Gruppo di Ricerca per le Applicazioni
della Statistica ai Problemi Ambientali

Working Paper GRASPA n. 4

Francesca Di Salvo Gianfranco Lovison

Parametric inference on composite samples with random weights

Il lavoro è frutto della collaborazione tra i due autori. Per quanto riguarda la stesura, Francesca Di Salvo ha curato i capitoli 3 e 4; Gianfranco Lovison ha curato i capitoli 1 e 2 e l'Appendice.

Lavoro svolto nell'ambito del progetto di interesse nazionale MURST 1998 "Metodi Statistici per l'Analisi dell'Ambiente e delle Interazioni Ambiente-Salute"

Parametric inference on composite samples with random weights

F. Di Salvo, G. Lovison
Istituto di Statistica
Università di Palermo - Italia

Abstract

Composite sampling involves taking a number of individual sampling units from a random variable X and mixing them before measuring the final composite sample. By this procedure a vector $\boldsymbol{x}$ of individual values and a matrix $\boldsymbol{W}$ of weights are produced (but never observed), whereas the observed data consist of the composited measurements $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$. The matrix of weights $\boldsymbol{W}$ can be either fixed or random, depending on the characteristics of the compositing procedures. The paper deals with the problems that arise in making parametric inference about the distribution of X using $\boldsymbol{y}$, when the weights are random. Attention is focussed on the case of Gamma distributed individual values $\boldsymbol{x}$ and Beta distributed weights $\boldsymbol{W}$, a situation of specific interest for environmental applications.

Keywords: Composite sampling, estimation, random weights, randomly weighted averages, environmental monitoring.

1 Introduction

Composite sampling involves taking a number of individual sampling units and mixing them before measuring, counting or otherwise analyzing the final composite sample so obtained. Mixing is usually physical, by a blender, homogenizer or other blending device, but other types of mixing (biological, mechanical, etc.) are possible. The basic idea behind composite sampling is to treat the process of sample selection (i.e., identification of sampling units to be included in the sample) as distinct from that of sample quantification (i.e., making measurements on the selected sample units). By this distinction, it is possible to achieve two conflicting goals simultaneously. On one hand, one can select a large

Correspondence address:
Gianfranco Lovison
Istituto di Statistica - Università di Palermo
Viale delle Scienze 90128 Palermo - Italia
Work supported by MURST 1998 grant.

number of sample units to ensure the desired sample support, so that the results of the analysis will be sufficiently reliable. On the other hand, by compositing several individual samples into fewer composites, one reduces the number of analytical measurements so that a study can be carried out within the available financial resources. This cost efficiency is giving composite sampling a prominent role in environmental studies, where the cost of analytical measurement is usually greater that that of sampling. A recent review on composite sampling is in Lovison, Gore, Patil (1994). The annotated bibliography of composite sampling (Boswell, Gore, Lovison, and Patil, 1995) may also be useful for a review of composite sampling methods and applications.

Composite sampling is used for drawing inference on the parameters of a (possibly multivariate) random variable in the population of interest or for classifying the individual sampling units as having or not a (possibly multivariate) trait. In this paper we concentrate on the use of composite samples for parametric inference on a continuous r.v.; terminology and notation largely follow those introduced by Lovison, Gore, Patil (1994). In Sec. 2 we set the stage by introducing the necessary notation and recalling some basic results concerning estimation with composite data. In Sec. 3 we propose likelihood-based methods for inference on composite data with random weights, concentrating on the case of Gamma distributed individual values and Beta distributed weights. Finally, section 4 is devoted to a discussion of the merits, limitations and possible extensions of the methods proposed.

2 Parametric inference with composite data

Let X be a continuous r.v., with p.d.f. $f_X(x;\boldsymbol{\theta})$, and suppose $m = ck$ individual sampling units are drawn at random from such a population and randomly allocated into c subsets. Each subset is then composited; there are therefore c composite samples of size k, and s subsamples are randomly selected from each. In the rest of this paper we assume that:

- the compositing procedure is *exclusive* (see Lovison, 1994), i.e. each individual sampling unit contributes to one and only one composite sample;

- the measurement error, made on the final composite sample, is absent or negligible compared with the other sources of variability.

Define:

$\boldsymbol{x}_j = [x_{j1}, x_{j2}, \ldots, x_{j\ell}, \ldots x_{jk}]'$ the vector of X-values taken up by the k individual sampling units forming the j-th composite, $j = 1, \ldots, c$; and:

$\boldsymbol{v}_{ji} = [v_{ji1}, v_{ji2}, \ldots, v_{ji\ell}, \ldots, v_{jik}]'$ the vector of physical sizes (volumes, masses, amounts, etc.) of the portions contributed by the k the individual sampling units to the i-th subsample from the j-th composite. In composite sampling analysis, we are mainly interested in the *relative physical size*, usually called weight and denoted by $w_{ji\ell} = v_{ji\ell}/sum_{\ell=1}^{k} v_{ji\ell}$, with which each individual sampling unit contributes to the i-th subsample from the j-th composite, $j = 1, \ldots, c$; $i = 1, \ldots, s$. Such weights can be arranged in a vector $\boldsymbol{w}_{ji} = [w_{ji1}, w_{ji2}, \ldots, w_{ji\ell}, \ldots, w_{jik}]'$. If the non-zero weights are arranged to be contiguous, i.e. to form a sub-(row)vector, and the elements of $\boldsymbol{x}$ are reordered accordingly, we can write:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{O}_{s\times k} & \vdots & \boldsymbol{O}_{s\times k} \\ \boldsymbol{O}_{s\times k} & \boldsymbol{W}_2 & \vdots & \boldsymbol{O}_{s\times k} \\ \ldots & \ldots & \boldsymbol{W}_j & \ldots \\ \boldsymbol{O}_{s\times k} & \boldsymbol{O}_{s\times k} & \vdots & \boldsymbol{W}_c \end{bmatrix}$$

where

$$\boldsymbol{W}_j = \begin{bmatrix} \boldsymbol{w}'_{j1} \\ \vdots \\ \boldsymbol{w}'_{ji} \\ \vdots \\ \boldsymbol{w}'_{js} \end{bmatrix}$$

is the $s \times k$ sub-matrix of non-zero weights with which the individual sampling units enter the s subsamples from the j-th composite sample. The block diagonal pattern of $\boldsymbol{W}$ is due to the assumption of exclusiveness, since if a compositing procedure is exclusive two rows $\boldsymbol{w}'_{ji}, \boldsymbol{w}'_{j'i'}$ of $\boldsymbol{W}$, pertaining to two different composite samples, $j \neq j'$, cannot overlap, i.e. thay cannot have a non-zero weight in the same column. Clearly, the m-vector $\boldsymbol{x}$ is patterned accordingly:

$$\boldsymbol{x} = [\boldsymbol{x}'_1 \ldots \boldsymbol{x}'_j \ldots \boldsymbol{x}'_c]'$$

