phi((x - \mu)/\sigma)/\sigma$ □and let $f_j = f_{\mu_j,\sigma_j^2}$ for $j = 0,1$. Assume also that $X \sim f_{\mu_j,\sigma_j^2}$. Before examining the tail probability of the density ratio, we need the tail probability of a normal distribution. Let Y □be a standard normal distribution with density ϕ□. It is well known that, for $y > 0$,

$$\frac{y}{1+y^2}\phi(y) < P(Y>y) < \frac{1}{y}\phi(y)$$

This implies that

$$\frac{y^2}{1+y^2} < \sqrt{2\pi}ye^{y^2/2}P(Y>y) < 1$$

and thus

$$\lim_{y\to\infty}\sqrt{2\pi}ye^{y^2/2}P(Y>y) \to 1 \qquad (4.5)$$

Let $\delta_j = \frac{\mu_1^j}{\sigma_1^2} - \frac{\mu_0^j}{\sigma_0^2}$ for $j = 1,2\left(\text{for } j=0 \text{ let } \delta_0 = \frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)$. The tail probability of the density ratio when $\sigma_1^2 > \sigma_0^2$ is given by

$$\begin{aligned}
&P_{\mu,\sigma^2}(Z > z) \\
&= P_{\mu,\sigma^2}\left(\frac{f_1(X)}{f_0(X)} > z\right) \\
&= P_{\mu,\sigma^2}\left(\frac{\left|X - \frac{\delta_1}{\delta_0}\right|}{\sigma} > \sqrt{\alpha\left(\delta_2 - \frac{\delta_1^2}{\delta_0} - \log(\beta)\right) + 2\alpha\log z}\right) \\
&= P\left(Y > -b + \sqrt{\alpha\left(\frac{(\mu_1-\mu_0)^2}{\sigma_1^2-\sigma_0^2} - \log(\beta)\right) + 2\alpha\log z}\right) \\
&\quad + P\left(Y > b + \sqrt{\alpha\left(\frac{(\mu_1-\mu_0)^2}{\sigma_1^2-\sigma_0^2} - \log(\beta)\right) + 2\alpha\log z}\right)
\end{aligned}$$

where

$$b = \frac{\left|\sigma_1^2(\mu-\mu_0) + \sigma_0^2(\mu_1-\mu)\right|}{\sigma\left(\sigma_1^2 - \sigma_0^2\right)}, \alpha = \frac{\sigma_0^2\sigma_1^2}{\sigma^2\left(\sigma_1^2 - \sigma_0^2\right)},$$

and $\beta = \sigma_0^2/\sigma_1^2$.

This expression, combined with (4.5), gives us the tail behavior, i.e., as $z \to \infty$,

$$\begin{cases}
z^\alpha\sqrt{2\alpha\log z}P_{\mu,\sigma^2}\left(\frac{f_1(X)}{f_0(X)} > z\right) \to c_1 \\
\qquad \text{when } b = 0, \text{and} \\
z^\alpha\sqrt{2\alpha\log z}\exp\left(-b\sqrt{2\alpha\log z}\right)P_{\mu,\sigma^2}\left(\frac{f_1(X)}{f_0(X)} > z\right) \to c_2 \\
\qquad \text{when } b > 0,
\end{cases}$$

where $c_1 = 2\frac{\exp\left[\frac{\alpha}{2}\left(\log(\beta) - \frac{(\mu_1-\mu_0)^2}{\sigma_1^2-\sigma_0^2}\right)\right]}{\sqrt{2\pi}}$ and

$$c_2 = \frac{\exp\left[-\frac{b^2}{2} + \frac{\alpha}{2}\left(\log(\beta) - \frac{(\mu_1-\mu_0)^2}{\sigma_1^2-\sigma_0^2}\right)\right]}{\sqrt{2\pi}}.$$

Therefore, when $\sigma_1^2 > \sigma_0^2$, the tail behavior coincides with those in Tables 5 and 6, which give the appropriate normalizing constants for $0 < \alpha < 2$ and $\alpha = 2$ respectively. When $\alpha > 2$, then the appropriate normalizing constant is $a_n = \sqrt{.5n\,\mathrm{var}_\theta(Z)}$.

The mean of the density ratio is

$$E_{\mu,\sigma^2}(Z) = \begin{cases}
e^{\left(\frac{1}{2}\left[\sigma^2\delta_1^2 + 2\mu\delta_1 - \delta_2\right]\right)} & \text{when } \sigma_1^2 = \sigma_0^2 \\
\sqrt{\frac{\alpha\beta}{\alpha-1}}e^{\left(\frac{1}{2\sigma^2}\left[\frac{\alpha}{\alpha-1}\left(\sigma^2\delta_1 + \mu\right)^2 - \mu^2 - \sigma^2\delta_2\right]\right)} & \\
\quad \text{when } \sigma_1^2 < \sigma_0^2 \text{ or } \left(\sigma_1^2 > \sigma_0^2 \text{ and } \alpha > 1\right) & \\
\infty & \text{when } \sigma_1^2 > \sigma_0^2 \text{ and } \alpha \le 1
\end{cases}$$

The variance is given by

$$\operatorname{var}_{\mu,\sigma^2}(Z) = \begin{cases} \left(e^{\sigma^2\delta_1^2} - 1\right)e^{\left(\sigma^2\delta_1^2 + 2\mu\delta_1 - \delta_2\right)} & \text{when } \sigma_1^2 = \sigma_0^2 \\ \sqrt{\dfrac{\alpha\beta^2}{\alpha-2}}\,e^{\left(\frac{1}{2\sigma^2}\left[\frac{\alpha}{\alpha-2}\left(2\sigma^2\delta_1+\mu\right)^2 - \mu^2\right] - \delta_2\right)} & \text{when } \sigma_1^2 < \sigma_0^2 \text{ or} \\ -\dfrac{\alpha\beta}{\alpha-1}\,e^{\left(\frac{1}{\sigma^2}\left[\frac{\alpha}{\alpha-1}\left(\sigma^2\delta_1+\mu\right)^2 - \mu^2\right] - \delta_2\right)} & \left(\sigma_1^2 > \sigma_0^2 \text{ and } \alpha > 2\right) \\ \infty & \text{when } \sigma_1^2 > \sigma_0^2 \text{ and } 1 < \alpha \le 2 \\ \text{undefined} & \text{when } \sigma_1^2 > \sigma_0^2 \text{ and } \alpha \le 1 \end{cases}$$

When $f_{\mu,\sigma^2} = f_0$ the mean reduces to $E_0(Z) = 1$ and the variance, which is I_0, reduces to

$$I_0 = \operatorname{var}_0(Z)$$

$$= \begin{cases} e^{\left([\mu_1 - \mu_0]^2 / \sigma_0^2\right)} - 1 & \text{when } \sigma_1^2 = \sigma_0^2 \\ \dfrac{\sigma_0^2}{\sigma_1\sqrt{2\sigma_0^2 - \sigma_1^2}}\,e^{\left(\frac{(\mu_1-\mu_0)^2}{2\sigma_0^2 - \sigma_1^2}\right)} - 1 & \text{when } \sigma_1^2 < \sigma_0^2 \\ \infty & \text{when } \sigma_1^2 \ge 2\sigma_0^2 \end{cases}$$

For $f_{\mu,\sigma^2} = f_1$ the mean becomes $E_1(Z) = I_0 + 1$ and the variance becomes

$$\operatorname{var}_1(Z) = \begin{cases} \left(e^{\left([\mu_1 - \mu_0]^2 / \sigma_0^2 \right)} - 1 \right) e^{\left(2[\mu_1 - \mu_0]^2 / \sigma_1^2 \right)} & \text{when } \sigma_1^2 = \sigma_0^2 \\ \dfrac{\sigma_0^3}{\sigma_1^2 \sqrt{2\left(1.5\sigma_0^2 - \sigma_1^2\right)}} e^{\left(1.5 \frac{(\mu_1 - \mu_0)^2}{1.5\sigma_0^2 - \sigma_1^2} \right)} & \text{when } \sigma_1^2 < 1.5\sigma_0^2 \\ -\dfrac{\sigma_0^4}{\sigma_1^2 \left(2\sigma_0^2 - \sigma_1^2\right)} e^{\left(2 \frac{(\mu_1 - \mu_0)^2}{2\sigma_0^2 - \sigma_1^2} \right)} & \\ \infty & \text{when } 1.5\sigma_0^2 \le \sigma_1^2 < 2\sigma_0^2 \\ \text{undefined} & \text{when } \sigma_1^2 \ge 2\sigma_0^2 \end{cases}$$

