Local Mixtures and Exponential Dispersion Models

by

Paul Marriott

Department of Statistics & Applied Probability,

National University of Singapore,

3 Science Drive 2,

Singapore 117543.

email: stapkm@nus.edu.sg

Summary

Exponential dispersion models are powerful tools for modelling. They are highly flexible yet they keep within a well understood inferential framework. This paper looks at mixtures of exponential models, in particular local mixtures. It considers the relationship between mixing and over-dispersion. By using geometric methods a new class of models is developed all of which are identifiable, flexible and interpretable. A powerful large sample inference theory, especially suited to families with boundaries, is developed which extends the good inferential properties of exponential dispersion families to the new class of mixture models.

Some key words: Affine geometry; Convex geometry; Differential geometry; Dispersion model; Mixture model; Multinomial approximation; Statistical manifold.

1 Introduction

Exponential dispersion models have proved to be very successful at increasing modelling flexibility while keeping within a well understood inferential framework. An excellent treatment of their theory and application can be found in Jørgensen (1997). Another way of enriching a simple parametric modelling framework is to consider mixing over parameters. This approach has many applications, see for example see Titterington *et al* (1985), Lindsay (1995) or McLachlan & Peel (2000). Marriott (2002, 2003) considers a restricted form of mixing, called local mixing, which allows a considerable increase in inferential tractability compared to general mixture models. This paper extends the theory of local mixing to general exponential models. It introduces a new class of models called true local mixtures whose members are flexible, identifiable, interpretable. Furthermore the class has a tractable asymptotic inference theory which means that these models are straightforward, if non-standard, inferentially.

This paper considers local mixtures for models of the exponential dispersion form

$$f_Z(z|\theta,\lambda) := \exp[\lambda\{z\theta - \kappa(\theta)\}]\nu_\lambda(z)$$

where $\nu_\lambda(z)$ is independent of θ, see Jørgensen (1997, page 72). However all results can also be easily extended to the additive form of dispersion models

$$\exp\{\theta z - \lambda\kappa(\theta)\}\nu_\lambda(z).$$

Throughout the measure $\nu_\lambda(z)$ is suppressed in the notation to aid clarity except when it becomes important in the analysis.

The paper is organized as follows. Section 2 introduces the idea of a true local mixture model and the examples used throughout. Section 3 discusses identification and parameterisation issues for these models, while §4 considers inference. In particular Theorem 4 describes a large sample asymptotic theory which makes the models inferentially extremely tractable, while simulation studies illustrate the relationship between local mixture models and other methods of dealing with over dispersion. All proofs can be found in the appendix.

2 Local mixture models

Consider the following mixtures

$$\int \exp[\lambda\{(\theta+\eta)z - \kappa(\theta+\eta)\}]dQ(\eta) \tag{1}$$

and

$$\int \exp\left(\lambda\{\theta(\mu+\eta)z - \kappa\{\theta(\mu+\eta)\}\}\right) dQ(\eta), \tag{2}$$

where Q is a distribution function. In (2) the mixing is over the mean value parameterisation where $\mu(\theta)$ is the expected value of Z under the parameters θ, λ.

The essential idea of a local mixture model is that the mixing is only responsible for a relatively small amount of the variation in the model. Loosely it assumes that the mixing distribution $Q(\eta)$ in (1) and (2) is close to a delta function and then applies a Laplace expansion. Using the expansions of Marriott (2002), under regularity, these integrals can be approximated by low dimensional parametric families. For the mixture given by integral (1) this approximation is

$$\begin{aligned} f_Z(z|\theta,\lambda)\Bigg(1 \quad &+ \quad \lambda\xi_1\left\{z - \frac{d}{d\theta}\kappa(\theta)\right\} \\ &+ \quad \lambda\xi_2\left[-\frac{d^2}{d\theta^2}\kappa(\theta) + \lambda z^2 - 2\,\lambda\, z\frac{d}{d\theta}\kappa(\theta) + \lambda\left\{\frac{d}{d\theta}\kappa(\theta)\right\}^2\right]\Bigg), \end{aligned} \tag{3}$$

with a positivity condition which ensures that (3) is a density given by

$${\xi_1}^2 - 4\,\xi_2 + 4\,{\xi_2}^2\lambda\,\frac{d^2}{d\theta^2}\kappa(\theta) < 0.$$

As shown in Marriott (2002) the local mixture expansion is invariant with respect to reparameterisation. Thus the same result as (3) is given by expanding integral (2) in the μ parameterisation. This expansion is denoted by $f_Z(z|\theta(\mu),\lambda,\nu_1,\nu_2)$ which is defined by

$$\begin{aligned} f_Z(z|\theta(\mu),\lambda)\Bigg[1 \quad &+ \quad \nu_1\left\{\frac{z-\mu}{V(\mu)}\right\} \\ &+ \quad \nu_2\left\{\frac{-V(\mu)+z^2-2\,\mu\, z+\mu^2-\frac{d}{d\mu}V(\mu)z+\frac{d}{d\mu}V(\mu)\mu}{V(\mu)^2}\right\}\Bigg] \end{aligned} \tag{4}$$

where $V(\mu)$ is the variance function for the $f_Z(z|\theta(\mu),\lambda)$ family, see Jørgensen (1997, pp48). The relationship between the parameters (ξ_1,ξ_2) and (ν_1,ν_2) can be calculated directly from the chain rule and is given by

$$\nu_1 = \xi_1\frac{\partial\mu}{\partial\theta} + \xi_2\frac{\partial^2\mu}{\partial\theta^2}, \nu_2 = \xi_2\left(\frac{\partial\mu}{\partial\theta}\right)^2.$$

In general there are two issues to consider when using such models. First identification, since it is not clear that the parameters $(\mu,\lambda,\nu_1,\nu_2)$ of (4) uniquely characterize densities. The second issue is the inferential implications of the boundary which are an intrinsic part of such models. This paper treats both these issues.

2.1 True local mixture models

Investigation of the geometry of the local mixture expansion reveals a surprising fact. Expansions of the form (3) and (4) need not themselves be mixture models. If a model is a mixture it must satisfy some natural inequalities in its moment structure. However, as is shown in the examples below, it is possible to find parameter values $(\mu,\lambda,\nu_1,\nu_2)$ which generate a proper density which fails to respect these natural mixture inequalities.

In order to see when the local mixture expansion can be interpreted as a mixture define the subclass of local mixture models, called true local mixture models, for which there exists a mixing distribution, $\tilde{Q}$, such that

$$f_Z\{z|\theta(\mu),\lambda,\nu_1,\nu_2\} = \int f_Z\{z|\theta(\mu+\eta),\lambda\}d\tilde{Q}(\eta). \qquad (5)$$

Note the differences between equation (5) and the expansion given by (4). First, equation (5) is not an asymptotic approximation, rather it is an equation which requires the existence of $\tilde{Q}$. Second, the point $\theta(\mu)$ need not be the mode of $\tilde{Q}$ as in the Laplace based expansion of Marriott (2002). Finally it is convenient to define a subclass of mixing distributions. Define a true local mixture model to be of width 2ϵ if the support of $\tilde{Q}$ lies in $[-\epsilon,\epsilon]$. This is a crude way of ensuring that the mixing distribution is 'small'.

2.2 Examples

In this section the examples are considered which are used throughout.

Example: Normal mixture family.

Consider a mixture over the mean parameter μ in a normal family with fixed variance, known to be 1. That is

$$f_Z(z|\mu, Q) := \int \phi(z|\mu+\eta, 1) dQ(\eta),$$

where Q is a localizing mixture family and $\phi(z|\mu, \sigma^2)$ is the normal density with mean μ and variance σ^2. The corresponding local mixture density is

$$f_Z(z|\mu, 1, \nu_1, \nu_2) = \phi(z|\mu, 1) \left[1 + \nu_1(z-\mu) + \nu_2\{(z-\mu)^2 - 1\}\right],$$

and the positivity constraint which ensures that the expansion is a density is given by

$$\nu_1^2 - 4\nu_2 + 4\nu_2^2 < 0. \tag{6}$$

Under $f_Z(z|\mu, Q)$, Z has a second moment greater than $1 + \{E_Q(\mu+\eta)\}^2$, however the moments of the corresponding local mixture model $f_Z(z|\mu, 1, \nu_1, \nu_2)$ are

$$\begin{aligned} E(Z) &= \mu + \nu_1, \\ E(Z^2) &= 1 + (\mu+\nu_1)^2 + 2\nu_2 - \nu_1^2. \end{aligned}$$

Thus the natural moment structure for mixing implies the inequality

$$2\nu_2 - \nu_1^2 \geq 0. \tag{7}$$

Inspection shows that there are parameter values which satisfy inequality (6) but not (7). Thus a necessary condition for $f_Z(z|\mu, 1, \nu_1, \nu_2)$ to be a true local mixture is that both (6) and (7) hold. It is shown later that this is also a sufficient condition for small enough mixing.

Example: Poisson mixture family.

