Ann. Inst. Statist. Math.
Vol. 52, No. 2, 215–230 (2000)

# MOMENT-BASED APPROXIMATIONS OF DISTRIBUTIONS USING MIXTURES: THEORY AND APPLICATIONS

BRUCE G. LINDSAY[1]*, RAMANI S. PILLA[2] AND PRASANTA BASAK[3]

[1] *Department of Statistics, 326 Classroom Building, Pennsylvania State University, University Park, PA-16802, U.S.A.*
[2] *National Institutes of Health, 6100 Executive Blvd., Room 7B-13, MSC 7510, Bethesda, MD 20892-7510, U.S.A.*
[3] *120 Eiche Library, The Pennsylvania State University, Altoona Campus, Altoona, PA-16601, U.S.A.*

(Received September 29, 1997; revised November 10, 1998)

**Abstract.** There are a number of cases where the moments of a distribution are easily obtained, but theoretical distributions are not available in closed form. This paper shows how to use moment methods to approximate a theoretical univariate distribution with mixtures of known distributions. The methods are illustrated with gamma mixtures. It is shown that for a certain class of mixture distributions, which include the normal and gamma mixture families, one can solve for a $p$-point mixing distribution such that the corresponding mixture has exactly the same first $2p$ moments as the targeted univariate distribution. The gamma mixture approximation to the distribution of a positive weighted sums of independent central $\chi^2$ variables is demonstrated and compared with a number of existing approximations. The numerical results show that the new approximation is generally superior to these alternatives.

*Key words and phrases*: Cumulants, cumulative distribution function, gamma mixtures, mixture distribution, moment matrix, $p$-point mixture, tail probability, weighted sums of chi-squares.

## 1. Introduction

There are a variety of statistical situations in which one can readily calculate or asymptotically approximate the higher order moments of a given target distribution, call it $H$, but the density and tail probabilities cannot be simply derived. An example is the distribution of weighted sums of independent central $\chi^2$ variables, each with one degree of freedom, i.e.,

$$S_n(\boldsymbol{d}) = \sum_{i=1}^{n} d_i W_i^2, \tag{1.1}$$

where $d_i$ are known positive weights and the $W_i$ are independent $N(0,1)$ random variables. The higher order cumulants of $S_n(\boldsymbol{d})$ are easily derived and one can then use recursion methods to calculate moments. (In Appendix A, we indicate how one can recursively solve for moments from cumulants and vice versa.)

We investigate here a general strategy of approximating a target distribution $H$, such as the distribution of (1.1), by a finite mixture of readily computed distributions,

* Dr. Lindsay's research was partially supported by the National Science Foundation under Grant DMS-9870193.

such as the gamma. The $p$-point mixture will be chosen so as to match the first $2p$ moments of $H$. The component densities of the mixture will be chosen both for ease of calculation and for theoretical considerations. For example, a mixture of normals is natural for a problem in which a central limit theorem holds.

We believe this approach is attractive for several reasons:

- There is some initial reason to think that matching moments will give a good approximation. For example, Lindsay and Roeder (1997) showed that if two distribution functions $H$ and $F$ have the same first $k$ moments, then they must cross each other at least $k$ times. Moreover, the possible error can be bounded by methods found in Lindsay and Basak (1995).
- Unlike some other approximation methods based on moments, such as Gram-Charlier series (Stuart and Ord (1987), p. 222), our approximation is a bona fide distribution function, with nonnegative density. And unlike yet other methods, such as saddle point approximations, it can be refined to arbitrary accuracy by taking enough terms.
- Some modern results on fitting by moments, derived in the context of method of moments estimators (Lindsay (1989*a*, 1989*b*) provide us with a set of theoretical and computational tools that make such approximations feasible.
- In matching the first $2p$ moments, we also match the first $2p$ cumulants, so there is a reason to expect good stochastic properties in certain central limit theorem-type problems.

A random variable $X$ or its distribution function $G(x)$ is said to have a *mixture distribution* relative to a parametric family of distributions $\{F_\theta : \theta \in \Omega\}$ if the distribution function has the form

$$G(x) = \int F_\theta(x) dQ(\theta),$$

where $Q$, the *mixing* or *latent distribution*, is a distribution on the parameter space $\Omega$. We consider here $Q$ to be a finite discrete distribution; in which case $G$ is called a finite mixture. In this case if $Q_p$ has $p$ support points $\theta_1, \ldots, \theta_p$ with probability masses $\pi_1, \ldots, \pi_p$ respectively, the *$p$-point mixture* distribution can be written as

$$G_p(x) = F_{Q_p}(x) = \sum_{j=1}^{p} \pi_j F_{\theta_j}(x).$$

In the families of mixtures we will be considering, there is an additional dispersion parameter $\lambda$ in each distribution, yielding mixtures of the form

$$G_p(x) = F_{Q_p,\lambda}(x) = \sum_{j=1}^{p} \pi_j F_{\theta_j,\lambda}(x). \tag{1.2}$$

Let $m_1 = m_1(H), \ldots, m_{2p} = m_{2p}(H)$ be the first $2p$ moments of a target distribution $H$. Our methodology will give us a $p$-point distribution $Q_p$ and a dispersion parameter $\lambda$ such that we match the moments of $H$ and $G_p$:

$$m_r = m_r(H) = m_r(F_{Q_p,\lambda}) \quad \text{for} \quad r = 1, \ldots, 2p. \tag{1.3}$$

Note that there are $2p$ free parameters in $F_{Q_p,\lambda}$; namely $\lambda, \theta_1, \ldots, \theta_p$ and $\pi_1, \ldots, \pi_{p-1}$, so we have the same number of equations as unknowns. Call $m_1, \ldots, m_{2p}$ the *target moments*. Thus we find a $p$-point mixture distribution $G_p(x)$ that matches the first $2p$ moments of the target distribution $H(x)$ and then use $G_p(x)$ to calculate the approximate tail probabilities.

Our strategy is based on a technique for fitting mixtures of normals that was developed by Lindsay (1989*b*). That theory is generalized here. (Some corrections are also made.) As an application, this paper focuses on fitting using mixtures of gamma distributions.

The paper starts with some numerical comparisons. Section 2 compares the mixture approximation to $S_n(\boldsymbol{d})$'s distribution with those given by Solomon and Stephens (1977) and Wood (1989). In these examples the exact cumulative distribution function (*CDF*) values are available and therefore one can compare the various approximations for their level of error. Section 3 develops the theory and methods for general mixtures while specializing the application to gamma mixtures. We note a key feature: at least in theory, able to fit an arbitrarily large number of moments and therefore as the number of moments increase, the sequence of approximations converges to the target distribution (assuming the distribution is determined by its moments). Section 4 discusses the computational issues. Lastly, Section 5 presents the conclusions of the paper.

## 2. Evaluation of several approximations

In this section, the gamma mixture approximation to the *CDF* of $S_n(\boldsymbol{d})$ is compared with the following approximations: Jensen-Solomon and the three-moment $\chi^2$ fit given by Solomon and Stephens (1977); three-parameter $F$, Satterthwaite-Welsh and Hall-Buckley-Eagleson given by Wood (1989).