5. SIMULATIONS

For the first set of simulations, background and contamination data are generated from exponential distributions with means $\theta_0 = 166.206$ and $\theta_1 = 592.922$, respectively, which are the estimated means for the 2-point mle for the Mining Data in the next section. Samples of sizes n = 100, 500, 1000 are generated, with 0, 1 and 5 percent of contamination (p = 0, 1.01, .05). With each sample we calculate $T_n = \sum_{i=1}^{n} \frac{f_1(X_i) - f_0(X_i)}{f_0(X_i)}$ and $S_n = \sum_{i=1}^{n} \frac{\hat{f}_1(X_i) - \hat{f}_0(X_i)}{\hat{f}_0(X_i)}$, where $\hat{f}_0$ and $\hat{f}_1$ are estimates of the densities based on the maximum likelihood estimators of θ_0 and θ_1. These estimates are used as if they were the true parameters and a normalizing constant and critical value are calculated based on these estimates. The process is repeated N = 10000 times and the number of rejections of the null hypothesis $H_0 : p = 0$ at the .05 level are recorded.

Following the results from Section 4.3.1, the variance of the terms in T_n corresponding to f_0 is infinite, but the tail behavior of the density ratio, under H_0, follows that of the first line of Table 5 with α = 592.922/(592.922-166:206) = 1.3895 and c = (166.206/592.922)$^{1.3895}$ =0.1708081. Hence, the normalizing constant is $a_n = \left(0.1708081 s_{1.3895} n\right)^{1/1.3895} = .4881128 n^{0.7196832}$. From Lemma A.1, the rejection region defined by $T_n / \left(.4881128 n^{0.7196832}\right) > 4.40186$ would reject the null hypothesis with probability 0.05 if there are no anomalies. A similar process is done to calculate the rejection region $S_n / a_n > d_{.05}$ in each sample, where the normalizing constant a_n and the critical value $d_{.05}$ change from sample to sample, and are calculated based on either the normal or the stable distribution[4]. The results of these simulations are shown

TABLE I

Asymptotic distribution theory for contamination models

Proportion of rejections of H_0, no anomalies, out of 10000 simulations with background and contamination generated from exponential distributions with means 166.206 and 592.922, respectively, with p = 0, 0.01, 0.05.

	Proportion of anomalies					
	Based on T_n			Based on S_n		
Sample sizes	0	0.01	0.05	0	0.01	0.05
100	0.0492	0.1646	0.5316	0.2853	0.3898	0.6340
500	0.0483	0.3836	0.9361	0.3248	0.5725	0.9252
1000	0.0544	0.5454	0.9935	0.3387	0.6898	0.9786

---

[4] *if* $\hat{\theta}_1 < 2\hat{\theta}_0$*, and these are assumed to be the actual parameters, the variance of the in Table 1.*

The simulations show that when using the true parameters to calculate T_n the proportion of samples that rejected the null hypothesis when it is true is about 0.05, as expected. Notice that with the background and contamination means fixed at θ_0 = 166.206 and θ_1 = 592.922, the power increases as the proportion of anomalies, *p*, increases and as the sample size increases. The exact asymptotic power for the LMP test can be calculated using Lemma 4.10 and the results from Section 4.3.1 and Table 5, with $\alpha = \theta_0 / (\theta_1 - \theta_0) = 0.3895003$ and $c = p(\theta_0/\theta_1)^{\alpha} = 0.6093393p$, which gives $(cs_{\alpha})^{1/\alpha} = 0.4473926p^{2.567392}$. Rejection occurs when $T_n / (.4881128n^{0.7196832}) > 4.40186$, which is equivalent to rejecting if $T_n / (0.4473926p^{2.567392}n^{2.567392}) > 4.802503 / (p^{2.567392}n^{1.847709})$

The left hand side of this inequality converges to a stable law with parameter α = 0.3895003. The power can be obtained for each value of *n* and *p*. For instance, if *n* = 100 and *p* = 0.05, the power is the probability that a value from a stable law is larger than 2.119844, that is, 0.510141. In the case of *n* = 500 and *p* = 0.01, the tail starts at 6.750576, which gives a power of 0.3517792. For *n* = 1000 and *p* = 0.05, the power is 0.996379. The simulations confirm these values. If estimates are used as if they were the true parameters, rejection occurs 28.5% of the time when there are no anomalies present (*p* = 0) and *n* = 100. This is troubling and indicates that false discovery is a serious problem in this case. This is not the case when *p* > 0 and Appendix B indicates the necessary adjustments that need to be made when estimating parameters. For the second set of simulations data are generated from normal distributions, where the background consists of standard normal variables $(\mu_0 = 0 \text{ and } \sigma_0^2 = 1)$ and the anomalies consist of a normal with mean $\mu_1 = 0$ and variance $\sigma_1^2 = 3$. Samples of sizes *n* = 100, 500, 1000 are generated, with 0, 1 and 5 percent of the observations being anomalies.

The results from Section 4.3.2 and Table 5 are used to determine the rejection region for each sample. Since the variance of f_1/f_0 is infinite under f_0, the tail behavior of the distribution of the ratio needs to be taken into consideration to see what stable law applies when $\mu = \mu_0 = 0$ and $\sigma^2 = \sigma_0^2 = 1$. This is described in Section 4.3.2 with

$$\alpha = 3/(3-1) = 1.5, \beta = 1/3, b = 0, c_1$$

$$= 2\beta^{\alpha/2} / \sqrt{2\pi} = 2\left(\tfrac{1}{3}\right)^{1.5/2} / \sqrt{2\pi} = 0.350025$$

The rejection region is now found using the results from Table V: reject the null hypothesis $H_0 : p = 0$ if

$$\frac{T_n}{0.7274158\left(n/\sqrt{\log n}\right)^{2/3}} > 3.824235$$

The results of these simulations are found in Table II. The LMP test seems somewhat conservative for *n* = 100, possibly because this sample size is too small to observe convergence to the stable law. This resulted in a test with somewhat poor power. For *n* = 500 and *n* = 1000, the test seems to perform better.

TABLE II

Asymptotic distribution theory for contamination models

Proportion of rejections of H_0, no anomalies, out of 10000 simulations with background and anomalies generated from normal distributions with mean 0 and variances 1 and 3, respectively

	Proportion of anomalies		
Sample sizes	0	0.01	0.05
100	0.0371	0.0788	0.2331
500	0.0447	0.1533	0.5649
1000	0.0408	0.2295	0.7885

The exact asymptotic power can be calculated for these tests by using Lemma 4.10, Table V and Section 4.3.2. Suppose the true proportion of anomalies p is positive $p > 0$. Then $P(Z_1 > z) = (1-p)P_0(Z_1 > z) + pP_1(Z_1 > z)$. Let $\alpha = 1/(3-1) = 0.5$ and $c = 2\left(\tfrac{1}{3}\right)^{0.5/2} p / \sqrt{2\pi} = 0.6062612p$ Thus, to get a stable law we need to normalize T_n by $0.2886751p^2\left(n/\sqrt{\log n}\right)^2$. A rearrangement of the rejection region (which was normalized originally by $0.7274158\left(n/\sqrt{\log n}\right)^{2/3}$) would result in rejections when

$$\frac{T_n}{0.2886751 p^2 \left(n/\sqrt{\log n}\right)^2} > 3.824235 \frac{0.7274158\left(n/\sqrt{\log n}\right)^{2/3}}{0.2886751 p^2 \left(n/\sqrt{\log n}\right)^2}$$

$$= \frac{9.636468(\log n)^{2/3}}{p^2 n^{4/3}}$$

The exact asymptotic power when n = 100 and $p = 0.5$ is the probability that a stable law variable with parameter 0.5 is greater than $\frac{9.636468\left(\log 100\right)^{2/3}}{.5^2 100^{4/3}} = 22.98661$. This probability is 0.1652201. Similarly, for n = 500 and proportions p = .01 and p = .05, the probabilities of rejecting the null hypothesis are 0.08789053 and 0.4189768, respectively, whereas the simulations estimated these numbers as 0.1533 and 0.5649, respectively.

