Simulating signed mixtures

Julien Stoehr, Christian P. Robert

Simulating signed mixtures

Christian P. Robert[1,3] and Julien Stoehr[1,2]

[1]CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, 75016 Paris, France
[2]UMR MIA Paris-Saclay, INRAE, AgroParisTech, Université Paris-Saclay, 91120 Palaiseau, France
[3]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Authors contributed equally to this work.

Abstract

Simulating mixtures of distributions with both positive and negative (signed) weights proves a challenge as standard simulation algorithms prove inefficient in handling the negative weights. In particular, the natural representation of mixture random variates as being associated with latent component indicators is no longer available. We propose an exact accept-reject algorithm for the general case of finite signed mixtures that relies on optimally pairing positive and negative components and designing a stratified sampling scheme on these pairs. We analyze the performances of our approach, relative to the inverse cdf approach, since the cdf of the targeted distribution remains available for signed mixtures of common distributions.

Keywords. Acceptance-rejection algorithm, Mixtures, Signed mixtures, Simulation, Inverse cdf, Quantile function

1 Introduction

1.1 Signed mixtures

Mixture distributions [Titterington et al., 1985] abound in the statistical literature as a ubiquitous tool to represent inhomogeneous populations and to enlarge the collection of common distributions [see, e.g., McLachlan and Peel, 2000]. The density function of such distributions writes as a linear combination of $K > 1$ base density functions f_k, $1 \leq k \leq K$, namely

$$\sum_{k=1}^{K} \omega_k f_k \,, \quad \text{where} \quad \begin{cases} \sum_{k=1}^{K} \omega_k = 1 \,, \\ \omega_1, \ldots, \omega_K > 0 \,. \end{cases} \tag{1}$$

It is then straightforward to simulate from (1) when the component density functions f_k are themselves easily simulated: a component index $1 \leq k \leq K$ is selected with probability ω_k and a realization from f_k is then produced. This simplicity explains why many simulation methods in the literature exploit an intermediary mixture construction to speed up the production of pseudo-random samples from more challenging distributions. For instance, Devroye [1985, Section XIV.4.5] points out that unimodal distributions can be written as countable mixtures of uniform distributions. Similarly, mixture distributions are often selected as proposals in MCMC algorithms [Robert and Casella, 2004, Cappé et al., 2008].

In this paper, we consider the more challenging setting of *signed* mixtures, namely the case when the mixture weights, $\omega_1, \ldots, \omega_K$, in (1) are *signed*, that is, when some of the ω_k's are *negative*. We define a *signed mixture density* as a linear combination of $P \geq 1$ positively weighted density functions f_k and of $N \geq 1$ negative weighted density functions g_k, with the constraint that the combination remains a properly defined probability density. Denoting by $\mathrm{supp}(h)$ the support of a real-valued function h, this implies that the joint support of the positively weighted density functions must contain the joint support of the negatively weighted density functions, namely

$$\bigcup_{1 \leq k \leq N} \mathrm{supp}(g_k) \subseteq \bigcup_{1 \leq k \leq P} \mathrm{supp}(f_k)\,.$$

A *signed mixture density* m thus writes as

$$m = \sum_{k=1}^{P} \omega_k^+ f_k - \sum_{k=1}^{N} \omega_k^- g_k\,, \quad \text{with} \quad \begin{cases} \sum_{k=1}^{P} \omega_k^+ - \sum_{k=1}^{N} \omega_k^- = 1\,, \\ \omega_1^+, \ldots, \omega_P^+, \omega_1^-, \ldots, \omega_N^- > 0 \end{cases} \tag{2}$$

and the constraint that m is a non-negative function. (This property is indeed sufficient to ensure m is a probability density.)

1.2 Simulation from signed mixtures

Density functions expressed as (2) present a significant challenge when considering the generic issue of simulating them and we are not aware of existing solutions to this problem. Indeed, a naïve solution consists in first simulating realizations from the associated mixture of positive weight components, which writes as

$$\sum_{k=1}^{P} \omega_k^+ f_k \Big/ \sum_{k=1}^{P} \omega_k^+\,, \tag{3}$$

when renormalized, and then using an accept-reject post-processing step [Bignami and De Matteis, 1971, Devroye, 1985, Robert and Casella, 2004] that subsamples values among these simulations. While formally correct, this approach may prove highly inefficient since the marginal probability of acceptance

$$1 \Big/ \sum_{k=1}^{P} \omega_k^+$$

can be arbitrarily close to zero. Furthermore, as already observed by Devroye [1985], checking for the acceptance condition is potentially costly if $K = P + N$ is large. The intuition behind the computational inefficiency of the standard accept-reject algorithm is that simulating from the positive weight components f_k is not necessarily producing values within regions of high probability for the actual distribution (2). Indeed, it is possible that the negative weight components $\omega_k^- g_k$ remove most of the mass attached to the $\omega_k^+ f_k$'s. Therefore high probability regions for m have no reasons in general to coïncide with high probability regions for all of the f_k's. By the same argument, resorting to an accept-reject method based on the mixture of the *negative* weight components is similarly inefficient. In addition, the so-called *series method* proposed by Devroye [1985, Section IV.5] is not well-suited for this target density since it requires a manageable functional upper bound on (2), namely one that can be simulated. Efficient alternatives are thus necessary and we herewith propose a generic solution.

1.3 Prevalence of signed mixtures

The motivation for considering *signed* mixtures and their simulation is many-fold. Besides approximations proposed for simulation reasons [Devroye, 1985], signed mixtures appear in series representations of density functions [Beaulieu, 1990, Delaigle and Hall, 2010, Hubalek and Kuznetsov, 2011] or as flexible modelling tools [Zhang and Zhang, 2005, Müller et al., 2012, Kroese et al., 2019, Loconte et al., 2024, the later using the denomination of *subtractive mixtures*]. The kernel conditional density estimators constructed by Schuster et al. [2020] open the possibility of signed mixtures, while Polson and Sokolov [2024] connect signed mixtures with the notion of negative probability appearing in quantum theory. Loconte et al. [2024] provides further references about the use of signed mixtures in machine learning, optimization, and signal processing.

Specific examples of probability density functions represented as signed series include the Raab-Green distribution, the Kolmogorov-Smirnov (test) distribution, and the Erdös-Kac distribution [Devroye, 1985, IV.5]. For instance, Elston and Glassy [1989] study the special case of Exponential signed mixtures

$$\sum_{k=1}^{P} \omega_k^+ \mathcal{E}(\lambda_k^+) - \sum_{k=1}^{N} \omega_k^- \mathcal{E}(\lambda_k^-)\,, \quad \text{such that} \quad \lambda_1^+, \ldots, \lambda_N^- > 0\,,$$

by exploiting a connection with *generalised Erlang distributions*, whose density functions are themselves linear combinations of multiple Exponential density functions with some negative coefficients. However, the complexity of their approach is of order $\mathcal{O}(2^{P+N})$, which calls for a more efficient alternative. Similarly, the bivariate Exponential distribution proposed by Gumbel [1960] is a signed mixture of bivariate Gamma distributions, whose efficient simulation is of direct interest in extreme value theory and copula representations.

A primary remark is that, when the sole purpose of the simulation is the approximation of integrals related with (2), the negativity of some weights is not directly a hindrance since

$$\int h(x)m(x)\mathrm{d}x = \int h(x) \sum_{k=1}^{P} \omega_k^+ f_k(x)\mathrm{d}x - \int h(x) \sum_{k=1}^{N} \omega_k^- g_k(x)\mathrm{d}x$$

holds. It thus suffices to produce simulations from both positive weight and negative weight mixtures. However, this solution may prove inefficient when K is large and when the supports of the positive and negative density functions strongly overlap. (The above decomposition also explains why the cdf of (2) can be computed, when considering $h(y) = \mathbb{I}_{(-\infty,x)}(y)$.)

1.4 A core decomposition for signed mixtures

This paper approaches the simulation from a signed mixture (2) by rewriting $m(\cdot)$ using a non-unique decomposition of the positive and negative weights and a rearrangement into three terms [1],

$$m = \sum_{k=1}^{K} \lambda_k \{a_k f_k - g_k\} + \sum_{i=1}^{P} r_i f_i - \sum_{j=1}^{N} s_j g_j, \quad \text{such that} \quad \begin{cases} \lambda_1, \ldots, \lambda_K > 0\,, \\ r_1, \ldots, r_P \geq 0\,, \\ s_1, \ldots, s_N \geq 0\,, \end{cases} \tag{4}$$

and for all $1 \leq k \leq K$, $\inf_{x\in\mathbb{R}}\{a_k f_k(x) - g_k(x)\} \geq 0$.

[1] In order to lighten the notational burden, (4) reuses notations such as f_k and g_k, with no exact correspondence with those found in (2). For instance, the number K of components in (4) may be equal to the product $P \times N$.