### 2.1 *Approximating the CDF of* $S_n(d)$

Several applications of $\chi^2$ tests lead to the problem of evaluation of probabilities involving $S_n(\boldsymbol{d})$; for example, the Pearson $X^2$ statistic and other goodness-of-fit statistics, of Cramer-von Mises type, based on the empirical distribution function (*EDF*) (see examples in Stephens (1986)). The exact probabilities are tabulated only for some special cases; for other cases, due to the well-known problems of numerical integration (Gabler and Wolff (1987)) it is complicated to compute them. Following are some of the references that have dealt with the problem of approximating $S_n(\boldsymbol{d})$. Imhof (1961) and Davies (1980) considered the numerical inversion of the characteristic function of $S_n(\boldsymbol{d})$. Imhof's method may be regarded as essentially exact and gives excellent results in both tails, as one would expect for such a tailor-made approximation. However, it is not easily implementable and when high accuracy is sought, Solomon and Stephens (1977) found it to be relatively expensive in computer time. Sheil and O'Muircheartaigh (1977) and Farebrother (1984) have exploited Ruben's (1962) $\chi^2$ mixture representation for $S_n(\boldsymbol{d})$ and produced algorithms for evaluating the *CDF* of $S_n(\boldsymbol{d})$. Davis (1977) has given a method for evaluating $S_n(\boldsymbol{d})$ based on a numerical solution of a differential equation and Oman and Zacks (1981) presented a mixture approximation to the *CDF* of $S_n(\boldsymbol{d})$. A recent article by Waller *et al.* (1995) reviews and discusses numerical inversion of the characteristic function as a tool for finding the distribution function of $S_n(\boldsymbol{d})$. As a data analysis tool the method is only limited by the numerical precision of both the evaluation of the characteristic function and the evaluation of the fast Fourier transform.

We conclude that the available methods in the literature seem to be computationally intensive or are designed for special cases. We believe that the mixture approach we present here has several strengths relative to these other methods: it is computationally simple and fast, it is highly accurate and it is general rather than being tailor made for any particular example.

We take $G_p(x)$ as a $p$-point gamma mixture and $H$ as the *CDF* of $S_n(\boldsymbol{d})$. The relevant theory for fitting $H$ with $G_p$ is presented in Section 3. We denote the mixture

approximation by $\widetilde{G}_p(x)$, where the fitted parameters $\widetilde{\theta}_j$, $\widetilde{\pi}_j$ and $\widetilde{\lambda}_p$ satisfy the equation (1.3).

2.2 *Gamma mixtures*

A gamma random variable $X$ with shape and scale parameters $\alpha$ and $\beta$ respectively can be written in terms of the parameters $(\alpha, \mu)$, where $\mu = (\alpha/\beta)$, giving probability density function

$$f(x;\alpha,\mu) = \frac{\left(\frac{\alpha}{\mu}\right)^{\alpha} x^{\alpha-1} \exp\left(-\frac{\alpha x}{\mu}\right)}{\Gamma(\alpha)} \quad \text{for} \quad \alpha > 0, \mu > 0, x > 0. \tag{2.1}$$

If in equation (1.2), we take $F_{\theta_j,\lambda}(x)$ to be the gamma distribution with $\theta_j = \mu_j$ and $\lambda = 1/\alpha$, then $F_{Q_p,\lambda}$ is a mixture of $p$ gamma distributions. We will denote this model by $X \sim \text{Gam}(Q_p, \lambda) \triangleq G_p$ (the symbol $\triangleq$ stands for definition). Note that to find the tail area $P[X > t]$ for such a mixture, we merely calculate $\sum \pi_j P[Y_j > t]$, where $Y_j \sim \text{Gam}(\alpha, \mu_j)$.

2.3 *Numerical illustrations*

In this section, we compare the mixture approximation $\widetilde{G}_p$ with other methods of approximating the *CDF* of $S_n(\boldsymbol{d})$. In order to facilitate this, we calculated the moment approximations at the same values of the weights and quantiles considered by Solomon and Stephens (1977) and Wood (1989). The "exact" *CDF* values of $S_n(\boldsymbol{d})$ that are labeled $H$ were reported in those papers as being calculated by numerical methods to high accuracy. The moment approximations were done using a FORTRAN program called Approximation Using Gamma Mixtures (*AUGM*) (Pilla (1995)).

Our first comparisons are with a set of methods considered by Solomon and Stephens (1977). Jensen and Solomon (1972) gave an approximation of the Wilson-Hilferty type, here denoted $J$, which takes $z = (S_n(\boldsymbol{d})/\theta_1)^h$, where $\theta_1$ is the mean of $S_n(\boldsymbol{d})$ and approximates $z$ by a normal distribution, where the mean and variance of the approximating distribution depend on the first three cumulants of $S_n(\boldsymbol{d})$. Solomon and Stephens (1977) gave a three-moment $\chi^2$ fit, here denoted $S$, that fits $S_n(\boldsymbol{d}) = Aw^r$, where $w$ has the $\chi^2_m$ distribution and $A, r$ and $m$ are determined by the first three moments of $S_n(\boldsymbol{d})$. Both the $J$ and $S$ approximations have the disadvantage that an iterative fit is necessary (numerical details omitted).

In Table 1, the exact *CDF* values of $S_n(\boldsymbol{d})(H)$ are compared with Jensen-Solomon $(J)$ and three-moment $\chi^2$ fit $(S)$ approximations considered by Solomon and Stephens (1977) as well as the $p$-point gamma mixture approximation $(\widetilde{G}_p)$ developed in this paper, with values of $p = 1, 2, 3, 4$. No method did well at the very smallest quantile, but the four point fit $\widetilde{G}_4$ was accurate to three digits for every other quantile and is by far the most accurate method.

Our next comparisons are with methods considered by Wood (1989). In the "*F*-method", a three-parameter $F$ distribution approximation to $S_n(\boldsymbol{d})$ is obtained by matching moments. The $F$ density has the form

$$f(x) = \frac{\beta^{\alpha_2} x^{\alpha_1 - 1} (\beta + x)^{-\alpha_1 - \alpha_2}}{B(\alpha_1, \alpha_2)} \quad \text{for} \quad x \geq 0,$$

where $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1 + \alpha_2)$ is the beta function and the parameters $\alpha_1$, $\alpha_2$ and $\beta$ are positive. Using standard formulae which relate moments to cumulants,

Table 1. Comparison of the Jensen-Solomon ($J$), Three-Moment $\chi^2$ Fit ($S$) and $p$-point mixture ($\widetilde{G}_p$) approximations at selected quantiles of $S_n(\boldsymbol{d}) = 2.5\,W_1^2 + 0.7\,W_2^2 + 0.4\,W_3^2 + 0.4\,W_4^2$ with the exact $CDF$ values ($H$).

| **Quantiles** | 0.500 | 1.000 | 3.000 | 3.600 | 8.500 | 11.00 | 14.00 |
|---|---|---|---|---|---|---|---|
| $H$ (True) | 0.0048 | 0.1385 | 0.5197 | 0.6005 | 0.8980 | 0.9452 | 0.9731 |
| $\widetilde{G}_4$ | 0.0453 | 0.1385 | 0.5197 | 0.6007 | 0.8979 | 0.9452 | 0.9732 |
| $J - H$ | 0.0597 | 0.0357 | 0.0218 | 0.0148 | −0.0036 | −0.0045 | −0.0045 |
| $S - H$ | 0.0486 | 0.0129 | −0.0044 | −0.0070 | −0.0005 | 0.0013 | 0.0014 |
| $\widetilde{G}_1 - H$ | 0.0905 | 0.0558 | −0.0074 | −0.0177 | −0.0104 | −0.0020 | 0.0020 |
| $\widetilde{G}_2 - H$ | 0.0499 | 0.0085 | −0.0079 | −0.0072 | 0.0025 | 0.0013 | 0.0001 |
| $\widetilde{G}_3 - H$ | 0.0422 | 0.0010 | −0.0015 | −0.0005 | 0.0002 | −0.0002 | −0.0001 |
| $\widetilde{G}_4 - H$ | 0.0405 | 0.0000 | 0.0000 | 0.0002 | −0.0001 | 0.0000 | 0.0001 |