Then:

$$y_{ji} = \Sigma_{\ell=1}^{k} w_{ji\ell} x_{j\ell} = \boldsymbol{w}'_{ji} \boldsymbol{x}_j$$

is the X-value measured (without, or with negligeable, error) on the i-th sub-sample from the j-th composite. If we arrange the y_{ji}'s in a vector $\boldsymbol{y}$ of dimension $n = cs$, these *composited measurements* can be written in matrix form as:

$$\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$$

Cast in this notation, the problem of parametric inference with composite data can be expressed as follows: we want to make inference on the parameter $\boldsymbol{\theta}$ of X (and, possibly, on the parameter of the distribution of the weights, if they are random as well), but we are bound to do it using only the composited measurements $\boldsymbol{y}$.

2.1 Fixed weights

Most of the literature about inference on composite data has concentrated on the case of fixed and uniform weights: $w_{ji\ell} = \frac{1}{k}$, $\forall \ell$, and particularly on the first two moment properties of the composited estimator of $E(X)$:

$$\bar{y} = \sum_{j=1}^{c} \sum_{i=1}^{s} y_{ji}$$

For a comprehensive review, reference should be made to Lovison, Gore, Patil (1994).

2.2 Random weights

There are two different situations which may lead to random weights:

(a) when the weights are generated by a truly random mechanism not under the researcher's control. An example of genuinely random mechanism generating the weights is the compositing of water samples taken with volumes proportional to flow (Schaeffer, Kerster and Janardan, 1980; Schaeffer et al., 1983);

(b) when the weights are fixed by the sampler, but the compositing apparatus fails to mix them thoroughly; then, if the composite is subsampled, the proportion in each subsample contributed by each compositing object varies from subsample to subsample according to some probability distribution. The less satisfactorily the compositing device blends the individual smapling units, the larger the variance of such a distribution will be. Imperfect physical mixing leading to random weights is more likely to happen when subsampling from discrete, countable materials (for example, kernels in a barge-load of corn, Elder, Thompson and Myers, 1980) and from bulk or viscous materials (wool, Brown and Fisher, 1972; coal, Duncan, 1962) than from liquid or gaseous materials (but Rohde, 1976, advocates the use of random weights in compositing water samples in order to estimate the density of phytoplankton, zooplankton or pollen grains).

There are two approaches in the literature to the problem of modeling the variability due to the compositing itself: the variance-components approach suggested by Duncan (1962) and the randomly weighted averages approach introduced by Brown and Fisher (1972), and developed by Rohde (1976), Elder (1977), and Elder, Thompson and Myers (1980). The work presented ih this paper follows the randomly weighted averages approach; a good discussion of the advantages of this approach when compared with the variance-components model is in Elder, Thompson and Myers (1980).

3 Parametric inference with random weights

If the weights are random, we must consider $\boldsymbol{w}_{ji}$ as the sample realization of a random variable W_j, with p.d.f. $f_{W_j}(\boldsymbol{w}_j;\boldsymbol{\psi})$. Theoretically, such random variables might be different from one sample to another, but, to keep things simple, we assume in the following that all vectors $\boldsymbol{w}ji$ are drawn independently from the same distribution W with p.d.f. $f_W(\boldsymbol{w};\boldsymbol{\psi})$. Under this assumption, along with that of exlusiveness, we may look at the whole compositing procedure as the union of c independent procedures, one for each composite, and write:

$$f_Y(\boldsymbol{y};\boldsymbol{\theta},\boldsymbol{\psi}) = \prod_j^c \prod_i^s f_Y(y_{ji};\boldsymbol{\theta},\boldsymbol{\psi}) \tag{1}$$

So, the basic task is to derive the distribution of $Y_{ji} = \sum_l^k w_{jil}x_{jl}$. The problem has been tackled by Lovison(1994) with respect to the first two moments of Y. Johnson, Kotz

(1992) address the same problem both in terms of moments and in term of full distribution characterization. Their paper is apparently the only reference on this problem in the literature; unfortunately, Johnson and Kotz only concentrate on a few distributions for X (Uniform, Normal, Cauchy) which give analytically treatable results for $f_Y(y_j i; \boldsymbol{\theta}, \boldsymbol{\psi})$. These distributions are hardly appropriate for the environmental applications of interest in this paper where, usually, X is a contaminant concentration, which is known to have a positively skewed distribution on a non-negative support.

To tackle the problem of deriving the distribution of Y we shall work in this paper with the simple case of composite sample size $k = 2$, when $Y = WX_1 + (1 - W)X_2$, and $s = 1$, i.e. the composite is either totally measured or only one subsample is drawn from it for laboratory measurement, and (1) simplifies to

$$f_Y(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_j^c f_Y(y_j; \boldsymbol{\theta}, \boldsymbol{\psi})$$

Some remarks about generalizations to $k > 2$ and $s > 1$ will be given in Sec.4.

Moreover, in order to focus on the type of composite data that arise in environmental applications, in what follows we shall assume:

$$X \sim Gamma(\alpha, \lambda), \quad W \sim Beta(\delta, \delta)$$

The assumption of a Gamma model for the distribution of a contaminant has a well established tradition; it satisfies the requirements of non-negative support and positive asymmetry; moreover, the two-parameters version offers great flexibility for the shape of the assumed distribution.

On the other hand, the choice of a one-parameter Beta (if $k = 2$) or Dirichlet (if $k > 2$) distribution for the random weights is somewhat standard. It satisfies the requirement that W be defined over the simplex, and has enough flexibility to cover most situations of interest, including the "uninformative" case of Rectangular weights ($\delta = 1$) and the degenerate cases of fixed weights ($\delta \to \infty$) and lack of mixability ($\delta \to 0$). Besides, Rohde (1976), using a characterization theorem of Fabius(1973), has given a physical justification to the use of Beta (or Dirichlet) weights in composite sampling applications. Notice that the assumption of one parameter Beta distribution implies that the distribution of weights is symmetric with $E(W) = 1/2$, i.e. that the weights are 'unbiased' (see Lovison, 1994).