Example 1. *Denoting $\varphi(\cdot;\mu,\sigma^2)$ the probability density function of the Normal distribution with mean μ and variance σ^2, consider the Normal signed mixture*

$$m(x) = 2\varphi(x;0,1) + 1.8\varphi(x;0.5,1) - \varphi(x;0.25,0.25) - 0.8\varphi(x;0.75,0.16).$$

A valid decomposition of this signed mixture as (4) is, for instance,

$$\begin{aligned} m(x) = 0.8\{2.25\varphi(x;0,1) - \varphi(x;0.75,0.16)\} + 0.9\{2\varphi(x;0.5,1) - \varphi(x;0.25,0.25)\} \\ + 0.2\varphi(x;0,1) - 0.1\varphi(x;0.25,0.25)\,. \end{aligned}$$

since it can be easily checked that the first two signed mixtures are positive functions.

The construction and optimization of (4) will be conducted in Section 3. The argument behind this representation (4) is that a generic signed mixture (4) can always be written[2] as a mixture of K two-component signed mixtures, the $\{a_k f_k - g_k\}$'s, plus potential positive and negative residual terms. Simulation-wise, the appeal attached to (4) is that those residuals have low probability mass and hence most of the draws from (4) correspond to the first sum in (4), whose simulation is straightforward. Indeed, this simulation proceeds by first selecting at random a component index k with probability proportional to λ_k and second generating from this component density $\{a_k f_k - g_k\}/(a_k - 1)$ by a naïve accept-reject approach when a_k is small enough, or by a more elaborate accept-reject method that is developed below, otherwise.

The plan of the paper is as follows. In Section 2, we construct a specific simulation method for two-component signed mixtures. Section 3 details how the pairing decomposition of (4) is chosen. Section 4 contains numerical experiments that compare different approaches of this simulation challenge. Technical details are postponed till Appendices A, B, C, D. Appendix A specifically focuses on two-component signed mixtures and elaborates in details examples of the signed mixtures of two Normal or two Gamma distributions.

1.5 Notations and conventions

Throughout the paper, we do not distinguish between the measures and their associated density functions. In what follows, the probability density function (pdf) of the signed mixture (with respect to the Lebesgue measure) is denoted by m. The positive and negative weight components are consistently referred to as f_i, $1 \le i \le P$ and g_j, $1 \le j \le N$, respectively, with indices omitted when there is no ambiguity. The positive part m^+ of a signed mixture corresponds to the mixture (3).

For a set $D \subseteq \mathbb{R}^d$, we denote by $\nu(D)$ the probability that a random variable with density ν belongs to D and by $|D|$ the volume of D, that is

$$|D| = \int \mathbb{1}_D(x)\mathrm{d}x.$$

We always assume that the cumulative distribution functions (cdf) of positive and negative weight components can be computed everywhere so that

$$m(D) = \sum_{k=1}^{P} \omega_k^+ f_k(D) - \sum_{k=1}^{N} \omega_k^- g_k(D)$$

is available.

[2] In some special cases from the literature, the signed mixtures already come paired as in (4).

2 Two-component signed mixtures

In this section, we address the specific case of a signed mixture with a sole positive and a sole negative weights. Given two distinct probability density functions f and g such that $\mathrm{supp}(g) \subseteq \mathrm{supp}(f)$, and

$$a^\star = \sup_{x \in \mathrm{supp}(f)} g(x)/f(x) < +\infty\,, \tag{5}$$

a *two-component signed mixture* of f and g is defined as

$$m = \frac{af - g}{a-1}\,, \quad \text{with} \quad a \geq a^\star. \tag{6}$$

Condition (5) ensures that m stands as a proper probability density when g has tails that are dominated by those of the positive component f. Note that $a^\star > 1$ as in generic accept-reject settings (see Appendix A, Lemma 3). The limiting case $a = a^\star$ corresponds to the minimal positive weight required to compensate the negative weight component, i.e., when the density function m reaches zero at some point of its support or asymptotically.

Vanilla sampling scheme As mentioned earlier, a natural, albeit naïve, method for sampling from (6) consists in an accept-reject algorithm with proposed values generated from the distribution f. Since

$$\sup_{x \in \mathrm{supp}(f)} m(x)/f(x) = \frac{a}{a-1}\,,$$

the proposed values x are accepted with probability

$$\frac{af(x) - g(x)}{af(x)}.$$

The average acceptance probability is equal to $(a-1)/a$, which makes the approach inefficient when $a - 1$ is near zero, i.e., when f and g are quite similar.

Stratified sampling scheme We can instead construct an alternative accept-reject scheme based on a piecewise upper bound on (6) towards yielding a higher acceptance probability on average. For this purpose, consider a partition $(D_0, \ldots, D_n)$ of $\mathrm{supp}(f)$ with the convention that D_0 contains the tails of f and potential subsets where the density f is unbounded. We assume that upper and lower bounds on both f and g, over all remaining elements of the partition, D_i, $1 \leq i \leq n$, exist and can be easily computed, so that the terms

$$h_i = a \sup_{x \in D_i} f(x) - \inf_{x \in D_i} g(x), \quad 1 \leq i \leq n$$

are available. These terms yield a rough upper bound on m on each D_i that can obviously be improved in the specific situation when direct access to the supremum of m on D_i is available. Tails are treated separately. Indeed, since the tail dominating component is necessarily attached to the positive part, af can then be used as an upper bound of m on D_0. The partition is therefore providing a direct and different upper bound on (6), that is, for all $x \in \mathrm{supp}(f)$,

$$m(x) \leq \frac{1}{a-1}\left\{ af(x)\mathbb{1}_{D_0}(x) + \sum_{i=1}^{n} h_i \mathbb{1}_{D_i}(x) \right\}.$$

Algorithm 1: Accept-reject method for two-component signed mixtures

Input: Partition $D_0, \ldots, D_n$, upper bounds $h_1, \ldots, h_n$.

sample k from $\mathcal{M}(m(D_0), \ldots, m(D_n))$;
repeat
  if $k = 0$ **then**
    sample x from f truncated to D_0;
    accept x with probability $\{af(x) - g(x)\}/\{af(x)\}$;
  end
  else
    sample x uniformly on D_k;
    accept x with probability $\{af(x) - g(x)\}/h_k$;
  end
until *accepting*;

This dominating function can be normalised into a proposal density, towards a novel accept-reject algorithm since sampling from this proposal is straightforward. Since this proposal is a special instance of a mixture distribution, a possible sampling strategy is as follows: one picks a partition element, that is a (latent) component index, at random according to the vector of probabilities of its components

$$\varrho = (af(D_0), h_1|D_1|, \ldots, h_n|D_n|) \Big/ \left\{ af(D_0) + \sum_{i=1}^{n} h_i|D_i| \right\}$$

and then simulates from f restricted to D_0 or uniformly on D_i, $1 \leq i \leq n$, respectively. Note that D_0 can be further decomposed towards making the simulation of the truncated distribution manageable in practice.

This strategy is however computationally sub-optimal since, in order to obtain one draw from m, it requires to sample on average (see Appendix A.2.1)

$$M = \frac{a}{a-1} f(D_0) + \frac{1}{a-1} \sum_{i=1}^{n} h_i|D_i|$$

(latent) component index variables, while only one is needed. Hence, we switch to a more efficient strategy. We follow a stratified sampling method (as detailed in Algorithm 1) that takes advantage of the partition structure as well as of the availability of the cdf of (6). First, the procedure selects a partition element D_k according to the signed mixture m, i.e., it draws the partition index k according to the Multinomial distribution $\mathcal{M}(m(D_0), \ldots, m(D_n))$. Once the partition element k is generated, the procedure samples a realization from the distribution m restricted to D_k. This is achieved by performing an accept-reject step that uses as proposal a truncated distribution if D_0 was picked, and a uniform distribution otherwise.

The stratified approach reduces the total number of simulated random variables, even though the average acceptance probability of Algorithm 1 remains the same as for the naïve accept-reject algorithm, that is $1/M$. While this modification might seem accessory, as the simulation method within a partition element remains unchanged, it offers the key advantage, of cutting the average computational budget of the proposition step from $3M$ to $1 + 2M$ random variables, in contrast to the initial strategy.

Obviously, the initial partition can easily be refined into smaller sets towards controlling the overall acceptance probability, as stated by the following result (Appendix A.2.2 details its proof).

Lemma 1. *Let* $\delta \in (1 - 1/a, 1)$ *and* $\varepsilon \in [0, (1-\delta)/\delta)$. *If*

$$g(D_0) = \frac{(a-1)\{1-\delta(\varepsilon+1)\}}{\delta}, \tag{7}$$

then there exists a partition $(D_1, \ldots, D_{n_\varepsilon})$ *of* $\mathrm{supp}(f) \setminus D_0$ *such that the average acceptance probability of Algorithm 1 is greater than* δ.

