AISM (2007) 59:111–134
DOI 10.1007/s10463-006-0095-z

Local mixtures of the exponential distribution

K. A. Anaya-Izquierdo · P. K. Marriott

Received: 6 June 2006 / Revised: 2 October 2006 /
Published online: 13 January 2007

Abstract A new class of local mixture models called *local scale mixture models* is introduced. This class is particularly suitable for the analysis of mixtures of the exponential distribution. The affine structure revealed by specific asymptotic expansions is the motivation for the construction of these models. They are shown to have very nice statistical properties which are exploited to make inferences in a straightforward way. The effect on inference of a new type of boundaries, called *soft boundaries*, is analyzed. A simple simulation study shows the applicability of this type of models.

Keywords Mixture model · Local mixtures · Laplace expansion · Scale dispersion models · Affine geometry

1 Introduction

Mixtures of the exponential distribution have received considerable attention in the statistical literature. For example, Jewell (1982) discusses a characterization of mixtures of Weibull distributions and also nonparametric maximum likelihood estimation of the mixing distribution. Keilson and Steutel (1974) discuss scale and power mixtures and applied their results to show that the squared coefficient of variation of the mixing distribution is a measure of distance in the space of mixtures of exponentials.

K. A. Anaya-Izquierdo (✉)
Department of Statistics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK
e-mail: k.anaya@open.ac.uk

P. K. Marriott
Department of Statistics and Actuarial Sciences, University of Waterloo, 200 University Ave.W., Waterloo, ON N2L 3G1, Canada

Mixtures of the exponential distribution are characterized as being completely monotone, see Jewell (1982) and Heckman and Walker (1990). This implies, in particular, that the density has to be monotone decreasing, so they do not exhibit multimodality.

Inference in mixture models is difficult in general. If the class of allowed mixing distributions is restricted in a statistically natural way, then inferences can be simplified without losing much of the flexibility of the resultant mixtures. Local mixture models have already proved their efficacy in achieving such task (see Critchley and Marriott, 2004; Marriott, 2003, 2002). A slightly different version of local mixture models called *local scale mixture models* is introduced here. This version is particularly suitable for the analysis of mixtures of the exponential distribution. For more general versions see Anaya-Izquierdo and Marriott (2006) and Anaya-Izquierdo (2006).

We are only interested in the case where the mixing distributions are continuous. For discrete mixtures, see McLachlan and Peel (2001).

2 Scale dispersion mixtures

Unless otherwise stated, $\mathcal{F}$ will denote the family of exponential densities parametrized by its mean μ, that is

$$\mathcal{F} := \left\{ f(x;\mu) = \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) \;:\; \mu > 0 \right\}.$$

We are interested in the following type of mixtures.

Definition 1 *The* family of scale dispersion mixtures of $\mathcal{F}$ *is defined as the family of densities of the form*

$$\begin{aligned} g(x;Q(\mu;m,\epsilon)) &= \int_0^\infty f(x;\mu)\,\mathrm{d}Q(\mu;m,\epsilon) \\ &= \int_0^\infty \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) a(\epsilon)\,\frac{1}{\mu}\exp\left(-\frac{d_0(\mu/m)}{2\epsilon}\right)\mathrm{d}\mu \qquad (1) \end{aligned}$$

when Q is a regular scale dispersion model, *that is, when the function $d_0(u)$ is smooth, nonnegative with $d_0(u) = 0$ if and only if $u = 1$ and there exist $\epsilon_0 > 0$, such that*

$$\frac{1}{a(\epsilon)} := \int_0^\infty \frac{1}{\mu}\exp\left(-\frac{d_0(\mu/m)}{2\epsilon}\right)\mathrm{d}\mu$$

is finite for $\epsilon \in (0, \epsilon_0)$.

The generality in this definition comes from the generality of the function $d_0(u)$. The function $d(\mu, m) := d_0(\mu/m)$ is called the unit deviance and the parameters m and ϵ represent *position* and *dispersion*, respectively. For details on dispersion models see Jorgensen (1997).

Example 1 The generalized inverse gaussian distribution described in Jorgensen (1997) is an example of a general class of scale dispersion models. The unit deviance in such case is given by

$$d_\beta(\mu; m) = 2\beta \log\left(\frac{m}{\mu}\right) + \frac{\mu}{m}(1+\beta) + \frac{m}{\mu}(1-\beta) - 2$$

where $m, \epsilon > 0$ and $\beta \in [-1, 1]$. This family defines a regular scale dispersion model for each fixed value of β. The values $\beta = \pm 1$ correspond (by taking the appropriate limits) to the gamma distribution and reciprocal gamma distribution, respectively, and the value $\beta = 0$ corresponds to the hyperbola distribution. The Q_β-mixture $g(x; Q_\beta)$ can be given in closed form by

$$\frac{\left(\frac{1+\beta}{1-\beta}\right)^{\frac{\beta}{2\epsilon}} K_{\frac{\beta}{\epsilon}-1}\left(\sqrt{\frac{(1+\beta)[2x\epsilon + m(1-\beta)]}{\epsilon^2 m}}\right)\left[\frac{2x\epsilon + m(1-\beta)}{m(1+\beta)}\right]^{\frac{\beta-\epsilon}{2\epsilon}}}{m K_{\frac{\beta}{\epsilon}}\left(\frac{\sqrt{1-\beta^2}}{\epsilon}\right)},$$

where $K_\nu(z)$ is the modified Bessel function of the third kind with index ν. This family contains the Pareto distribution family of the second kind, when $\beta = -1$, which is well known for being a heavy tailed family, see Embrechts et al. (1997).

Example 2 As another example of a scale dispersion model consider the lognormal distribution. The unit deviance in such case is given by $d(\mu; m) = \log^2(\mu/m)$. To the best of our knowledge, there is no closed form expression for the corresponding mixture density.

3 Local scale mixture models

In this section we describe a class of models which captures the behavior of scale dispersion mixtures of $\mathcal{F}$ particularly when the dispersion parameter ϵ is small.

Definition 2 *The* local scale mixture model *of order d of the exponential distribution $\mathcal{F}$ is defined as the parametric family*

$$\mathcal{G}_{\mathcal{F}} = \left\{ g(x; \mu, \gamma) = f(x; \mu) + \sum_{k=2}^{d} \frac{\mu^k \gamma_k}{k!} f^{(k)}(x; \mu) \; : \; \mu > 0 \, , \, \gamma \in \Gamma_\mu \right\}$$

if for every $\mu > 0$ we have that $\Gamma_\mu \subset \mathbb{R}^{d-1}$ is nonempty for all $\mu > 0$.

For each μ, Γ_μ is the largest set of γ values for which $g(x;\mu,\gamma)$ is nonnegative for all $x > 0$. The set Γ_μ is convex for all $\mu > 0$. The boundary $\partial(\Gamma_\mu)$ is called the *hard boundary* of $g(x;\mu,\gamma)$ at μ. Note that μ is a scale parameter for $\mathcal{F}$ so μ is also a scale parameter for $g(x;\mu,\gamma)$. Therefore Γ_μ does not depend on μ and we will just write Γ.

The motivation for the specific form of the densities in $\mathcal{G}_\mathcal{F}$ is rather technical and also not crucial for the development of the rest of the paper, therefore some details of it are placed in the appendix. The most important thing to remark at this point, is the fact that local scale mixture models of the exponential distribution are motivated by the asymptotic expansion in Theorem 3 of the appendix.

In general, local scale mixture models can be considered only as generalizations of the exponential distribution $\mathcal{F}$. It is not the aim of local scale mixture models to approximate scale dispersion mixtures in any analytical sense. Our philosophy is that we can capture some information of scale dispersion mixing structure in the data by modeling using local scale mixture models of even order.

Under such philosophy, the interpretation of these models is the following. A local scale mixture model of the exponential distribution $\mathcal{F}$ of even order d and $\mu = m$, mimic the behavior of some scale dispersion mixture (with mean m and small ϵ) in the sense of the asymptotic expansion of Theorem 3 with $d = 2r$. Moreover, the parameters γ_i play the same role as the normalized central moments of the mixing distribution, which clearly only depend on ϵ. In this respect we will call such parameters *pseudo-moments*. In particular γ_2 plays the same role as the squared coefficient of variation of the mixing distribution.