3.1 Moments method estimation

The result of Johnson, Kotz (1990) on the moments of a randomly weighted average of two i.i.d. random variables:

$$E(Y^r) = \mu_r(Y) = \sum_{j=1}^{r} \binom{r}{j} E[W^j(1-W)^{r-j}\mu_j(X)\mu_{r-j}(X) \tag{2}$$

suggests the method of moments as a natural approach to estimation of α, β, δ. In the case of Gamma r.v.'s randomly weighted by Beta weights, formula (2) yealds, for the first three moments of Y:

$$\begin{aligned} \mu_1(Y) &= \frac{\alpha}{\lambda} \\ \mu_2(Y) &= \frac{\alpha^2}{\lambda^2} + \frac{\delta+1}{2\delta+1}\frac{\alpha}{\lambda^2} \\ \mu_3(Y) &= \frac{\alpha^3}{\lambda^3} + 3\frac{\delta+1}{2\delta+1}\frac{\alpha^2}{\lambda^3} + \frac{\delta+2}{2\delta+1}\frac{\alpha}{\lambda^3} \end{aligned}$$

Substituting $\mu_1(Y), \mu_3(Y), \mu_3(Y)$ with M_1, M_2, M_3, the first three sample ordinary moments calculated from n composited measurements, we obtain the system of moments equations:

$$\begin{cases} M_1 = \frac{\tilde{\alpha}}{\tilde{\lambda}} \\ M_2 = \frac{\tilde{\alpha}^2}{\tilde{\lambda}^2} + \frac{\tilde{\delta}+1}{2\tilde{\delta}+1}\frac{\tilde{\alpha}}{\tilde{\lambda}^2} \\ M_3 = \frac{\tilde{\alpha}^3}{\tilde{\lambda}^3} + 3\frac{\tilde{\delta}+1}{2\tilde{\delta}+1}\frac{\tilde{\alpha}^2}{\tilde{\lambda}^3} + \frac{\tilde{\delta}+2}{2\tilde{\delta}+1}\frac{\tilde{\alpha}}{\tilde{\lambda}^3} \end{cases}$$

Solving by substitution gives:

$$\begin{aligned} \tilde{\alpha} &= \frac{M_1^2}{(M_2 - M_1^2)}\frac{\tilde{\delta}+1}{2\tilde{\delta}+1} \\ \tilde{\lambda} &= \frac{M_1}{(M_2 - M_1^2)}\frac{\tilde{\delta}+1}{2\tilde{\delta}+1} \\ \frac{\tilde{\delta}}{(2\tilde{\delta}+1)^2} &= \frac{M_1M_3 - 2M_2^2 + M_1^2M_2}{(M_2 - M_1^2)^2} \end{aligned} \tag{3}$$

which implies that moment estimators of α, β, δ can be found by solving for $\tilde{\delta}$ the quadratic equation:

$$c(M_1, M_2, M_3)\tilde{\delta}^2 + [2c(M_1, M_2, M_3) - 1]\tilde{\delta} + c(M_1, M_2, M_3) = 0$$

where: $c(M_1, M_2, M_3) = \frac{M_1 M_3 - 2M_2^2 + M_1^2 M_2}{(M_2 - M_1^2)^2}$.

Unfortunately this is only apparently a nice solution to our estimation problem. In fact, closer examination of the discriminant:

$$[2c(M_1, M_2, M_3) - 1]^2 - 4[c(M_1, M_2, M_3)]^2 = 1 - 4c(M_1, M_2, M_3)$$

reveals that admissible, i.e. real and positive, solutions for $\tilde{\delta}$ can be found only when $0 \leq c(M_1, M_2, M_3) \leq 1/4$, a very narrow interval, often missed in real datasets. Besides, even within this admissibility interval there are two solutions, inversely related, i.e. such that $\tilde{\delta}_1 = 1/\tilde{\delta}_2$, each giving in turn a different estimator for (α, λ). Therefore, moments estimators are not only rarely admissible, but also not unique when they exist.

In conclusion, the only merit of the method of moments seems to be its ease of computation. As such, at best it can provide reasonable starting values for α and λ within other, more suitable estimation procedures, as long as a sensible guess can be made about the value of δ. It is in this role that it will be used in the numerical evaluation of maximum likelihood estimators in the next section.

3.2 Likelihood based inference

3.2.1 Derivation of the distribution of the random mixture Y

To derive the distribution of Y in this case, we begin by establishing the conditional distribution of $Y \mid W$, $f_{Y|W}(y \mid w; \alpha, \lambda, \delta)$. For any $w \in (0,1)$, the distributions of wX_1 and $(1-w)X_2$ are again Gamma with the scale parameter multiplied by $1/w$ and $1/(1-w)$ respectively:

$$f_{wX_1}(wx_1; \alpha, \lambda) = \frac{\left(\frac{\lambda}{w}\right)^{\alpha} z_1^{\alpha-1} \exp\left[-\frac{\lambda}{w} z_1\right]}{\Gamma(\alpha)} \quad (4)$$

$$f_{(1-w)X_2}((1-w)x_2; \alpha, \lambda) = \frac{\left(\frac{\lambda}{1-w}\right)^{\alpha} z_2^{\alpha-1} \exp\left[-\frac{\lambda}{(1-w)} z_2\right]}{\Gamma(\alpha)} \quad (5)$$

Moreover, it is easy to show that they are still independent; hence the p.d.f. of $Y \mid w = wX_1 + (1-w)X_2$ can be obtained as a special case of 'weighted convolution' of two i.i.d. Gamma distributions, with $w_1 = w$ and $w_2 = (1-w)$, along the line of derivation developed in the Appendix:

$$\begin{aligned} f_{Y|W}(y \mid w; \alpha, \lambda) &= \frac{\left(\frac{\lambda}{w}\right)^{\alpha}\left(\frac{\lambda}{1-w}\right)^{\alpha} y^{2\alpha-1}\exp\left[-\frac{\lambda}{1-w}y\right]}{\Gamma(\alpha)}\left[y\lambda 2\left(\frac{w-\frac{1}{2}}{w(1-w)}\right)\right]^{\frac{1}{2}-\alpha} \times \\ & \exp\left[y\lambda\left(\frac{w-\frac{1}{2}}{w(1-w)}\right)\right]\sqrt{\pi}\mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{\lambda y\left(w-\frac{1}{2}\right)}{w(1-w)}\right) \qquad (6) \end{aligned}$$

where $\mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{\lambda y\left(w-\frac{1}{2}\right)}{w(1-w)}\right)$ is the modified Bessel function of the first type of order $(\alpha-\frac{1}{2})$ and argument $\left(\frac{\lambda y\left(w-\frac{1}{2}\right)}{w(1-w)}\right)$.