The constraint on parameters δ and ε ensures that (i) $g(D_0) \in (0,1)$ and (ii) Algorithm 1 outperforms the vanilla method in terms of average acceptance probability. The result also provides a heuristic on how to build the partition to achieve this. If we aim at an overall acceptance probability of δ, we first build D_0 so it satisfies (7) for a user-specified tolerance ε. The purpose of this threshold is twofold: it informs on the average acceptance probability for a countably infinite partition, namely $\delta/(1-\delta\varepsilon)$ (see Appendix A.2.2), and, more interestingly, on the largest error possible when approximating $1-m(D_0)$ by the upper Riemann sum

$$\frac{1}{a-1}\sum_{i=1}^{n_\varepsilon} h_i |D_i|.$$

Note that this error can be larger than 1 when the targeted acceptance probability is lower than 0.5. With a positive tolerance level[3], the stratified approach leaves room for improving performances. The mass of D_0 with respect to g decreases with ε. Choosing ε close to $(1-\delta)/\delta$ allows for larger errors but this requires to partition a larger domain. Conversely, choosing ε close to 0 involves partitioning a smaller domain but requires a possibly larger cardinal of the partition. Once D_0 is set, we recursively divide $\mathrm{supp}(f) \setminus D_0$ to find a suitable $(D_1, \ldots, D_{n_\varepsilon})$, that is till the upper Riemann sum approximates $1 - m(D_0)$ with an error less than ε:

$$\frac{1}{a-1}\sum_{i=1}^{n_\varepsilon} h_i|D_i| - 1 + m(D_0) \leq \varepsilon \quad \text{i.e.} \quad \frac{1}{a-1}\sum_{i=1}^{n_\varepsilon} h_i|D_i| - \frac{1}{\delta} + \frac{a f(D_0)}{(a-1)} < 0.$$

Various recursive processes can be used to achieve this stopping rule. In Section 4, we started with a partition based on equally spaced points, and we then recursively refined every partition element D_i for which

$$\frac{1}{a-1} h_i |D_i| - m(D_i) > \frac{\varepsilon}{n_\varepsilon}.$$

For a practical implementation of building such a partition, we refer the reader to the example in Appendix A.3.3.

3 Pairing mechanism

For a generic signed mixture (2), it is rarely the case that the density m naturally appears in the format (4). We thus propose a method to construct a pairing of positive and negative (weight) components and a residual mixture towards a representation of the mixture as (4) that improves the average acceptance probability.

For a given signed mixture (2), denote E the set of all acceptable pairs of positive and negative weight component indices, i.e., such that we can define a two-component signed mixture from the associated density functions, namely

$$E = \left\{ (i,j), \begin{array}{l} 1 \leq i \leq P \\ 1 \leq j \leq N \end{array} \middle| \ \mathrm{supp}(g_j) \subseteq \mathrm{supp}(f_i) \quad \text{and} \quad a_{ij}^\star = \sup_{x \in \mathrm{supp}(f_i)} \frac{g_j(x)}{f_i(x)} < +\infty \right\}.$$

[3]Note that setting $\epsilon = 0$ serves no practical purpose, as it means having a countably infinite partition.

The set E is always non-empty since, otherwise, the signed mixture m could not constitute a proper probability density. Subsequently, E_i^+ and E_j^- will denote the sets of pairs that contain the positive component i and the negative component j, respectively.

A *pairing* refers to a set of two-component signed mixtures that can be constructed from mixture m, and is defined as a subset $F \subset E$, and a collection of weights $(\omega_{ij}^+, \omega_{ij}^-)_{(i,j)\in F}$ that satisfy the following constraints

$$\forall (i,j) \in F, \quad \omega_{ij}^+ - a_{ij}^\star \omega_{ij}^- \geq 0, \tag{8}$$

$$\forall i,\ 1 \leq i \leq P, \quad \sum_{(i,j)\in E_i^+ \cap F} \omega_{ij}^+ \leq \omega_i^+, \tag{9}$$

$$\forall j,\ 1 \leq j \leq N, \quad \sum_{(i,j)\in E_j^- \cap F} \omega_{ij}^- \leq \omega_j^-. \tag{10}$$

The constraint (8) ensures that the weights associated with the pair (i,j) define a two-component signed mixture that is positive everywhere. Constraints (9) and (10) guarantee that when we gather the two-component signed mixtures, the overall weight does not exceed the total weight in m.

A pairing is associated with a residual mixture

$$\sum_{i=1}^P r_i f_i - \sum_{j=1}^N s_j g_j, \quad \text{where} \quad r_i = \omega_i^+ - \sum_{(i,j)\in E_i^+ \cap F} \omega_{ij}^+ \quad \text{and} \quad s_j = \omega_j^- - \sum_{(i,j)\in E_j^- \cap F} \omega_{ij}^-.$$

The decomposition of m associated with the pairing thus writes as

$$\sum_{(i,j)\in F} (\omega_{ij}^+ f_i - \omega_{ij}^- g_j) + \sum_{i=1}^P r_i f_i - \sum_{j=1}^N s_j g_j. \tag{11}$$

Sampling from m can hereby be achieved by proposing a sample from the mixture made of the two-component signed mixtures and of the positive weight components, namely

$$\pi = \sum_{(i,j)\in F} \frac{\omega_{ij}^+ - \omega_{ij}^-}{C} \left(\frac{\omega_{ij}^+ f_i - \omega_{ij}^- g_j}{\omega_{ij}^+ - \omega_{ij}^-} \right) + \sum_{i=1}^P \frac{r_i}{C} f_i, \quad \text{where} \quad C = \sum_{i=1}^P \omega_i^+ - \sum_{(i,j)\in F} \omega_{ij}^-,$$

and by accepting the resulting simulation x with probability $m(x)/C\pi(x)$. Sampling from π proceeds as for any standard (unsigned) mixture distribution, albeit requiring an extra accept-reject step when sampling from the component of π that corresponds to pairs $(i,j) \in F$ (see Algorithm 2).

If sampling each pair $(i,j) \in F$ relies on the vanilla approach, the overall procedure resumes to sampling by proposing from the mixture m^+ (3). Indeed, one sample from m requires C samples from π on average, and to get one sample from π we need to propose

$$\sum_{(i,j)\in F} \frac{\omega_{ij}^+ - \omega_{ij}^-}{C} \frac{\omega_{ij}^+}{\omega_{ij}^+ - \omega_{ij}^-} + \sum_{i=1}^P \frac{r_i}{C} = \frac{1}{C} \sum_{i=1}^P \omega_i^+$$

random variables on average. However, improving the acceptance probability to sample from at least one of the two-component signed mixtures involved in the decomposition (11) is enough to improve the performance of the sampling method, as shown by the following result (whose proof is given in Appendix B.1).

Algorithm 2: Accept-reject method for general signed mixtures

Input: A pairing F.

compute the vector of probabilities prob $\propto \left(\omega_{ij}^+ - \omega_{ij}^-, r_\ell\right)_{(i,j)\in F, 1\leq \ell \leq P}$;

repeat

  sample k according to prob;

  if *k is associated with a pair $(i,j) \in F$* **then**

    sample x from the two-component signed mixture with an accept-reject scheme;

  end

  else

    sample x from f_k;

  end

  accept x with probability $m(x)/\{C\pi(x)\}$.

until *accepting*;

Lemma 2. *Consider $\delta \in (0,1)$ and a pairing F for m. Assume that we sample from each pair $(i,j) \in F$, using*

1. *the vanilla sampling scheme if $(1-\delta)\omega_{ij}^+ - \omega_{ij}^- \geq 0$, that is if the average acceptance probability of the vanilla scheme associated to the pair is larger than δ,*
2. *a piecewise sampling scheme that guarantees an average acceptance probability greater than δ, otherwise.*

Then Algorithm 2 requires on average less than

$$\sum_{i=1}^{P} \omega_i^+ + \frac{1}{\delta} \sum_{(i,j)\in F} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}}$$

proposed random variables to sample once from m.

A direct consequence of this result is to define the optimal pairing scheme (in terms of the number of proposed samples) as the one that minimizes the objective function

$$\sum_{(i,j)\in F} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}}.$$

The solution to this optimization problem is equivalent to minimizing the objective function

$$\sum_{(i,j)\in E} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\} \tag{12}$$

under the linear constraints (8), (9) and (10) (refer to the justification in Appendix B.2). The optimal pairing solution can thus be found by an optimization algorithm targeting the above objective, such as the simplex method [Dantzig, 1963].