In a sense, we are turning the nonparametric problem of estimating the whole mixing distribution to the parametric one of estimating its moments which, as argued in the appendix, very much determine the mixing distribution itself.

The geometry of a local scale mixture model is that of a -1 *affine fibre bundle* and is easily understood in terms of the affine space

$$\langle X, V, +\rangle = \left\langle \left\{ f : \int f(x)\,\mathrm{d}x = 1 \right\}, \left\{ s : \int s(x)\,\mathrm{d}x = 0 \right\}, + \right\rangle.$$

The space of positive densities is only a convex subset and $+$ is the usual addition operator between real-valued functions. The affine structure of $\langle X, V, +\rangle$ agrees with Amari's -1 geometry. See Amari and Nagaoka (2000). For each fixed $\mu > 0$ local scale mixture models are -1 flat with respect to this affine space.

Note that local scale mixture models are simple reparametrizations of the usual version of local mixture models (see Critchley and Marriott, 2004; Marriott, 2003, 2002). The explicit dependence on μ of the parameters of the derivative terms is introduced to give them a clear and easy interpretation in terms of the mixing distribution. From the geometric point of view, the fiber bundle structure is also taken into account in the sense that the dependence of each fiber on μ is explicitly given. The absence of the first derivative term in

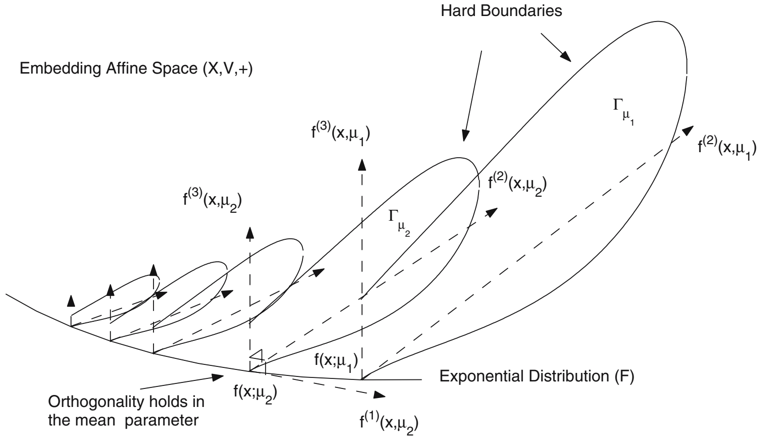

Fig. 1 Geometry of local mixtures

this new version is now crucial for many statistical reasons including identifiability (see Anaya-Izquierdo and Marriott, 2006). The geometric construction of a local scale mixture model is explained graphically in Fig. 1.

As local scale mixture models are designed to perform like true mixtures, it is natural to ask the question of whether they can be exact mixtures themselves.

Definition 3 *A local scale mixture density* $g(x; \mu, \gamma)$ *is a* True Local Mixture density *if it lies inside the convex hull of the family* $\mathcal{F}$ *in the affine space* $\langle X, V, +\rangle$.

The following Theorem gives a characterization of general mixtures of the exponential distribution. Therefore, it gives necessary and sufficient conditions for a local scale mixture model to be a true local mixture model.

Theorem 1 *Let* $S(x)$ *be an absolutely continuous survival function such that* $S(0) = 1$. *Then,*

$$S(x) = \int_0^\infty \exp\left(-\frac{x}{\mu}\right) \mathrm{d}Q(\mu)$$

for some proper probability distribution Q *if and only if*

$$(-1)^k \frac{\partial^k S(x)}{\partial x^k} \geq 0\,, \tag{2}$$

for all $x \geq 0$ *and all* $k \in \mathbb{N}$.

Proof See Feller (1970) p 439. □

4 Estimation in local scale mixture models

In this section we propose to use likelihood and moment based inference for the statistical analysis of local scale mixture models. First, asymptotic likelihood inference is straightforward due to the following result.

Theorem 2 *Let $\mathcal{F}$ be the exponential distribution parametrized by its mean μ. Then the local scale mixture model of $\mathcal{F}$ of order d has the following properties:*

1. *The model is identified in all its parameters $(\mu, \gamma_2, \ldots, \gamma_d)$*
2. *For each fixed $\mu > 0$, the log-likelihood of the parameter $\gamma = (\gamma_2, \ldots, \gamma_d)$ is concave over its convex domain Γ_μ.*
3. *The parameters $(\mu, \gamma_2, \ldots, \gamma_d)$ are Fisher orthogonal at $(\mu, 0, \ldots, 0)$ for all $\mu > 0$.*

Proof See Anaya-Izquierdo and Marriott (2006).

If the parameter of interest is the mean μ then asymptotic inferences can be performed using the profile log-likelihood for μ. Using property 2, the problem of finding the maximum likelihood estimator of the parameter γ, for each fixed $\mu > 0$, is a well defined and known in nonlinear programming. That is, finding the maximum of a concave function over a convex set. Moreover, using property 3, the maximum likelihood estimator obtained for μ will be asymptotically independent of the nuisance parameter γ giving a clean inferential separation.

On the other hand, the first four central moments of a local scale mixture model are given by

$$\begin{aligned} E[X;\mu,\gamma] &= \mu \\ V[X;\mu,\gamma] &= \mu^2\left[1+2\gamma_2\right] \\ E[(X-\mu)^3;\mu,\gamma] &= 2\mu^3\left[1+6\gamma_2+3\gamma_3\right] \\ E[(X-\mu)^4;\mu,\gamma] &= 9\mu^4\left[1+84/9\,\gamma_2+72/9\,\gamma_3+24/9\,\gamma_4\right] \end{aligned} \qquad (3)$$

Therefore, simple moment based estimators for (μ, γ) can be obtained from these expressions. Strong similarities with some of the results of Lindsay (1989) arise. Note that the behavior of the moments is that of inflating the corresponding moments of the exponential distribution. These expressions give an idea of why local scale mixture models are particularly effective in detecting mixture structure through moments.

5 Modeling

To keep the presentation focused, we only study the two particular local scale mixture models when $d = 2$ and $d = 4$, but much of the spirit of our analyzes can be extended to more general models.

The local scale mixture model of order $d = 2$ can be written as

$$g(x;\mu,\gamma_2) = \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right)\left[1+\gamma_2\left[1-\frac{2x}{\mu}+\frac{x^2}{2\mu^2}\right]\right]. \tag{4}$$

The parameter space for γ_2 is simply given by $\Gamma = [0,1]$. The hard boundary $\{0,1\}$ has a nice and simple interpretation, it says that the local scale mixture model is a proper density when the pseudo squared coefficient of variation γ_2 is on the interval $[0,1]$. Keilson and Steutel (1974) show that the squared coefficient of variation of the mixing distribution in the exponential case is a formal distance between a mixture of $\mathcal{F}$ and the unmixed $\mathcal{F}$. Assume, for example, that the mixing model is a gamma distribution with mean m and dispersion ϵ. The squared coefficient of variation is ϵ. If $\epsilon \geq 1$, this mixing family has a unique mode at zero, but if $0 < \epsilon < 1$ then the mode is positive and the density shrinks to m as $\epsilon \to 0$. Being inside the hard boundary, in this case means that the local mixture is only going to be able to model the behavior of this mixture when the squared coefficient of variation of the mixing distribution is small and in particular, less than one.

More generally, we can interpret the hard boundary for this local scale mixture model as follows. This model is going to be able to model the behavior of the scale dispersion mixture when the variance function of the mixing distribution (in this case is approximately $\mu^2\gamma_2$) is smaller than the variance function of the exponential distribution, that is $\mu^2\gamma_2 \leq \mu^2$ for all $\mu > 0$. As any variance has to be nonnegative, this gives the other inequality. Obviously, we are assuming the mean of the exponential distribution model is also the mean of the mixing distribution. Finally, note the variance of the local scale mixture model is bounded in the following way: $\mu^2 \leq V[X;\mu,\gamma_2] \leq 3\mu^2$.