6. DATA EXAMPLES

Following the analysis from Grego et al. (1990) of the mining accident data, Figure 1 has the gradient functions for the 2 and 3-point mixture mle's where the mixing is over the mean of an exponential distribution and Figure 2 has the assignment function for the second component in the 3-point mle. The estimates of the means and mixing proportions are given in Table III. The gradient plot indicates that the 2-point mle is not the global mle but the 3-point is. The assignment function indicates a distinct difference in the first 53 times and rest of the times. Further analysis by Grego et al indicates that the first 53 are well fit by a single exponential and the rest by a 3- point mixture.

TABLE III

Asymptotic distribution theory for contamination models

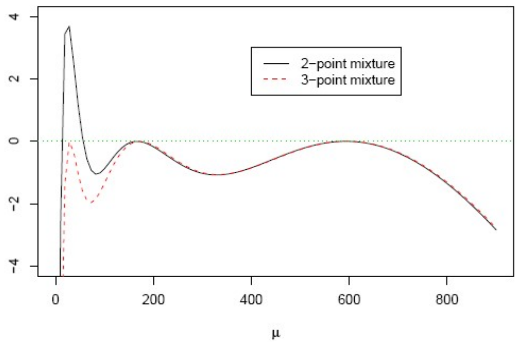
Maximum likelihood estimates for the mining data

	$\mu_1(p_1)$	$\mu_2(p_2)$	$\mu_3(p_3)$
2-point mle	592.922 (.175149)	166.206 (.824851)	
3-point mle	595.495 (.171379)	171.587 (.805528)	29.0972 (.023093)

FIGURE 1

Asymptotic distribution theory for contamination models

Gradient plots of a 2- and 3-point mixtures (mle) of exponentials for the Mining Data



For the mining data we will use an exponential with mean 171.587 as f_0 and a 2-point mixed exponential with means 595.495 and 29.0972 and mixing proportions proportional to .171379 and .023093, respectively, as f_1. That is, $f_1 = f_{Q_1}$ where Q_1 has point masses at 595.495 and 29.0972 with mixing proportions .881253 and .118747 and the family, $\{f_\mu\}$, being mixed over is the exponential with its mean parameterization. These are assumed as the true parameters and Lemma 4.8 along with Table V can be used to calculate critical values for the LMP test statistic. The LMP test statistic, T_n, assuming all the parameters are known, is then given by

$$T_n = 0.8812528 \sum_{i=1}^{n} \left(\frac{\frac{1}{595.495} e^{-X_i/595.495}}{\frac{1}{171.587} e^{-X_i/171.587}} - 1 \right)$$

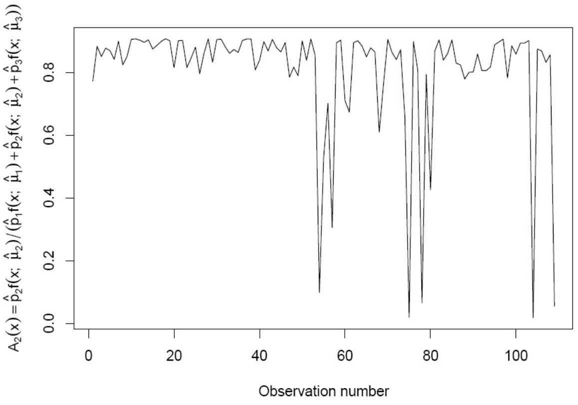
$$+ 0.1187472 \sum_{i=1}^{n} \left(\frac{\frac{1}{29.0972} e^{-X_i/29.0972}}{\frac{1}{171.587} e^{-X_i/171.587}} - 1 \right)$$

Under the null hypothesis, i.e., $X_i \sim f_0$, the terms in the first sum have infinite variance whereas the terms in the second sum have finite variance (see Appendix 4.3.1 for details). Using the notation of Appendix 4.3.1 with $\theta_0 = 171.587$ and $\theta_1 = 595.495$, let $\alpha = 595.495/(595.495\text{-}171.587) = 1.404774$, $\mathrm{c} = (171.587/595.498)^{1.404774} = 0.1741281$ and $a_n = 0.5052435 n^{0.7118582}$. If T_n is normalized by a_n, the second sum will quickly converge to zero as $n \to \infty$. The first sum converges in distribution to a stable law with parameter $\alpha = 1.404774$. Therefore $T_n/(.8812528 a_n)$ converges in distribution to the same stable law.

FIGURE 2

Asymptotic distribution theory for contamination models

Assignment function for the second component of the 3-point mixture (mle) of exponentials for the Mining Data



For the mining data, $T_n = 574.871$ and $T_n / (.8812528 a_n) = 45.77311$. Using Table IV for $\alpha = 1.4$ we can see that the p-value is between .005 and .001. The actual p-value is 0.002102145 (calculated computationally with $\alpha = 1.404774$). This indicates that there is strong evidence that some observations come from f_1. Note that parameters in both f_0 and f_1 are being estimated based on the 3-pt global mle. These estimates have to be taken into consideration in using the LMP test statistic to determine if spurious observations are present. As pointed out in the appendix, it would be impossible to estimate f_1 if $p = 0$, and hence the distribution of the LMP is not clear in this case. If $p > 0$, then we only need to check the regularity conditions discussed in Lemma B.1 and Remark B.2. For the exponential distribution these conditions reduce to the finiteness of the first three moments. We now illustrate some of these ideas using gene expression data. The approach here will start by the assignment function to identify possible anomalies (expressed genes) to get a pooled model. After that, we do the LMP test. Efron (2007) compared prostate data of $m_1 = 50$ non-tumor subjects with $m_2 = 52$ tumor patients for each of n = 6033 genes (see Singh et al., 2002). For each gene they perform a two-sample t-test to compare the mean gene-expression between cancer and noncancer subjects. Let t_i for $i = 1,\ldots,n$ denote the test statistics used for each gene. For genes that have the same mean expression values for both groups ti will follow a central t-distribution with $m_1 + m_2 - 2 = 100$ degrees of freedom. Efron (2007) defines $z_i = \varphi^{-1}\left(F_{100}(t_i)\right)$, where F_v denotes the cumulative distribution function (cdf) of a t-distribution with v degrees of freedom and φ denotes the cdf of a standard normal distribution.

Then the distribution of z_i is standard normal for those genes that have the same mean expression for both groups of subjects. Efron then fits the mixture $f = (1-p) f_0 + p f_1$ as follows. Suppose f is a 7-parameter exponential family and estimate this density from the z-values, obtaining $\hat{f}$. Suppose f_0 is the standard normal density and estimate p by using $\log\left((1-p) f_0\right)$ as a "quadratic approximation" of $\hat{f}$. From this he estimates the assignment function A_0 (false Discovery rate). These calculations can be done using the R package locfdr. He discovered 51 genes using false discovery rate, declaring an anomaly when $A_0 < 0.2$.

Another approach for this data is to work directly with the t-values. Let μ_{i1} and μ_{i2} be the mean expressions of gene i for tumor and non-tumor subjects respectively. Suppose the variance of the gene-expression for gene i is σ_i^2 is the same for both groups. Then t_i follows a central t-distribution if $\mu_{i1} = \mu_{i2}$. When the means are diferent t_i follows a non-central t-distribution with non-centrality parameter $\delta_i = \frac{\mu_{i1} - \mu_{i2}}{\sqrt{\sigma_i^2\left(\frac{1}{m_1} + \frac{1}{m_2}\right)}}$. In either case the degrees of freedom are $m_1 + m_2 - 2$.