Consider the family of Poisson distributions parameterised by θ

$$Po(z|\theta) = \frac{1}{z!} \exp\{z\theta - \exp(\theta)\},$$

where $z \in I\!N$. The local mixture family for this case has the form

$$Po(z|\theta)\left\{1+\xi_1\,\left(z-e^{\theta}\right)+\xi_2\,\left(-e^{\theta}+z^2-2\,ze^{\theta}+e^{2\,\theta}\right)\right\},$$

with a positivity condition sufficient to ensure that this a density given by

$${\xi_1}^2+4\,{\xi_2}^2e^{\theta}-4\,\xi_2<0.$$

As with the normal case this condition is not sufficient to ensure that the local mixture is a true local mixture.

Example: Binomial mixture family.

The binomial family,

$$Bi(z|\pi,n)=\frac{n!\,\pi^z\,(1-\pi)^{n-z}}{z!\,(n-z)!},$$

has a local mixture expansion $f_Z(z|\pi,n,\nu_1,\nu_2)$ given by

$$Bi(z|\pi,n)\left\{1+\nu_1\frac{(z-\pi\,n)}{\pi\,(1-\pi)}+\nu_2\frac{(z^2-z+2\,z\pi-2\,z\pi\,n+\pi^2n^2-\pi^2n)}{\pi^2\,(\pi-1)^2}\right\}.$$

The positivity condition checks if there are any integers between 0 and n for which the local mixture expression is negative. Thus there are n linear constraints of the form

$$1+\nu_1\frac{(z-\pi\,n)}{\pi\,(1-\pi)}+\nu_2\frac{(z^2-z+2\,z\pi-2\,z\pi\,n+\pi^2n^2-\pi^2n)}{\pi^2\,(\pi-1)^2}\geq 0$$

for all $z \in \{0,\cdots,n\}$. As with the other examples these conditions do not ensure that the local mixture is a true local mixture. Conditions which do ensure this are given later.

3 Identification and reparametrisation

3.1 Visualising the geometry

The key geometric idea in Marriott (2002) is that a local mixture family is an embedded finite dimensional manifold with boundary. Hence it has good geometric properties which can be exploited for inference. The geometric structure of the family is discovered by considering its embedding in an infinite dimensional affine space $(X_{\text{Mix}}, V_{\text{Mix}})$, where for all sufficiently smooth, square integrable $f(z)$

$$X_{\text{Mix}} = \left\{ f(z) | \int f(z) d\nu = 1 \right\}, V_{\text{Mix}} = \left\{ f(z) | \int f(z) d\nu = 0 \right\},$$

see Marriott (2002) for details. The geometry of affine spaces is simple and tractable, hence all calculations are made relative to this space.

Rather than go through the formal geometric arguments it might be more helpful to consider the model visually. This might seem to be difficult since the affine space $(X_{\text{Mix}}, V_{\text{Mix}})$ is infinite dimensional, but a lot of information can be gathered by taking finite dimensional affine projections of this space. To do this consider the following result.

Theorem 1 *Define, for any integers* (n_1, n_2, n_3) *for which the corresponding integrals converge, the map*

$$\begin{aligned} (X_{\text{Mix}}, V_{\text{Mix}}) &\rightarrow \mathbb{R}^3 \\ f_Z(z) &\mapsto (E_f(Z^{n_1}), E_f(Z^{n_2}), E_f(Z^{n_3})). \end{aligned}$$

This map has the property that finite dimensional affine subspaces in $(X_{\text{Mix}}, V_{\text{Mix}})$ *map to finite dimensional affine subspaces in* $\mathbb{R}^3$.

Proof *See Appendix*

This theorem allows finite dimensional projections of $(X_{\text{Mix}}, V_{\text{Mix}})$ to be taken. These projections respect the geometric structure. If a line is straight in $(X_{\text{Mix}}, V_{\text{Mix}})$ it will automatically be straight in the projection. Furthermore, a point which lies in a convex hull in $(X_{\text{Mix}}, V_{\text{Mix}})$ will lie in the convex hull of the image in $\mathbb{R}^3$. Thus, as far as possible, the visual geometry in the plot respects the geometry in the infinite dimensional space.

Example: Normal mixture family.

Figure 1 shows how the example of local mixtures of the normal family can be visualised. The curve in Fig. 1 is the image of the family $\phi(z|\mu, 1)$ in a three dimensional representation given by the first three non-central moments. The two arrows are the tangent and mixture curvature vectors for this family.

3.2 Orthogonal parameterisations

Theorem 9 of Marriott (2002) shows that any member of the local mixture family can be written as an affine combination of the tangent and second derivative vectors, translated so that the origin is moved to $f_Z(z|\theta(\mu), \lambda)$. It is a member of the affine space

$$A(\mu) := \left\langle \frac{\partial}{\partial \mu} f_Z(z|\theta(\mu), \lambda), \frac{\partial^2}{\partial \mu^2} f_Z(z|\theta(\mu), \lambda) \right\rangle_{f_Z(z|\theta(\mu),\lambda)},$$

where the notation $\langle v_1, v_2 \rangle_x$ denotes the affine space through x spanned by v_1, v_2. The positivity condition for any family defines a convex subset of this affine plane within which lie all the local mixture models which are densities. Define the convex subset of the affine space $A(\mu)$ by

$$F(\mu) := \{ f_Z(z|\mu, \lambda, \nu_1, \nu_2) \in A(\mu) | f_Z(z|\mu, \lambda, \nu_1, \nu_2) \text{ is a density } \}.$$

The full local mixture family is the union of all $F(\mu)$ as μ varies.

Example: Normal mixture family.

For the normal example one of the sets $F(\mu)$ is shown in Fig. 1 and denoted by I. The union of all such families is shown in Fig. 2. It can be seen that local mixtures are represented by the three dimensional interior of a smooth two dimensional surface.

As is seen visually in Fig. 1, and is generally true, the tangent space to $F(\mu)$ and the curve $f_Z(z|\theta(\mu), \lambda)$ share a common tangent direction, $\frac{\partial}{\partial \mu} f_Z(z|\theta(\mu), \lambda)$. This is both geometrically and statistically inconvenient since it introduces problems with identification and produces a singularity in the parameterisation. Hence it is natural to reparametrise the model $f_Z(z|\mu, \lambda, \nu_1, \nu_2)$ by $(\mu', \lambda', \nu_1', \nu_2')$ so that: (i) $\nu_1' = \nu_2' = 0$ corresponds to the family $f_Z(z|\mu, \lambda)$ in the natural way, and (ii) that for fixed μ' and λ' the two-dimensional sub-manifold defined by $\{f_Z(z|\mu',\lambda',\nu_1',\nu_2')\}$ is orthogonal to $\frac{\partial}{\partial\mu}f_Z(z|\mu,\lambda,0,0)$.

For any exponential dispersion family the directions $\frac{\partial}{\partial\eta}$ and $\frac{\partial}{\partial\mu}$ are Fisher orthogonal if

$$\int \frac{(z-\mu)}{V(\mu)}\frac{\partial f_Z}{\partial \eta}\{z|\mu(\eta),\lambda,\nu_1(\eta),\nu_2(\eta)\}\,dz = 0$$

This holds if $E(Z)$ is a constant in the $\frac{\partial}{\partial\eta}$ direction. Since the first moment for Z with density $f_Z(z|\mu,\lambda,\nu_1,\nu_2)$ is $\mu+\nu_1$ it is required to find a parameterisation $(\mu',\lambda',\nu_1',\nu_2')$ such that $\mu'+\nu_1' =$ a constant, while respecting condition (i). This can always be easily achieved.

Example: Normal mixture family.

A parameterisation which achieves this for the normal family is given by $\mu' = \mu-\nu_1, \lambda_1' = \lambda_1, \nu_1' = \nu_1, \nu_2' = \nu_2$, with the positivity constraint

$$\nu'^2_1 - 4\nu'_2 + 4\nu'^2_2 < 0.$$

This is illustrated visually by Fig. 3. Here the space of local mixture models, shown in Fig. 2, is 'sliced' Fisher orthogonally to the family $\phi(z|\mu,1)$ and a single 'slice' is shown, denoted by II. Note that the intersection of the slice and $\phi(z|\mu,1)$ lies in the interior of II.

Having considered the geometric structure of local mixture models, now consider the subfamily of true local mixture models. This is characterised by being the intersection of the convex hull of the set $f_Z(z|\theta(\mu),\lambda)$, with λ being fixed and known, with the finite dimensional manifold which is the space of local mixtures.

Example: Normal mixture family.

In Fig. 3 the subset of points in the orthogonal slice II which are true local mixtures, and have mixing distributions $\tilde{Q}$ of width smaller than 2ϵ, is shown by the subregion III. This is the subset of II which lies in the convex hull of a compact subset of the image of $\phi(z|\mu,1)$.