Let us now check for which parameters values, the local scale mixture model (4) is a true local mixture model. The answer is the empty set as the following theorem shows.

Corollary 1 *Local scale mixture model* (4) *is not a true local mixture for any value of its parameters.*

Proof We need to check that the survival function satisfies conditions (2) of the theorem. Note that, for $k = 1$, the condition is just the nonnegativity condition that defines the hard boundary and therefore is always satisfied. It is easy to check that the other conditions are equivalent to

$$0 \leq \gamma_2 \leq \frac{2}{k+1} \quad k = 2,3,\ldots.$$

So, the model can never be a genuine mixture as the only parameter values for that to happen are any $\mu > 0$ but $\gamma_2 = 0$, which correspond to a degenerate mixing distribution. □

As this theorem states that the local scale mixture model (4) cannot be represented as a mixture of $\mathcal{F}$ exactly, then it makes sense to restrict its parameter values to satisfy relevant necessary conditions for being an exact mixture. See for example Lindsay, (1995) for a range of such conditions. This lead us to the following definition.

Definition 4 *A* soft boundary *for a local scale mixture model $\mathcal{G}_{\mathcal{F}}$ is the boundary of any set contained in* Γ.

Example 3 The $k = 2$ condition in (2) implies that mixtures of exponentials must have a non-increasing density. In Fig. 2 we plot some densities of model (4) for a fixed value of μ. As can be seen from the plots, some of them have a small bump and therefore are not monotone. This happens for values of γ_2 close to 1. Then, restricting the local scale mixture model using the soft boundary imposed by the $k = 2$ condition on Theorem 1, allows only for non-increasing densities. That is, by using the parameter space $[0, 2/3]$ for γ_2. Then here $2/3$ is an example of a soft boundary. We can further restrict the natural parameter space by using a larger k, but the interpretation of the resulting boundary is not easy.

If $g(x; Q)$ is a mixture of $\mathcal{F}$ with mean $\bar{\mu}$ then $R(x) = g(x; Q)/f(x; \bar{\mu}) - 1$ is convex (see Shaked, 1980) and has the sign sequence $(+, -, +)$ as x transverses the real axis. For model (4) we have

$$R(x) = \frac{\mu^2 \gamma_2}{2} \left(\frac{f^{(2)}(x; \mu)}{f(x; \mu)} \right) = \gamma_2 \left(1 - \frac{2x}{\mu} + \frac{x^2}{2\mu^2} \right)$$

which is clearly convex for all $\mu > 0$ and $\gamma \in [0, 1]$. In fact, we can explicitly see the sign changes in Fig. 2. The unmixed exponential density is plotted for reference as a solid line.

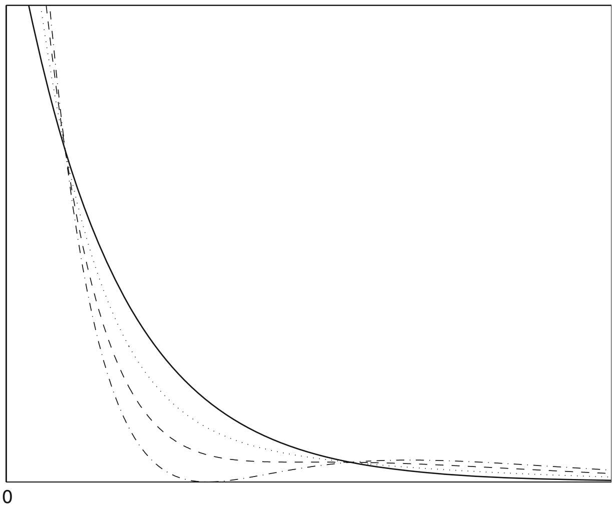

Fig. 2 Some densities of model (4) (*solid* is the unmixed exponential)

Moreover, if $g(x;Q)$ is a mixture of $\mathcal{F}$ then we must have

$$E_g[X^r] \geq E_f[X^r], \qquad r \geq 2,$$

when g is a mixture with the same mean as f. This is true in this case as the function x^k for $k \geq 1$ is convex on the positive real line. For model (4) we have

$$E_g[X^r] = \mu^r \left[r! + \frac{r(r-1)}{2}\gamma_2 \right],$$

so the previous set of inequalities translates to $\gamma_2 \geq 0$ which is always true. So, for the local mixture model (4), that set of moments is always bigger than the corresponding set of the unmixed model.

Model (4) has been implicitly used in many papers related to testing for the presence of mixing in the exponential model. Some relevant references are Mosler and Seidel (2003), Jaggia (1997), Chang and Suchindran (1997) and Kiefer (1984). Model (4) is implicitly used to construct a dispersion score test statistic. For example, Kiefer assumes a mixture model of the form

$$g(x;Q_1) = \int \tilde{f}(x;\varphi + u)\mathrm{d}Q_1(u),$$

where Q_1 has mean zero and $\tilde{f}(x;\phi) = f(x;e^{-\phi})$. That is, ϕ is the logarithm of the reciprocal of the mean. Jaggia assumes a mixture of the form

$$g(x;Q_2) = \int \tilde{\tilde{f}}(x;\vartheta\, v)\mathrm{d}Q_2(v),$$

where now Q_2 has mean one and $\tilde{\tilde{f}}(x;\theta) = f(x;1/\theta)$. That is, θ is the rate parameter. Using a Taylor expansion argument they obtain the following approximations to $g(x;Q_1)$ and $g(x;Q_2)$, respectively,

$$\tilde{f}(x;\varphi) + \frac{v_1}{2}\tilde{f}^{(2)}(x;\varphi) = \tilde{f}(x;\varphi)\left[1 + \frac{v_1}{2}\left\{1 - 3xe^{\varphi} + x^2e^{2\varphi}\right\}\right]$$

$$\tilde{\tilde{f}}(x;\vartheta) + \frac{v_2}{2}\tilde{\tilde{f}}^{(2)}(x;\vartheta) = \tilde{\tilde{f}}(x;\vartheta)\left[1 + \frac{v_2}{2}\left\{\frac{\vartheta x^2 - 2x}{\vartheta}\right\}\right],$$

where $v_1 := Var_{Q_1}[\phi]$ and $v_2 := Var_{Q_2}[\theta]$. These expressions define local mixture models closely related to (4). They correspond to simple reparametrizations of a local scale mixture model as in Definition 2 but where the family $\mathcal{F}$ is reparametrized using the log reciprocal mean and the rate parameter, respectively.

Using these approximations, Kiefer (1984) and Jaggia (1997), obtain the corresponding dispersion score statistics, which are given by

$$DS_1(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\left[1 - 3x_i e^{\hat{\varphi}} + x_i^2 e^{2\hat{\varphi}}\right] = \frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\bar{x}^2} - 1,$$

$$DS_2(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\hat{\vartheta}x_i^2 - 2x_i}{\hat{\vartheta}}\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 - \bar{x}^2$$

where $\hat{\varphi}$ and $\hat{\vartheta}$ are the maximum likelihood estimates of φ and ϑ, respectively, under the assumption of no mixing, that is, under the assumption that each observation in the sample follows an exponential distribution with unknown mean $e^{-\varphi}$ and $1/\vartheta$, respectively.

Both statistics have a simple interpretation. $DS_1(\boldsymbol{x})$ is the relative difference between the sample variance and the variance under the model, and $DS_2(\boldsymbol{x})$ is the absolute difference between the sample variance and the variance under the model. So, when the sample variance exceeds the variance under the model we have empirical evidence of mixing. If fact, if $\mathcal{F}$ is extended to be an exponential family expressed in its natural parametrization, then the dispersion score always has the form

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 - V(\bar{x}),$$

where V is the variance function of $\mathcal{F}$. Lindsay (1989) shows that in the case when $\mathcal{F}$ is a natural exponential family with quadratic variance function, it is more informative to use the mean parametrization in the following sense. Lindsay shows that, for any known μ_0 and $k = 1, 2, \ldots$

$$\widehat{m}_{0,k} = \frac{1}{n}\sum_{i=1}^{n}\frac{k!}{a_k}P_k(x_i; \mu_0)$$

is an unbiased estimator of

$$m_{0,k} := E_Q[(\mu - \mu_0)^k]$$

for any mixing distribution Q and some constants a_k. The functions $P_k(x_i; \mu_0)$ are the orthogonal polynomials associated with the natural exponential family. In our exponential case we have

$$\widehat{m}_{0,2} = \frac{1}{2n}\sum_{i=1}^{n}\left[2\mu_0^2 - 4\mu_0 x_i + x_i^2\right]$$

is an unbiased estimator of $E_Q[(\mu - \mu_0)^2]$. Since we want to estimate the variance, substituting μ_0 by $\bar{x}$ we obtain $DS_2(\boldsymbol{x})/2$. So, the dispersion score $DS_2(\boldsymbol{x})$ can be regarded as an estimator of the variance of the mixing distribution.