Assume that all non-centrality parameters have the same magnitude, i.e.,$|\delta_i| = \delta$. For simplicity also assume that half of the non-centrality parameters are positive and half are negative. If the proportion of t-values that follow a non-central t-distribution is p, then the distribution of each t_i is $f = (1-p) f_0 + p f_1$ where f_0 denotes the density of central t-distribution and $f_1 = .5 g_\delta + .5 g_{-\delta}$ denotes the density of the genes with different mean expression and g_δ denotes a t-distribution with non-centrality parameter δ. We shall comment on this choice of f_1 at the end of this section. This model only has two parameters to estimate p and δ. Maximum likelihood estimation (mle) could be accomplished using the Expectation-Maximization (EM) algorithm, however it is hard to work with the density of the t-distribution. Instead, we use a modified version of the EM-algorithm with an ad-hoc M-step. The idea is that for each gene, the mle of δ_i is simply t_i by the invariance property of the mle (replace μ_{i1} and μ_{i2} with their mle's, the mle σ_i^2 is approximated by its unbiased version, the pooled variance estimate). One ad-hoc estimate of δ could be the average of $|t_i|$. A better estimate uses the weighted average $\sum_i w_i |t_i| / \sum_i w_i$ where w_i is the posterior probability of coming from f_1 given t_i. For the M-step this weighted average was used, calculating the w_i for one iteration using the estimates from the previous iteration as true parameters. For the prostate data, the proportion of t-values with $|t_i| > 2$ and the average of the absolute t-values (i.e., $\left(|t_1| + \ldots + |t_n|\right)/n$) were used as initial values for p and δ, respectively. The iterations were stopped when the change in both δ and p was no greater then 10^{-8}. Convergence was attained in 174 iterations, giving $\hat{\delta} = 2.473228$ and $\hat{p} = 0{:}04612997$.

FIGURE 3

Asymptotic distribution theory for contamination models

Histogram of *t* - values corresponding to the prostate gene-expression data

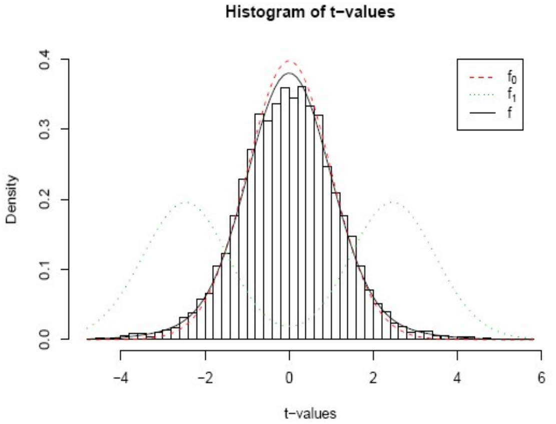


Figure 3 shows a histogram of the *t*-values with the estimated densities superimposed. It can be seen that the central t-distribution f_0 (dashed line) does not fit the *t*-values very well since the t-values have a heavier tail. The mixture of the two non-central t-distributions with parameters δ and $-\delta$, f_1 (dotted line), help to explain the tails. When these two distributions f_0 and f_1 are mixed with p as the proportion for f_1, then the fitted distribution f (solid line) fits the histogram quite nicely using only two parameters (compared to fitting 8 parameters). To do the LMP test, we need to explore the distribution of the density ratio f_1/f_0, and this is quite hard to do with the non-central t-distribution. To work around this, suppose that the variance of the ratio is finite and just use the regular central limit theorem. A random sample of one million values from a central t-distribution with 100 degrees of freedom was generated and the ratio f_1/f_0 was calculated for each value, where the sample variance was 102.2010. This estimate of the variance of f_1/f_0 is assumed to be the true variance. For the prostate data $T_n = \sum_i \frac{f_1(t_i) - f_0(t_i)}{f_0(t_i)} = 27186.5$ and assume that the variance of f_1/f_0 is 102.201. Then, if all observations are from f_0 and the observations were independent then $T_n / \sqrt{n\sigma^2} = 34.62255$. When compared to the quantiles of a standard normal distribution, this value indicates very strong evidence that some of the genes have different mean expression values for tumor and non-tumor patients. Next, we created the sets D_i, as in (2.2), and calculated their empirical posterior probabilities using an expression similar to (2.1), i.e., $P\left(D_i \middle| t_1, \ldots, t_n\right) \propto \left(1 - p^*\right)^{n-i} p^{*i} \prod_{j \in D_i^c} f_0\left(t_j\right) \prod_{j \in D_i} f_1\left(t_j\right)$. Figure 4 shows the log of these posterior probabilities, giving a maximum at 139, indicating that 139 genes have significant difference in their expression number. Since Efron (2007) found 51 anomalies, the LMP test is used to verify the hypothesis $H_0 : p = p_0 = 51/6033 = 0.00845$. In this case the test statistic is

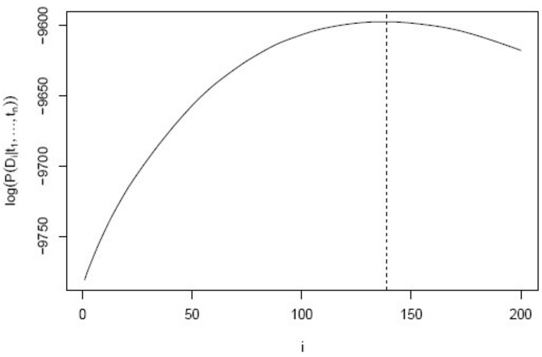
$$T_n = \sum_{i=1}^{n} \frac{f_1 - f_0}{\left(1 - p_0\right) f_0 + p_0 f_1} = 5081.583$$

Since T_n is the sum of bounded r.v.'s (Proposition 4.1), the regular central limit theorem can be used to decide on a rejection region. As before, we simulated one million samples from a central *t* and another million from the noncentral t with noncentrality parameter $\delta = 2.473228$ (using $-\delta$ gives the same variance). Then we calculated the average of these two, weighted by $1 - p_0$ and p_0 respectively. This gives a sample variance for $\left(f_1 - f_0\right)/\left(\left(1 - p_0\right) f_0 + p_0 f_1\right)$ of $\sigma^2 = 18.20424$. So, $T_n/\sqrt{n\sigma^2} = 15.33367$, which is quite significant when compared to quantiles of a standard normal distribution. Therefore, we concluded that the proportion of anomalies is greater than 0.00845.

FIGURE 4

Asymptotic distribution theory for contamination models

Log-probability of the number of "anomalies" corresponding to the prostate gene-expression data



Since the use of *t*-values gave 139 anomalies, we now repeat the exercise from the previous paragraph to test the hypothesis $H_0 : p = p_0 = 139/6033 = 0.02304$. The estimated variance of $\left(f_1 - f_0\right)/\left(\left(1 - p_0\right) f_0 + p_0 f_1\right)$ is 9.111389. The test statistic is T_n = 1779.638, which normalized gives $T_n/\sqrt{n\sigma^2} = 7.590539$. This indicates that the proportion of anomalies is greater than 0.02304. This suggests a better method to identity anomalies is needed, possibly one based on a cutoff for $\left(f_1 - f_0\right)/f_0$.

It is worthwhile to mention that the independence assumption between genes may not be realistic. With regard to the choice of $f_1 = .5g_\delta + .5g_{-\delta}$, note that for this data one is only interested in determining what genes have different expression numbers and not the direction of the difference. Thus, one could consider $\left|t_i\right|$. For this symmetric situation f_1 seems appropriate to model the anomalies and what we would recommend for future data analysis. This Choice worked well for the original data set, but the biological justification for equal proportions of positive or negatively expressed genes is not clear to us.

Appendices

The first appendix simply states some stable distribution results used in Section 4, while the second appendix covers the asymptotics for the standard case when the parameters of a mixture model are in the interior of the parameter space.