As remarked in the Appendix, this conditional distribution is connected with both the so-called 'Bessel function distributions' (Johnson, Kotz, Balakhrishnan, 1994, pp.50-52) and the distribution proposed by Sim(1992) for the convolution of Gamma r.v.'s with different scale parameters.

After some algebra, the joint distribution $f_{Y,W}(y, w; \alpha, \lambda, \delta) = f_{Y|W}(y \mid w; \alpha, \lambda) f_W(w; \delta)$ can be written as:

$$f_{Y,W}(y, w; \alpha, \lambda, \delta) = \frac{\lambda^{\alpha+\frac{1}{2}}\sqrt{\pi}\left(\frac{y}{2}\right)^{\alpha-\frac{1}{2}}}{\mathrm{B}(\delta,\delta)\Gamma(\alpha)} \frac{\left(w-\frac{1}{2}\right)^{\alpha-\frac{1}{2}}[w(1-w)]^{\delta-\frac{3}{2}}\mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{\lambda y\left(w-\frac{1}{2}\right)}{w(1-w)}\right)}{\exp\left(\frac{\lambda y}{2w(1-w)}\right)}$$

Finally, by integrating out w, we obtain the marginal distribution of Y:

$$f_Y(y; \alpha, \lambda, \delta) = \frac{\lambda^{\alpha+\frac{1}{2}}\sqrt{\pi}\left(\frac{y}{2}\right)^{\alpha-\frac{1}{2}}}{\mathrm{B}(\delta,\delta)\Gamma(\alpha)} \int_0^1 \frac{\left(w-\frac{1}{2}\right)^{\alpha-\frac{1}{2}}[w(1-w)]^{\delta-\frac{3}{2}}\mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{\lambda y\left(w-\frac{1}{2}\right)}{w(1-w)}\right)}{\exp\left(\frac{\lambda y}{2w(1-w)}\right)} dw \qquad (7)$$

The integral in (7) cannot be found analytically; therefore numerical integration is required. We employed the `NIntegrate` function in Mathematica 4.0 to obtain numerical evaluations of (7). Figg. 1,2,3 give some examples of the form of (7), for different values of the parameters α, λ, δ.

Figure 1: Random mixture of G(1,1) with Beta(δ), $\delta \to 0, \delta = 0.5, 1, 2, \delta \to \infty$

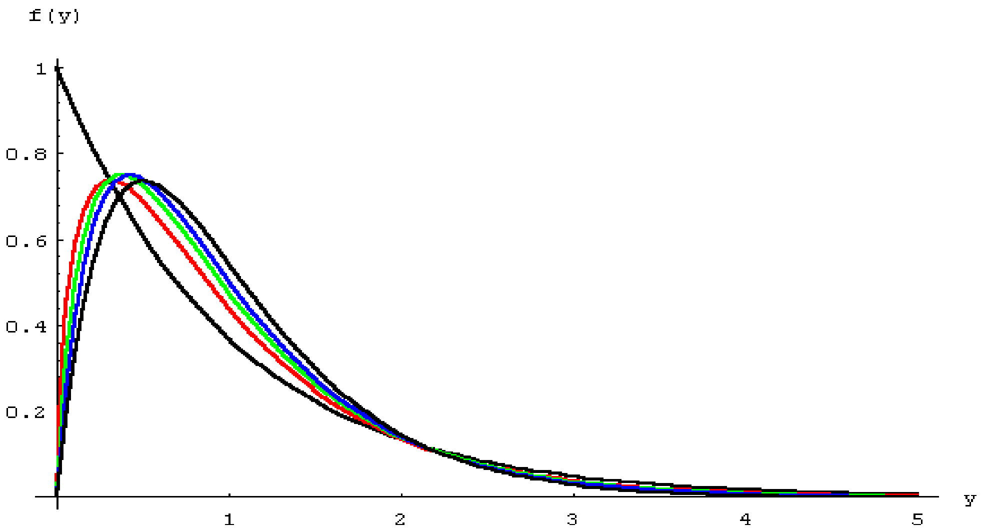


Figure 2: Random mixture of G(3,2) with Beta(δ), $\delta \to 0, \delta = 0.5, 1, 2, \delta \to \infty$