Table 2. Comparison of the $F(F)$, Satterthwaite-Welsh ($SW$), Hall-Buckley-Eagleson ($HBE$) and $p$-point mixture ($\widetilde{G}_p$) approximations at selected quantiles of $S_n(\boldsymbol{d}) = 0.5W_1^2 + \sum_{i=2}^{6} 0.1W_i^2$ with the exact $CDF$ values ($H$).

| **Quantiles** | 0.113 | 0.165 | 0.221 | 0.301 | 0.485 | 1.271 | 1.946 | 2.500 | 3.086 | 3.883 |
|---|---|---|---|---|---|---|---|---|---|---|
| $H$ (True) | 0.010 | 0.025 | 0.050 | 0.100 | 0.250 | 0.750 | 0.900 | 0.950 | 0.975 | 0.990 |
| $\widetilde{G}_4$ | 0.010 | 0.025 | 0.051 | 0.100 | 0.250 | 0.750 | 0.900 | 0.950 | 0.975 | 0.990 |
| $F - H$ | 0.006 | 0.011 | 0.015 | 0.017 | 0.010 | −0.011 | −0.001 | 0.002 | 0.002 | 0.001 |
| $HBE - H$ | −0.010 | −0.025 | −0.050 | −0.100 | 0.027 | 0.002 | −0.006 | −0.003 | −0.001 | 0.000 |
| $SW - H$ | 0.027 | 0.040 | 0.050 | 0.055 | 0.037 | −0.032 | −0.013 | −0.001 | 0.003 | 0.003 |
| $\widetilde{G}_1 - H$ | 0.027 | 0.040 | 0.050 | 0.055 | 0.037 | −0.032 | −0.013 | −0.001 | 0.003 | 0.003 |
| $\widetilde{G}_2 - H$ | 0.004 | 0.007 | 0.009 | 0.008 | 0.001 | −0.002 | 0.003 | 0.002 | 0.000 | −0.001 |
| $\widetilde{G}_3 - H$ | 0.000 | 0.001 | 0.002 | 0.001 | −0.001 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 |
| $\widetilde{G}_4 - H$ | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Wood expresses $\alpha_1$, $\alpha_2$ and $\beta$ in terms of the first three cumulants. The idea is to find the $r$-th ($r = 1, 2, 3$) cumulant of $S_n(\boldsymbol{d})$, $\kappa_r(\boldsymbol{d})$ and substitute them in the re-expressed formulae of $\alpha_1$, $\alpha_2$ and $\beta$. Finally, use the $F$ distribution with these parameter values to approximate $S_n(\boldsymbol{d})$.

Based on the earlier work of Hall (1983), an approximation was suggested by Buckley and Eagleson (1988) and is here denoted $HBE$. Finally, the Satterthwaite-Welsh ($SW$) approximation is the same as our one point gamma approximation.

In Table 2, the exact $CDF$ values of $S_n(\boldsymbol{d})$ ($H$) are compared with three-parameter $F$ ($F$), Satterthwaite-Welsh ($SW$) and Hall-Buckley-Eagleson ($HBE$) approximations given by Wood (1989) and the approximation $\widetilde{G}_p$. Table 2 clearly shows the superiority of approximation $\widetilde{G}_p$ over the others, with $\widetilde{G}_4$ having three digit accuracy at all but one quantile. Also note the shrinking of the error, namely $H - \widetilde{G}_p$, as the number of support points ($p$) increases.

Computation of $\widetilde{G}_4$ in Table 2 required a total of 0.36 seconds of CPU time on SUN SPARC 10. Compared to the other methods, it seems clear that our method is relatively inexpensive to compute.

Table 3 gives a summary of how the different methods performed over all the quantiles for the four distributions given by Solomon and Stephens (1977) and the 14 distri-

Table 3. Comparisons of the Summary statistics for all the 18 distributions over all the quantiles.

| Data Source | Method | % of times accurate to 3 digits | Average Absolute Error |
|---|---|---|---|
| | $J$ | 17.86 | 0.00957 |
| Solomon | $S$ | 32.14 | 0.00349 |
| | $\widetilde{G}_4$ | 89.29 | 0.00240 |
| | $F$ | 17.81 | 0.00703 |
| Wood | $HBE$ | 15.99 | 0.01534 |
| | $SW$ | 4.84 | 0.02035 |
| | $\widetilde{G}_4$ | 79.43 | 0.00042 |

butions given by Wood (1989). The $\widetilde{G}_4$ approximation shows an excellent performance when compared with other methods. Column 3 in the table gives the percentage of times each method was accurate to three digits to the true ($H$) value. The last column gives the average absolute error incurred by using the given approximation. (For more details, see Lindsay *et al.* (1995).)

## 3. Fitting mixtures of distributions

In this section we will present the general mathematical developments that enable one to construct $\widetilde{G}_p$ when using a family of mixtures. We start by showing how the highly nonlinear equations in (1.3) can be solved with straightforward procedures for selected families of distributions.

### 3.1 *Deriving unbiased moment estimators*

Lindsay (1989*b*) proposed methods that use method of moments to estimate the normal mixture parameters; that is, the $p$-point mixing distribution, $Q_p$, and the dispersion parameter, $\sigma_p^2$. That paper dealt only with mixtures of normals and concentrated on the problem of estimating the parameters of a mixture distribution from the sample moments. We show here how one can extend those methods by replacing the normal model with another family $F_{\theta,\lambda}$ that satisfies certain properties described below.

In the families $F_{\theta,\lambda}$ we will consider here, we assume that for each $r$ there exists a polynomial $\delta_r(x,\lambda)$ of degree $r$ in $x$, written $\sum_{j=0}^{r} a_j^{(r)}(\lambda)x^j$, such that

$$E_{\theta,\lambda}[\delta_r(X,\lambda)] = \theta^r. \tag{3.1}$$

*Example* 1. If we use the gamma density in equation (2.1), set $\theta = \mu$ and $\lambda = 1/\alpha$, then

$$\delta_r(x,\lambda) = \frac{X^r}{(1+\lambda)(1+2\lambda)\cdots(1+(r-1)\lambda)} \tag{3.2}$$

satisfies $E_{\theta,\lambda}[\delta_r(X,\lambda)] = \mu^r$ as desired.