Figure 3 illustrates a number of important features of the geometry. First the intersection of $\phi(z|\mu,1)$ with the space of true local mixtures lies on the boundary of this space. Any inference procedure must take this into account. Second note that the boundary of III at the intersection is not smooth. If possible mixing distributions, $\tilde{Q}$, are restricted to having a fixed compact support then the tangent space of III is a tangent cone. The 'angle' that is made at the vertex of this cone is a function of the size of the compact support of $\tilde{Q}$. The inferential consequences of this are considered in §4. Finally the union of such spaces as III across different values of μ is shown in Fig. 4. The image of $\phi(z|\mu, 1)$ lies on the 'edge' of this surface. Also note its non-smooth boundary.

3.3 Identification

The following results generalise and formalise the visual intuition of the previous example. It shows that, in general, true local mixture models are identified in all parameters $(\mu, \lambda, \nu_1, \nu_2)$ for sufficiently small ν_1 and ν_2. This follows from a general proof, Theorem 2, and then two special cases, the normal and Poisson families.

Theorem 2 *Assume that the exponential dispersion model of the form*

$$f_Z(z|\mu, \lambda) := \nu_\lambda(z) \exp\left(\lambda\left[\theta(\mu)z - \kappa\{\theta(\mu)\}\right]\right)$$

has a smooth variance function $V(\mu)$ such that

$$2 + \frac{\partial^2}{\partial \mu^2} V(\mu) > 0$$

for all μ, and that $f_Z(z|\mu, \lambda)$ is identified in μ and λ.

(i) For fixed, known λ, the space of true local mixture models $f_Z(z|\mu, \lambda, \nu_1, \nu_2)$ given by equation (4) is a manifold with boundary and, locally to $\nu_1 = \nu_2 = 0$, is diffeomorphic to $I\!R^2 \times I\!R^+$ where $I\!R^+ := \{x \in I\!R | x \geq 0\}$. The boundary is given by the condition

$$\nu_2 - \frac{\nu_1^2}{2} \geq 0.$$

Hence the space of true local mixtures is, locally to $\nu_1 = \nu_2 = 0$, identified in the unknown parameters μ, ν_1, ν_2.

(ii) For free, unknown λ, if

$$\frac{\partial}{\partial\lambda}\log\nu_\lambda(z) \tag{8}$$

is not a polynomial in z then the space of true local mixture models is, locally to $\nu_1 = \nu_2 = 0$, identified in all its parameters $(\mu, \lambda, \nu_1, \nu_2)$.

Proof *See Appendix.*

Note that the condition on the variance function of Theorem 2 appies to many exponential dispersion families including the normal, Poisson, binomial, gamma, extreme stable, compound Poisson, and many others, see Jørgensen (1997, page 130). However, while the condition in Theorem 2(ii) on the normalising factor $\nu_\lambda(z)$ is quite general and applies to the binomial, gamma, negative binomial and many other families, Jørgensen (1997, pp 85-91), there are two very important special cases for which this condition does not hold. These are the normal and Poisson families. These cases are treated here separately as they require a rather more detailed analysis.

Example: Normal mixture family.

The well known fact that a mixture of normal distributions can itself be normal, in particular that

$$\int \phi(z|\mu+\eta,\sigma_1^2)\phi(\eta|0,\sigma_2^2)d\eta = \phi(z|\mu,\sigma_1^2+\sigma_2^2) \tag{9}$$

makes one suspect that the identification issue for this family must be quite delicate. Direct calculation shows that the condition in Theorem 2 (ii) does not hold. Despite this, as is shown in the Appendix, locally to $\nu_1 = \nu_2 = 0$ the local mixture family is fully identified in $(\mu, \lambda, \nu_1, \nu_2)$.

Example: Poisson mixture family.

Jørgensen (1997, p 90) shows how to write the Poisson family as an additive exponential dispersion model,

$$\frac{\lambda^z}{z!}\exp(\theta z - \lambda\exp(\theta)),$$

but also shows that this is the only family for which θ and λ are not identified, since $\mu = \lambda \exp(\theta)$. Hence this family falls outside the regularity conditions of Theorem 2. The relevant result for this family is that the true local mixture model is identified for the three parameters μ, ν_1, ν_2, locally to $\nu_1 = \nu_2 = 0$.

To be consistent with the idea of local mixing it is interesting to see what effect putting structure on the possible mixing distribution $\tilde{Q}$ has on the geometry. It is natural to ask that $\tilde{Q}$ is small in some sense. This can be done in many ways and one method is considered here and in more detail in §4.

For a true local mixture $f_Z(z|\mu, \lambda, \nu_1, \nu_2)$ there exists a mixing distribution $\tilde{Q}$ such that $f_Z(z|\mu, \lambda, \nu_1, \nu_2) = \int f_Z(z|\mu+\eta, \lambda) d\tilde{Q}(\eta)$. By calculation of moments and applying Taylor's theorem the following identities follow easily

$$\begin{aligned} \nu_1 &= E_{\tilde{Q}}(\eta), && (10) \\ \nu_2 &= \frac{1}{2} E_{\tilde{Q}}(\eta^2) + \frac{1}{3!} \frac{V'''(\mu^*)}{2 + V''(\mu)} E_{\tilde{Q}}(\eta^3), && (11) \end{aligned}$$

where $'$ denotes differentiation with respect to μ and μ^* is some value in the interval $(\mu, \mu + \nu_1)$.

If the mixing distribution has finite support then, following Kumar (2002), there are natural inequalities on all moments hence, from equations (10)-(11), there are restrictions on the ν_1 and ν_2 parameters for each μ, λ. In particular simple calculations show that for fixed μ and λ,

$$E_{\tilde{Q}}(\eta^2) \leq \nu_1^2 + \epsilon \nu_1.$$

Such a restriction is shown, for the normal example, in Fig. 5 on the space A_3 defined in the proof of Theorem 2 and in the region III of Fig. 3. The shaded areas show the subset of true local mixtures for which $\tilde{Q}$ has width 2ϵ. The following theorem formalises this observation.

Theorem 3 *Assume that an exponential dispersion model $f_Z(z|\mu, \lambda)$ satisfies the conditions of Theorem 2 and that $\tilde{Q}$, defined by equation (5), has finite support and is of width 2ϵ.*

(i) For fixed, known λ the sub-set of true local mixture models is a manifold with boundary and locally to $\nu_1 = \nu_2 = 0$ is diffeomorphic to $I\!R \times I\!R^+ \times I\!R^+$.

(ii) For free λ if the conditions of Theorem 2 holds, or if the family is normal, the family of true local mixtures, locally to $\nu_1 = \nu_2 = 0$, is diffeomorphic to $I\!R^2 \times I\!R^+ \times I\!R^+$.

Proof *See Appendix.*

4 Inference

4.1 Boundaries and inference

As motivation for the basic principle of this section consider inference on ρ in the simplest type of mixture model

$$\rho f(z) + (1-\rho)g(z) \tag{12}$$

where $f(z)$ and $g(z)$ are known. Hall and Titteringon (1984) consider this problem when f and g are unknown and the analysis below follows their approach in the simpler case.

For this problem there are two types of boundary for ρ. First ρ has the interpretation of being a probability, thus $0 \leq \rho \leq 1$. Denote the extreme values 0 and 1 here as soft boundaries. Second since the expression $g(z) + \rho(f(z) - g(z))$ integrates to one for all ρ, (12) is a density if and only if it is non-negative for all z. Let the set ρ for which (12) is a density be given by $[\rho_{\min}, \rho_{\max}]$. Denote the boundary of $[\rho_{\min}, \rho_{\max}]$ as a hard boundary. It is immediate that the soft boundary lies always inside the hard boundary.

Figure 6 illustrates a log-likelihood for such a model. For this example the hard boundary is $\rho = 0, 2$ while the soft is $\rho = 0, 1$. In the figure, and in general, the singularities of the log-likelihood occur at, or after, the hard boundary. However near the soft boundary point, $\rho = 1$, the log-likelihood is uniformly asymptotically quadratic, giving rise under regularity to an asymptotic truncated normal posterior.

The critical condition for asymptotic truncated posterior normality is that the soft boundary dominates the inference problem, either because it is stricter, or because the hard boundary is in an inferentially unimportant region of the parameter space. Figure 6 illustrates both of these points. The soft boundary point $\rho = 1$ is stricter than the hard boundary point $\rho = 2$, and although $\rho = 0$ is both hard and soft it lies a long way in the tail of the posterior distribution. For this example then a good approximation to a truncated normal posterior is expected.

This discussion can be generalised to true local mixture models. For this class the boundaries are again of two forms. The hard boundaries come from the positivity conditions on the expansions (3) or (4). The soft boundaries come from restricting to true local mixtures which lie within the convex hull of $f_Z(z|\mu, \lambda)$. If the bounds from the second of these conditions are stricter and bounded away from the positivity bounds then approximately truncated quadratic log-likelihoods and truncated normal posteriors should be expected. In general since a true local mixture satisfies equation (5), the soft boundary lies inside the hard positivity boundary. Detailed analysis of the examples in this paper shows that asymptotically the soft boundary is strictly inside the hard boundary and the hard boundary is inferenicially asymptotically unimportant.

The following proof follows Hall and Titterington (1984) by embedding the true local mixture model in a multinomial family in order to investigate the log-likelihood approximation.