At this point, it is convenient to consider the following reparametrization of Model (4). Define $\eta_2 = \mu^2\gamma_2$. Clearly, η_2 will play the same role as the variance of the mixing distribution. Simple moment estimators of μ and η_2 can be obtained, namely

$$\hat{\mu}^{\text{mom}} = \bar{x}$$
$$\widehat{\eta_2}^{\text{mom}} = \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 - \bar{x}^2\right] = \frac{DS_2(\boldsymbol{x})}{2}.$$

Clearly, we also have

$$\widehat{\gamma_2}^{\text{mom}} = \frac{1}{2}\left[\frac{1}{n}\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\bar{x}^2} - 1\right] = \frac{DS_1(\boldsymbol{x})}{2}.$$

It is now clear that $DS_1(\boldsymbol{x})$ can be considered as a simple moment estimator of the squared of the coefficient of variation of the mixing distribution. However, we can restrict the values of this estimator to be inside the interval $[0, 2/3]$ although this is not part of the definition of the dispersion score. Moreover, Darling (1953) shows that

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\bar{x}^2} = n[DS_1(\boldsymbol{x}) + 1]$$

defines a (right sided) locally most powerful test against mixtures of exponentials and also derives its asymptotic distribution under the hypothesis of no mixing. See also O'Reilly and Stephens (1982) for other tests of fit of the exponential distribution.

Now, the local scale mixture model of order $d = 4$ can be written as

$$g(x; \mu, \gamma_2, \gamma_3, \gamma_4) = \frac{1}{\mu}\exp\left(-\frac{x}{\mu}\right) p_4(x; \mu, \gamma_2, \gamma_3, \gamma_4), \tag{5}$$

where $p_4(x; \mu, \gamma_2, \gamma_3, \gamma_4)$ is the quartic polynomial

$$\left(\frac{\gamma_4}{24}\right)\frac{x^4}{\mu^4} + \left(\frac{\gamma_3}{6} - \frac{2\gamma_4}{3}\right)\frac{x^3}{\mu^3} + \left(\frac{\gamma_2}{2} - \frac{3\gamma_3}{2} + 3\gamma_4\right)\frac{x^2}{\mu^2}$$
$$+ (-2\gamma_2 + 3\gamma_3 - 4\gamma_4)\frac{x}{\mu} + (1 + \gamma_2 - \gamma_3 + \gamma_4).$$

It is straightforward to prove that this model is also a true local mixture model only if $\gamma_2 = \gamma_3 = \gamma_4 = 0$. In fact, it is a general result that local scale mixture models of the exponential distribution are never true local mixtures.

The parameter space for $(\gamma_2, \gamma_3, \gamma_4)$ contains negative values of γ_2. So, the first soft boundary we are going to impose is the obvious one $\gamma_2 \geq 0$, since the squared coefficient of variation of any distribution must be positive. In model (4) we got this boundary for free.

From expansions (11) at the end of the proof of Theorem 3, we observe that

$$\frac{E_Q[(\mu - E_Q[\mu])^4]}{(E_Q[\mu])^4} \sim 3 \left[\frac{E_Q[(\mu - E_Q[\mu])^2]}{(E_Q[\mu])^2} \right]^2 + O(\epsilon^3).$$

So, the normalized moments behave like that for small ϵ. We can therefore restrict the parameter values by

$$\gamma_4 = 3\,\gamma_2^2. \tag{6}$$

It is interesting to note that this restriction is forcing the pseudo coefficient of kurtosis to be zero, as in the normal distribution. Clearly, for each μ, the resulting model is a curved model embedded in model (5). Note this restriction automatically makes γ_4 positive as desired.

The parameter space for (γ_2, γ_3) can be easily characterized using the results of Ulrich and Watson, (1994). The hard boundary for the parameters (γ_2, γ_3) is plotted with a thick line in Fig. 3. Also in Fig. 3 we plotted the $k = 2$ soft boundary described by Theorem 1. Recall this boundary implies monotone decreasing densities.

The parametrization in terms of (γ_2, γ_3) has a drawback, as it takes into account values which corresponds to small variance but relatively high skewness of the mixing distribution. One such case is indicated with a cross in Fig. 3. Clearly, those parameters values are not compatible with the small mixing assumption.

Note that a mixing distribution can be small in two quite different ways. It is convenient to denote the distinction with two separate classes. The first will be called *Laplace type mixing* where the mixing distribution is of small variance and unimodal. The second type will be called *contamination type mixing* where, although the variance is small, there can be a small proportion of the realized values a long way from the mean. Such a class can have more than one mode and in general show high skewness. For small enough ϵ, proper dispersion models are of the Laplace type because they are asymptotically normal (see Jorgensen, 1997). We have distinguished this contamination type because empirically we want to avoid this latter kind of mixing in our models by imposing soft boundaries such as the following.

Consider the following reparametrization of the model defined by (6)

$$\begin{aligned} \gamma_2 &= \alpha \\ \gamma_3 &= (3 - \beta)\,\alpha^2. \end{aligned} \tag{7}$$

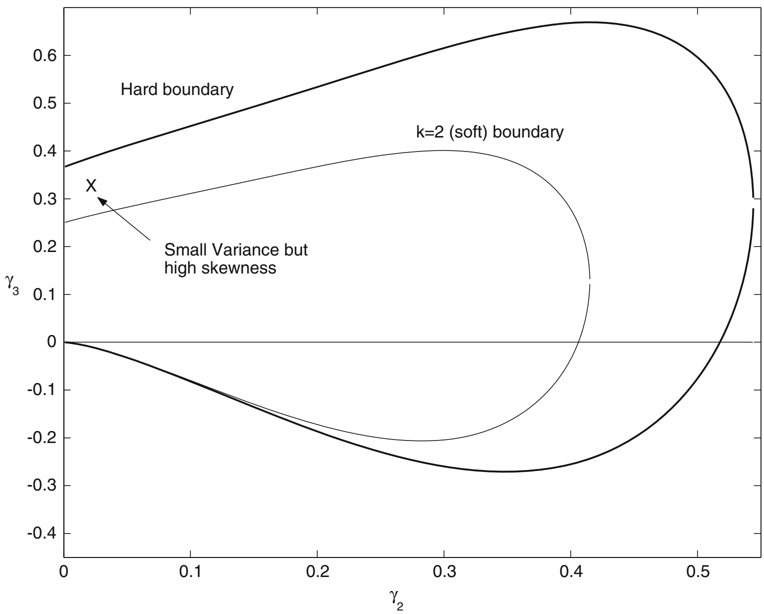

Fig. 3 Boundaries of model (5) with $\gamma_4 = 3\gamma_2^2$

Note that, for each fixed value of μ and β, the induced subfamily is a curved subfamily embedded in model (5) via the mapping

$$\alpha \mapsto (\alpha, (3-\beta)\alpha^2, 3\alpha^2).$$

Note that this is exactly mimicking the behavior (up to $O(\epsilon^3)$) of the normalized moments of a scale dispersion mixing model for small ϵ according to the expansions in (11).

Those subfamilies are plotted with dashed curves for a range of values of β in Fig. 4. Then it is clear from the figure that restricting the values of the parameters β to a specific interval of the form $[C_1, C_2]$ will avoid contamination type mixing distributions. One important fact is that each of these curved mixture families is converging very slowly to the boundary $\gamma_2 = 0$, so not very extreme values of C_1 and C_2 are enough to capture the local behavior of a big range of scale dispersion mixing models.