From equation (3.1) it follows that if $Q_p$ is a mixing distribution on the parameter $\theta$ of a such a family and $X$ has the corresponding mixture distribution $F_{Q,\lambda}$, then $\delta_r(X,\lambda)$ is an unbiased estimator of the $r$-th moment of $\theta$, $m_r(Q)$. That is, the expectation of

$\delta_r(X, \lambda)$ under the mixture model, $F_{Q,\lambda}$, is given by

$$E_{Q,\lambda}[\delta_r(X,\lambda)] = \iint \delta_r(x,\lambda) dF_{\theta,\lambda}(x) dQ(\theta) \tag{3.3}$$
$$= \int E_{\theta,\lambda}[\delta_r(X,\lambda)] dQ(\theta)$$
$$= \int \theta^r dQ(\theta) = m_r(Q).$$

One can also write $m_r(Q) = E_{Q,\lambda}[\delta_r(X,\lambda)]$ as

$$E_{Q,\lambda}\left[\sum_j a_j^{(r)}(\lambda) X^j\right] = \sum_j a_j^{(r)}(\lambda) E_{Q,\lambda}[X^j]$$
$$= \sum_j a_j^{(r)}(\lambda) m_j(F_{Q,\lambda}).$$

We thus have

$$\sum_{j=0}^{r} a_j^{(r)}(\lambda) m_j(F_{Q,\lambda}) = m_r(Q).$$

This implies that if we wish to find a $p$-point distribution $Q_p$ and dispersion parameter $\lambda$ to match the target moments, that is, satisfying $m_r(F_{Q_p,\lambda}) = m_r(H) = m_r$ for $r = 1, \ldots, 2p$, we need to solve for $\lambda$ and $Q_p$ in the system of equations

$$\sum_{j=0}^{r} a_j^{(r)}(\lambda) m_j = m_r(Q_p), \qquad r = 1, \ldots, 2p.$$

In this equation $m_1, \ldots, m_{2p}$ are the target moments and the coefficients $a_j^{(r)}$ are determined by the family $F_{\theta,\lambda}$. For an arbitrary $\lambda$, let $\sum_j a_j^{(r)}(\lambda) m_j$ be denoted by $\delta_r^*(\lambda)$, so that the equations we need to solve for $\lambda$ and $Q_p$ are:

$$\delta_r^*(\lambda) = m_r(Q_p), \qquad r = 1, \ldots, 2p-1. \tag{3.4}$$

*Example* 1. (*continued*) In the gamma model, we have for $\theta = \mu$ and $\delta_r$ as in (3.2), $E_{\theta,\lambda}[\delta_r(X,\lambda)] = \theta^r$ and so in the mixture model, where $\theta \sim Q_p$, we have

$$E_{Q_p,\lambda}[\delta_r(X,\lambda)] = \int \theta^r dQ(\theta) = m_r(Q_p).$$

After inserting the target moments $m_r$ into the left side, we find the goal is to solve for $\lambda$ and $Q_p$ in $\delta_r^*(\lambda) = m_r(Q_p)$, where

$$\delta_r^*(\lambda) = \frac{1}{1(1+\lambda)(1+2\lambda)\cdots(1+(r-1)\lambda)} \cdot m_r. \tag{3.5}$$

The equations (3.4) are solved in two steps. First, we know that one can solve these equations for a distribution $Q_p$ only if $\lambda$ is chosen such that the numbers $\delta_1^*(\lambda), \ldots, \delta_{2p}^*(\lambda)$ are exactly the sequence of moments to some distribution $Q_p$ with exactly $p$-points of

support. However, we can use moment theory to solve for a unique $\widetilde{\lambda}_p$ that makes this true, without ever determining $Q_p$.

Then, in the second step, we find the $p$-point distribution $\widetilde{Q}_p$ by solving

$$\delta_r^*(\widetilde{\lambda}_p) = m_r(Q_p), \qquad r = 1, \ldots, 2p-1$$

using standard techniques from moment theory.

3.2 *Solving for* $\lambda$

We start by showing how $\widetilde{\lambda}_p$ can be determined. The techniques we use are heavily dependent on the use of certain moment matrices. For a given sequence of numbers, $m_1, \ldots, m_{2p}$, let

$$\boldsymbol{M}_p = \begin{pmatrix} 1 & m_1 & m_2 & \cdots & m_p \\ m_1 & m_2 & m_3 & \cdots & m_{p+1} \\ m_2 & m_3 & m_4 & \cdots & m_{p+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m_p & m_{p+1} & m_{p+2} & \cdots & m_{2p} \end{pmatrix}$$

be the $p$-th *moment matrix*. The following theorem (see, for example, Lindsay (1989*b*)) establishes the relevant properties of moment sequences to determine whether a moment solution exists.

THEOREM 3.1. (a) *A sequence of numbers* $1, m_1, m_2, \ldots, m_{2p}$ *are the moments of a distribution with exactly* $p$ *points of support if and only if* $\det \boldsymbol{M}_j > 0$ *for all* $j = 1, \ldots, (p-1)$ *and* $\det \boldsymbol{M}_p = 0$. (b) *If the sequence of numbers* $1, m_1, m_2, \ldots, m_{2p-2}$ *satisfies* $\det \boldsymbol{M}_j > 0$ *for all* $j = 1, \ldots, (p-1)$ *and* $m_{2p-1}$ *is any scalar, then* $\exists$ *a unique* $p$*-point distribution with exactly those initial* $(2p-1)$ *moments.*

To proceed further, we will need two assumptions on the function $\delta_r^*(\lambda)$.

ASSUMPTION A1. $\delta_r^*(\lambda)$ is continuous in $\lambda$ for $\lambda \in [0, \infty)$.

ASSUMPTION A2. $\delta_r^*(0) = m_r(H)$.

*Example* 1. (*continued*) In equation (1.2), if we take $F_{\theta_j,\lambda}(x)$ as gamma distribution with mean $\theta_j = \mu_j$ and variance $\lambda = 1/\alpha$, then the preceding assumptions are satisfied for the gamma family.

We next define the pseudo-moment matrix

$$\Delta_p(\lambda) = \{\delta_{j+k}^*(\lambda)\}_{\substack{j=0,\ldots,p \\ k=0,\ldots,p}} \quad \text{with} \quad \delta_0^* = 1. \tag{3.6}$$

Note that at $\lambda = 0$, from Assumption A2 we get $\Delta_p(0) = \boldsymbol{M}_p(H)$, the moment matrix for $H$. Also, if $H$ is a true mixture, $F(Q, \lambda_0)$, then, $\delta_r^*(\lambda_0) = m_r(Q)$ by equation (3.4), so $\Delta_p(\lambda_0) = \boldsymbol{M}_p(Q)$, the moment matrix for $Q$.

*Example* 1. (continued) For the gamma distribution the unbiased estimators of $m_r(Q) = \int \mu^r dQ(\mu)$ for $r = 1, \ldots, 2p$ are obtained using equation (3.5). Thus, for

example, if $H$ is a Gam$(Q, \lambda)$ mixture, then

$$m_1(Q) = E(X) = m_1(H),$$
$$m_2(Q) = \frac{1}{1+\lambda}E(X^2) = \frac{1}{1+\lambda}m_2(H),$$
$$m_3(Q) = \frac{1}{(1+\lambda)(1+2\lambda)}E(X^3) = \frac{1}{(1+\lambda)(1+2\lambda)}m_3(H),$$

and

$$m_4(Q) = \frac{1}{(1+\lambda)(1+2\lambda)(1+3\lambda)}E(X^4) = \frac{1}{(1+\lambda)(1+2\lambda)(1+3\lambda)}m_4(H). \tag{3.7}$$

For any given distribution $H$, the pseudo-moment $\delta_r^*(\lambda)$ is the corresponding right hand side of the above equations and so the *second pseudo-moment matrix* $\Delta_2(\lambda)$, using equation (3.6), is then given by

$$\Delta_2(\lambda) = \begin{pmatrix} 1 & \delta_1^*(\lambda) & \delta_2^*(\lambda) \\ \delta_1^*(\lambda) & \delta_2^*(\lambda) & \delta_3^*(\lambda) \\ \delta_2^*(\lambda) & \delta_3^*(\lambda) & \delta_4^*(\lambda) \end{pmatrix}.$$

Our objective is to find $\widetilde{\lambda}_2$ such that this is the moment matrix for a $p = 2$ point distribution $\widetilde{Q}_2$.