A Generalized central limit theorems

The lemmas in this section are well known results which give us the limiting distributions of $\sum Z_i$, properly normalized (see Geluk and de Haan, 2000, and the references therein). A stable distribution with parameter α, $0 < \alpha \le 2$, is defined by its characteristic function $\varphi_\alpha\left(t\right)$ given by

$$\varphi_\alpha\left(t\right) = \exp\left\{-t^\alpha\left[1 + i\operatorname{sign}(t)\tan\left(\frac{\alpha\pi}{2}\right)\left(\left|t\right|^{1-\alpha} - 1\right)\right]\right\} + it\tan\left(\frac{\alpha\pi}{2}\right) \tag{A.1a}$$

when $\alpha \ne 1$ and by

$$\varphi_1\left(t\right) = \exp\left\{-t\left[1 + i\operatorname{sign}(t)\frac{2}{\pi}\log\left|t\right|\right] + i2t\Gamma'(1)/\pi\right\} \tag{A.1b}$$

when $\alpha = 1$. $\Gamma'\left(1\right)$ is the derivative of the gamma function evaluated at 1 (the negative of Euler's constant). For $\alpha = 2$ the stable distribution becomes a normal distribution with variance 2.

To calculate quantiles and probabilities for this distribution one could use the R functions qstable and pstable (from package fBasics) setting the parameters as alpha = α, beta = 1, gamma = 1, delta = $\tan\left(\alpha\pi/2\right)$ and pm = 0. In the case of $\alpha = 1$ the parameter delta should be set to $2\Gamma'\left(1\right)/\pi \approx -0.3674669$. Table IV shows some quantiles for selected values of α.

Let $Z_1, \ldots, Z_n$ be iid with $P\left(Z_1 > a\right) = 1$ for some finite *a*. We say that Z_1 is in the domain of attraction of an α-stable law, denoted by $Z_1 \in D_\alpha$, if there exist real sequences a_n and μ_n for which

$$\frac{1}{a_n}\sum_{i=1}^{n}\left(Z_i - \mu_n\right)$$

converges in law to a stable distribution with parameter α. The following two lemmas give necessary and sufficient conditions for $Z_1 \in D_\alpha$.

TABLE IV
Asymptotic distribution theory for contamination models
Quantiles of the stable distribution

	Right tail probabilities				
α	0.1	0.05	0.01	0.005	0.001
1	6.7612	13.6373	65.653	129.7645	NA
1.05	-6.3069	-0.5517	40.1627	87.9868	445.4734
1.1	-0.5213	4.355	36.8106	73.2907	330.9228
1.15	1.1151	5.2897	31.5672	59.932	250.189
1.2	1.7651	5.3706	26.929	49.3496	192.8381
1.25	2.0504	5.1874	23.0733	41.0467	151.2448
1.3	2.1718	4.9181	19.8967	34.4791	120.4426
1.35	2.2118	4.628	17.2687	29.2202	97.1814
1.4	2.2089	4.3431	15.0761	24.9537	79.2942
1.45	2.1832	4.074	13.2282	21.4463	65.3058
1.5	2.1457	3.8242	11.6541	18.5251	54.1916
1.55	2.1028	3.5946	10.2983	16.0605	45.2258
1.6	2.0584	3.3845	9.1171	13.9535	37.8839
1.65	2.0147	3.1931	8.0759	12.1273	31.7795
1.7	1.9733	3.0195	7.1466	10.5209	26.6205
1.75	1.9352	2.8631	6.3068	9.0842	22.179
1.8	1.9011	2.7234	5.5394	7.774	18.2669
1.85	1.8716	2.6002	4.8357	6.5525	14.7128
1.9	1.8468	2.4931	4.2054	5.3942	11.3245
1.95	1.827	2.402	3.6825	4.361	7.787
2	1.8124	2.3262	3.29	3.6428	4.3702

Lemma A.1. For $0 < \alpha < 2$, $Z_1 \in D_\alpha$ if and only if $R(z) = P(Z_1 > z)$ is regularly varying with index $-\alpha$, i.e.,

$$\lim_{t \to \infty} \frac{R(tz)}{R(t)} = z^{-\alpha} \qquad \text{(A.2)}$$

What are suitable choices for μ_n and a_n? When $\alpha > 1$, the regular variation condition ensures that $E(Z_1)$ exists and is finite, so an appropriate choice of μ_n when $1 < \alpha < 2$ is $\mu_n = E(Z_1)$. When $0 < \alpha < 1$ a fitting option is $\mu_n = 0$. For $\alpha = 1$ a suitable centering sequence is

$$\mu_n = \int_{\min(a,0)}^{0} P(Z_1 \le t)\,dt + \int_{\max(0,a)}^{a_n} P(Z_1 > t)\,dt\,. \qquad \text{(A.3)}$$

For an appropriate a_n, let $s_\alpha = \Gamma(1-\alpha)\cos(\alpha\pi/2)$ for $0 < \alpha < 1$ and $s_\alpha = \dfrac{\pi}{2\Gamma(\alpha)\sin(\alpha\pi/2)}$ for $1 \le \alpha < 2$. A fitting sequence a_n of normalizing constants would satisfy the condition

$$\lim_{n \to \infty} n s_\alpha P(Z_1 > a_n) = 1 \qquad \text{(A.4)}$$

Table V shows suitable choices for a_n given certain tail behaviors. Condition A.4 also makes a_n an appropriate normalizing constant for $M_n = \max\{Z_1,\dots,Z_n\}$ since

$$P(M_n / a_n \le m) \to e^{-m^{-\alpha}/s_\alpha} \qquad \text{(A.5)}$$

i.e., the Fréchet distribution.

TABLE V
Asymptotic distribution theory for contamination models
Normalizing constants for some specific tail behaviors
$(0 < c < \infty, b > 0, 0 < \alpha < 2)$

Tail behavior as $z \to \infty$	Normalizing constant
$z^\alpha P(Z_1 > z) \to c$	$a_n = (c s_\alpha n)^{1/\alpha}$
$z^\alpha \sqrt{2\alpha \log z}\, P(Z_1 > z) \to c$	$a_n = \left(\frac{c s_\alpha n}{\sqrt{2 \log n}}\right)^{1/\alpha}$
$z^\alpha \sqrt{2\alpha \log z} \exp\left(-b\sqrt{2\alpha \log z}\right) P(Z_1 > z) \to c$	$a_n = \left(\frac{c s_\alpha n}{\sqrt{2 \log n}} \exp\left[b^2 + b\sqrt{2 \log n}\right]\right)^{1/\alpha}$

Corollary A.2. If $Z_1 \in D_\alpha$ for $1 < \alpha < 2$, then

$$\left(\frac{s_\alpha}{s_{\alpha/2}}\right)^{2/\alpha} \frac{1}{a_n^2} \sum_{i=1}^{n} Z_i^2$$

converges in distribution to a stable law with parameter $\alpha/2$.

For $\alpha/2$, the normal case, let $\mu = E(Z_1)$ and $h(t) = \int_\mu^t (z-\mu) P(Z_1 > z)\,dz - \int_a^\mu (z-\mu) P(Z_1 \le z)\,dz$ and let a_n be a sequence of real numbers satisfying

$$\lim_{n \to \infty} \frac{n h(a_n)}{a_n^2} = 1 \qquad \text{(A.6)}$$

Note that if the variance of Z_1 exists and is finite then $h(t) \to \operatorname{var}(Z_1)/2$ as $t \to \infty$, in which case $a_n = \sqrt{.5n\operatorname{var}(Z_1)}$.

Lemma A.3. $h(t)$ is slowly varying, i.e., regularly varying of order 0, (therefore $\mu = E(Z_1) < \infty$) if and only if

$$\frac{1}{a_n}\sum_{i=1}^{n}\left(Z_i - \mu\right)$$

converges in law to a normal distribution with mean 0 and variance 2.

If $h(t)$ is slowly varying, we say that Z_1 is in the domain of attraction of a 2-stable law (normal distribution with variance 2). Table 6 gives suitable choices for certain tail behaviors with slowly varying $h(t)$.