For instance, the interval $[C_1, C_2] = [-1, 1]$ covers the behavior of the generalized inverse gaussian distribution. This family satisfies $d_0^{(3)} = 2\beta - 6$. Also, from (11) we observe that

$$\frac{E_{Q_\beta}[(\mu - E_{Q_\beta}[\mu])^3]}{(E_{Q_\beta}[\mu])^3} \sim -\frac{d_0^{(3)}}{2}\left[\frac{E_{Q_\beta}[(\mu - E_{Q_\beta}[\mu])^2]}{(E_{Q_\beta}[\mu])^2}\right]^2 + O(\epsilon^3),$$

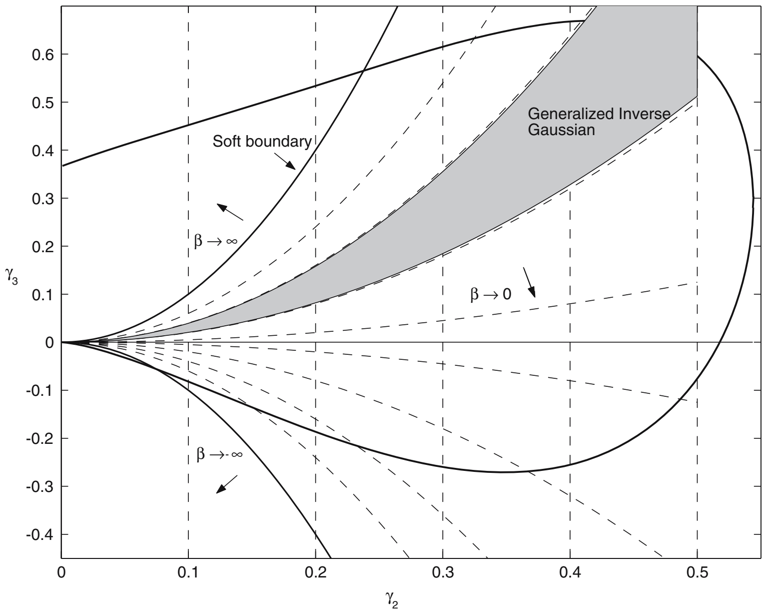

Fig. 4 Visual representation of reparametrization (7)

which justifies the particular form of parametrization (7) above. But parametrization (7) is just a matter of convenience. Actually, a single value of $d_0^{(3)}$ might correspond to different distributions. For example, if $\beta = 0$, the value of $d_0^{(3)} = -6$ corresponds to either the lognormal (which is not included in the generalized inverse gaussian family) or the hyperbola distribution.

Finally, we can alternatively impose soft boundaries by using, for example, the inequality obtained by Klaassen et al. (2000), relating the normalized moments of a unimodal distribution. For the model defined by (6), this inequality results in the cusp

$$\gamma_3^2 \leq \gamma_4\gamma_2 - \frac{3}{2}\gamma_2^3 = \frac{3}{2}\gamma_2^3. \tag{8}$$

This soft boundary also avoids contamination type mixing distributions.

6 Simulation study

For illustration purposes, we focus here only on the local scale mixture model defined by (6). As a simple simulation exercise, we generated 10,000 independent replications of a random sample of size $n = 1,000$ from a scale dispersion mixture with Reciprocal Gamma mixing distribution with mean equal to 5 and squared coefficient of variation $\gamma_2 = 0.1$ and $\gamma_2 = 0.25$.

For small to moderate sample sizes, there will typically be insufficient information in the data to estimate the parameters γ_i (large samples be necessary

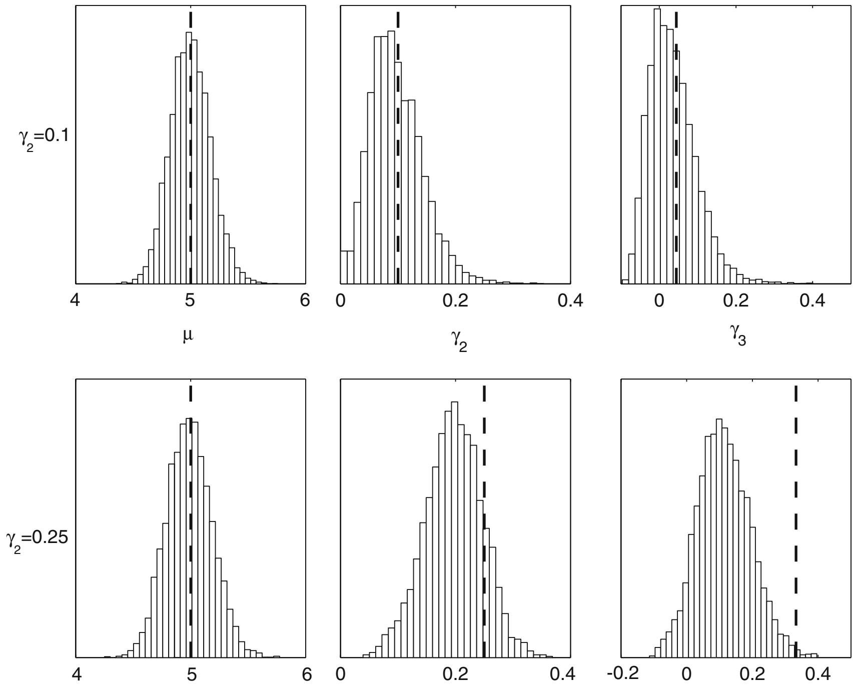

Fig. 5 MLE's for $(\mu, \gamma_2, \gamma_3)$

for that). But this is not very important if interest is centered, for example, in the mean of the mixture distribution.

Figure 5 shows the histograms of the maximum likelihood estimators of μ, γ_2 and γ_3 under model (6) in both situations. The correct value of the parameter is plotted with a dashed vertical line. It is clear that the local scale mixture model *only makes sense as an approximation of a genuine scale dispersion mixture when the mixing distribution has small squared coefficient of variation* γ_2. Otherwise, it can be considered simply as a generalization of the base family $\mathcal{F}$. The local scale mixture model is clearly trying to fit a mixing distribution with smaller coefficient of variation and smaller third normalized moment in the case where $\gamma_2 = 0.25$ and this produces the biases shown in Fig. 5.

This underestimation property is not a big disadvantage of local scale mixture models as generally the information obtained about the pseudo squared coefficient of variation is enough (as shown below) to draw some conclusions about the presence of mixing, which is usually of interest in applications.

Also note that, when the true $\gamma_2 = 0.25$, inferences about μ remain essentially the same. That is, even though the local scale mixture is underestimating the dispersion structure, inferences about the mean does not change. This is a result of the orthogonality property in Theorem 2, as well as the fact that the parameter μ (the mean of the unmixed negative exponential) retains its meaning under the local scale mixture model [recall first equation in (3)].