We stress that Algorithm 2 does not necessarily achieve an overall average acceptance probability of δ for the optimal pairing. Indeed, the average number of proposed random variables for the pairing writes as

$$\sum_{i=1}^{P} \omega_i^+ + \frac{1}{\delta} \sum_{(i,j)\in F} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\} = \frac{1}{\delta} + \left(1 - \frac{1}{\delta}\right) \sum_{i=1}^{P} r_i + \frac{1}{\delta} \sum_{j=1}^{N} s_j.$$

It is then lower than $1/\delta$ only when

$$\sum_{j=1}^{N} s_j \leq (1-\delta) \sum_{i=1}^{P} r_i.$$

For instance, when an optimal pairing has no positive weight residuals, attaining exactly the targeted probability δ is then achieved solely if we have no negative residuals as well. Even though we control the acceptance probability when sampling from a pair, the reject step towards getting samples of m by simulating from π degrades the overall performances. Conversely, if there are no negative weight residuals, Algorithm 2 achieves a higher acceptance probability than the user-specified rate δ. This setting does not involve a reject step to get from π to m. In that case, we do control sampling performances for each pair and each positive weight residual can be simulated exactly.

4 Comparison experiments

In this section, we examine the performances of three methods that return simulations from arbitrary signed mixture distributions m, namely

1. the vanilla scheme corresponding to the accept-reject method based on the positive part of m,
2. the stratified scheme we proposed for acceptance probabilities $\delta \in \{0.4, 0.6, 0.8\}$ and tolerance levels $\varepsilon \in \{0.1, 0.2, 0.5, 1\}$ compatible with δ,
3. a numerical inversion of the cumulative distribution function associated with m for a precision of 10^{-10} (see Appendix C).

Each method is run to get $n \in \{10, 10^2, 10^3, 10^4\}$ samples from m. For a given sample size n, we report the proportion $\widehat{\delta}_n$ of accepted proposed variables. Its theoretical value is denoted δ for both the vanilla and the stratified schemes. We also detail the relative efficiency of a method A compared to a method B, defined as the ratio of the running time of B by the running time of A. A relative efficiency larger than 1 indicates that A outperforms B in terms of computational budget. We focus on the relative efficiency $\mathcal{R}_n$ of our method compared to the vanilla approach and the relative efficiency $\mathcal{Q}_n$ of accept-reject based methods compared to the numerical inversion of the cdf.

While the above construction is as generic as possible, we run the comparison on special instances of signed mixtures of exponential families distributions, namely f_k and g_k are both either Normal or Gamma distributions. Both families enjoy an explicit condition for (5) to hold and hence define a proper two-component signed mixture of f_k and g_k (see Appendix A.3). We also provide details on how to build the partition $D_0, \ldots, D_{n_\varepsilon}$ for such a two-component signed mixture in Appendix A.3.3. For each family, we consider two kinds of numerical experiments.

4.1 Alternating signed mixtures

The first comparison is provided for a particular signed mixture that writes as the alternating sum

$$m \propto \sum_{k=1}^{K} \left(\frac{a_k^\star}{a_k^\star - 1} \right) \left(\frac{a_k^\star f_k - g_k}{a_k^\star - 1}, \right) \tag{13}$$

where each term involves the two-component signed mixture (6) of f_k and g_k for the minimal positive weight possible. Such a signed mixture exhibits a natural pairing structure where

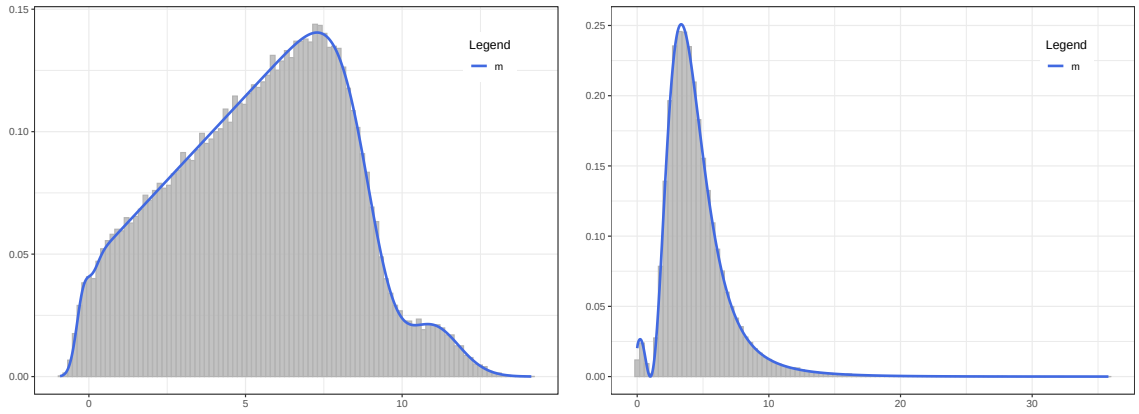

Figure 1: Histogram of 10^5 samples from an alternating signed mixture (13) with Normal distributions (14) (left) and Gamma distributions (15) (right).

the weight of each pair in the overall signed mixture is inversely proportional to the average acceptance probability of the pair. There exists at least one solution to the optimisation problem that comes with no residual mixture.

Signed mixture of Normal distributions A two-component Normal signed mixture can only be defined when the variance of the positive weight component is strictly greater than the variance of the negative one (see Appendix A.3.1). We thus consider the signed mixture (13) with $K = 51$ and for all k, $1 \leq k \leq 51$,

$$\begin{aligned} f_k &\equiv \mathcal{N}(\mu_k, \sigma_k^2) \\ g_k &\equiv \mathcal{N}(\mu_k + 0.01, (\sigma_k - 0.01)^2) \end{aligned} \quad \text{with} \quad \begin{cases} \mu_k = 0.2\,(k-1), \\ \sigma_k = 0.25 + 0.015\,(k-1), \end{cases} \tag{14}$$

and

$$a_k^\star = \frac{\sigma_k}{\sigma_k - 0.1} \exp\left\{\frac{0.01}{4\sigma_k - 0.02}\right\}.$$

Signed mixture of Gamma distributions A two-component Gamma signed mixture can only be defined when the shape and rate of the positive weight component are lower, respectively strictly lower, than the shape and rate of the negative one (see Appendix A.3.2). As an example of the signed mixture (13), we consider a setting with $K = 41$ and for all k, $1 \leq k \leq 41$,

$$\begin{aligned} f_k &\equiv \Gamma(\alpha_k, \beta_k) \\ g_k &\equiv \Gamma(\alpha_k + 0.01, \beta_k + 0.01) \end{aligned} \quad \text{with} \quad \begin{cases} \alpha_k = 1 + 0.1\,(k-1), \\ \beta_k = 0.25 + 0.04375\,(k-1), \end{cases} \tag{15}$$

and

$$a_k^\star = \frac{\Gamma(\alpha_k)}{\Gamma(\alpha_k + 0.1(k-1))} \left(\frac{\beta_k}{\beta_k + 0.01}\right)^{0.01(k-1)} \exp\left\{0.01(1-k)\right\}.$$

Comments Table 1 displays the results for both families. In both examples, the simplex method retrieves the natural pairing associated with the alternating sum form (13) for all $\delta \in \{0.4, 0.6, 0.8\}$. The stratified method overall outperforms both the vanilla method and the numerical inversion of the cdf, regardless of the selected acceptance probability δ and

the tolerance level ε. Unless simulating a dozen variables, our method is between 1.6 and 90 times faster than the vanilla method while the reduction in computation time is smaller when compared to the numerical inverse of the cdf but can still go up to a factor 22. In general, for a given acceptance probability δ, increasing the tolerance level ε results in a lower computational cost of our stratified method, supporting the hypothesis that integration error prevails when building the partition. Conversely, the higher the acceptance probability, the higher the cost of our method. This pattern directly results from the construction of the partition $D_0, \dots, D_{n_\varepsilon}$, where a higher acceptance probability implies a larger domain to partition and a smaller tolerance requires finer partition elements. Lastly, the computational benefit increases with the number of variables simulated, as the cost of both the simplex method and the computation of the partition becomes negligible in front of the cost of sampling random variables.

Table 1: Sampling performances for alternating signed mixtures (13) of Normal distributions (14) and Gamma distributions (15).