TABLE VI

Asymptotic distribution theory for contamination models

Normalizing constants for some specific tail behaviors of distributions in the domain of attraction of the normal law

$\left(0 < c < \infty, b > 0\right)$

Tail behavior as $z \to \infty$	Normalizing constant
$z^2 P(Z_1 > z) \to c$	$a_n = \sqrt{.5cn\log n}$
$2z^2\sqrt{\log z}P(Z_1 > z) \to c$	$a_n = \sqrt{cn\sqrt{.5\log n}}$
$2z^2\sqrt{\log z}\exp\left(-2b\sqrt{\log z}\right)P(Z_1 > z) \to c$	$a_n = \sqrt{\frac{c}{2b}n\exp\left[b^2 + 2b\sqrt{.5\log n}\right]}$

B Asymptotics for the MLE of a *k*-point mixture

The next lemma states the asymptotic distribution for the *k*-point MLE of the mixing distribution *Q* when *Q* is a discrete probability measure on Θ with k distinct mass points, $\theta_1, \ldots, \theta_k$, and respective masses, $p_1, \ldots, p_k$. Here we assume that $\theta_1, \ldots, \theta_k$ are in the interior of Θ and that all the masses are positive and less than 1. The following notation will be used. Let $\underline{\theta} = \left(\theta_1, \ldots, \theta_k\right)$, $\underline{p} = \left(p_1, \ldots, p_{k-1}\right)$ and $\underline{\eta} = \left(\eta_1, \ldots, \eta_{2k-1}\right) = \left(\theta_1, \ldots, \theta_k, p_1, \ldots, p_{k-1}\right)$. Then,

$$f\left(x;\underline{\eta}\right) = f\left(x;\underline{\theta}, p\right) \equiv f_Q(x) = \sum_{i=1}^{k} p_i f_{\theta_i}(x)$$

$$= f_{\theta_k}(x) + \sum_{i=1}^{k-1} p_i\left(f_{\theta_i}(x) - f_{\theta_k}(x)\right)$$

Let $X_1, \ldots, X_n$, be iid from $f\left(x;\underline{\eta}\right)$ and let $\hat{\underline{\eta}}$ denote the mle based on $X_1, \ldots, X_n$. For $i, j = 1, \ldots, 2k-1$ let

$$I_{ij}\left(\underline{\eta}\right) = \operatorname{cov}_{\underline{\eta}}\left(\frac{\partial}{\partial\eta_i}\log f\left(X_1;\underline{\eta}\right), \frac{\partial}{\partial\eta_j}\log f\left(X_1;\underline{\eta}\right)\right)$$

and let $I\left(\underline{\eta}\right)$ denote the information matrix whose $i^{th} - j^{th}$ entry is $I_{ij}\left(\underline{\eta}\right)$.

Lemma B.1. Under suitable regularity conditions,

$$\sqrt{n}\left(\hat{\underline{\eta}}_n - \underline{\eta}\right) \xrightarrow{D} MN\left(\underline{0}, I^{-1}\left(\underline{\eta}\right)\right)$$

Proof. See Lehmann (1983, Section 6.4) regarding suitable regularity conditions and a proof under those conditions.

Remark B.2. The "suitable regularity conditions" alluded to in Lemma B.1 involve the usual differentiability assumptions on $f\left(x;\underline{v}\right)$ and passing derivatives through expectations as well as the usual assumptions of positive definiteness of the information matrix. These hold for the examples we consider here. The main regularity condition that we need to verify is that

$$\left|\frac{\partial^3 \log f\left(x;\underline{v}\right)}{\partial v_a \partial v_b \partial v_c}\right| \le M_{abc}(x); a, b, c = 1, 2, \ldots, 2k-1$$

where $E\left(M_{abc}(X_1)\right) < \infty$

For mixtures these third order derivatives have been derived in a separate document (See http://people.clemson.edu/～veraf/docs/ThirdDerivativeEquations.pdf) If $p_i > 0$ for $i = 1, \ldots, k$, then some of the quantities envolved in these third order partial derivatives are bounded (see Proposition B.3 below). Therefore, a suficient regularity condition is that the absolute value of functions such as

$$\frac{\frac{\partial^3}{\partial\theta_i^3}f_{\theta_i}(x)}{f_{\theta_i}(x)}, \frac{\frac{\partial^2}{\partial\theta_i^2}f_{\theta_i}(x)}{f_{\theta_i}(x)}, \frac{\frac{\partial}{\partial\theta_j}f_{\theta_j}(x)}{f_{\theta_j}(x)},$$

$$\frac{\frac{\partial}{\partial\theta_i}f_{\theta_i}(x)}{f_{\theta_i}(x)}\frac{\frac{\partial}{\partial\theta_j}f_{\theta_j}(x)}{f_{\theta_j}(x)}\frac{\frac{\partial}{\partial\theta_h}f_{\theta_h}(x)}{f_{\theta_h}(x)}, \frac{\frac{\partial^2}{\partial\theta_i^2}f_{\theta_i}(x)}{f_{\theta_i}(x)},$$

$\frac{\frac{\partial}{\partial\theta_i}f_{\theta_i}(x)}{f_{\theta_i}(x)}\frac{\frac{\partial}{\partial\theta_j}f_{\theta_j}(x)}{f_{\theta_j}(x)}, \frac{\frac{\partial}{\partial\theta_i}f_{\theta_i}(x)}{f_{\theta_i}(x)}$ are bounded by functions of *X* with finite expectation.

The next proposition shows that the derivative of $\log f\left(x;\underline{\theta},\underline{p}\right)$ with respect to p_i is bounded.

Proposition B.3.

$$\frac{f_{\theta_i}(x)-f_{\theta_k}(x)}{p_1 f_{\theta_1}(x)+\dots+p_k f_{\theta_k}(x)}, i=1,\dots,k-1$$

is bounded if $p_j>0$ for $j=1,\dots,k$ (hence all its moments are finite).

Proof. Notice that

$$\frac{f_{\theta_i}(x)-f_{\theta_k}(x)}{p_1 f_{\theta_1}(x)+\dots+p_k f_{\theta_k}(x)}=\frac{1}{p_i}\left(\frac{f_{\theta_i}(x)}{f_{\theta_i}(x)+\sum_{j\neq i}\frac{p_j}{p_i}f_{\theta_j}(x)}\right)$$
$$-\frac{1}{p_k}\left(\frac{f_{\theta_k}(x)}{f_{\theta_k}(x)+\sum_{j\neq k}\frac{p_j}{p_k}f_{\theta_j}(x)}\right)$$

Therefore,

$$\left|\frac{f_{\theta_i}(x)-f_{\theta_k}(x)}{p_1 f_{\theta_1}(x)+\dots+p_k f_{\theta_k}(x)}\right|\leq\frac{1}{p_i}+\frac{1}{p_k}$$

The essence of the proof of Lemma B.1 is a consequence of the following two Lemmas.

Lemma B.4.

$$\sqrt{n}\left(\underline{\eta}-\hat{\underline{\eta}}_n\right)=\frac{1}{\sqrt{n}}J_n\left(\underline{\eta}\right)\left(\frac{H_n\left(\underline{\eta}\right)}{n}\right)^{-1}+R_n$$

Where

$$R_n\xrightarrow{P}0, J_n\left(\underline{\eta}\right)=\left[\sum_{i=1}^{n}\frac{\partial\log f\left(X_i;\underline{\eta}\right)}{\partial\eta_j}:j=1,2,\dots,2k-1\right]$$

, and

$$H_n\left(\underline{\eta}\right)=\left[\sum_{i=1}^{n}\frac{\partial^2\log f\left(X_i;\underline{\eta}\right)}{\partial\eta_j\partial\eta_l}:j,l=1,2,\dots,2k-1\right]$$

Since $\frac{H_n\left(\underline{\eta}\right)}{n}\xrightarrow{P}-I\left(\underline{\eta}\right)$ it is immediate that.