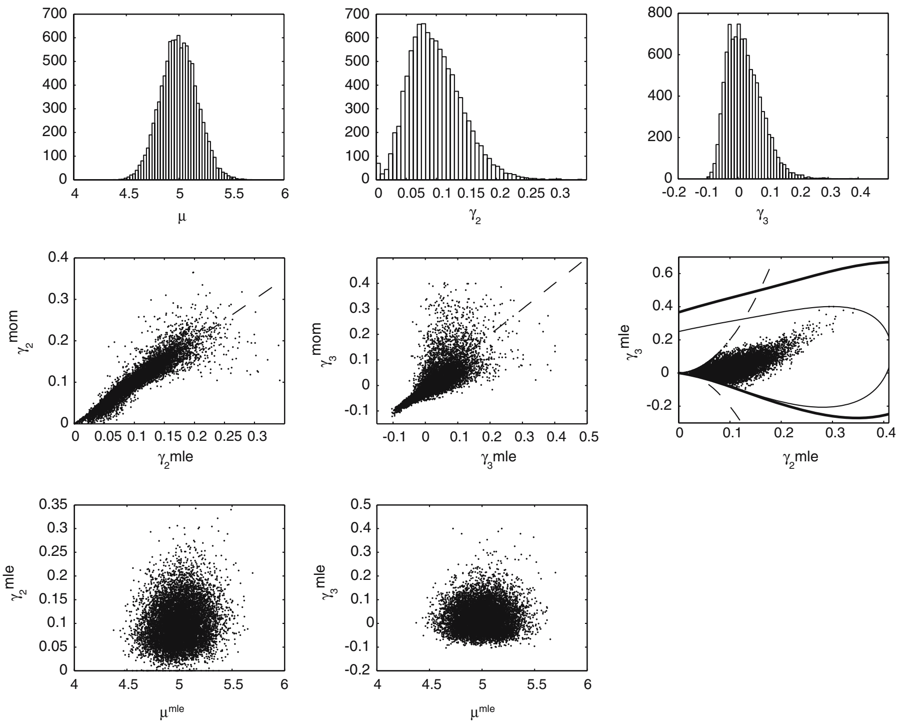

Fig. 6 Gamma mixing with $\mu = 5$ and $\gamma_2 = 0.1$. Sample size 1,000

We also generated 10,000 independent replications of a random sample of size $n = 1{,}000$ from scale dispersion mixtures with Gamma mixing distributions of mean $\mu = 5$ and squared coefficient of variation $\gamma_2 = 0.1$. Some results of both sets of simulations are presented in Figs. 6 and 7.

In the first row of each figure, we show the scatter plots of the estimators $(\hat{\gamma}_2^{\text{mle}}, \hat{\gamma}_2^{\text{mom}})$ and $(\hat{\gamma}_3^{\text{mle}}, \hat{\gamma}_3^{\text{mom}})$. We found a good agreement between both estimators only for small values. Recall that $\hat{\gamma}_2^{\text{mom}}$ is equivalent to a score test statistic. This means that, that the mle's of γ_2 and γ_3 contain valuable information in testing for the presence of mixing. In comparison to traditional application of score tests statistics, our estimators have the advantage of incorporating information about the skewness and possibly higher order moments of the mixing distribution.

Also, in the same row, is the scatter plot of $(\hat{\gamma}_2^{\text{mle}}, \hat{\gamma}_3^{\text{mle}})$. To explain this plot we first need to mention how we calculated the mle's. We judiciously chose as soft boundary for the parameter β the interval $[C_1, C_2] = [-20, 20]$. For example, the generalized inverse Gaussian family is, by far, contained in such an interval. In practice, we will be required to fix an interval of that form if we really want to model local mixtures in a meaningful way, so the interval we have chosen seems reasonable.

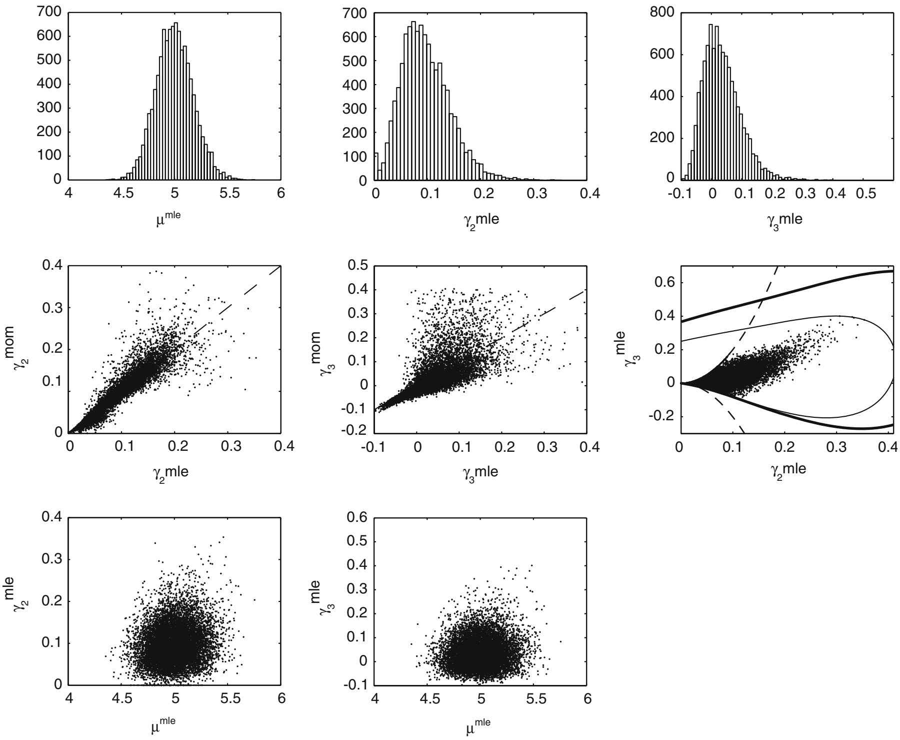

Fig. 7 Reciprocal gamma mixing with $\mu = 5$ and $\gamma_2 = 0.1$. Sample size 1,000

We plotted with a thick line the hard boundary for (γ_2, γ_3) and with a dotted line the $k = 2$ boundary. With a dashed line is plotted the $-20 \leq \beta \leq 20$ boundary. The joint distribution of the mle of (γ_2, γ_3) can reasonably be well approximated by a distribution with elliptic contours. It is clear also, that there exists a positive correlation between both estimators. This is just a consequence of the fact that, under the mixing distribution, there exists a relationship of the same kind between the squared coefficient of variation and the third normalized moment for small values of the dispersion parameter ϵ. The points that stick into the soft boundaries correspond to samples which are discordant with the small mixing assumption, for example, that have large upper order statistics and therefore represent evidence of contamination type mixing. In this sense, those soft boundaries can be used as a simple diagnostic for detecting non-local mixing.

In the last row of each figure we present the scatter plots of $(\hat{\mu}^{\text{mle}}, \hat{\gamma}_2^{\text{mle}})$ and $(\hat{\mu}^{\text{mle}}, \hat{\gamma}_3^{\text{mle}})$. We found evidence of elliptically contoured joint distributions and approximate orthogonality between the estimators. Finally, we mention that we found similar results to the previous simulations using other local scale mixture models not presented in this paper.

7 Discussion

The essential idea of a local mixture model is that the mixing is only responsible for a relatively small amount of variation in the model. In many practical applications, the unmixed model $\mathcal{F}$ explains most of the variation and the mixing only adds a small component to improve the fit of the baseline model.

In the case of local scale mixture models of the exponential distribution, the implicit assumption of a scale dispersion mixing, seems reasonable from this point of view and also, because most of the behavior of a general mixture of exponentials is essentially determined by the coefficient of variation of the mixing distribution (Keilson and Steutel, 1974). The locality implicit in our models should always has to be in mind when trying to make more global conclusions about the mixture structure. For an extension of local mixture models, to cover global properties of mixtures, see the paper by Marriott in this volume.

In this paper, we showed that simple likelihood and moment inference methods can give considerable insight in the analysis of mixtures of the exponential distribution. Alternatively, Bayesian inference can be easily implemented to give also very informative inference summaries. See Marriott (2002) and Anaya-Izquierdo and Marriott (2006).

Local scale mixtures of the exponential distribution are related with other typical models to analyze positive data. For example, the relationship with the Weibull distribution can be understood within the framework of example 1 applying an appropriate limit. See chapter 4 of Anaya-Izquierdo (2006) for more details. A less clear relation with the gamma distribution arises, for example, when one notes that the model in expression (4) can be written as an affine combination of three different models: a exponential distribution with mean μ, a gamma distribution with mean 2μ and a gamma distribution with mean 3μ.

Finally, our results can be modified to handle natural exponential families with quadratic variance function other than the exponential distribution. See Anaya-Izquierdo and Marriott (2006).

Acknowledgments Part of this work was undertaken while the authors were visiting the Institute of Statistics and Decision Sciences, Duke University. The authors would like to thank two anonymous referees and also Professor Shinto Eguchi for many helpful suggestions. K. Anaya-Izquierdo would also like to thank Universidad Nacional Autónoma de México and Consejo Nacional de Ciencia y Tecnología México for their financial support.