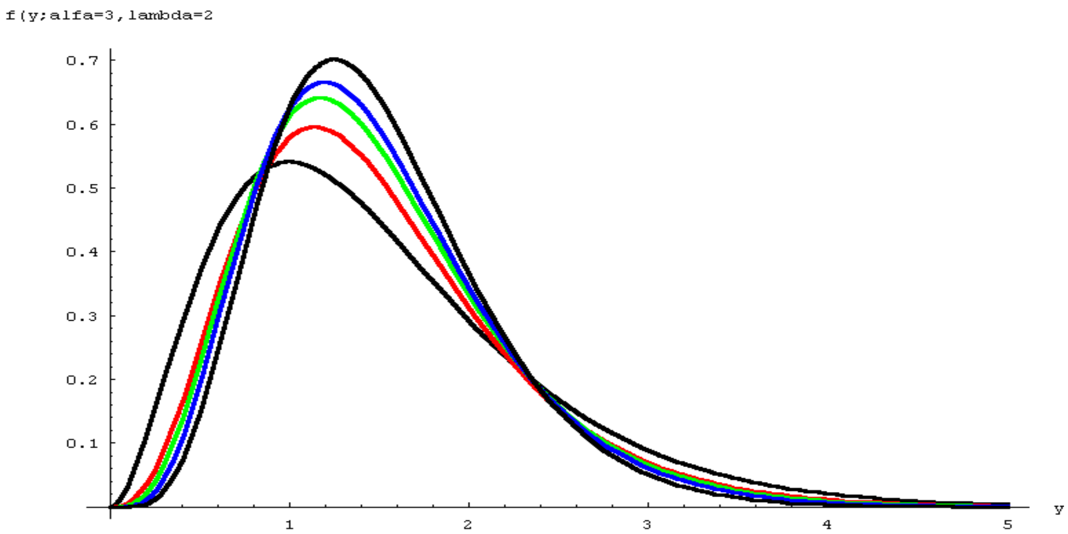
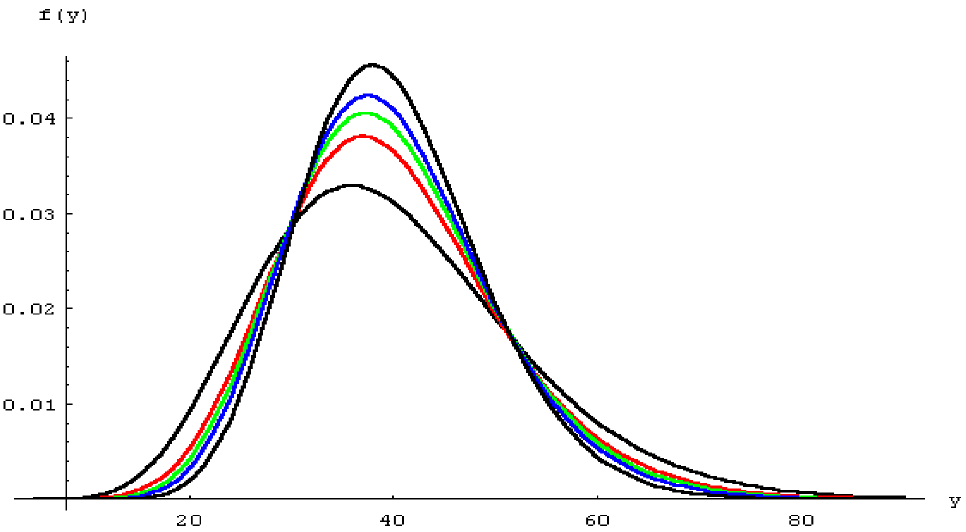


Figure 3: Random mixture of G(10,0.25) with Beta(δ), $\delta \to 0, \delta = 0.5, 1, 2, \delta \to \infty$



Inspection of Figg. 1-3 suggests an interesting limiting behaviour of the random mixture Y with respect to δ. As $\delta \to 0$, the distribution of Y tends to the original Gamma(α, λ) distribution of each one of the two components X_1 and X_2. As $\delta \to \infty$, the distribution of the random mixture Y tends to Gamma$(2\alpha, 2\lambda)$, the distribution of the mean $\bar{X}$ of X_1 and X_2.

This intuitive result can be justified in a more theoretical way; a sketch of proof is as follows.

As $\delta \to 0$, the Beta(δ, δ) distribution tends to a (degenerate) two-points distribution, with two equal probability masses concentrated at 0 and 1; as a consequence the conditional distribution (6) tends to either one of the original Gamma(α, λ) distributions.

As $\delta \to \infty$, the Beta(δ, δ) distribution tends to a (degenerate) distribution with all probability mass at $E[Beta(\delta, \delta)] = 1/2$. As a consequence (6) tends to the distribution of $\frac{1}{2}X_1 + \frac{1}{2}X_2$.

Summarizing, the distribution of Y is 'sandwiched' between the Gamma(α, λ) distribution of X and the Gamma$(2\alpha, 2\lambda)$ distribution of the mean $\bar{X}$ of X_1 and X_2. This implies that, with a composite sample size $k = 2$, we cannot expect much variation in the distribution of Y, since the distribution of the mean of a random sample of size 2 does not differ dramatically from the distribution of the individual sample values. As a consequence, we also expect low information in a sample of composited data $\boldsymbol{y}$ about the parameter δ.

3.2.2 Maximum likelihood estimation through Simulated Annealing

Let $\boldsymbol{y} = [y_1, y_2, \ldots, y_n]'$ be a vector of n independent composited measurements drawn from (7). Then the likelihood function for $(\alpha, \lambda, \delta)$ is given by:

$$
\begin{aligned}
L(\alpha, \lambda, \delta \mid \boldsymbol{y}) &= \left[\frac{\lambda^{\alpha+\frac{1}{2}}\sqrt{\pi}}{\mathrm{B}(\delta,\delta)\Gamma(\alpha)} \right]^n \times \\
& \prod_{i=1}^{n} \left(\frac{y_i}{2}\right)^{\alpha-\frac{1}{2}} \int_0^1 \frac{\left(w-\frac{1}{2}\right)^{\alpha-\frac{1}{2}} [w(1-w)]^{\delta-\frac{3}{2}} \mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{\lambda y_i \left(w-\frac{1}{2}\right)}{w(1-w)}\right)}{\exp\left(\frac{\lambda y_i}{2w(1-w)}\right)} dw \quad (8)
\end{aligned}
$$

Clearly, given the complexity of this function, analytical maximization is unfeasible; we have therefore resorted to numerical methods for finding the MLE's of $(\alpha, \lambda, \delta)$. Since standard numerical optimization strategies are not appropriate either, we have tackled the problem by an iterative stochastic algorithm, the Simulated Annealing Algorithm (S.A.A.), which is frequently used in statistics for complex optimization problems, including derivation of maximum likelihood estimators.

The Simulated Annealing Algorithm is based on the analogy with an algorithm used in Statistical Physics for simulation of the annealing of a solid to its minimum energy state. An attractive feature of S.A.A. is the advantage of being generally applicable; a detailed description of necessary and sufficient conditions to ensure asymptotic convergence properties is in Aarts E.H.L., Laarhoven P.J.M. Van (1989). Here we just notice that to ensure convergence to the global maximum, the choice of the basic ingredients for implementation, a vector of starting values of parameters and a stopping rule, must be strongly dependent on the shape of the specific likelihood function under investigation. For computational convenience we work with the transformation $-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})$, and consequently we deal with the minimization of an objective function having many local minima.

The standard S.A.A. for minimizing a function is a random walk, evolving through the following steps:

1. at some value of a control parameter T, an initial point is picked up within the parameter space;

2. from this starting point a new one is produced in the neighbourhood, by selecting a random distance and direction within the parameter space; then the function is evaluated at the new point;

3. the point reached at Step 2 is accepted by using the Metropolis criterion: if the objective function is decreased at the new point, this is accepted; otherwise, we move to the new point with transition probability:

$$p_{old,new} = \exp\left[\frac{(-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})_{new} - (-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})_{old})}{T}\right]$$

   i.e. if and only if a standard uniform random number u satisfies the condition $u \leq p$.