Lemma B.5.

$$\sqrt{n}\left(\hat{\underline{\eta}}_n-\underline{\eta}\right)=\frac{1}{\sqrt{n}}J_n\left(\underline{\eta}\right)\left(I\left(\underline{\eta}\right)\right)^{-1}+Q_n$$

Where $Q_n\xrightarrow{P}0$.

This representation for $\hat{\underline{\eta}}_n$ will be needed in the final lemma.

Remark B.6. Note that if $n^{-.5}J_n\left(\underline{\eta}\right)$ converges in law to a multivariate normal distribution then $n^{-q}J_n\left(\underline{\eta}\right)$ converges in probability to 0 for $q>.5$, but convergence may be slow.

We now determine the asymptotic distribution of the LMP test statistic when parameters in the statistic are estimated by mle's. To do this we adopt the following notation. Let $\underline{\theta}=\left(\underline{\theta}_0,\underline{\theta}_1\right)$ where $\underline{\theta}_0=(\theta_{01},\dots,\theta_{0l})$ indicates the parameters governing the background and $\underline{\theta}_1=(\theta_{11},\dots,\theta_{1m})$ denotes the parameters activating the spurious observations. The vector $\underline{p}=\left(\underline{p}_0,\underline{p}_1\right)$ denotes the vector of mixing proportions and $p=\sum_{j=1}^{m}p_{1j}$ denotes the proportion assigned to the spurious distributions. Then $Q=\left(1-p\right)Q_0+pQ_1$. The *k*-point mle of Q is $\hat{Q}=(1-\hat{p})\hat{Q}_0+\hat{p}\hat{Q}_0$ where $\hat{Q}_0$ and $\hat{Q}_1$ are the discrete probability measures putting masses at the mle's of $\underline{\theta}=\left(\underline{\theta}_0,\underline{\theta}_1\right)$ and the respective vector of mle masses are $\hat{\underline{m}}=\left(\frac{\hat{\underline{p}}_0}{1-\hat{p}},\frac{\hat{\underline{p}}_1}{\hat{p}}\right)$. If Q_0 and Q_1 were known, then the LMP test statistic for testing $H_0:p=0$ versus $H_1:p>0$ would be given by

$$T_n=\sum_{i=1}^{n}\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_i\right)-1\right)\equiv\sum_{i=1}^{n}R\left(\underline{v}\right)\left(X_i\right)$$

where $\underline{v}=\left(\underline{\theta}_0,\underline{\theta}_1,\underline{p}_0,\left(p_{11},\dots,p_{1m-1}\right)\right)$. This suggest using

$$S_n=\sum_{i=1}^{n}\left(\frac{f_{\hat{Q}_1}}{f_{\hat{Q}_0}}\left(X_i\right)-1\right)=\sum_{i=1}^{n}R\left(\hat{\underline{v}}\right)\left(X_i\right)$$

as the test statistic for the more general formulation of the problem. For each *x*, let $G\left(\underline{v}\right)(x)$ and $H\left(\underline{v}\right)(x)$ denote the gradient vector and Hessian matrix of $R\left(\underline{v}\right)(x)$ at $\underline{v}$ and assume that the entries in the Hessian matrix satisfies $\left|H_{ab}\left(\underline{\eta}'\right)(x)\right|\leq N_{ab}(x)$ for $a,b=1,2,\dots,2k-1$ and for all $\underline{\eta}'$ in an open neighborhood of $\underline{\eta}$. Then a second order Taylor's series expansion gives

$$R\left(\hat{\underline{v}}\right)(x)=R\left(\underline{v}\right)(x)+G\left(\underline{v}\right)(x)\bullet\left(\hat{\underline{v}}-\underline{v}\right)^t$$
$$+\frac{1}{2}\sum_{a,b}\left(\hat{v}_a-v_a\right)\left(\hat{v}_b-v_b\right)\gamma_{ab}\left(x\right)N_{ab}\left(x\right)$$

where $\left|\gamma_{ab}\left(x\right)\right|\leq1$ and • is the inner product. So,

$$\frac{1}{\sqrt{n}}\left(S_n - T_n\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{R(\underline{\hat{v}})(X_i) - R(\underline{v})(X_i)\}$$

$$= \frac{1}{n}\sum_{i=1}^{n} G(\underline{v})(X_i)\bullet\sqrt{n}(\underline{\hat{v}}-\underline{v})^t$$

$$+\frac{1}{2}\sum_{a,b}\sqrt{n}\left(\hat{v}_a - v_a\right)\left(\hat{v}_b - v_b\right)\frac{1}{n}\sum_{i=1}^{n}\gamma_{ab}(X_i)N_{ab}(X_i)$$

$$\equiv G_n + R_n$$

If $E\left|G_a\left(\underline{v}\right)\left(X_i\right)\right|$ and $E\left|N_{ab}\left(X_i\right)\right|$ are finite for all *a* and *b*, then

$$\frac{1}{\sqrt{n}}S_n = \frac{1}{\sqrt{n}}T_n + E\left(G(\underline{v})(X_1)\right)\bullet\sqrt{n}\left(\underline{\hat{v}} - \underline{v}\right)^t + R_n$$

where $R_n \xrightarrow{P} 0$. Since $\underline{\hat{v}}$ is just a relabeling of $\underline{\hat{\eta}}$, it follows from Corollary B.5 that

$$\frac{1}{\sqrt{n}}S_n = \frac{1}{\sqrt{n}}T_n + E\left(G(\underline{v})(X_1)\right)\bullet\left(\frac{1}{\sqrt{n}}J_n\left(\underline{v}\right)\left(I(\underline{v})\right)^{-1}\right)^t + R'_n \tag{B.1}$$

where $R'_n \xrightarrow{P} 0$.

Next, we develop the distribution theory for S_n assuming that $I_0 < \infty$ and $W_0 < 1$. Recall that

$T_n = \sum_{i=1}^{n}\frac{f_{Q_1}}{f_{Q_0}}\left(X_i\right)$ and

$J_n\left(\underline{\eta}\right) = \left[\sum_{i=1}^{n}\frac{\partial\log f\left(X_i;\underline{\eta}\right)}{\partial\eta_j}: j=1,2,...,2k-1\right]$ where

$E_{\underline{v}}\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_1\right)\right) \equiv \mu\left(\underline{v}\right)$ and

$E_{\underline{v}}\left(\frac{\partial\log f\left(X_1;\underline{v}\right)}{\partial v_j}=0\right)$ for $j = 1,2,...,2k-1$. Let

$$c_{0,0}\left(\underline{v}\right) = \text{var}_{\underline{v}}\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_1\right)\right),$$

$$c_{0,j}\left(\underline{v}\right) = c_{j,0}\left(\underline{v}\right) = \text{cov}\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_1\right), \frac{\partial\log f\left(X_1;\underline{v}\right)}{\partial v_j}\right)$$

for $j = 1,2,...,2k-1$ and $c_{i,j}\left(\underline{v}\right) = I_{i,j}\left(\underline{v}\right)$ for $i,j \neq 0$. The matrix $C\left(\underline{v}\right)$ of the $c_{i,j}\left(\underline{v}\right)$'s is the covariance matrix of the row vector $\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_1\right), \frac{\partial\log f\left(X_1;\underline{v}\right)}{\partial v_j}; j=1,2,...,2k-1\right) \equiv \left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_1\right), \underline{V}\right)$. From (B.1) we need the covariance matrix of $\left(\frac{f_{Q_1}}{f_{Q_0}}\left(X_i\right), \underline{V}I^{-1}\left(\underline{v}\right)\right)$. A straight-forward calculation shows that this matrix is $\sum\left(\underline{v}\right)$ of $\sigma_{i,j}\left(\underline{v}\right)$'s where $\sigma_{0,0}\left(\underline{v}\right) = c_{0,0}\left(\underline{v}\right), \sigma_{i,j}\left(\underline{v}\right) = I_{i,j}^{-1}\left(\underline{v}\right)$ for $i,j \neq 0$, and the row vector $\underline{\sigma}_0 = \left(\sigma_{0,j}(\underline{v}); j=1,2,...,2k-1\right) = \left(c_{0,j}(\underline{v}); j=1,2,...,2k-1\right)I^{-1}\left(\underline{v}\right)$ determines $\sigma_{0,j}\left(\underline{v}\right) = \sigma_{j,0}\left(\underline{v}\right)$ for $j = 1,2,...,2k-1$. Thus, by the multivariate central limit theorem,

Lemma B.7.

$$\frac{1}{\sqrt{n}}\left(T_n - n\mu\left(\underline{v}\right), J_n\left(\underline{v}\right)I^{-1}\left(\underline{v}\right)\right) \xrightarrow{D} MN\left(\underline{0}, \sum\left(\underline{v}\right)\right)$$

.