Theorem 4 *Let $f_Z(z|\mu, \lambda, \nu_1, \nu_2)$ be a true local mixture model which either satisfies the conditions of Theorem 2, or is a normal or Poisson family. Asymptotically, in sample size, in any region strictly inside the hard boundary, the log-likelihood is uniformly quadratic on $(\mu', \lambda, \alpha, \beta)$ in a shrinking neighbourhood of the mode, where*

$$\mu' = \mu + \nu_1, \alpha = 2\nu_2 - \nu_1^2, \beta = \frac{\nu_1^3}{3} - \nu_1\nu_2.$$

Furthermore, assuming any prior chosen is continuous then the posterior for $(\mu', \lambda, \alpha, \beta)$ will be an asymptotically truncated normal distribution.

Proof See Appendix.

4.2 Examples

This section looks at examples of inference in a true local mixture model. In order to keep the presentation focused it concentrates on over-dispersion and mixing in the binomial example. The comparison of the normal and binomial cases illustrates the fundamental differences between continuous and discrete distributional theory. For an extensive treatment of the normal based examples see Marriott (2002, 2003).

There is a large literature regarding the problem of over-dispersion with binomial models and this section looks at a number of approaches. The geometric approach taken here has strong links with that of Lindsay (1995) and in particular with the related work of Wood (1999). Wood looks at the problem of learning about the mixing distribution, Q. In contrast to this a practitioner might be interested in estimating $\mu = E(Z)$, under an over-dispersed binomial model. Two approaches are of interest here, quasi-likelihood and direct modelling through, for example, the beta-binomial model. For the first of these approaches see Cox (1983), McCullagh (1983) and Firth (1987) and references therein. For the second approach see Crowder (1978). Finally one might be interested in testing for over-dispersion, again see Cox (1983). This section demonstrates that for each of these possible inferential questions the true local mixture approach is powerful, insightful and computationally straightforward.

For computations with true local mixture models the Markov chain Monte Carlo algorithm is use here. This is not the only possible approach, for example Marriott (2003) uses a simple closed form geometrically based estimator, while Critchley and Marriott (2003) use moment based methods. However the richness of the output of Markov chain Monte Carlo gives the clearest illustration of the local mixture approach. Furthermore Theorem 4 indicates that it is to be expected that the performance of the algorithm will be extremely good. Finally, the algorithm allows exploration of the effect of different prior assumptions on the form, in particular the size, of the mixing distribution, and of the effect of the hard and soft boundaries.

Wood (1999) uses a geometric framework very similar to the one in §3.1. The binomial family $Bi(z|\pi, n)$ is embedded in the simplex,

$$T_n = \left\{ (x_0, x_1, \ldots, x_n) \mid \sum_{i=1}^{n} x_i = 1, x_i \geq 0 \text{ for all } i \right\},$$

in $I\!R^{n+1}$ by the mapping

$$(Bi(0|\pi, n), Bi(1|\pi, n), \ldots, Bi(n|\pi, n)).$$

It is immediate that the affine space $(X_{\text{Mix}}, V_{\text{Mix}})$ used in this paper is isomorphic to the hyperplane in $I\!R^{n+1}$ which contains T_n. Thus for the binomial example the affine geometry of this paper and Wood's are identical.

Consider an example motivated by Example 1 of Wood (1999, p. 1715). In this example data was generated from a $Bi(z|\pi, n)$ distribution where $n = 10$ and π was drawn from a distribution with mean 0.5 and a small standard deviation. Using a large simulated dataset Wood shows that the mixing distribution can be estimated from such data. Such an example can be thought of as a local mixture. Using Wood's example as motivation consider a much smaller dataset, with sample size 50, generated from his fitted distribution. Here π comes from the discrete distribution with support at $(0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.53)$ with probabilities $(0.0116, 0.0881, 0.1430, 0.1759, 0.1865, 0.1745, 0.1394, 0.0810)$. Since the sample size is so much smaller than Wood's it is unrealistic to expect that the full mixing distribution can be estimated. Rather the local mixture methodology instead estimates the parameters in $f_Z(z|\pi, \nu_1, \nu_2)$.

Figure 7 shows the log likelihood function for this dataset for the slice of the space of true local mixtures which is orthogonal to $Bi(z|\pi, n)$ at the sample mean. Figure 7 (a) shows the log likelihood in the (ν_1, ν_2) parameterisation while 7 (b) shows the log likelihood contours for the (α, β) parameterisation. This parameterisation,

$$(\alpha, \beta) := \left(\nu_2 - \frac{\nu_1^2}{2}, \frac{\nu_1^3}{3} - \nu_1 \nu_2 \right).$$

was used in Theorems 2 and 4 and shown in Fig. 5. In Fig. 7 (a) the hard boundary given by the intersection of the slice with the boundary of T_n can clearly be seen. The soft boundary, given by the true local mixture condition, is also plotted. In panel (b) this second boundary is given by $\alpha = 0$. As predicted by Theorem 4 the log-likelihood contours for the (α, β) parameterisation are well approximated by a quadratic function, and the soft boundary $\alpha > 0$ defining true local mixture models, is inferentially important. The hard boundaries can also been seen, but these are inferentially not important in this dataset. In Fig. 7 (c) and (d) the posterior distributions for α and β, calculated by Markov chain Monte Carlo using vague priors, are shown. As predicted by Theorem 4 truncated normal distributions seem to be very good approximations to these posteriors.

Consider now the question of estimating $E(z)$. As shown by Cox (1983) the sample mean is an unbiased estimator in this example. Cox further analysed this problem using an asymptotic, in sample size N, expansion where the variance of the mixing distribution is of order $N^{-1/2}$. Such an expansion is completely consistent with the local mixing hypothesis, and to first order gives a similar expansion to equation (A2). Furthermore the method of quasi-likelihood generates the sample mean as an estimator. Firth (1987) investigated the efficiency of the quasi-likelihood estimate and notes that it depends on the skewness of the mixing distribution. To investigate this using the true local mixture method a mixing distribution with a large amount of skewness was chosen. Figure 8 shows the result of an analysis for a sample of size 50. The skewness in the data, Fig. 8 (a), reflects the skewness in the mixing distribution. Figure 8 (b) shows the posterior distribution of μ under the true local mixture model, calculated by Markov chain Monte Carlo. The solid vertical line is the value of the sample mean of the data. The dotted vertical lines show the 95% confidence interval calculated using the beta-binomial model. It can be seen that this estimate is significantly under estimating the posterior mode, while giving a good estimate of the posterior variance. As might be expected from Firth (1987) the skewness of the mixing distribution is causing a problem for the quasi-likelihood and beta-binomial estimates. Panels (c) and (d) show the posterior distributions of α and β. As expected the truncated normal approximations to these distributions seem to hold very well. Furthermore the distribution of β, which essentially is measuring the skewness of the mixing distribution, is significantly above 0. These posterior distributions contain the evidence in the data for the existence of mixing and thus can be used as the basis for tests of over dispersion.

Finally consider the effect of different assumptions on the size of the mixing distribution. This is illustrated in Fig. 9. The sample, shown in Fig. 9 (a), is now very highly skewed. Two Markov chain Monte Carlo runs are made, first without an explicit restriction on the mixing distribution second with the assumption that Q has support a small compact region. For both cases the Markov chain sample for α and β is shown in Fig. 9 (b), with the crosses represented the restricted mixing distribution. The corresponding posterior for the μ parameter is shown in Fig. 9 (c). The solid line being the density estimate for the unrestricted case, while the dashed being the density estimate for the restricted case. It can be seen that the high skewness in the data has resulted in the sample mean, illustrated by a solid vertical line, being considerably away from the posterior mode in the unrestricted case. However in the restricted case the mode and the mean agree well. Further the posterior variance for the restricted case agrees well with that from the beta-binomal model whose 95% confidence limits are given by the heavily dashed vertical lines. For reference the lighter dash lines are the binomial models 95% confidence limits. The good agreement between the restricted mixing and the beta binomial infernece follows from the fact that both concentrate attension on the inflation of the variance in the data and not higher order moments. In the trye local mixing model the variance increase corresponds to the $\frac{\partial}{\partial\alpha}$ direction which lies exactly in the direction of the segment of parameter space allowed by the restricted mixing.

Acknowledgments

Part of this work was undertaken while the author was a Visiting Associate Professor at the Institute of Statistics and Decision Sciences, Duke University. The author would also like to thank Frank Critchley for many helpful discussions.

Appendix: Proofs of Theorems

Proof of Theorem 1

Any finite dimensional affine subspace in $(X_{\mathrm{Mix}}, V_{\mathrm{Mix}})$ can be defined by

$$f + \sum_i \lambda_i v_i,$$

where $f \in X_{\mathrm{Mix}}, v_i \in V_{\mathrm{Mix}}$. The image of this space will therefore be in $I\!R^3$,

$$\left(E_{f+\sum_i \lambda_i v_i}(Z^{n_1}), E_{f+\sum_i \lambda_i v_i}(Z^{n_2}), E_{f+\sum_i \lambda_i v_i}(Z^{n_3})\right)$$

Expanding gives

$$\left(E_f(Z^{n_1}), E_f(Z^{n_2}), E_f(Z^{n_3})\right) + \sum_i \lambda_i \left(E_{v_i}(Z^{n_1}), E_{v_i}(Z^{n_2}), E_{v_i}(Z^{n_3})\right).$$

which is an affine space, when all integrals exist. $\square$

Proof of Theorem 2.