The first step of our strategy is to let $\widetilde{\lambda}_p$ (where the subscript denotes dependence on $p$) be the smallest nonnegative root, if it exists, of $\det[\Delta_p(\lambda)] = 0$. Our claim is that $\delta_1^*(\widetilde{\lambda}), \ldots, \delta_{2p}^*(\widetilde{\lambda})$ is then the moment sequence for some $p$-point distribution $Q_p$.

We start by studying the properties of roots defined in this way. The following lemma shows that we can generally assume that the root $\widetilde{\lambda}_p$ is strictly positive.

LEMMA 3.1. *The root $\widetilde{\lambda}_p = 0$ if and only if $H$ has $p$ or fewer points of support.*

PROOF. Note that $\widetilde{\lambda}_p = 0$ if and only if $\det \Delta_p(0) = 0$. However, $\det \Delta_p(0) = 0 = \det \boldsymbol{M}_p(H)$ and the latter is zero if and only if $H$ has $p$ or fewer points of support. □

Given the positivity of $\widetilde{\lambda}_p$, we can make the following strong statement about the pseudo-moment matrix.

LEMMA 3.2. *If $\widetilde{\lambda}_p > 0$, then $\Delta_p(\lambda)$ is positive definite for $\lambda \in [0, \widetilde{\lambda}_p)$.*

PROOF. The smallest eigenvalue of a symmetric matrix is a continuous function of the matrix entries, so by Assumption A1 the smallest eigenvalue of $\Delta_p(\lambda)$, say $e_1(\lambda)$, is a continuous function of $\lambda$. Note that by the hypothesis $\Delta_p(0) = \boldsymbol{M}_p(H)$, which is positive definite, so $e_1(0) > 0$. We have $\det \Delta_p(\lambda) = \prod_{j=0}^{p} e_j(\lambda)$, the product of its eigenvalues, so the determinant can not be zero in $[0, \widetilde{\lambda}_p)$ unless an eigenvalue is zero. But if no eigenvalue crosses or touches zero, all remain positive. Thus $\Delta_p(\lambda)$ is positive definite on the given range. □

The next proposition indicates the nested structure of the $\lambda$-roots. It is very useful in computations as it enables a simple root search strategy.

PROPOSITION 3.1. *If $\widetilde{\lambda}_1$ exists finitely, then every $\widetilde{\lambda}_p$ exists and we have*

$$\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_{p-1} \geq \widetilde{\lambda}_p \geq \cdots.$$

PROOF. The proof is inductive on $p$. Suppose $\widetilde{\lambda}_{p-1}$ exists finitely. Then $\Delta_{p-1}(\widetilde{\lambda}_{p-1})$ has a zero determinant and so is singular. Hence there exists a nonzero vector $\boldsymbol{x}$ such that $\Delta_{p-1}(\widetilde{\lambda}_{p-1})\boldsymbol{x} = 0$. It follows that

$$(\boldsymbol{x}' \ 0)\Delta_p(\widetilde{\lambda}_{p-1})\begin{pmatrix}\boldsymbol{x}\\ 0\end{pmatrix} = \boldsymbol{x}'\Delta_{p-1}(\widetilde{\lambda}_{p-1})\boldsymbol{x} = 0.$$

Since $(\boldsymbol{x}' \ 0)$ is a nonzero vector, it follows that $\Delta_p(\widetilde{\lambda}_{p-1})$ cannot be positive definite. Lemma 3.2 then implies that $\widetilde{\lambda}_p$ exists and $\widetilde{\lambda}_{p-1} \notin [0, \widetilde{\lambda}_p)$. □

Finally, we establish our claim.

PROPOSITION 3.2. (a) *If $\widetilde{\lambda}_{p-1} > \widetilde{\lambda}_p$, then $\exists$ a unique $p$-point mixture $\widetilde{Q}_p$ satisfying $\delta_r^*(\widetilde{\lambda}_p) = m_r(\widetilde{Q}_p)$ for $r = 1, \ldots, 2p$.* (b) *If $\widetilde{\lambda}_{p-1} = \widetilde{\lambda}_p$, then there does not exist a $p$-point solution.*

PROOF. Since $\widetilde{\lambda}_{p-1} > \widetilde{\lambda}_p$, $\Delta_{p-1}(\widetilde{\lambda}_p)$ is positive definite by Lemma 3.2. Hence, by Theorem 3.1 there exists a unique $p$-point mixture $\widetilde{Q}_p$ satisfying $\delta_r^*(\widetilde{\lambda}_p) = m_r(\widetilde{Q}_p)$ for $r = 1, \ldots, 2p-1$. Moreover, we claim $\delta_{2p}^*(\widetilde{\lambda}_p) = m_{2p}(\widetilde{Q}_p)$. This follows because the $2p$-th moment of a $p$-point distribution is uniquely determined by the preceding $2p-1$ moments, as follows from the relationship $\det \boldsymbol{M}_p = 0$. If $\widetilde{\lambda}_{p-1} = \widetilde{\lambda}_p$, then $\Delta_{p-1}(\widetilde{\lambda}_p)$ is not positive definite and hence does not represent the moment matrix for a distribution $Q$ with $p$ points of support. □

*Remark* 1. In Lindsay ((1989*b*), Theorem 5C), the analogue of this result was stated incorrectly. Instead of the correct condition "if $\widetilde{\lambda}_{p-1} > \widetilde{\lambda}_p$", it was stated — " if $\det \Delta_p(\lambda)$ has a root of order 1 at $\lambda = \widetilde{\lambda}_p$".

3.3 *Solving for $\widetilde{Q}_p$*

We now describe how to reconstruct $\widetilde{Q}_p$ from its moments, $1, \delta_p^*(\widetilde{\lambda}_p), \ldots, \delta_{2p}^*(\widetilde{\lambda}_p)$. We start by finding the support points. These can be determined using the following Theorem (Lindsay (1989*b*)). Define the polynomial:

$$\widetilde{S}_p(\widetilde{\lambda}_p, t) = \det\begin{bmatrix} 1 & \delta_1^*(\widetilde{\lambda}_p) & \cdots & \delta_{p-1}^*(\widetilde{\lambda}_p) & 1 \\ \delta_1^*(\widetilde{\lambda}_p) & \delta_2^*(\widetilde{\lambda}_p) & \cdots & \delta_p^*(\widetilde{\lambda}_p) & t \\ \delta_2^*(\widetilde{\lambda}_p) & \delta_3^*(\widetilde{\lambda}_p) & \cdots & \delta_{p+1}^*(\widetilde{\lambda}_p) & t^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \delta_p^*(\widetilde{\lambda}_p) & \delta_{p+1}^*(\widetilde{\lambda}_p) & \cdots & \delta_{2p-1}^*(\widetilde{\lambda}_p) & t^p \end{bmatrix}. \tag{3.8}$$

THEOREM 3.2. *If there exists a solution $\widetilde{Q}_p$ to the moment equations, then it has as its support points $\tilde{\mu}_1, \ldots, \tilde{\mu}_p$ the roots of $\widetilde{S}_p(\widetilde{\lambda}_p, t) = 0$.*

See Section 4 for an algebraic simplification that we used in computing the roots of the polynomial $\widetilde{S}_p(\widetilde{\lambda}_p, t)$.