4. the starting point is updated if the new point is accepted and the parameter control is reduced by some multiplicative constant $k, 0 < k < 1$.

5. steps 2-4 are repeated until the stopping rule

$$0 < (-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})_{new} - (-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})_{old}) < \epsilon$$

   is satisfied for some small, predefined ϵ.

Due to the highly irregular behaviour of the likelihood, we have found the random path in $\Re^{3+}$ of the standard S.A.A. to perform unsatisfactorily in the search for an approximation of the global minimum. To overcome this difficulty, we have implemented a modified algorithm, which we call Alternating Simulated Annealing Algorithm (A.S.A.A.), consisting of two phases:

Phase 1) minimize the objective function for α and λ, holding δ fixed, by using steps 2 - 5 of the S.A.A. and starting from some tentative estimates $\tilde{\alpha}$ and $\tilde{\lambda}$ with fixed δ;

Phase 2) minimize the objective function for δ, again using steps 2 - 5 of the S.A.A. and starting from the provisional estimate found at Phase 1 for α and λ.

Phases 1 and 2 alternate, and the algorithm is terminated when one of the following criteria is satisfied:

either

(i) a prescribed number ν of successive steps (we have used $\nu = 30$) fails to produce an acceptable new point;

or

(ii) from one stage to the next there is no appreciable decrease of the objective function.

The procedure requires an initial value of δ. We have chosen to repeat the A.S.A.A. three times, starting from $\delta = 0.5, \delta = 1$ and $\delta = 2$, and employing these values to obtain moment estimates of α and λ (see formula (3)) as starting values for Phase 1 of the algorithm. The rationale behind this choice is that, as mentioned in Sec. 3, values of δ less than 1, equal to 1 and greater than 1 are qualitatively different, corresponding to completely different levels of mixability of the material.

Finally, the best solution among all those provided by the different starting values is taken as the maximum likelihood estimate of the vector $(\alpha, \lambda, \delta)$.

It should be underlined that this modified version of the algorithm returns a local minimum depending on the starting values, and there is no assurance that a global optimum is found in a finite time. However, using more than one starting vector of initial estimates contributes to improve the performance of the search. To this end, the probability mechanism used for generating new points is an important factor. The control parameter T rules the transition probability; its initial value must suitably depend on the magnitude of the objective function, because it must be selected in order to inhibit, but not to eliminate, the opportunity for $-\log L(\alpha, \lambda, \delta \mid \boldsymbol{y})$ to increase and jump from a local minimum to another. Suitable *ad hoc* choices must also be made concerning other settings of the procedure (like the constant k, the distributions of the random distances and directions, etc.) so that the search procedure for estimation of δ identifies a smaller region to be explored and a more rigorous acceptance - rejection rule is imposed for the estimation of α and λ.

We intend to investigate the features of parametric model (7) and the performance of the modified algorithm by a simulation experiment: this will entail the simulation of the sampling distribution of the Maximum Likelihood Estimators of $(\alpha, \lambda, \delta)$. To this end, we set up a simulation plan involving several triplets of parameters, in order to cover various situation of practical interest, and different samples size. In particular we have fixed samples size of $n = 5$ and $n = 10$ composited measurements and values of the parameters of $\alpha = 0.5, 1, 4, 8$; $\lambda = 0.5, 1, 1.5, 5$; $\delta = 0.5, 1, 2$.

In the procedure implemented in Mathematica, for each set of parameters we simulate $m = 200$ samples; the data are simulated from the distribution of composite measurements, generating random numbers from the Gamma distribution $\mathrm{G}(\alpha, \lambda)$ and combining them with weights randomly generated from a Beta (δ, δ). For each sample the maximum likelihood estimates of the parameters are calculated by applying the A.S.A.A. described above. The sampling distribution of the maximum likelihood estimators is finally estimated through the observed frequency distribution of the 200 estimates and used to evaluate the performance of the estimation procedure.

3.2.3 Preliminary results

As an illustration of the more extended simulation study, we present the results from the simulation for the set $(\alpha = 4.0, \lambda = 1.5, \delta = 0.5)$ and $n = 5$. Figures 4, 5 and 6 show the observed frequency distributions of the maximum likelihood estimates. The distributions of estimates of α, λ (see Figure 4 and 5) present common characteristics, the most noticeable being the bias and skewness of these distributions. This result is not surprising, since it is common knowledge that the maximum likelihood estimators sampling distributions of the parameters α, λ in the Gamma density function are skewed, with bias a monotone decreasing function of the sample size; with small samples, like that used in this simulation, this behaviour is confirmed in the randomly weighted Gamma distribution.