	Stratified									
ε	0.1	0.2	0.5	1.0	0.1	0.2	0.5	0.1	0.2	Vanilla
NORMAL SIGNED MIXTURE										
δ	0.4				0.6			0.8		0.018
$\widehat{\delta}_{10}$	0.156	0.256	0.500	0.769	0.588	0.769	0.833	0.278	1.000	0.017
$\mathcal{R}_{10}$	1.876	2.136	**2.153**	2.125	1.604	1.720	**1.738**	1.128	**1.208**	1.000
$\mathcal{Q}_{10}$	2.892	3.293	**3.319**	3.276	2.473	2.652	**2.680**	1.738	**1.862**	1.542
$\widehat{\delta}_{10^2}$	0.407	0.592	0.417	0.633	0.389	0.637	0.645	0.820	0.714	0.018
$\mathcal{R}_{10^2}$	10.70	11.90	10.93	**12.71**	10.49	**11.69**	11.66	8.374	**9.995**	1.000
$\mathcal{Q}_{10^2}$	2.213	2.461	2.262	**2.630**	2.170	**2.417**	2.413	1.732	**2.067**	0.207
$\widehat{\delta}_{10^3}$	0.422	0.442	0.480	0.657	0.615	0.607	0.737	0.810	0.868	0.017
$\mathcal{R}_{10^3}$	26.05	27.25	27.76	**27.94**	11.39	**19.94**	16.08	22.25	**23.70**	1.000
$\mathcal{Q}_{10^3}$	4.521	4.730	4.817	**4.849**	1.976	**3.461**	2.791	3.861	**4.114**	0.174
$\widehat{\delta}_{10^4}$	0.411	0.426	0.499	0.640	0.604	0.641	0.776	0.827	0.855	0.018
$\mathcal{R}_{10^4}$	**38.05**	36.51	37.02	37.04	37.59	38.65	**38.84**	36.96	**37.96**	1.000
$\mathcal{Q}_{10^4}$	**6.233**	5.980	6.063	6.068	6.157	6.330	**6.362**	6.053	**6.218**	0.164
GAMMA SIGNED MIXTURE										
δ	0.4				0.6			0.8		0.008
$\widehat{\delta}_{10}$	0.909	0.769	1.000	1.000	0.833	1.000	1.000	1.000	0.714	0.006
$\mathcal{R}_{10}$	0.674	0.715	0.783	0.790	0.426	0.428	0.428	0.386	0.471	**1.000**
$\mathcal{Q}_{10}$	1.731	1.836	2.010	2.028	1.095	1.100	1.098	0.990	1.209	**2.568**
$\widehat{\delta}_{10^2}$	0.435	0.408	0.595	0.495	0.752	0.498	0.820	0.990	0.962	0.010
$\mathcal{R}_{10^2}$	2.606	2.422	3.240	**3.744**	2.029	**2.109**	1.606	1.740	**2.178**	1.000
$\mathcal{Q}_{10^2}$	1.613	1.499	2.005	**2.317**	1.256	**1.305**	0.994	1.077	**1.348**	0.619
$\widehat{\delta}_{10^3}$	0.418	0.385	0.487	0.648	0.646	0.520	0.640	0.884	0.858	0.009
$\mathcal{R}_{10^3}$	20.11	20.51	25.82	**27.13**	17.28	18.34	**22.77**	15.23	**19.45**	1.000
$\mathcal{Q}_{10^3}$	5.728	5.844	7.355	**7.727**	4.923	5.225	**6.487**	4.339	**5.540**	0.285
$\widehat{\delta}_{10^4}$	0.420	0.426	0.456	0.523	0.581	0.644	0.602	0.828	0.852	0.009
$\mathcal{R}_{10^4}$	54.03	56.36	70.68	**90.18**	61.65	64.92	**77.01**	66.44	**75.97**	1.000
$\mathcal{Q}_{10^4}$	13.44	14.02	17.59	**22.44**	15.34	16.15	**19.16**	16.53	**18.90**	0.249

δ, $\widehat{\delta}_n$: theoretical average acceptance probability of the method and its estimated value for a n-sample. $\mathcal{R}_n, \mathcal{Q}_n$: relative efficiency for a n-sample of the sampling method compared respectively to the vanilla method and the numerical inversion of the cdf.

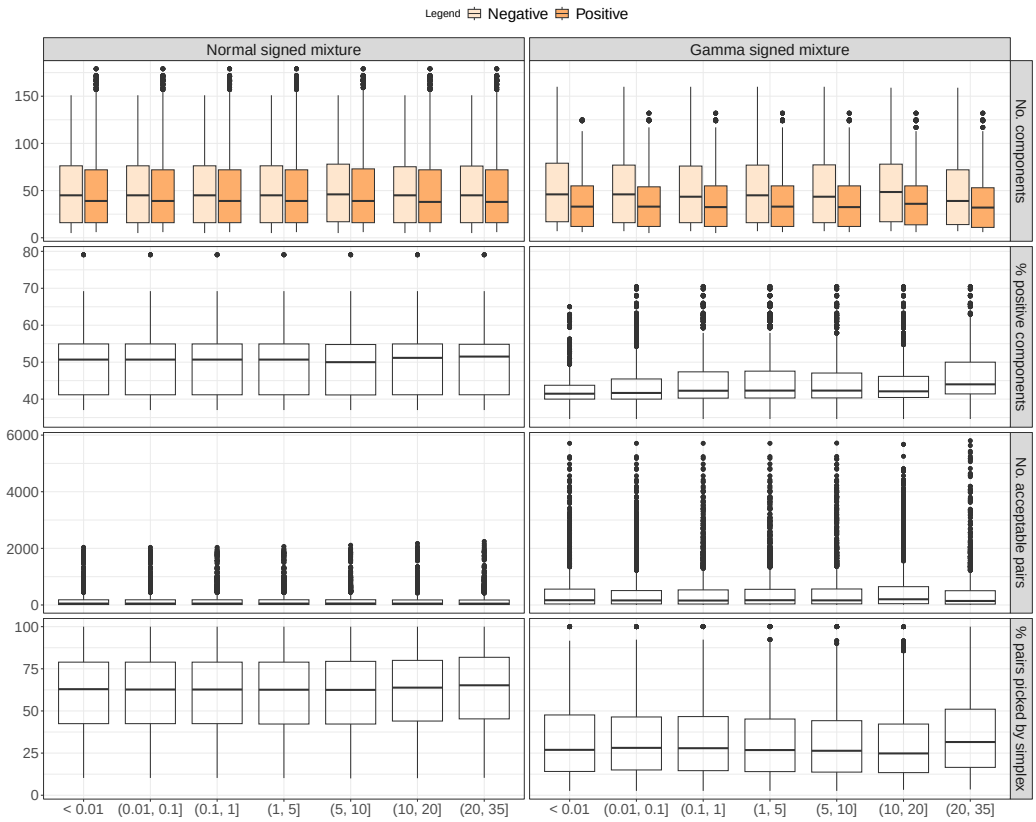

Figure 2: Summaries per vanilla average acceptance probability categories (x-axis in %) of the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (left) and Gamma distributions (right): number of positive and negative weight components (top row), proportion of positive weight components in the model (second row), number of acceptable pairs in the model (third row) and proportion of acceptable pairs selected by the simplex algorithm (bottom row).

4.2 Randomly generated signed mixtures

The second comparison is based on a collection of 2,800 randomly generated signed mixtures (see Appendix D) with a wide range of variety from the number of components to the average acceptance probability of the vanilla method. Table 2 details the distribution of the models into 7 categories depending on the acceptance probability of the vanilla method. The aim was to have models with arbitrary low vanilla acceptance probability in order to challenge our approach in situations where the vanilla method may perform extremely poorly. Models considered also encompass a few components up to a hundred with varying proportions of positive and negative weight components, ensuring then real diversity in the complexity of models (see Figure 2).

Table 2: Repartition of the 2,800 randomly generated signed mixtures of Normal distributions and Gamma distributions according to the average acceptance probability δ of the vanilla accept-reject method.

δ (in %)	$\leq 10^{-2}$	$(10^{-2}, 0.1]$	$(0.1, 1]$	$(1, 5]$	$(5, 10]$	$(10, 20]$	$(20, 35]$
Normal distributions	400	400	400	400	393	404	403
Gamma distributions	287	505	408	400	420	480	300

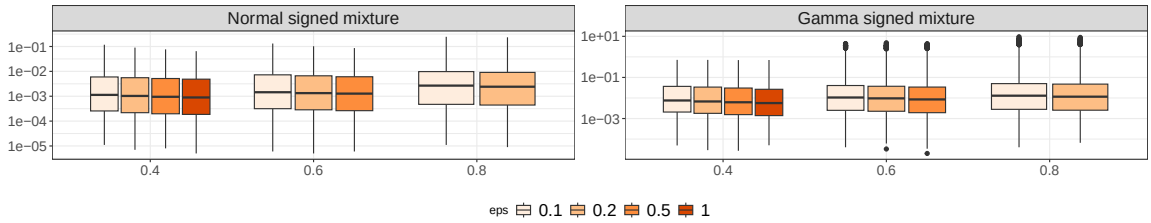

Figure 3: Running time (in sec.) of the stratified method with respect to user-specified acceptance probability δ (x-axis) and tolerance level ϵ (`eps`) for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (left) and Gamma distributions (right).