For the second step in reconstructing $\tilde{Q}_p$, once we have found the roots $\tilde{\mu}_1, \ldots, \tilde{\mu}_p$ to the polynomial $\tilde{S}_p(\tilde{\lambda}_p, t)$, we solve for the masses $\tilde{\pi}_j$ at each support point $\tilde{\mu}_j$ by solving the linear system of equations

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tilde{\mu}_1 & \tilde{\mu}_2 & \cdots & \tilde{\mu}_p \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mu}_1^{p-1} & \tilde{\mu}_2^{p-1} & \cdots & \tilde{\mu}_p^{p-1} \end{bmatrix} \begin{bmatrix} \tilde{\pi}_1 \\ \tilde{\pi}_2 \\ \vdots \\ \tilde{\pi}_p \end{bmatrix} = \begin{bmatrix} 1 \\ \delta_1^*(\tilde{\lambda}_p) \\ \vdots \\ \delta_{p-1}^*(\tilde{\lambda}_p) \end{bmatrix}. \tag{3.9}$$

The matrix on the left is nonsingular, being a Vandermonde matrix, so that there is a unique solution to these linear equations.

In the following example we illustrate the above methodology using the two point gamma mixture fit.

*Example* 1. (*continued*). Given a solution $\tilde{\lambda}_2$ to $\det \Delta_2(\lambda) = 0$, we can solve for $\tilde{Q}_2$ as follows. The two support points are calculated by solving the quadratic equation

$$\det \begin{pmatrix} 1 & \tilde{m}_1 & 1 \\ \tilde{m}_1 & \tilde{m}_2 & t \\ \tilde{m}_2 & \tilde{m}_3 & t^2 \end{pmatrix} = 0, \tag{3.10}$$

where $\tilde{m}_r$'s are the $\delta_r^*(\tilde{\lambda}_2)$'s given in equation (3.7). The solutions to equation (3.10) are given by

$$t = \frac{(\tilde{m}_3 - \tilde{m}_1\tilde{m}_2) \pm \sqrt{(\tilde{m}_3 - \tilde{m}_1\tilde{m}_2)^2 - 4(\tilde{m}_2 - \tilde{m}_1^2)(\tilde{m}_1\tilde{m}_3 - \tilde{m}_2^2)}}{2(\tilde{m}_2 - \tilde{m}_1^2)}.$$

Call the two solutions $\tilde{t}_1 = \tilde{\mu}_1$ and $\tilde{t}_2 = \tilde{\mu}_2$ respectively. We then solve for the masses $\tilde{\pi}_1$ and $\tilde{\pi}_2$ using the full rank equations

$$\begin{pmatrix} 1 & 1 \\ \tilde{\mu}_1 & \tilde{\mu}_2 \end{pmatrix} \begin{pmatrix} \tilde{\pi}_1 \\ \tilde{\pi}_2 \end{pmatrix} = \begin{pmatrix} 1 \\ \tilde{m}_1 \end{pmatrix},$$

giving us

$$\tilde{\pi}_1 = \frac{\tilde{\mu}_2 - \tilde{m}_1}{\tilde{\mu}_2 - \tilde{\mu}_1}$$

and

$$\tilde{\pi}_2 = 1 - \tilde{\pi}_1.$$

### 3.4 *Uniqueness and root finding issues*

Here we will discuss the uniqueness of the fitted mixture distribution $G_p$. That is, given a sequence $m_1, \ldots, m_{2p}$ of moments that can be fit by some $p$-point $(\tilde{Q}_p, \tilde{\lambda}_p)$ mixture, is it possible that there exists another mixture $(\tilde{Q}_p^*, \tilde{\lambda}_p^*)$ that generates exactly the same moment sequence? The following assumption gives us a natural sufficient condition that will ensure the uniqueness of $(\tilde{Q}_p, \tilde{\lambda}_p)$ as generated from $m_1, \ldots, m_{2p}$.

ASSUMPTION A3. For every $\theta \in \Theta$ and for every $0 \leq \lambda < \lambda_0$ there exists a mixing distribution $Q_\theta$ with infinite support $\ni F_{\theta,\lambda_0} = F_{Q_\theta,\lambda}$.

This assumption is verified for the gamma example in Appendix B. Note that Assumption A3 implies that if $Q$ is any mixing distribution on $\theta$, then for any $\lambda < \lambda_0$, we can find $Q^\star$ with infinite support satisfying $F_{Q,\lambda_0} = F_{Q^\star,\lambda}$. This is done by setting

$$Q^\star(A) = \int_A dQ^\star(\phi) = \int_A dQ_\theta(\phi) dQ(\theta). \tag{3.11}$$

PROPOSITION 3.3. *Suppose the pseudo moments $\delta_p^\star(\lambda)$ exist as above and that Assumption* A3 *holds. Then for every p-point distribution $\widetilde{Q}_p$ and $\lambda_0$, the values $(\widetilde{Q}_p, \lambda_0)$ are uniquely determined by the first 2p target moments, $m_1, \ldots, m_{2p}$.*

PROOF. For any $0 < \lambda < \lambda_0$, let $Q^\star$ be the mixing distribution such that $F(Q^\star, \lambda)$ is distributionally equivalent to $F(\widetilde{Q}_p, \lambda_0)$. Then $\delta_1^\star(\lambda), \ldots, \delta_{2p}^\star(\lambda)$ are the moments of $Q^\star$ and so the pseudo-moment matrix $\Delta_p(\lambda)$ is the moment matrix for a distribution with infinite support. It therefore is positive definite with positive determinant. It follows that $\lambda_0 = \widetilde{\lambda}_p$ must be the first nonnegative root of $\det[\Delta_p(\lambda)]$ and so is uniquely determined.

However, knowing $\lambda_0$ then uniquely determines $\delta_1(\lambda_0), \ldots, \delta_{2p}(\lambda_0)$, the first $2p$ moments of $\widetilde{Q}_p$, which in turn uniquely determines $\widetilde{Q}_p$, since it has $p$-points of support. □

The following proposition gives us important practical information for the design of algorithms for finding $\widetilde{\lambda}_p$.

PROPOSITION 3.4. *Suppose that $\widetilde{\lambda}_{p-1} > \widetilde{\lambda}_p$ and that Assumption* A3 *holds. We have* (i) $\det \Delta_p(\lambda) > 0$ *for* $(0, \widetilde{\lambda}_p)$ *and* (ii) $\det \Delta_p(\lambda) < 0$ *for* $\lambda \in (\widetilde{\lambda}_p, \widetilde{\lambda}_{p-1})$. *Therefore $\widetilde{\lambda}_p$ is the unique root of* $\det \Delta_p(\lambda)$ *in* $[0, \widetilde{\lambda}_{p-1})$.

PROOF. (i) Follows from the Lemma 3.2. (ii) Suppose for contradiction of (ii) that there exists a root $\lambda^\star$ of $\det \Delta_p(\lambda)$ that is in $(\widetilde{\lambda}_p, \widetilde{\lambda}_{p-1})$. For any $\lambda \in (\widetilde{\lambda}_p, \widetilde{\lambda}_{p-1})$, $\Delta_{p-1}(\lambda)$ is positive definite by Lemma 3.2 and so $\Delta_p(\lambda^\star)$ satisfies the conditions needed to be the moment matrix of a $p$-point distribution $Q_p^\star$. Hence $(Q_p^\star, \lambda^\star)$ and $(Q_p, \lambda)$ represent two formal solutions to the moment equations, which is a contradiction to the preceding proposition and so root $\lambda^\star$ cannot exist.