First consider the following lemma.

Lemma If the Fisher information at μ is non zero for the family

$$f_Z(z|\mu, \lambda) := e^{\lambda\,\theta(\mu)\,z - \lambda\,\kappa\{\theta(\mu)\}},$$

and if $l \neq k$, then

$$\frac{\partial^l}{\partial \mu^l} f_Z(z|\mu, \lambda) \text{ and } \frac{\partial^k}{\partial \mu^k} f_Z(z|\mu, \lambda)$$

are linearly independent as functions of z.

Proof . First consider $\frac{\partial^l}{\partial \theta^l} f_Z(z|\theta, \lambda)$ which has the form $P_l(z) f_Z(z|\theta, \lambda)$ where $P_l(z)$ is a polynomial in z of order l with leading coefficient λ^l. It is immediate that

$$\frac{\partial^l}{\partial \theta^l} f_Z(z|\theta, \lambda) \text{ and } \frac{\partial^k}{\partial \theta^k} f_Z(z|\theta, \lambda)$$

are linearly independent as functions of z when $l \neq k$.

The derivative $\frac{\partial^l}{\partial \mu^l} f_Z(z|\mu,\lambda)$ can be calculated from the chain rule and has the form $P_l(z) f_Z(z|\mu,\lambda)$, now with the leading term having the coefficient $\lambda^l \left(\frac{\partial \theta}{\partial \mu}\right)^l$. This is non zero if the Fisher information is non zero and the result follows. □

(i) Having proved the lemma now assume that $f_Z(z|\mu,\lambda)$ satisfies the conditions of the theorem. Any local mixture model with mean μ_0 has the form

$$f_Z(z|\mu_0-\nu_1,\lambda)+\nu_1\frac{\partial}{\partial\mu}f_Z(z|\mu_0-\nu_1,\lambda)+\nu_2\frac{\partial^2}{\partial\mu^2}f_Z(z|\mu_0-\nu_1,\lambda) \quad \text{(A1)}$$

whose tangent directions $\frac{\partial}{\partial\nu_1}$ and $\frac{\partial}{\partial\nu_2}$ are orthogonal to $\frac{\partial}{\partial\mu}$ when $\nu_1=\nu_2=0$. Hence it is sufficient to examine the identification of expression (A1) for ν_1 and ν_2 and a fixed mean, μ_0.

All local mixtures lie in the affine space $(X_{\mathrm{Mix}}, V_{\mathrm{Mix}})$, thus can be written as

$$f_Z(z|\mu_0,\lambda)+V(z)$$

where $V(z)\in V_{\mathrm{Mix}}$. Let $V_{\mathrm{power}}(\mu_0)$ be the vector subspace of V_{Mix} defined by the set of all formal power series

$$\sum_{k=1}^{\infty} C(k)\frac{\partial^k}{\partial\mu^k}f_Z(z|\mu_0,\lambda)$$

where $C(k)$ is independent of z, and let $X_{\mathrm{power}}(\mu_0)$ be the set of functions of the form

$$X_{\mathrm{power}} := \left\{ f_Z(z|\mu_0,\lambda)+v \mid v\in V_{\mathrm{power}}(\mu_0)\right\}.$$

By expanding (A1) by Taylor's theorem it follows that locally to $f_Z(z|\mu_0,\lambda)$ it has the form

$$f_Z(z|\mu_0,\lambda)+\left(\nu_2-\frac{\nu_1^2}{2}\right)\frac{\partial^2}{\partial\mu^2}f_Z(z|\mu_0,\lambda)+\left(\frac{\nu_1^3}{3}-\nu_1\nu_2\right)\frac{\partial^3}{\partial\mu^3}f_Z(z|\mu_0,\lambda)+\sum_{l\geq 4}C_l\frac{\partial^l}{\partial\mu^l}f_Z(z|\mu_0,\lambda). \quad \text{(A2)}$$

and lies in the affine space $(X_{\mathrm{Power}}(\mu_0), V_{\mathrm{Power}}(\mu_0))$. By the lemma the higher order terms in the sum are linearly independent of the first three terms. Hence there exists a well defined affine map from $(X_{\mathrm{Power}}(\mu_0), V_{\mathrm{Power}}(\mu_0))$ into the two dimensional affine space

$$A_3(\mu,\lambda) := \left\langle \frac{\partial^2}{\partial\mu^2}f_Z(z|\mu_0,\lambda), \frac{\partial^3}{\partial\mu^3}f_Z(z|\mu_0,\lambda)\right\rangle_{f_Z(z|\mu_0,\lambda)}$$

defined by dropping the higher order terms. To prove identification in $(X_{\text{Mix}}, V_{\text{Mix}})$ it is sufficient to prove it in this finite dimensional subspace. Using the obvious coordinate system, the image of the local mixtures is given by

$$(\alpha, \beta) := \left(\nu_2 - \frac{\nu_1^2}{2}, \frac{\nu_1^3}{3} - \nu_1\nu_2\right)$$

and Fig. 5 shows this image directly.

Under the condition on the variance function and by using Jensen's inequality, any true local mixture $f_Z(z|\mu_0, \lambda, \nu_1, \nu_2)$ satisfies the moment inequality

$$E_{f_Z(z|\mu_0,\lambda)}(Z^2) \leq E_{f_Z(z|\mu_0-\nu_1,\lambda,\nu_1,\nu_2)}(Z^2).$$

Expanding the right hand side as a series in ν_1, ν_2 gives

$$\mu_0^2 + V(\mu_0) + \left\{2 + \frac{\partial^2}{\partial\mu^2}V(\mu)\right\}\left\{\nu_2 - \frac{\nu_1^2}{2}\right\} + \cdots.$$

Thus, for sufficiently small, ν_1, ν_2 it is a necessary condition for being a true local mixture that

$$\nu_2 - \frac{\nu_1^2}{2} > 0. \tag{A3}$$

In Fig. 5 this region is shown by the half plane, $\alpha > 0$. Analysis shows that in this region the coordinate system (α, β) is non singular, and has a point singularity only on $\alpha = 0$ at the origin $\nu_2 = 0$.

To show that the condition is sufficient to give a true local mixture consider the convex hull of the image of $f_Z(z|\mu, \lambda)$ in the affine space defined the first three deriviatives

$$\left\langle \frac{\partial}{\partial\mu}f_Z(z|\mu_0, \lambda), \frac{\partial^2}{\partial\mu^2}f_Z(z|\mu_0, \lambda), \frac{\partial^3}{\partial\mu^3}f_Z(z|\mu_0, \lambda)\right\rangle_{f_Z(z|\mu_0,\lambda)}.$$

At μ_0 the image is given by truncating the Taylor expansion to

$$f_Z(z|\mu_0, \lambda) + \sum_{i=1}^{3}\frac{1}{i!}(\mu - \mu_0)^i\frac{\partial^i}{\partial\mu^i}f_Z(z|\mu_0, \lambda).$$

Any two component mixture

$$\pi f_Z(z|\mu_0 + \mu_1, \lambda) + (1 - \pi)f_Z(z|\mu_0 + \mu_2, \lambda)$$

which satisfies the constraint that its mean is μ_0 has an image in the (α, β) parameterisation, defined above, given by $(-\mu_1\mu_2, -(\mu_1+\mu_2)\mu_1\mu_2)$. Hence since the projection of the convex hull of $f_Z(z|\mu,\lambda)$ in $(X_{\text{Mix}}, V_{\text{Mix}})$ is onto the convex hull of the projection of $f_Z(z|\mu,\lambda)$ in $A_3(\mu,\lambda)$ it is clear that that true local mixtures, in the (α,β) parameterisation, spans $\mathbb{R}^+ \times \mathbb{R}$.

The space of tangent vectors for $f_Z(z|\mu,\lambda,\nu_1,\nu_2)$ at $\nu_1 = \nu_2 = 0$ is therefore an orthogonal sum of the tangent space $\frac{\partial}{\partial\mu}$ and that of $\frac{\partial}{\partial\nu_1}$ and $\frac{\partial}{\partial\nu_2}$ which has the form $\mathbb{R} \times \mathbb{R}^+$. Locally to $\nu_1 = \nu_2 = 0$ the space of true local mixture models is a manifold with boundary and is locally diffeomorphic to $\mathbb{R}^2 \times \mathbb{R}^+$.