Comments The running time of our method does not depend significantly on the user-specified acceptance probability nor the tolerance level (see Figure 3). However, we can point out a consistent pattern regarding the influence of both δ and ε. Allowing a larger tolerance level leads to a reduced cost since it implies building a partition with less elements. However, opting for a larger acceptance probability happens to increase the running time. In such settings, we end up with a larger domain to partition and a tolerance level restricted to a smaller range. Hence this results in increasing the number of partition terms, as we aim at a more precise piecewise approximation of the signed mixture. Our method is not designed to efficiently achieve acceptance probability arbitrary close to 1. Instead, users can benefit from reasonably lowering the acceptance probability δ. Obviously, this holds as long as δ remains larger than the vanilla acceptance probability and the simulation cost does not exceed the advantage of the stratification.

The relative efficiency of our stratified solution compared to the vanilla ranges from around 10^{-5} to 10^{5} and unsurprisingly decreases with the vanilla average acceptance probability (see Figure 4, top row). The stratified approach far outperforms the vanilla method on challenging situations, that is when an accept-reject from the positive part would lead to an average acceptance probability lower than 1%, a domination found even for very small samples. For a hundred samples, sampling from the positive part of the signed mixture becomes equivalent to, if not better than, the stratified solution when the vanilla average acceptance probability exceeds 5%. For larger sample sizes, the relative efficiency remains in general larger than 1. Furthermore, we point out that the median running time of our method for a given sample size is quite stable across the different categories of vanilla acceptance probabilities and mostly lower than the second (see Figure 4, second row). In comparison, the median running time of the vanilla method strongly depends on its associated acceptance probability (see Figure 4, third row). This asymmetry means that in situations where the vanilla method performs better, the actual computational benefit is of a negligible scale. Conversely, our method presents a reduction of the simulation cost that is more than substantial in challenging settings, cutting the cost for instance from a few minutes to less than a second.

In the stratified scheme, we have better control of the simulation cost, even in the presence of negative weight residuals (see Figure 7), due to the acceptance probability constraint on each pair. This explains the general median stability we observe on Figure 4 regardless of the overall weight of the positive part in the model. The major elements of influence are the computation of the partition and of the pairing using the simplex method. Regarding the partition, we already observed that it does not alter strongly the computational cost of our solution and hence the relative efficiency compared to the vanilla method, but it can be further confirmed with Figure 8 in Appendix. As for the pairing step, Figure 5 illustrates the influence of the number $|E|$ of acceptable pairs on the computational budget. Namely, the cost of our approach increases as the number of pairs increases, and the method becomes less

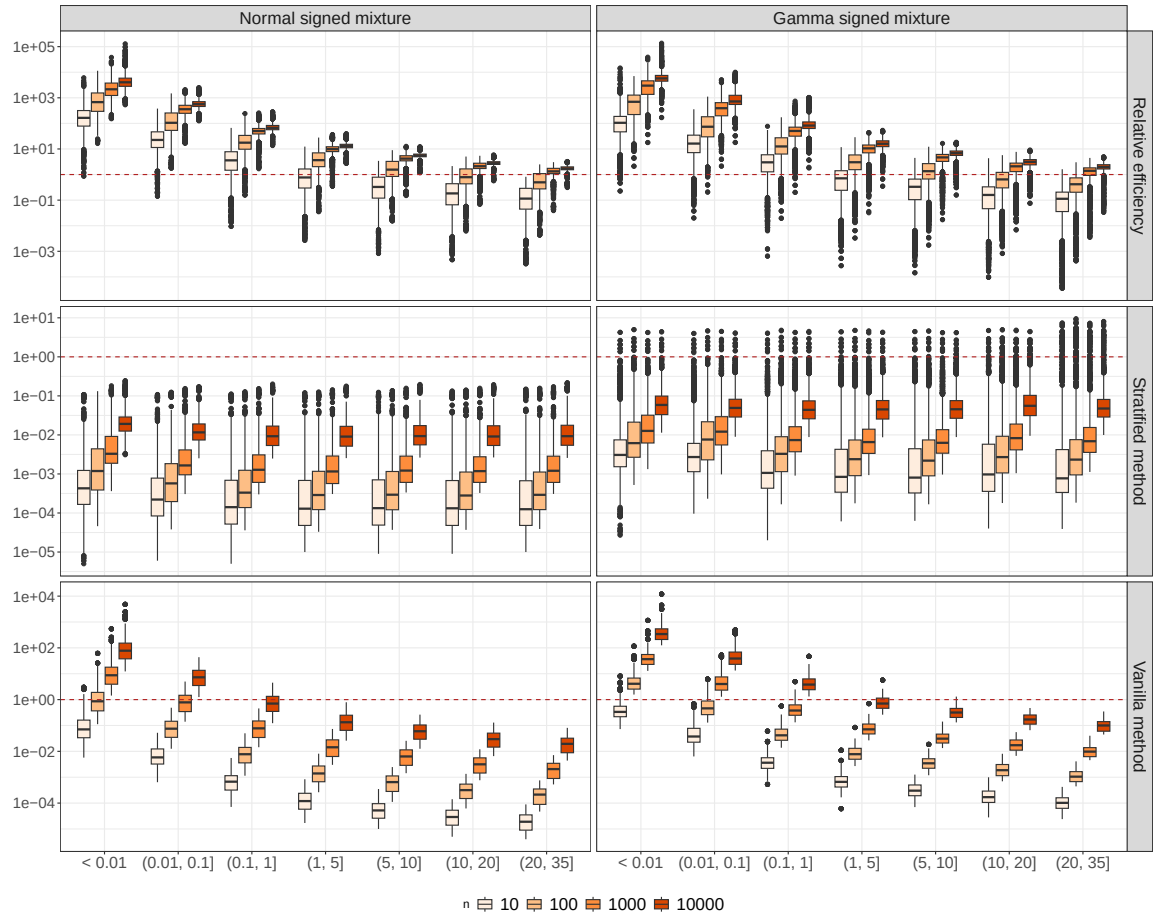

Figure 4: Time performances of accept-reject based methods per vanilla average acceptance probability categories (x-axis in %) and number of draws n for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (left) and Gamma distributions (right): relative efficiency of the stratified method compared to the vanilla method (top row), running time (in sec.) of, respectively, the stratified method (second row) and the vanilla method (bottom row).

competitive than the vanilla approach. Indeed, the simplex algorithm is then used to solve an optimization problem involving $2|E|$ variables and $|E| + N + P$ constraints. For a moderate number of samples, the efficiency of our solution is reduced when the model contains over a thousand acceptable pairs. In this regime, the simplex may prove more time-consuming than simulating even numerous random variables.

Computing a numerical inverse of the cdf does not exhibit a practical advantage over our accept-reject based method from a computational perspective (see Figure 6). Indeed, the median relative efficiency of our method compared to the numerical inverse is close to 1, if not greater. Additionally, the numerical inverse solution only generates samples from an approximate probability measure. As shown in the bottom row of Figure 6, this surrogate quantile function is solely beneficial compared to the vanilla method when the latter exhibits low acceptance probability. Yet, our approach is specifically designed to provide an efficient and exact solution in such a setting.

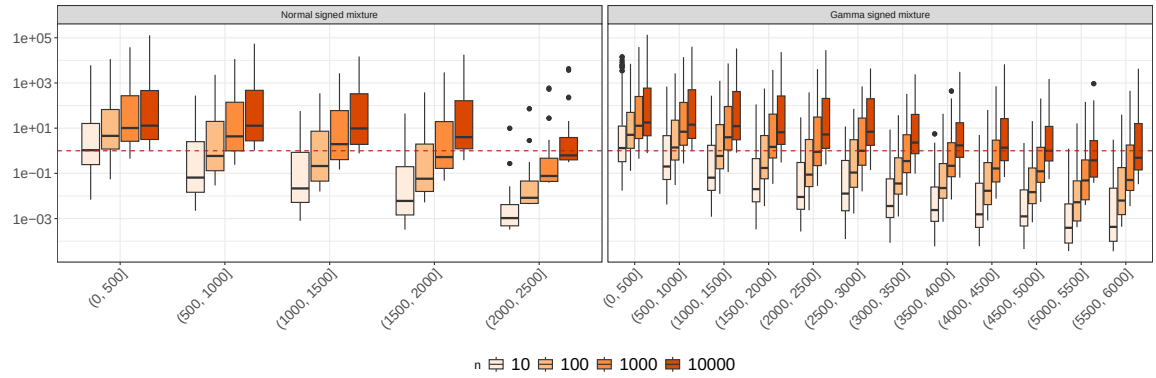

Figure 5: Relative efficiency of the stratified method compared to the vanilla method with respect to the number of acceptable pairs (x-axis) and the number of draws n, for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (left) and Gamma distributions (right).