Since there are no roots in $(\widetilde{\lambda}_p, \widetilde{\lambda}_{p-1})$, $\det \Delta_p(\lambda) > 0$ is strictly positive or negative on that range. Suppose for contradiction it is positive. Then on this range $\Delta_p(\lambda)$ is positive definite and hence represents the moment matrix for a distribution $Q_\lambda$ with more than $p$-points of support. It follows using (3.11) that there exists $(Q_\infty^\star, \widetilde{\lambda}_p)$ with infinite support generating the same $G$ distribution as $(Q_\lambda, \lambda)$. However, the moments of $Q_\infty^\star$ must then be $\delta_1^\star(\widetilde{\lambda}_p), \ldots, \delta_{2p}^\star(\widetilde{\lambda}_p)$, which satisfy $\det \Delta_p(\widetilde{\lambda}_p) = 0$, contradicting that $Q_\infty^\star$ has infinite support. We conclude that $\det \Delta_p(\lambda)$ must be negative on $(\widetilde{\lambda}_p, \widetilde{\lambda}_{p-1})$. □

*Remark* 2. One technical difficulty standing in the way of a simple general theory is as follows: Suppose that we find $\widetilde{\lambda}_p$ such that $\Delta_p(\widetilde{\lambda}_p)$ is a moment matrix for $p$-point distribution $\widetilde{Q}_p$. Then there is no general guarantee that $\widetilde{Q}_p$ has all its support points within the $\theta$-parameter space. If it does, we will say it is a *proper* solution. This is not a problem in the mixtures of normals case, as the parameter space is $(-\infty, \infty)$, so all the solutions are necessarily proper, but it could be an issue in other cases. We will not tackle the technical problems involving the existence of proper solutions here. The reader can refer to Lindsay ((1989*b*) Section 2.3) for some of the issues.

## 4. Computational issues

This section describes some of the computational issues involved in finding the $\widetilde{\lambda}_p$ and the roots $\widetilde{\mu}_1, \ldots, \widetilde{\mu}_p$. It is most efficient to proceed sequentially. Thus first we find $\widetilde{\lambda}_1$. In the gamma case, $\widetilde{\lambda}_1$ is the unique root of the equation:

$$\det \begin{pmatrix} 1 & m_1(H) \\ m_1(H) & \frac{1}{(1+\lambda)} m_2(H) \end{pmatrix} = 0.$$

In order to find $\widetilde{\lambda}_2$, the unique root of $\det \Delta_2(\lambda)$ in $[0, \widetilde{\lambda}_1)$, we used a bisection algorithm. Bisection simply divides the interval repeatedly in half, selecting at each stage the half-interval in which the sign change occurs. Given the sign-change behavior, it is simple, effective, easy to program and guaranteed to converge. We note there is some danger of facing the situation $\widetilde{\lambda}_{p-1} = \widetilde{\lambda}_p$ because of numerical inaccuracy. To avoid this difficulty it is recommended to carry out the computations in double precision.

Next, to find the roots $\widetilde{\mu}_1, \ldots, \widetilde{\mu}_p$ to the polynomial equation $\widetilde{S}_p(\widetilde{\lambda}_p, t) = 0$, one needs to find the right hand side of equation (3.8). For the sake of programming simplification, the following trick was used. We replaced the last column of the matrix on the right hand side of equation (3.8) with $\delta_r^*(\lambda)$ for $r = p, \ldots, 2p$. Notice that the resulting matrix is identical to the $\boldsymbol{M}_p$ matrix with estimated moments as elements. Denote this matrix by $\widetilde{\boldsymbol{M}}_p$. Next, we found the eigenvalues and eigenvectors of $\widetilde{\boldsymbol{M}}_p$. Thus $\widetilde{\boldsymbol{M}}_p = ADA^T$, where $D$ is a diagonal matrix of eigenvalues. One can easily obtain $\widetilde{\boldsymbol{M}}_p^{-1}$ by taking the reciprocal of the corresponding elements of the $D$ matrix. Hence finding $\widetilde{\boldsymbol{M}}_p^{-1} = AD^{-1}A^T$ is a computationally straightforward task. Notice that the elements of the last row of $\widetilde{\boldsymbol{M}}_p^{-1}$ are the coefficients of the support point equation $\widetilde{S}_p(\widetilde{\lambda}_p, t) = 0$, up to proportionality. However, the constant of proportionality can be ignored for it can be cancelled out of the equation. At this stage, solving the equation for the roots can be easily accomplished with a simple root finding subroutine. We used *zrhqr* subroutine (see Press *et al.* (1992)) since the polynomial has real roots.

Lastly, solving the equation (3.9) for the masses $\widetilde{\pi}_j$ at each support point is accomplished with a simple subroutine that takes advantage of the special structure of the Vandermonde matrix. We used *vander* subroutine that solves the Vandermonde linear system (see Press *et al.* (1992)).

## 5. Conclusions

We have shown that by using the method of moments, one can create families of distributional approximations that are highly flexible and straightforward to fit. We have shown that if one uses mixtures of gammas, one can in this fashion reasonably approximate the distribution of a linear combination of chi-squared variables. Although the theory looks rather complicated, the method is relatively straightforward to program (the FORTRAN program *AUGM* is available from author Pilla). In theory one can use this methodology to fit an arbitrary number of moments. However, increasing the number of mixture components can lead to computational accuracy problems due to the use of ill-conditioned moment matrices. Our experience with the examples considered in this article was that after $p$ reached 6 or 7, the moment method would fail to generate a $p$-point approximation due to $\widetilde{\lambda}_{p-1} = \widetilde{\lambda}_p$ (see Proposition 3.2(b)). Indeed, up to this value of $p$, the estimates tended to improve but when we tried to push the method beyond this point by fitting with a larger $p$, there was no change in $\widetilde{\lambda}$. This indicates that one

cannot fit a higher order mixture to improve the approximation. In light of this, we offer a practical recommendation that one use the largest value of $p$ possible; that is proceed sequentially and stop when $\widetilde{\lambda}_{p-1} = \widetilde{\lambda}_p$.

## Acknowledgements

We would like to thank the referee for helpful comments and suggestions.

## Appendix A. Relation between moments and cumulants

The $r$-th moment about the origin of a random variable $X$, or equivalently of its distribution $H(x)$ denoted by $m_r$ is defined as $m_r(H) = \int x^r dH(x)$ and the corresponding $r$-th moment about the mean ($r$-th central moment) denoted by $\mu_r$ is defined as $\mu_r(H) = \int (x - m_1)^r dH(x)$. The moment generating function (mgf) of $X$ is given by

$$M(t) = \int_{-\infty}^{\infty} \exp(tx) dH(x).$$

The cumulant generating function (cgf) $K(t)$ of $X$ is defined by natural logarithm of the mgf, i.e.,

$$K(t) = \log M(t).$$

Both these generating functions can be expanded as a Taylor series.

Given a list of cumulants (or moments) one can obtain the other list by the recursion

$$m_r = \kappa_r + \binom{r-1}{1} \kappa_{r-1} m_1 + \binom{r-1}{2} \kappa_{r-2} m_2 + \cdots + \binom{r-1}{r-1} \kappa_1 m_{r-1},$$

which can be obtained from the following mathematical relation

$$K'(t) M(t) = M'(t).$$

For example, $\kappa_1 = m_1$, $\kappa_2 = \mu_2$, $\kappa_3 = \mu_3$, $\kappa_4 = \mu_4 - \mu_2^2$. Such a recursion, simplifies the calculation of the moments of a convolution, as the cumulants are easily calculated because they add over independent observations. For example, the cumulants of a weighted sums of independent central $\chi_1^2$ variates,

$$S_n(\boldsymbol{d}) = d_1 W_1^2 + d_2 W_2^2 + \cdots + d_n W_n^2$$

are $\kappa_r = \sum_{i=1}^n d_j^r \kappa_r(\chi_1^2) = \sum_{i=1}^n d_j^r 2^{r-1}(r-1)!$ for $r = 1, 2, \ldots$. Also, it can be easily shown that $\mu_r(dX + b) = d^r \mu_r(X)$ and $\kappa_r(dX + b) = d^r \kappa_r(X)$ for $r \geq 2$, where $d$ and $b$ are any real constants.