(ii) If λ is now considered a free parameter the identification of 4 parameters needs to be considered. It is sufficient to show that the tangent vector

$$\frac{\partial}{\partial\lambda} f_Z(z|\mu,\lambda)$$

does not lie in the space spanned by $\frac{\partial}{\partial\mu}, \frac{\partial}{\partial\nu_1}, \frac{\partial}{\partial\nu_2}$ of $f_Z(z|\mu,\lambda,\nu_1,\nu_2)$ at $\nu_1 = \nu_2 = 0$. By direct calculation it can be seen that each element of this space can be written as $P(z)f_Z(z|\mu,\lambda)$ where $P(z)$ is a polynomial in z. Since, under the assumption of the theorem the tangent vector

$$\frac{\partial}{\partial\lambda} f_Z(z|\mu,\lambda) = \left[\frac{\partial}{\partial\lambda}\log\{\nu_\lambda(z)\} + \theta z - \kappa(\theta)\right] f_Z(z|\mu,\lambda)$$

is not of this form it follows that it does not lie in the tangent space of the other parameters. Hence locally to $\nu_1 = \nu_2 = 0$ the family $f_Z(z|\mu,\lambda,\nu_1,\nu_2)$ is identified. □

Proof of identification in normal example.

As in the proof of Theorem 2 consider for some value μ_0 the subset of local mixture models which have $E(z) = \mu_0$. As above the image of this set in the affine space

$$A_4(\mu,\sigma^2) := \left\langle \frac{\partial^2}{\partial\mu^2} f_Z(z|\mu_0,\sigma^2), \frac{\partial^3}{\partial\mu^3} f_Z(z|\mu_0,\sigma^2), \frac{\partial^4}{\partial\mu^4} f_Z(z|\mu_0,\sigma^2) \right\rangle_{f_Z(z|\mu_0,\sigma^2)}$$

is

$$f_Z(z|\mu_0,\sigma^2) \quad + \quad \left(\nu_2 - \frac{\nu_1^2}{2}\right)\frac{\partial^2}{\partial\mu^2} f_Z(z|\mu_0,\sigma^2) + \left(\frac{\nu_1^3}{3} - \nu_1\nu_2\right)\frac{\partial^3}{\partial\mu^3} f_Z(z|\mu_0,\sigma^2)$$

$$+ \quad \left(\frac{\nu_1^2\nu_2}{2} - \frac{\nu_1^4}{8}\right)\frac{\partial^4}{\partial\mu^4} f_Z(z|\mu_0, \sigma^2).$$

By expanding this as a Taylor series around σ_0^2 it can be seen that the projection into A_4 is well defined and given by

$$\begin{aligned} f_Z(z|\mu_0, \sigma_0^2) \quad &+ \quad \left(\frac{\delta}{2} + \nu_2 - \frac{\nu_1^2}{2}\right)\frac{\partial^2}{\partial\mu^2} f_Z(z|\mu_0, \sigma_0^2) \\ &+ \quad \left(\frac{\nu_1^3}{3} - \nu_1\nu_2\right)\frac{\partial^3}{\partial\mu^3} f_Z(z|\mu_0, \sigma_0^2) \\ &+ \quad \left(\frac{\delta^2}{8} + \frac{\delta}{2}(\nu_2 - \frac{\nu_1^2}{2}) + \frac{\nu_1^2\nu_2}{2} - \frac{\nu_1^4}{8}\right)\frac{\partial^4}{\partial\mu^4} f_Z(z|\mu_0, \sigma_0^2) \end{aligned}$$

where $\sigma^2 = \sigma_0^2 + \delta$. In the obvious coordinates this can be written as

$$\left(\alpha + \frac{\delta}{2}, \beta, \frac{\delta^2}{8} + \frac{\delta}{2}\alpha - \frac{1}{4}C(\alpha,\beta)^2\alpha - \frac{3}{4}C(\alpha,\beta)\beta\right)$$

where when $\alpha \geq 0$ $C(\alpha, \beta)$ is a well-defined function. The map from α, β, δ to this space can easily be shown to be one to one when $\alpha \geq 0$. □

Proof of Theorem 3.

As shown in the proof of Theorem 2 the subset of the space of true local mixtures orthogonal to $\frac{\partial}{\partial\mu}$ at μ_0 is, locally to $\nu_1 = \nu_2 = 0$, diffeomorphic to a subset of $A_3(\mu_0, \lambda)$. If the mixing distribution is of width 2ϵ this is diffeomorphic to the convex hull of (x^2, x^3) for $|x| \leq \epsilon$. The result follows immediately. □

Proof of Theorem 4

Standard proofs of the asymptotic normality of the posterior, for example Walker (1969), can almost be applied directly, except for the fact that the parameter space is not an open subset of $I\!R^4$. When the conditions of Theorem 2 apply, or the family is normal or Poisson, then the true local mixture family is locally diffeomorphic to a closed segment

$$\mathcal{S} = \{(x, y, z, w) \in I\!R^4 | z, w \geq 0\},$$

in particular the boundary is typically inferentially important.

In order to return the inference to a more regular setting it is convenient to embed the model in a larger one where standard results apply. Following Hall and Titterington (1984) consider approximating the log-likelihood for a true local mixture family by a multinomial approximation determined by the probabilities $(\pi_1, \cdots, \pi_M)$, where the number of bins, M, in the multinomial model grows at a rate $N^{1/4}$, where N is the sample size. In such a family the log-likelihood is uniformally quadratic in any neighbourhood strictly bounded away from the hard boundaries $\pi_i = 0, 1$.

The embedding is given by

$$(\mu', \lambda, \alpha, \beta) \to \pi_i(\mu', \lambda, \alpha, \beta) = \int_{D_i} f_Z(z|\mu', \lambda, \alpha, \beta) dz.$$

The approximation of the log-likelihood relies on

$$\log \{f_Z(z_i|\mu', \lambda, \alpha, \beta)|D_j|\} - \log \left\{ \int_{D_j} f_Z(z|\mu', \lambda, \alpha, \beta) dz \right\} \to 0 \qquad \text{(A4)}$$

as $N \to \infty$, where $z_i \in D_j$. By assumption the region of parameter space of interest lies strictly inside the positivity boundary, thus $f_Z(z_i|\mu', \lambda, \alpha, \beta)$ is strictly bounded away from zero hence the contribution to the log-likelihood remains finite for all z_i and the convergence in (A4) applies.

The posterior on $(\mu', \lambda, \alpha, \beta)$ is approximated by the posterior in $\mathbb{R}^M$ conditionally on being on the image of $(\mu', \lambda, \alpha, \beta)$. Since this image is diffeomorphic to $\mathcal{S}$ it is approximately a truncated normal. It follows by standard arguments that the posterior is asymptotically truncated normal. $\square$

References

Cox, D.R., (1983). Some remarks on overdispersion, *Biometrika*, **70**, 1, pp 269-74.

Critchley, F. & Marriott, P. (2003). Data-informed influence analysis, Under review for *Biometrika*.

Crowder, M.J., (1978). Beta-binomial anova for proportions, *Appl. Statist.* **27**, No. 1, pp. 34-37.

Firth, D., (1987), On the efficiency of quasi-likelihood estimation, *Biometrika*, **74**, 2, pp 233-45.

Hall, P. & Titterington, D.M., (1984). Efficient nonparmetric estimation of mixture proportions, *J.R.Statist. Soc.* **B** 46, No 3 pp. 465-473.

Jørgensen, B., (1997). *The Theory of Dispersion Models*, Chapman and Hall: London.

Kumar, P., (2002). Moment inequalities of a random variable defined over a finite interval, *J. Inequalities in Pure and Applied Mathematics*, **3** 3, Art 41.

Lindsay, B.G., (1995). *Mixture models: Theory, Geometry, and Applications*, Hayward CA: Institute of Mathematical Sciences.

Marriott, P., (2002). On the local geometry of mixture models, *Biometrika*, **89**, 1.

Marriott, P., (2003). On the geometry of measurement error models, *Biometrika* **90**, 3.

McCullagh, P., (1983). Quasi-likelihood functions, *Annals of Statist.*, **11**, 1, pp 59-67.

McLachlan, G.J. & Peel, D., (2000). *Finite Mixture models*, London: Wiley.

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

Walker, A.M., (1969). On the asymptotic behaviour of posterior distributions. *J.R. Statist. Soc.* B, **31**, 80-88.

Wood, G.R., (1999). Binomial mixtures: geometric estimation of the mixing distribution, *Annals of Statistics*, Vol. 27, 5, 1706-1721

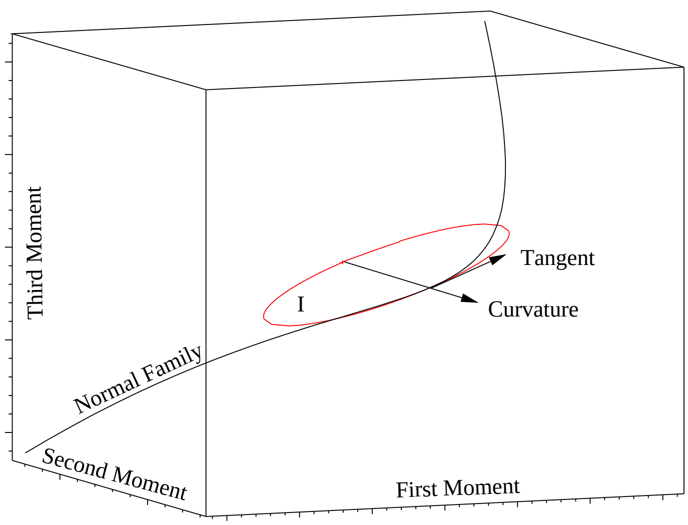

Figure 1: The image of $N(\mu, 1)$ and the space of local mixtures at a single point μ.