Figure 4: Empirical distribution for the 200 estimates of α

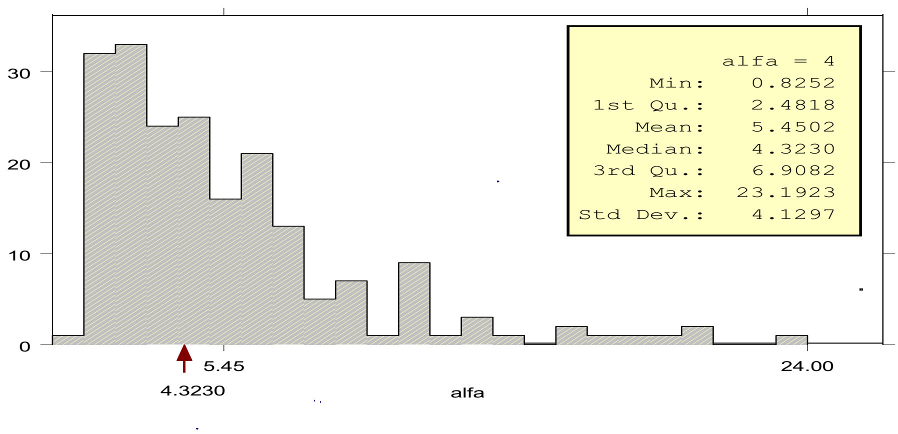


Figure 5: Empirical distribution for the 200 estimates of λ

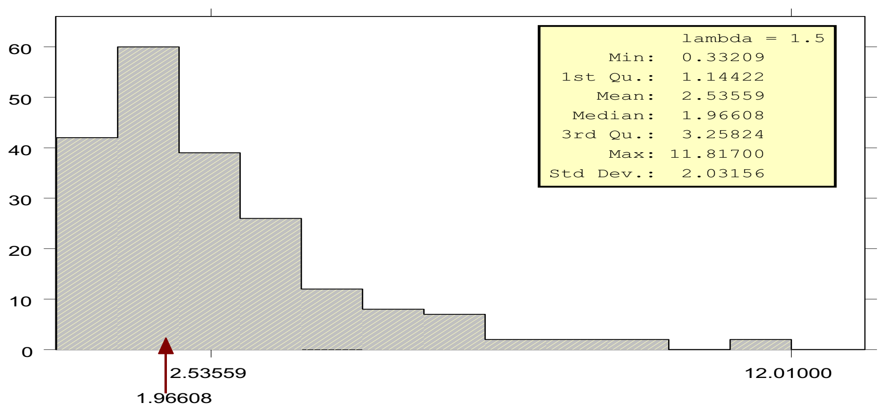
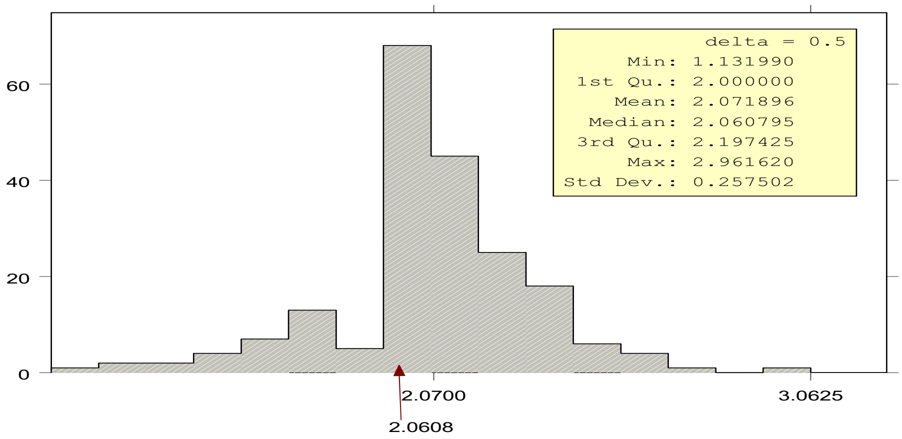


Figure 6: Empirical distribution for the 200 estimates of δ



More information about the asymptotic behaviour of ML estimators might be collected simulating their distribution at increasing sample sizes, but this is not of practical interest, since in the context of composite sample one typically deals with small samples.

As for the sampling distribution of the ML estimator of δ, Figure 6 bears out what we already anticipated, on theoretical grounds, in Sec. 3.2.1 about the amount of information on the distribution of the weights carried by composites of two Gamma distributions: the likelihood is basically unable to discriminate among the three different processes corresponding to weights generated randomly from a Beta r.v. with parameter $0 < \delta < 1, \delta = 1, \delta > 1$.

4 Discussion

The results presented in the previous section are a tentative example of the way we approach parametric model (7); obviously, to make an exhaustive analysis, more simulation experiments would be required. This will be done in a forthcoming, more complete paper containing results for the different sets of parameters in Table 1. Besides, the paper will also include:

- a study of the behaviour of the conditional parametric distribution (5) obtained when the weights are fixed and known, to evaluate the behaviour of the ML method, when one has sound conjectures about the process of compositing;

- the case of composite samples with $k > 2$ individual units for each composite, when we expect the likelihood to be able to better discriminate between different values of δ;

- the case of $s > 1$ subsamples from each composite. It is worth recalling that, in general, several subsamples can be taken from each composite sample and several measurements can be made on each subsample. Although this possibility is rarely exploited, owing to cost or time constraints, it has been shown to improve the statistical properties of the composite sample data in the fixed weights case (see, for example, Lovison, Gore, Patil, 1994). In the random weights case, the availability of more than one sub-sample from each composite entails the availability of (approximately) independent replicates from the conditional distribution of the random mixture given the (unobserved) weights. This suggests an EM-type algorithm for the estimation of such distribution, an approach that we intend to develop in the future.

References

[1] Aarts E.H.L., Laarhoven P.J.M. van(1989), Simulated Annealing: an introduction. *Statistica Neerlandica*, **43**, 1, 31–52.

[2] Boswell M.T., Gore S., Lovison G., Patil G.P.(1995), Composite sampling: an annotated bibliography. *Environmental and Ecological Statistics*, **21**, 113–120.

[3] Elder R. S., Thompson W. O., Myers R. H. (1980), Properties of composite sampling procedures *Technometrics*, **22**, 179–186.

[4] Johnson N.L., Kotz S. (1990), Randomly weighted averages: some aspects and extensions. *The American Statistician*, **44**, 245–249.

[5] Johnson N.L., Kotz S., Balakhrishnan N. (1994), *Univariate continuous distributions*, 2nd ed. Wiley Interscience, New York.

[6] Schaeffer D. J., Kerster H. W., Janardan K. G. (1980) Grab versus composite sampling: A primer for the manager and engineer. *Environmental Management*, **4**.

[7] Schaeffer D. J., Kerster H. W., Bauer D. R., Rees K., McCormick S.(1983) Composite samples overestimate waste loads. *Water Pollution Control Federation Journal*, **55**, 1387–1392.

[8] Lovison G. (1993), A unified linear model for estimation with composite sample data. In: Patil G.P., Rao C.R.(eds.) *Multivariate Environmental Statistics.* Elsevier Science.

[9] Lovison G., Gore S., Patil G.P.(1994), Design and analysis of composite samples. In: Patil G.P., Rao C.R.(eds.) *Handbook of Statistics, Vol.12 : Environmental Statistics.* Elsevier Science.

[10] Rohde C.A. (1976), Composite sampling. *Biometrics*, **32**, 273–282.

[11] Sim, C.H.(1992), Point processes with correlated gamma interarrival times. *Statistics & Probability Letters*, **15**, 135–141.