A Appendix on asymptotic expansions

Here we describe the asymptotic behavior of scale dispersion mixtures of $\mathcal{F}$ when $\epsilon \downarrow 0$. To find an asymptotic expansion of (1) it is tempting to expand the density $f(x;\mu)$ in a Taylor series around $\mu = m$ and then perform termwise integration, but such procedure is only justified when the integrand is exponentially decaying. In this respect, note that any scale dispersion mixture can be expressed as the ratio of two integrals of the form

$$I_{x,m}(\epsilon) = \int_0^\infty H(x,\mu)\exp\left(-\frac{h_m(\mu)}{\epsilon}\right)\mathrm{d}\mu \tag{9}$$

for some positive functions H and h_m. The important issue is that, for each m, the function $h_m(\mu)$ has an absolute minimum at $\mu = m$. This structure is particularly suitable for the application of Laplace's Method. Laplace's method gives an asymptotic series expansion for $I_{x,m}(\epsilon)$ valid when $\epsilon \downarrow 0$. For details on Laplace's method see Wong (2001).

So, we can always obtain an asymptotic series expansion for each numerator and denominator and then divide those to obtain an expansion for any scale dispersion mixture as the following theorem shows. Here, $f^{(i)}(x;m)$ denotes the derivatives of $f(x;\mu)$ with respect to μ evaluated at $\mu = m$.

Theorem 3 *Let $\mathcal{F} = \{f(x;\mu) : \mu > 0\}$ be the exponential distribution. Then, for any $r \in \mathbb{N}$, the scale dispersion mixture $g(x;Q(\cdot,m,\epsilon))$ has the following asymptotic expansion:*

$$f(x;M_1(m,\epsilon)) + \sum_{i=2}^{2r}\frac{[M_1(m,\epsilon)]^i N_i(\epsilon)}{i!} f^{(i)}(x;M_1(m,\epsilon)) + O_{x,m}(\epsilon^{r+1}) \tag{10}$$

as $\epsilon \downarrow 0$. The functions $M_1(m,\epsilon)$ and $N_i(\epsilon)$ satisfy

$$E_Q[\mu] \sim M_1(m,\epsilon) = m\left[1 + O(\epsilon)\right],$$

$$N_i(\epsilon) = O(\epsilon^{u(i)}),$$

$$\frac{E_Q[(\mu - E_Q[\mu])^i]}{(E_Q[\mu])^i} \sim N_i(\epsilon) + O(\epsilon^{r+1}), \quad i = 2,3,\ldots,2r$$

as $\epsilon \downarrow 0$ and where $u(i) = \lfloor (i+1)/2 \rfloor$. Here $\lfloor x \rfloor$ means rounding towards $-\infty$.

Proof This is just a slightly different version of one of the Theorems in Anaya-Izquierdo and Marriott (2006). We will only prove the case $r = 2$. Using Laplace's method in both numerator and denominator of $g(x;Q(\mu,m,\epsilon))$ and then dividing the two series, one obtains, (after a considerable amount of algebra),

$$g(x;Q(\mu;m,\epsilon)) \sim f(x;m) + \sum_{i=1}^{4} m^i A_i^*(\epsilon) f^{(i)}(x;m) + O_{x,m}(\epsilon^3)$$

as $\epsilon \downarrow 0$ where

$$A_1^*(\epsilon) = -\epsilon\left(1 + \frac{d_0^{(3)}}{4}\right) + \epsilon^2\left(\frac{d_0^{(3)}d_0^{(4)}}{6} - 2 + \frac{d_0^{(4)}}{4} - \frac{d_0^{(5)}}{16} - \frac{3d_0^{(3)}}{4} - \frac{5[d_0^{(3)}]^3}{64} - \frac{[d_0^{(3)}]^2}{4}\right)$$

$$A_2^*(\epsilon) = \frac{\epsilon}{2} + \epsilon^2\left(\frac{5[d_0^{(3)}]^2}{32} + 1 - \frac{d_0^{(4)}}{8} + \frac{d_0^{(3)}}{2}\right)$$

$$A_3^*(\epsilon) = -\epsilon^2\left(\frac{1}{2} + \frac{5d_0^{(3)}}{24}\right)$$

$$A_4^*(\epsilon) = \frac{\epsilon^2}{8},$$

and $d_0^{(i)}$ for $i = 3, 4, 5$ are the derivatives of $d_0(u)$ evaluated at $u = 1$. Now we define

$$M_1^*(m, \epsilon) := m\left[1 + \delta(\epsilon)\right]$$

$$M_2(\epsilon) := \frac{\epsilon}{2} + \epsilon^2\left[\frac{[d_0^{(3)}]^2}{8} + \frac{1}{2} - \frac{d_0^{(4)}}{8} + \frac{d_0^{(3)}}{4}\right]$$

$$M_3(\epsilon) := -\frac{\epsilon^2 d_0^{(3)}}{12}$$

$$M_4(\epsilon) := \frac{\epsilon^2}{8},$$

where $\delta(\epsilon) := A_1^*(\epsilon) + O(\epsilon^3)$. Using Taylor's Theorem on $f(x; M_1^*(m, \epsilon))$ and $f^{(i)}(x; M_1^*(m, \epsilon))$, we obtain that for small $\delta(\epsilon)$

$$f(x; M_1^*(m, \epsilon)) + \sum_{i=2}^{4} m^i M_i(\epsilon) f^{(i)}(x; M_1^*(m, \epsilon))$$

$$= f(x; m[1 + \delta(\epsilon)]) + \sum_{i=2}^{4} m^i M_i(\epsilon) f^{(i)}(x; m[1 + \delta(\epsilon)])$$

$$\sim f(x; m) + \sum_{i=1}^{4} m^i A_i^*(\epsilon) f^{(i)}(x; m) + O_{x,m}(\epsilon^3).$$

Now

$$f(x;M_1^*(m,\epsilon)) + \sum_{i=2}^{4} m^i M_i(\epsilon) f^{(i)}(x;M_1^*(m,\epsilon))$$
$$= f(x;M_1^*(m,\epsilon)) + \sum_{i=2}^{4} \left[M_1^*(m,\epsilon)\right]^i \frac{M_i(\epsilon)}{\left[1+\delta(\epsilon)\right]^{i}} f^{(i)}(x;M_1^*(m,\epsilon)) .$$

By defining

$$N_i^*(\epsilon) := \frac{i!\, M_i(\epsilon)}{\left[1+\delta(\epsilon)\right]^{i}},$$

it is easy to check that simple expansions of $N_i^*(\epsilon)$ for small ϵ, coincide (up to $O(\epsilon^3)$) with the following expansions,

$$\frac{E_Q[(\theta - E_Q[\theta])^2]}{(E_Q[\theta])^2} \sim \epsilon + \left[\frac{12 + 4d_0^{(3)} - d_0^{(4)} + [d_0^{(3)}]^2}{4}\right] \epsilon^2 + O(\epsilon^3)$$

$$\frac{E_Q[(\theta - E_Q[\theta])^3]}{(E_Q[\theta])^3} \sim -\frac{d_0^{(3)}}{2} \epsilon^2 + O(\epsilon^3)$$

$$\frac{E_Q[(\theta - E_Q[\theta])^4]}{(E_Q[\theta])^4} \sim 3\epsilon^2 + O(\epsilon^3). \tag{11}$$

□

Under a reparametrization of $\mathcal{F}$, the scale dispersion structure of the mixing distribution can be lost but the dispersion structure is preserved, see Jorgensen (1997) for details. In such a case, an appropriate variant of Theorem 3 holds. See Anaya-Izquierdo and Marriott (2006).

We are not very concerned about the behavior of the remainder terms, because it is not our aim to use these expansions to approximate scale dispersion mixtures in any analytical sense. We are only interested in the behavior of scale dispersion mixtures when ϵ is small. So, according to Theorem 3, up to an specific asymptotic order, the asymptotic behavior of this kind of mixtures depends on

1. the behavior of $f(x;\mu)$ near the mean of the mixing distribution through its higher order derivatives and
2. the mixing distribution only through the set $\{M_1(m,\epsilon), N_2(\epsilon), \ldots, N_{2r}(\epsilon)\}$ which is related to its moments.