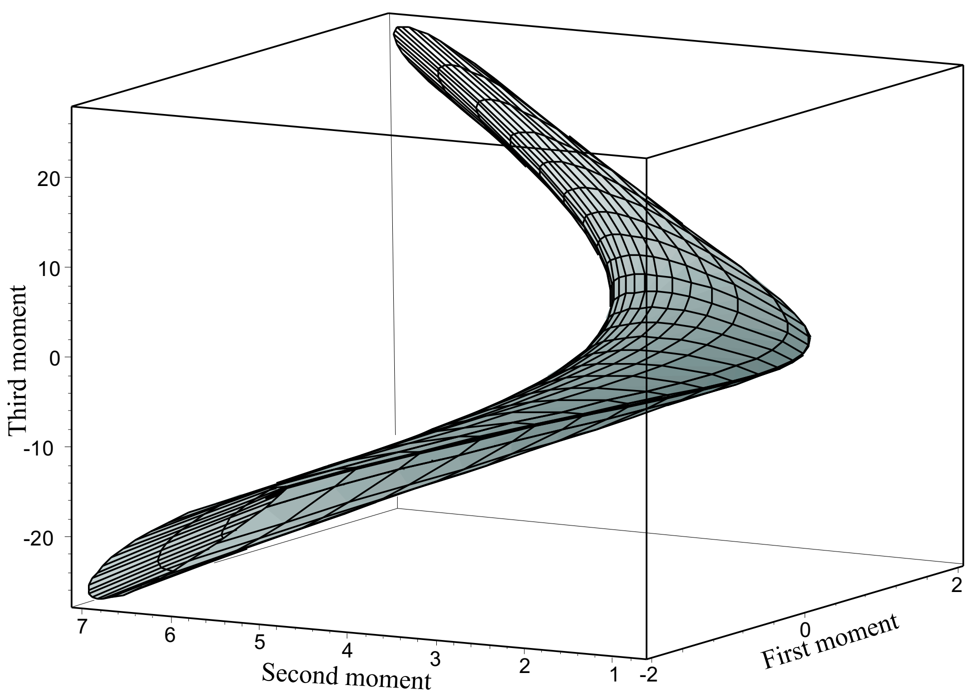

Figure 2: The local mixture models for the normal$(\mu, 1)$ family

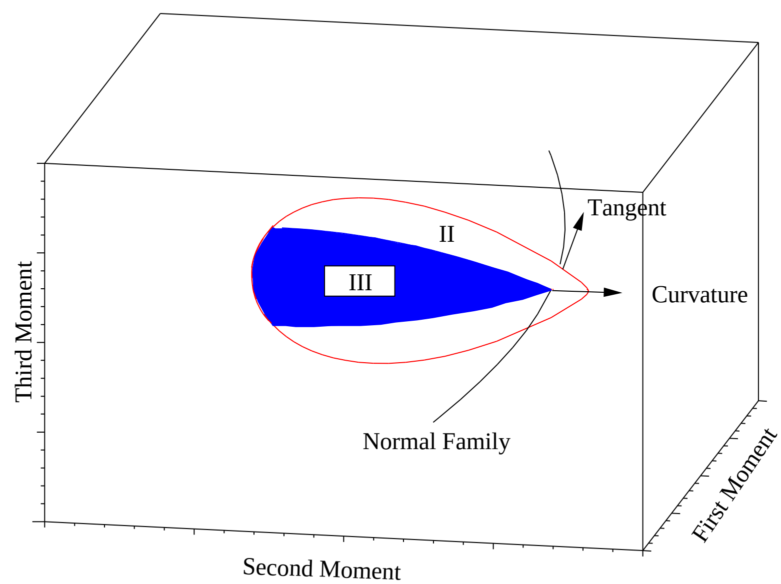

Figure 3: The image of $N(\mu, 1)$ and the orthogonal slice of the space of local mixture models. Region II shows the space of local mixtures with a fixed mean, while the subregion III shows the true local mixtures whose mixing distribution has a fixed compact support.

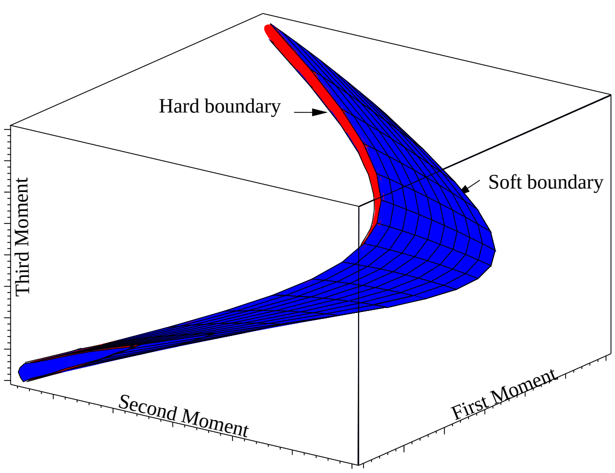

Figure 4: Three dimensional projection of the true local mixture model based $N(\mu, 1)$. Both the hard and soft boundaries are illustrated.

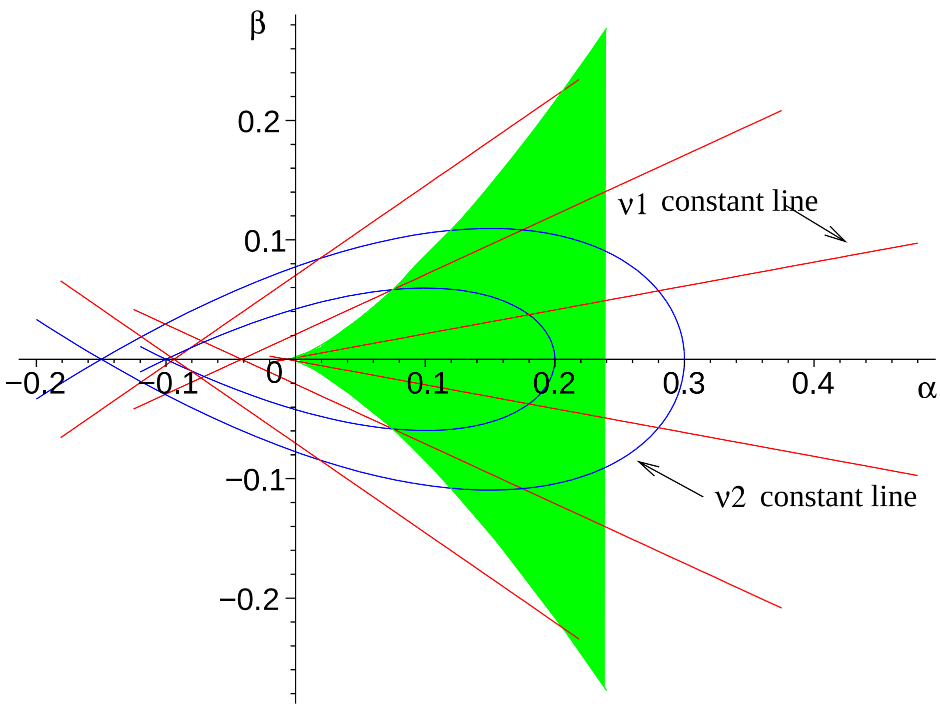

Figure 5: The ν_1, ν_2-parametrisation on a fibre for the true local mixtures. Also shown is the region generated by mixtures of width 2ϵ

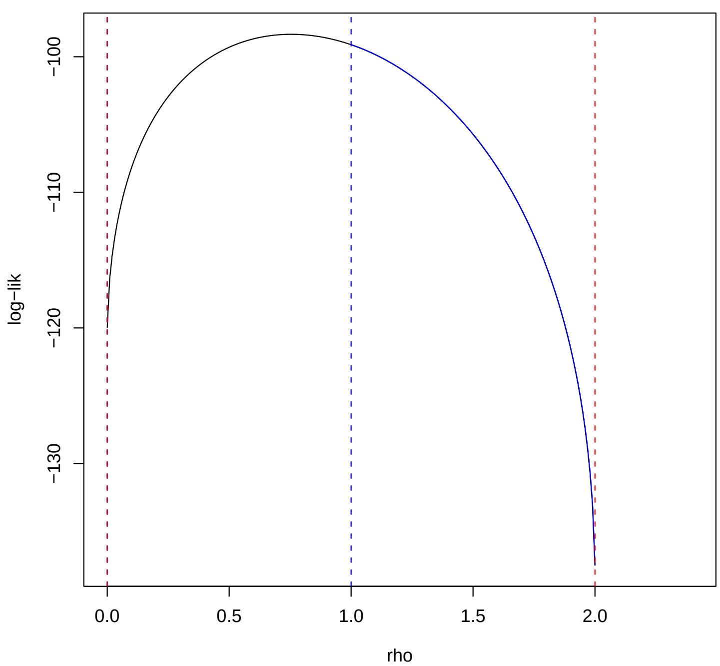

Figure 6: The log-likelihhood for a simple mixture.