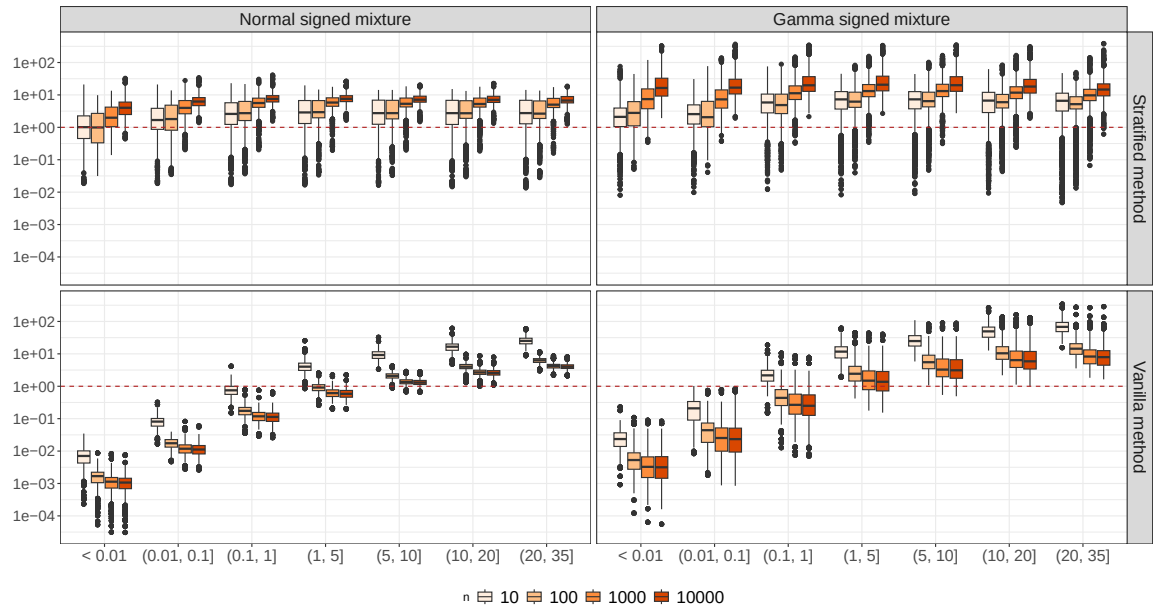

Figure 6: Relative efficiency of, respectively, the stratified method (top row) and the vanilla method (bottom row) compared to the numerical inverse cdf per vanilla average acceptance probability categories (x-axis in %), and number of draws n for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (left) and Gamma distributions (right).

5 Conclusions

The challenge of simulating a signed mixture (2) surprisingly differs from the standard simulation of an unsigned mixture in that the negative components of (2) have no natural association with a latent variable. It thus proves impossible to directly eliminate simulations that issue from these negative terms, i.e., to formalize a negative version of accept-reject and one has to resort to more rudimentary approaches. As discussed above, sampling from a signed mixture using only the positive part of the density may prove cumbersome, especially when the weight of the latter is small. While elementary, our alternative approach achieves noticeably superior computational performances by combining a simplex step towards identifying an efficient decomposition of the model into a well-balanced set of two-component mixtures,

and a piecewise constant approximation of these two-component distributions. Controlling a lower bound on the average acceptance probability ensures steady performance, regardless of the overall weight of the positive part. Furthermore, this alternative performs most satisfactorily relative to the inverse cdf approach, a feat explained in part by the necessity to numerically invert the cdf, even in cases when the quantile function of both positive and negative components is known.

Acknowledgements

A discussion with Murray Pollock (University of Newcastle) was instrumental in sparkling our interest in the matter. The first author has been partly supported by a senior chair (2016-2021) from l'Institut Universitaire de France, by a Prairie chair from the Agence Nationale de la Recherche (ANR-19-P3IA-0001) and by the European Union under the (2023-2030) ERC Synergy grant 101071601 (OCEAN). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Code availability

The code used throughout the paper is available at https://github.com/jstoehr/negmix.

References

N. Beaulieu. An infinite series for the computation of the complementary probability distribution function of a sum of independent random variables and its application to the sum of Rayleigh random variables. *IEEE Transactions on Communications*, 38(9):1463–1474, 1990. doi: 10.1109/26.61387.

A. Bignami and A. De Matteis. A Note on Sampling from Combinations of Distributions. *IMA Journal of Applied Mathematics*, 8(1):80–81, 1971.

O. Cappé, R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert. Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18(4):447–459, 2008.

B. G. Dantzig. *Linear Programming and Extensions*. Princeton University Press, Princeton, 1963.

A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38(2):1171–1193, 2010.

L. Devroye. *Non-Uniform Random Variate Generation*. Springer-Verlag, New-York, 1985.

D. A. Elston and C. A. Glassy. Simulating from a mixture of exponential distributions with some negatively weighted components. *Journal of Statistical Computation and Simulation*, 33(1):1–9, 1989.

E. J. Gumbel. Bivariate exponential distributions. *Journal of the American Statistical Association*, 55:698–707, 1960.

F. Hubalek and A. Kuznetsov. A convergent series representation for the density of the supremum of a stable process. *Electronic Communications in Probability*, 16:84–95, 2011. doi: 10.1214/ECP.v16-1601.

D. Kroese, Z. Botev, T. Taimre, and R. Vaisman. *Data Science and Machine Learning: Mathematical and Statistical Methods*. Chapman & Hall/CRC Machine Learning & Pattern Recognition. CRC Press, New York, 2019. ISBN 9781000730777. URL https://books.google.fr/books?id=F7zADwAAQBAJ.

L. Loconte, A. M. Sladek, S. Mengel, M. Trapp, A. Solin, N. Gillis, and A. Vergari. Subtractive mixture models via squaring: Representation and learning, 2024. URL https://arxiv.org/abs/2310.00724.

G. J. McLachlan and D. Peel. *Finite Mixture Models*. J. Wiley, New York, 2000.

P. Müller, S. Ali-Löytty, M. Dashti, H. Nurminen, and R. Piché. Gaussian mixture filter allowing negative weights and its application to positioning using signal strength measurements. In *2012 9th Workshop on Positioning, Navigation and Communication*, pages 71–76, 03 2012. ISBN 978-1-4673-1437-4. doi: 10.1109/WPNC.2012.6268741.

N. Polson and V. Sokolov. Negative probability, 2024. URL https://arxiv.org/abs/2405.03043.

C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, second edition, 2004.

I. Schuster, M. Mollenhauer, S. Klus, and K. Muandet. Kernel conditional density operators. In S. Chiappa and R. Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 993–1004, Brookline, MA, 2020. PMLR.

D. Titterington, A. Smith, and U. Makov. *Statistical Analysis of Finite Mixture Distributions*. wiley, New York, 1985.

B. Zhang and C. Zhang. Finite mixture models with negative components. In P. Perner and A. Imiya, editors, *Machine Learning and Data Mining in Pattern Recognition*, pages 31–41, Berlin Heidelberg, 2005. Springer. ISBN 978-3-540-31891-0.

Appendix A Two-component signed mixtures

A.1 Lower bound property

Lemma 3. *Assuming two separate probability density functions f and g such that g is absolutely continuous with respect to f, then*

$$a^\star = \sup_{\mathrm{supp}(f)} \frac{g}{f} > 1.$$

Proof. Let assume $a^\star \leq 1$ and denote $E = \{x \in \mathrm{supp}(f) \mid f(x) = g(x)\}$. We have for all $x \in \mathrm{supp}(f) \setminus E$, $g(x) < f(x)$ and

$$\int_E g(x)\mathrm{d}x = \int_E f(x)\mathrm{d}x \quad \text{and} \quad \int_{\mathrm{supp}(f)\setminus E} g(x)\mathrm{d}x < \int_{\mathrm{supp}(f)\setminus E} f(x)\mathrm{d}x.$$

Since $\mathrm{supp}(g) \subseteq \mathrm{supp}(f)$, we thus have

$$1 = \int_{\mathrm{supp}(f)} g(x)\mathrm{d}x < \int_E f(x)\mathrm{d}x + \int_{\mathrm{supp}(f)\setminus E} f(x)\mathrm{d}x = 1.$$

Reductio ad absurdum complete. □

A.2 Results on stratified sampling scheme

A.2.1 Average acceptance probability of Algorithm 1

Behaviour in the tails The distribution m restricted to D_0 satisfies

$$\frac{1}{m(D_0)} m(x) \mathbb{1}_{D_0}(x) \leq \frac{af(D_0)}{(a-1)m(D_0)} \frac{1}{f(D_0)} f(x) \mathbb{1}_{D_0}(x).$$

To get one sample from m restricted to D_0, we need on average

$$M_0 = \frac{af(D_0)}{(a-1)m(D_0)}$$

samples from the distribution f truncated to D_0.