The asymptotic distribution of S_n is immediate from Lemma B.7 and identity (B.1).

Lemma B.8.

Let $\underline{a} = \left(1, E\left(G(\underline{v})(X_1)\right)\right)$. Then,

$$\frac{1}{\sqrt{n}}\left(S_n - n\mu(\underline{v})\right) \xrightarrow{D} N\left(0, \underline{a}\sum(\underline{v})\underline{a}^t\right)$$

Next, we develop the distribution theory for S_n when $Z_1 = \left(f_{Q_1}(X_1) - f_{Q_0}(X_1)\right)/f_{Q_0}\left(X_1\right)$ has var $\left(Z_1\right) = \infty$. We assume that Z_1 is in the domain of attraction of an α-stable law for some $0 < \alpha \leq 2$ under $\underline{v}$. Let a_n be defined as in Appendix A. Note that $n/a_n^2 \to 0$ since var$\left(Z_1\right) = \infty$.

Identity (B.1) can be rewritten as

$$\frac{S_n}{a_n} = \frac{T_n}{a_n} + \frac{\sqrt{n}}{a_n}E\left(G(\underline{v})(X_1)\right)\bullet\left(\frac{1}{\sqrt{n}}J_n(\underline{v})\left(I(\underline{v})\right)^{-1}\right)^t + R'_n$$

For the next lemma, let $b_n = nE_p\left(Z_1\right)$ if $1 < \alpha \leq 2$, $b_n = \mu_n$ (as in (A.3)) if $\alpha = 1$, and $b_n = 0$ if $0 < \alpha < 1$.

Lemma B.9. If Z_1 is in the domain of attraction of an α-stable law, then $\frac{S_n - b_n}{a_n}$ converges in distribution to a stable law with parameter α.

Proof. The result follows since $\sqrt{n}/a_n \to 0$.

We now close Appendix B by considering the consistency of the mle at Q_0. To do this, we

consider the parameter space *P* of all discrete probability measures on Θ with at most *k* mass points. Endow *P* with the Levy metric and note that with this metric *P* is closed. Also assume that $f\left(x,\theta\right)$ is continuous in θ. Thus, $f_Q\left(x\right)$ is continuous in Q in the Levy metric. With this framework, one can apply Wald (1949) proof with minor modifications (in particular, see Section 4 of that paper) to show that the *k*-point mle converges almost surely to Q_0 in the Levy metric, d_L, when Q_0 obtains. Wald's parameter space (denoted by Ω in his paper while points in Ω are denoted there by θ not ω) is a subset of a Cartesian product space but his proof holds for the parameter space *P* considered here under his Assumptions 1, 2,
4-6 with his θ's replaced by our Q's.

Regarding Wald's assumptions with regard to the problem here, note that his Assumptions 3 and 8 hold since $f\left(x,\theta\right)$ is continuous in θ (supremums of lower semi-continuous functions are lower semicontinuous, and hence, measurable). Also Assumption 7 holds since *P* is closed. Note that the family $\left\{f\left(x,\theta\right)\right\}$ being identifiable does not imply that $\left\{f_Q\left(x\right)\right\}$ is identifable. E.g, mixtures over *p* of *n*-trial binomials with $k > 2n-1$ have an infinite number of representations. The limiting condition in Assumption 5 for $\left\{f\left(x,\theta\right)\right\}$ implies that $f\left(x,\theta\right)$ goes to zero as the points in the support of *Q* goes to infity or minus infity. This limiting condition lets Wald truncate the parameter pace to a bounded set which is then compact because of Assumption 7. Here the limiting condition implies tightness of a subset *P*′ of *P*, and hence, compactness of this subset. The essence of Wald's proof is to use the compactness to get construct a finite open cover, $I_0, I_1, \dots, I_m$ for P'. Without loss of generality, assume $Q_0 \in I_0$ but not in I_j for $j = 1, 2, \dots, m$. This cover, defined for $\rho > 0$; is done in such a way that

1. $d_L\left(Q_1,Q_2\right) < \rho$ whenever Q_1 and Q_2 are in the same I_j,

2.
$$E_{Q_0}\sup\left\{\log f_Q(x):Q\in I_0\right\} > E_{Q_0}\sup\left\{\log f_Q(x):Q\in I_j\right\}$$
for $j > 0$,

3.
$$E_{Q_0}\sup\left\{\log f_Q(x):Q\in I_0\right\}$$
$$> E_{Q_0}\sup\left\{\log f_Q(x):\text{ for } Q \text{ouside the cover}\right\}$$

for *Q* outside the cover g.

It follows from the strong law and (1-3) that $P\left(\lim d\left(\hat{Q}_n, Q_0\right) < \rho\right) = 1$ where $\hat{Q}_n$ is the *k*-point mle. Since ρ is arbitrary, $\hat{Q}_n$ converges almost surely to Q_0.

7. ACKNOWLEDGEMENTS

This work was initiated when all three authors were involved in the Anomaly Detection Working Group at the Statistical and Applied Mathematical Sciences Institute (SAMSI). The first author was a post-doctoral fellow at SAMSI and the National Institute of Statistical Science (NISS). The authors would like to thank SAMSI and NISS for a stimulating environment and support while there. In particular, support from SAMSI for all three authors was funded by NSF under Grant DMS-0112069 and support from NISS for the first author was funded under grant EIA0131884. In addition, the third author was partially supported by NSF Grants DMS 0243594 and DMS 0805809.

REFERENCES AND ELECTRONIC

[1]. **CHERNOFF, H.** (1954). On the distribution of the likelihood ratio. Annals of Mathematical Statistics 25 (3), 573-578.

[2]. **EFRON, B.** (2007). Size, power and false discovery rates. Annals of Statistics 35 (4), 1351-1377.

[3]. **FERGUSON, T. S.** (1967). Mathematical Statistics: A Decision Theoretical Approach. Academic Press, NY.

[4]. **GELUK, J. L. AND L. DE HAAN** (2000). Stable probability distributions and their domains of attraction: a direct approach. Probability and Mathematical Statistics 20 (1), 169-188.

[5]. **GREGO, J., H.-L. HSI, AND J. D. LYNCH** (1990). A strategy for analyzing mixed and pooled exponentials. Applied Stochastic Models and Data Analysis 6 (1), 59-70.

[6]. **HUANG, K.** (1963). Statistical mechanics. Wiley, NY.

[7]. **LEHMANN, E. L.** (1983). Theory of Point Estimation. Probability and mathematical statistics. Wiley, NY.

[8]. **LINDSAY, B. G.** (1983a). The geometry of mixture likelihoods: A general theory. Annals of Statistics 11 (1), 86-94.

[9]. **LINDSAY, B. G.** (1983b). The geometry of mixture likelihoods, part ii: The exponential family. Annals of Statistics 11 (3), 783-792.

[10]. **SINGH, D., P. G. FEBBO, K. ROSS, D. G. JACKSON, J. MANOLA, C. LADD, P. TAMAYO, A. A. RENSHAW, A. V. D'AMICO, J. P. RICHIE, E. S. LANDER, M. LODA, P. W. KANTOFF, T. R. GOLUB, AND W. R. SELLERS** (2002). Gene expression correlates of clinical prostate cancer behavior. Cancer Cell 1, 203-209.

[11]. **WALD, A.** (1949). Note on the consistency of the maximum likelihood estimate. The Annals of Mathematical Statistics 20 (4), 595-6