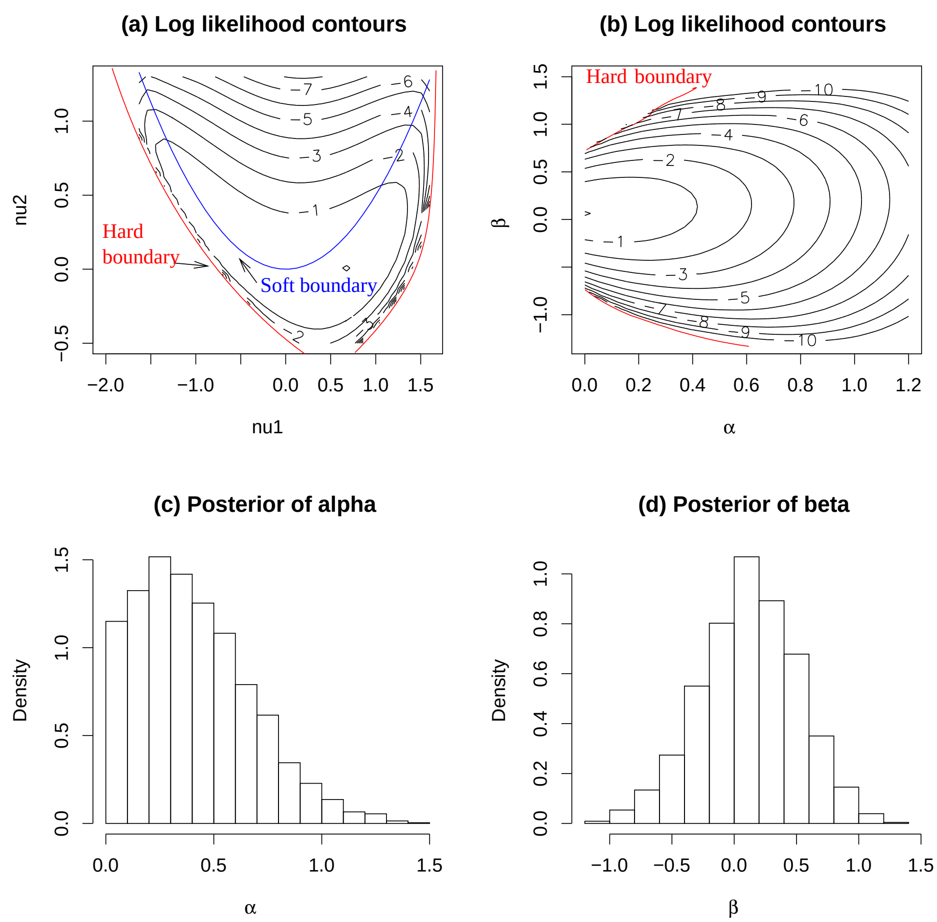

Figure 7: (a) Log likelihood contours in ν_1, ν_2-parameters for orthogonal slice of local mixture models. (b) Log likelihood contours in α, β-parameters for orthogonal slice of true local mixture models. (c) posterior distribution for α (d) posterior distribution for β

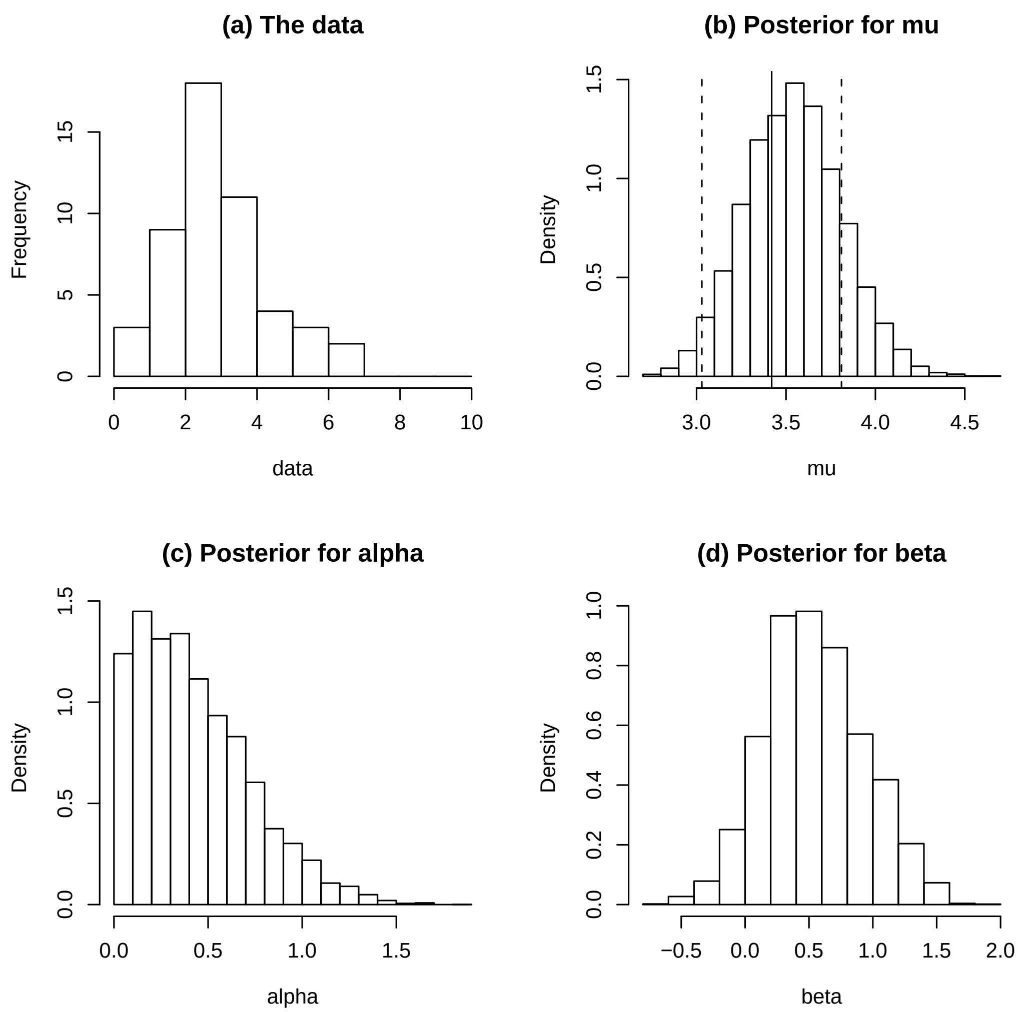

Figure 8: (a) The data (b) The posterior distribution for μ with the vertical line showing the value of the sample mean (c) the posterior distribution for α (d) the posterior distribution for β

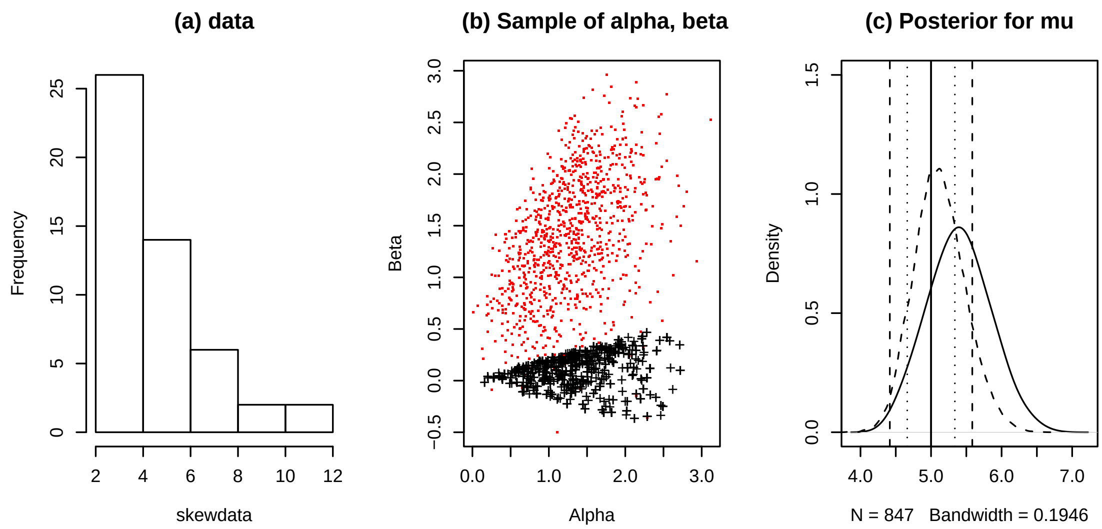

Figure 9: (a) The data (b) The Markov chain Monte Carlo sample for α and β. In the unrestricted case the points are illustrated with dots, in the restricted case with crosses. (c) Density estimates for the the posterior distributions for μ, solid line unrestricted case, dash line restricted case.