The second point makes sense as when the mixing distribution is unimodal and sufficiently concentrated, then it can be very much determined by its first few moments. See Johnson and Rogers (1951) and Janson (1988). Laplace's expansions are modal expansions, but here we modify them to obtain a mean centered expansion. This explains the absence of the first derivative term. Details are in the proof. The pairing of asymptotic orders is explained by the scale dispersion mixing structure. In any scale dispersion model, the difference between the mean and the position m and also the coefficient of variation are of order ϵ. The third and fourth normalized moments are of order ϵ^2 and so on.

It is well known that the moment structure of any mixture is determined by the moment structure of the mixing distribution. A relevant reference being Lindsay (1989). The linear structure of the expansions, given by Theorem 3, gives a simple description of how the moments of the mixing distributions affects the moments of the observed random variable X. To clarify ideas, consider the following example.

Example 4 (Example 1 revisited) In the case where the mixing distribution is the reciprocal Gamma ($\beta = -1$), we obtain the following expansions by applying Theorem 3. For simplicity in notation, here we suppress the dependence of $M_1, f^{(i)}$ and the remainders on x, m and ϵ. We also write g_β for $g(x; Q_\beta)$. For the mean we obtain $M_1 \sim m\left[1 + \epsilon + \epsilon^2 + O(\epsilon^3)\right]$ and when $r = 1, 2, 3$ we obtain, respectively,

$$g_{-1} \sim f + M_1^2 \left[\frac{\epsilon}{2!}\right] f^{(2)} + O(\epsilon^2)$$

$$g_{-1} \sim f + M_1^2 \left[\frac{\epsilon + 2\epsilon^2}{2!}\right] f^{(2)} + M_1^3 \left[\frac{4\epsilon^2}{3!}\right] f^{(3)} + M_1^4 \left[\frac{3\epsilon^2}{4!}\right] f^{(4)} + O(\epsilon^3)$$

$$g_{-1} \sim f + M_1^2 \left[\frac{\epsilon + 2\epsilon^2 + 4\epsilon^3}{2!}\right] f^{(2)} + M_1^3 \left[\frac{4\epsilon^2 + 20\epsilon^3}{3!}\right] f^{(3)}$$
$$+ M_1^4 \left[\frac{3\epsilon^2 + 42\epsilon^3}{4!}\right] f^{(4)} + M_1^5 \left[\frac{40\epsilon^3}{5!}\right] f^{(5)} + M_1^6 \left[\frac{15\epsilon^3}{6!}\right] f^{(6)} + O(\epsilon^4).$$

For the Hyperbola case ($\beta = 0$), we obtain $M_1 \sim m\left[1 + \epsilon/2 - \epsilon^2/8 + O(\epsilon^3)\right]$ and for $r = 1, 2, 3$, respectively.

$$g_0 \sim f + M_1^2 \left[\frac{\epsilon}{2!}\right] f^{(2)} + O(\epsilon^2)$$

$$g_0 \sim f + M_1^2 \left[\frac{\epsilon}{2!}\right] f^{(2)} + M_1^3 \left[\frac{3\epsilon^2}{3!}\right] f^{(3)} + M_1^4 \left[\frac{3\epsilon^2}{4!}\right] f^{(4)} + O(\epsilon^3)$$

$$g_0 \sim f + M_1^2 \left[\frac{\epsilon - 3/8\epsilon^3}{2!}\right] f^{(2)} + M_1^3 \left[\frac{3\epsilon^2 - \epsilon^3/2}{3!}\right] f^{(3)}$$
$$+ M_1^4 \left[\frac{3\epsilon^2 + 15\epsilon^3}{4!}\right] f^{(4)} + M_1^5 \left[\frac{30\epsilon^3}{5!}\right] f^{(5)} + M_1^6 \left[\frac{15\epsilon^3}{6!}\right] f^{(6)} + O(\epsilon^4)$$

What we can learn from this example is that (for sufficiently small ϵ) the asymptotic expansions given by Theorem 3 can be very similar to each other (with the same value of r), even if the mixing distribution is different. Moreover, a small value of r can describe well the asymptotic behavior of the mixture because the higher order moments become negligible for a big value of r.

References

Amari, S., Nagaoka, H. (2000). *Methods of information geometry*. American Mathematical Society.

Anaya-Izquierdo, K.A. (2006). Statistical and geometrical analysis of local mxiture models and a proposal of some new tests of fit for censored data. PhD Thesis. Universidad Nacional Autónoma de México.

Anaya-Izquierdo, K.A., Marriott, P.K. (2006). Local mixture models of exponential families (Submitted).

Chang, H.-Y., Suchindran, C.M. (1997). Testing overdispersion in data with censoring using the mixture of exponential families. *Communications in Statistics: Theory and Methods*, *26*, 2945–2966.

Critchley, F., Marriott, P. (2004). Data-informed influence analysis. *Biometrika*, *91*, 125–140.

Darling, D.A. (1953). On a class of problems related to the random division of an interval. *Annals of Mathematical Statistics*, *24*, 239–253.

Embrechts, P., Kluppelberg, C., Mikosch, T. (1997). *Modelling extremal events: For insurance and finance*. Berlin Heidelberg New York: Springer.

Feller, W. (1970). *An Introduction to Probability Theory and it Applications*. London: Wiley.

Heckman, J.J., Robb, R., Walker, J.R. (1990). Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments. *Journal of the American Statistical Association*, *85*, 582–589.

Jaggia, S. (1997). Alternative forms of the score test for heterogeneity in a censored exponential model. *The Review of Economics and Statistics*, *79*(2), 340–343.

Janson, S. (1988). Normal convergence by higher order semiinvariants with applications to sums of dependent random variables and random graphs. *Annals of Applied Probability*, *16*(1), 305–312

Jewell, N.P. (1982). Mixtures of exponential distributions. *The Annals of Statistics*, *10*, 479–484.

Johnson, N.L., Rogers, C.A. (1951). The moment problem for unimodal distributions. *Annals of Mathematical Statistics*, *22*, 432–439.

Jorgensen, B. (1997). *The theory of dispersion models*. London: Chapman & Hall.

Keilson, J., Steutel, F.W. (1974). Mixtures of distributions, moment inequalities and measures of exponentiality and normality. *The Annals of Probability*, *2*, 112–130.

Kiefer, N.M. (1984). A simple test for heterogeneity in exponential models of duration. *Journal of Labor Economics*, 3–4, 539–549.

Klaassen, C.A.J., Mokveld, P.J., van Es, B. (2000). Squared skewness minus kurtosis bounded by 186/125 for unimodal distributions. *Statistics & Probability Letters*, *50*, 131–135.

Lindsay, Bruce G. (1989). Moment matrices: applications in mixtures. *The Annals of Statistics*, *17*(2), 722–740.

Lindsay, B.G. (1995). *Mixture models: theory, geometry, and applications*. Institute of Mathematical Statistics.

Marriott, P. (2002). On the local geometry of mixture models. *Biometrika*, *89*(1), 77–93.

Marriott, P. (2003). On the geometry of measurement error models. *Biometrika*, *90*(3), 567–576.

McLachlan, G.J., Peel, D. (2001). *Finite mixture models*. London: Wiley.

Mosler, K., Seidel, W. (2001). Testing for Homogeneity in an Exponential Mixture Model. *Australian & New Zealand Journal of Statistics*, *43*(2), 231–247.

O'Reilly, F., Stephens, M.A. (1982). Characterizations and goodness-of-fit tests. *Journal of the Royal Statistical Society, Series B*, *44*, 353–360.

Shaked, M. (2003). On mixtures from exponential families. *Journal of the Royal Statistical Society, Series B*, *42*, 192–198.

Ulrich, G., Watson, L.T. (1994). Positivity conditions for quartic polynomials. *SIAM Journal on Scientific Computing*, *15*(3), 528–544.

Wong, R. (2001). *Asymptotic approximations of integrals* (*Classics in applied mathematics*). Philadelphia: SIAM.