When we match moments to order $2p$ we also fit cumulants to the same order. For example, in the normal mixture model, $\sum \pi_j N(\mu_j, \sigma^2)$,

$$\kappa_r(H) = \kappa_r(Q) + \kappa_r(N),$$

the second term on the right hand side being zero for $r > 2$. Hence, when we match cumulants, $\kappa_r(Q)$ equals $\kappa_r(H)$ for $r > 2$.

## Appendix B. Nested mixing structure of gamma distributions

In order to apply the Proposition 3.3, "identifiability", we need to verify the Assumption A3. We want to know if for any given $\alpha$, $\mu_0$ and any $\delta > 0$, with $\alpha_0 = \alpha - \delta > 0$, there exists a continuous mixing distribution $Q$ such that

$$\int f(x;\alpha,\mu)dQ(\mu) = f(x;\alpha_0,\mu_0). \tag{B.1}$$

The following proposition clarifies the assumption.

PROPOSITION B.1. *$f(x;\alpha_0,\mu_0)$ can be written as a mean mixture of gammas, i.e., as in equation* (B.1).

PROOF. Let $\eta = \alpha/\mu$ and $\eta_0 = \alpha_0/\mu_0$. We need to show that the mixing distribution $Q \propto (\eta-\eta_0)^{\delta-1}\eta^{-\alpha}I\{\eta > \eta_0\}d\eta$. That is, $dQ = k(\eta-\eta_0)^{\delta-1}\eta^{-\alpha}I\{\eta > \eta_0\}d\eta$ with $k$ as the constant of proportionality. Thus we will show that $dQ$ satisfies equation

$$\int_0^\infty \eta^\alpha \exp(-\eta x)dQ\left(\frac{\alpha}{\eta}\right) = c\exp(-\eta_0 x)x^{-\delta}, \tag{B.2}$$

where $c = \Gamma(\alpha)(\eta_0)^{\alpha_0}/\Gamma(\alpha_0)$ is a constant. The left hand side of equation (B.2) becomes

$$\begin{aligned}\int_0^\infty k\exp(-\eta x)(\eta-\eta_0)^{\delta-1}I\{\eta>\eta_0\}d\eta &= k\int_{\eta_0}^\infty \exp(-\eta x)(\eta-\eta_0)^{\delta-1}d\eta,\\ &= k\exp(-\eta_0 x)\int_0^\infty \exp(-\xi x)\xi^{\delta-1}d\xi,\\ &= k\exp(-\eta_0 x)x^{-\delta}\Gamma(\delta).\end{aligned} \tag{B.3}$$

Simplification follows by taking $\eta - \eta_0 = \xi$. Equation (B.3) is valid since $\delta > -1$ and $x > 0$. If we let $k = \frac{c}{\Gamma(\delta)}$, then (B.3) becomes $cx^{-\delta}\exp(-\eta_0 x)$. Thus the mixing distribution is given by $dQ = \frac{c}{\Gamma(\delta)}(\eta-\eta_0)^{\delta-1}I\{\eta>\eta_0\}d\eta$. □

The following corollary shows the nested mixing structure of gamma distributions.

COROLLARY B.1. $\{\text{Gamma}(\alpha+\delta,Q)\} \subset \{\text{Gamma}(\alpha,Q)\}$ *for any* $\delta \ni \alpha > \delta > 0$.

Assumption A3 and Proposition B.1 give the identifiability of $(\widetilde{Q}_p, \widetilde{\lambda}_p)$ in the gamma mixtures.

## REFERENCES

Buckley, M. J. and Eagleson, G. K. (1988). An approximation to the distribution of quadratic forms in normal random variables, *Austral. J. Statist.*, **30A**, 150–159.

Davies, R. B. (1980). The distribution of a linear combination of independent chi-squared random variables, *Applied Statistics*, **29**, 323–333.

Davis, A. W. (1977). A differential equation approach to linear combinations of independent chi-squares, *J. Amer. Statist. Assoc.*, **72**, 212–214.

Farebrother R. W. (1984). The distribution of positive linear combination of independent chi-squared random variables, *Applied Statistics*, **33**, 332–339.

Gabler, S. and Wolff, C. (1987). A quick and easy approximation to the distribution of a sum of weighted chi-square variables, *Statistische Hefte*, **28**, 317–325.

Hall, P. (1983). Chi-squared approximation to the distribution of a sum of independent random variables, *Ann. Probab.*, **11**, 1028–1036.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables, *Biometrika*, **48**, 417–426.

Jensen, D. R. and Solomon, H. (1972). A Gaussian approximation to the distribution of a definite quadratic form, *J. Amer. Statist. Assoc.*, **67**, 898–902.

Lindsay, B. G. (1989*a*). On the determinants of moment matrices, *Ann. Statist*, **17**, 711–721.

Lindsay, B. G. (1989*b*). Moment matrices: Applications in mixtures, *Ann. Statist.*, **17**, 722–740.

Lindsay, B. G. and Basak, P. (1995). Moments determine the tail of a distribution (but not much else), Tech. Report, No. 95-7, Center for Likelihood Studies, Department of Statistics, The Pennsylvania State University, University Park, PA.

Lindsay, B. G. and Roeder, K. (1997). Moment-based oscillation properties of mixture models, *Ann. Statist.*, **25**, 378–386.

Lindsay, B. G., Pilla, R. S. and Basak, P. (1995) Moment-based approximations of distributions using mixtures: theory and applications, Tech. Report, No. 95-9, Center for Likelihood Studies, Department of Statistics, The Pennsylvania State University, University Park, PA.

Oman, S. D. and Zacks, S. (1981). A mixture approximation to the distribution of a weighted sum of chi-squared variables, *J. Statist. Comput. Simulation*, **13**, 215–224.

Pilla, R. S. (1995). *Approximation Using Gamma Mixtures (AUGM): A special version of the FORTRAN program*, Department of Statistics, The Pennsylvania State University, University Park, PA-16802 (Program is available by sending an email to pillar @ helix-nih-gov).

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed., Press Syndicate of the University of Cambridge, Cambridge.

Ruben, H. (1962). Probability concent of regions under spherical distributions, IV: The distribution of homogeneous and non homogeneous quadratic functions of normal variables, *Ann. Math. Statist.*, **33**, 542–570.

Sheil, J. and O'Muircheartaigh, I. (1977). The distribution of non-negative quadratic forms in normal variables, *Applied Statistics*, **26**, 92–98.

Solomon, H. and Stephens, M. A. (1977). Distribution of a sum of weighted chi-square variables, *J. Amer. Statist. Assoc.*, **72**, 881–885.

Stephens, M. A. (1986). Tests based on EDF statistics, Chapter 4, *Goodness-of-Fit Techniques*, Marcel & Decker, New York.

Stuart, A. and Ord, J. K. (1987). *Kendall's Advanced Theory of Statistics, Vol. 1, Distribution Theory*, Fifth Edition, Oxford University Press, New York.

Waller, L. A., Turnbull, B. W. and Hardin, J. M. (1995). Obtaining distribution functions by numerical inversion of characteristic functions with application, *Amer. Statist*, **49**, 346–350.

Wood, A. T. A. (1989). An $F$ approximation to the distribution of a linear combination of chi-squared variables, *Commun. Statist. Simulation Comput.*, **18**, 1439–1456.