Appendix: On the weighted convolution of two Gamma distributions

Let X_1, X_2 be indipendently distributed as $G(\alpha, \lambda)$ r.v.'s and w_1, w_2 be two real positive numbers. We want to derive the distribution of the 'weighted convolution':

$$Y = w_1 X_1 + w_2 X_2. \tag{9}$$

To do so, we first observe that, for any given $w \in (0,1)$, the distributions of $Z_1 = w_1 X_1$ and $Z_2 = w_2 X_2$ are again Gamma with the scale parameter multiplied by $\frac{1}{w_1}$ and $\frac{1}{w_2}$ respectively:

$$f_{wX_1}(z_1; \alpha, \lambda) = \frac{\left(\frac{\lambda}{w}\right)^{\alpha} z_1^{\alpha-1} \exp\left[-\frac{\lambda}{w} z_1\right]}{\Gamma(\alpha)} \tag{10}$$

$$f_{(1-w)X_2}((1-w)x_2; \alpha, \lambda) = \frac{\left(\frac{\lambda}{1-w}\right)^{\alpha} z_2^{\alpha-1} \exp\left[-\frac{\lambda}{1-w} z_2\right]}{\Gamma(\alpha)} \tag{11}$$

Moreover, it is easy to show that they are still independent; hence the p.d.f. of $Y = Z_1 + Z_2 = wX_1 + w_2 X_2$ can be obtained as the (standard) convolution of (10) and (11):

$$f_{Y|W}(y \mid w; \alpha, \lambda) = \frac{\left(\frac{\lambda}{w_1}\right)^{\alpha} \left(\frac{\lambda}{w_2}\right)^{\alpha}}{[\Gamma(\alpha)]^2} \int_0^y z_1^{\alpha-1} (y - z_1)^{\alpha-1} \exp\left[-\frac{\lambda}{w_1} z_1\right] \exp\left[-\frac{\lambda}{w_2}(y - z_1)\right] dz_1$$

which, by the change of variable $t = \frac{z_1}{y}$, gives:

$$f_{Y|W}(y \mid w; \alpha, \lambda) = \frac{\left(\frac{\lambda}{w_1}\right)^{\alpha} \left(\frac{\lambda}{w_2}\right)^{\alpha} y^{2\alpha-1} \exp\left[-\frac{\lambda}{w_2} y\right]}{[\Gamma(\alpha)]^2} \int_0^1 t^{\alpha-1} (1-t)^{\alpha-1} \exp\left[y\left(\frac{\lambda}{w_2} - \frac{\lambda}{w_1}\right) t\right] dt \tag{12}$$

On using a result given by Abramowitz, Stegun (1964, p.505):

$$\int_0^1 t^{\alpha-1} (1-t)^{\alpha-1} \exp\left[y\left(\frac{\lambda}{w_2} - \frac{\lambda}{w_1}\right) t\right] dt = \frac{[\Gamma(\alpha)]^2}{\Gamma(2\alpha)} M\left[\alpha, 2\alpha, y\left(\frac{\lambda}{w_2} - \frac{\lambda}{w_1}\right)\right] \tag{13}$$

where: $M[a, b, c]$ is the confluent hypergeometric function with arguments a, b, c.

By representing $M\left[\alpha, 2\alpha, y\left(\frac{\lambda}{w_1} - \frac{\lambda}{w_2}\right)\right]$ as an infinite series:

$$\int_0^1 t^{\alpha-1} (1-t)^{\alpha-1} \exp\left[y\left(\frac{\lambda}{w_2} - \frac{\lambda}{w_1}\right) t\right] dt = \frac{[\Gamma(\alpha)]^2}{\Gamma(2\alpha)} \sum_{r=0}^{\infty} \frac{(\alpha)_r}{(2\alpha)_r r!} \left[y\left(\frac{\lambda}{w_2} - \frac{\lambda}{w_1}\right)\right]^r \tag{14}$$

where $(a)_r$ is the Pochhammer symbol: $(a)_r = a(a+1)(a+2)\dots(a+r-1) = \frac{\Gamma(a+r)}{\Gamma(a)}$, we can write:

$$f_{Y|W}(y \mid w;\alpha,\lambda) = \frac{\left(\frac{\lambda}{w_1}\right)^{\alpha}\left(\frac{\lambda}{w_2}\right)^{\alpha} y^{2\alpha-1}\exp\left[\frac{-\lambda}{w_2}y\right]}{\Gamma(\alpha)} \times \sum_{r=0}^{\infty}\frac{\Gamma(\alpha+r)}{\Gamma(2\alpha+r)r!}\left[y\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right]^r \quad (15)$$

which is a special case of the distributon proposed by Sim(1992) for the convolution of Gamma r.v.'s with different scale parameters.

By using another result given by Abramowitz, Stegun (1964, p.509), which connects confluent hypergeometric functions and Bessel functions in special cases, we can re-write $M\left[\alpha, 2\alpha, y\left(\frac{\lambda}{w_1}-\frac{\lambda}{w_2}\right)\right]$ as:

$$\begin{aligned} M\left[\alpha,2\alpha,y\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right] &= \Gamma(1+(\alpha-\frac{1}{2}))\exp\left[\frac{y}{2}\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right]\times \\ &\left[\frac{y}{4}\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right]^{\frac{1}{2}-\alpha} \mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{y}{2}\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right) \end{aligned} \quad (16)$$

where $\mathrm{I}_{(a)}[z]$ is the modified Bessel function of the first type of order a and argument z.

Substituting () into (), and noticing that $\Gamma(\alpha+\frac{1}{2}) = \frac{\Gamma(2\alpha)\sqrt{\pi}}{\Gamma(\alpha)2^{2\alpha-1}}$ for the duplication formula of Gamma functions, we obtain:

$$\begin{aligned} f_{Y|W}(y \mid w;\alpha,\lambda) &= \frac{\left(\frac{\lambda}{w_1}\right)^{\alpha}\left(\frac{\lambda}{w_2}\right)^{\alpha} y^{2\alpha-1}\exp\left[-\frac{\lambda}{w_2}y\right]}{\Gamma(\alpha)}\left[y\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right]^{\frac{1}{2}-\alpha}\times \\ &\exp\left[\frac{y}{2}\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right]\sqrt{\pi}\mathrm{I}_{(\alpha-\frac{1}{2})}\left(\frac{y}{2}\left(\frac{\lambda}{w_2}-\frac{\lambda}{w_1}\right)\right) \end{aligned} \quad (17)$$

This form is connected with the so-called 'Bessel function distributions' (Johnson, Kotz, Balakhrishnan, 1994, pp.50-52), which have been introduced by McKay(19) and motivated by Baqtthacharya () for the 'weighted convolution' of Chi-square distributions.

In passing we observe that the derivations (a) and (b) show an equivalence relationship between the family of distributions introduced by Sim(1992) and the 'Bessel function distributions' which appears to have been overlooked in the literature.