Behaviour in $D_1, \dots D_n$ The distribution m restricted to D_i satisfies

$$\frac{1}{m(D_i)} m(x) \mathbb{1}_{D_i}(x) \leq \frac{h_i |D_i|}{(a-1)m(D_i)} \frac{1}{|D_i|} \mathbb{1}_{D_i}(x).$$

To get one sample from m restricted to D_i, we need on average

$$M_i = \frac{h_i |D_i|}{(a-1)m(D_i)}$$

samples from the uniform distribution on D_i.

Global behaviour To get one sample from m, we need to propose on average

$$\sum_{i=0}^{n} m(D_i) M_i = \frac{a}{a-1} f(D_0) + \frac{1}{a-1} \sum_{i=1}^{n} h_i |D_i| = M$$

random variables.

Remark 1. *Sampling from the distribution f restricted to D_0 is not necessarily straightforward and might require an accept-reject scheme as well. Both methods based on piecewise proposals have nevertheless still the same acceptance probability on average. If we need N_0 samples from a proposal to get one sample from f restricted to D_0, Algorithm 1 then requires simulating*

$$\tilde{M} = \frac{a}{a-1} f(D_0) N_0 + \frac{1}{a-1} \sum_{i=1}^{n} h_i |D_i|$$

random variables. Conversely, sampling from the dominating piecewise function would require

$$M \left\{ \frac{af(D_0)}{M(a-1)} N_0 + \frac{1}{M(a-1)} \sum_{i=1}^{n} h_i |D_i| \right\} = \tilde{M}$$

random variables.

A.2.2 Proof of Lemma 1

Proof. We have

$$\frac{1}{a-1}\sum_{i=1}^{n} h_i|D_i| \underset{n\to+\infty}{\longrightarrow} \int_{\mathrm{supp}(f)\setminus D_0} m(x)\mathrm{d}x = 1 - \frac{af(D_0) - g(D_0)}{a-1}.$$

Hence, for all $\varepsilon > 0$, there exists n_ε, such that for all $n \geq n_\varepsilon$

$$\left| M - \frac{a - 1 + g(D_0)}{a-1} \right| \leq \varepsilon.$$

Given $\varepsilon \in [0, (1-\delta)/\delta)$, if

$$g(D_0) = \frac{(a-1)\{1 - \delta(\varepsilon+1)\}}{\delta},$$

then

$$\left| M - \frac{1}{\delta} + \varepsilon \right| \leq \varepsilon.$$

This leads to $1/M \geq \delta$. □

Remark 2. *Under the assumption of Lemma 1, we have*

$$M \underset{n\to+\infty}{\longrightarrow} \frac{1}{\delta} - \varepsilon.$$

Remark 3. *A direct consequence of Lemma 1 is that, if we pick the partition of* $\mathrm{supp}(f)\setminus D_0$ *such that*

$$1 - m(D_0) + \varepsilon = \frac{1}{a-1}\sum_{i=1}^{n_\varepsilon} h_i|D_i|,$$

then using the assumption on $g(D_0)$ *we get*

$$M = \frac{a}{a-1} f(D_0) + 1 - m(D_0) + \varepsilon = 1 + \frac{1}{a-1} g(D_0) + \varepsilon = \frac{1}{\delta}.$$

A.3 Exponential families examples

Assume that, within the context of Section 2, the terms f and g are both distributions from the same exponential family

$$\mathcal{F} = \left\{ c(\theta)h(x)\exp\{\eta(\theta)^\top T(x)\}\,;\, x \in \mathbb{R}^d,\, \theta \in \Theta \subseteq \mathbb{R}^q \right\}.$$

A pairing of f and g, parametrized respectively by θ^+ and θ^-, into a two-component signed mixture is thus possible if

$$a^\star = \sup_{x\in\mathrm{supp}(f)} \{\eta(\theta^-) - \eta(\theta^+)\}^\top T(x) < +\infty. \tag{16}$$

A.3.1 Example of Normal distributions

Let $f \equiv \mathcal{N}(\mu^+, \sigma_+^2)$ and $g \equiv \mathcal{N}(\mu^-, \sigma_-^2)$. Since

$$-\frac{(x-\mu^-)^2}{2\sigma_-^2} + \frac{(x-\mu^+)^2}{2\sigma_+^2} \underset{\pm\infty}{\sim} -x^2\left(\frac{\sigma_+^2 - \sigma_-^2}{2\sigma_-^2\sigma_+^2}\right)$$

condition (16) is fulfilled if $\sigma_-^2 < \sigma_+^2$ (or if $\mu^+ = \mu^-$ and $\sigma_-^2 = \sigma_+^2$ which is of no interest). Assuming $\sigma_-^2 < \sigma_+^2$, critical points are then solution of

$$\frac{\mu^-}{\sigma_-^2} - \frac{\mu^+}{\sigma_+^2} + 2x\left(-\frac{1}{2\sigma_-^2} + \frac{1}{2\sigma_+^2}\right) = 0.$$

We derive a global maximum at

$$x^\star = \frac{\mu^-\sigma_+^2 - \mu^+\sigma_-^2}{\sigma_+^2 - \sigma_-^2}.$$

Then

$$a^\star = \frac{\sigma_+}{\sigma_-}\exp\left\{\frac{(\mu^+ - \mu^-)^2}{2(\sigma_+^2 - \sigma_-^2)}\right\}.$$

Monotonicity of a two-component Normal signed mixture Assume $f \equiv \mathcal{N}(0,1)$ and $g \equiv \mathcal{N}(\mu, \sigma^2)$, with $\mu \geq 0$ and $\sigma < 1$. The signed mixture m has at most 3 extreme values. More specifically, it admits

(i) a unique global maximum in $(-\infty, 0]$, if

$$a \geq \sup_{x>\mu} \frac{(x-\mu)g(x)}{\sigma^2 x f(x)}\,;$$

(ii) a local maximum in $(-\infty, 0]$, a local minimum and a local maximum in $[\mu, +\infty)$, otherwise.

We have for all $x \in \mathbb{R}$

$$m'(x) = \frac{f(x)}{a-1}\left\{\psi(x) - ax\right\}, \quad \text{where} \quad \psi : x \mapsto \frac{(x-\mu)g(x)}{\sigma^2 f(x)}.$$

The number of solutions to $m'(x) = 0$ then depends on the number of intersection points between ψ and $x \mapsto ax$. The assumption on two-component signed mixtures imposes $g(x)/f(x) \underset{x\to\pm\infty}{\longrightarrow} 0$. Since it happens at exponential speed, we also have $\psi(x) \underset{x\to\pm\infty}{\longrightarrow} 0$. On the other hand, for all $x \in \mathbb{R}$,

$$\psi'(x) = \left\{(\sigma^2-1)x^2 + x(2\mu - \mu\sigma^2) + \sigma^2 - \mu^2\right\}\frac{g(x)}{\sigma^4 f(x)}.$$

A straightforward computation shows that the equation $\psi'(x) = 0$ has two distinct solutions and thus ψ has a global minimum and a global maximum, respectively at

$$x_1 = \frac{\mu}{2} + \frac{\mu - \sigma\sqrt{\mu^2\sigma^2 + 4 - 4\sigma^2}}{2(1-\sigma^2)} \quad \text{and} \quad x_2 = \frac{\mu}{2} + \frac{\mu + \sigma\sqrt{\mu^2\sigma^2 + 4 - 4\sigma^2}}{2(1-\sigma^2)}.$$

Moreover, since

$$\psi''(x) = \frac{g(x)}{\sigma^6 f(x)}\left\{(\sigma^2-1)^2 x^3 + Q_2(x)\right\}$$

where $Q_2(x)$ is a univariate polynomial of degree 2, ψ changes convexity solely one time in $[x_1, x_2]$. Note that $\psi''(\mu) = 2\mu g(\mu)/\{\sigma^2 f(\mu)\} \geq 0$ and thus the change of convexity happens between μ and x_2. Functions ψ and $x \mapsto ax$ have then at most 3 intersection points.

If $\boldsymbol{\mu}$= 0, we have a first obvious solution: $x = 0$. Since ψ is an odd function when $\mu = 0$, the latter solution is unique if

$$a \geq \psi'(0) = \frac{1}{\sigma^3} = \sup_{x>0} \frac{g(x)}{\sigma^2 f(x)}.$$

It is the unique global maximum for m, which thus has the same monotonicity as f. Otherwise, it is a local minimum and we have two local maxima corresponding to the intersection points solution of

$$\exp\left\{\frac{x^2}{2\sigma^2}(\sigma^2 - 1)\right\} = a\sigma^3,$$

that is $\pm 2\sigma^2 \log(a\sigma^3)/(1 - \sigma^2)$.

If $\boldsymbol{\mu}$> 0, we do not have a closed form for the critical points. However ψ is a non-positive function on $(-\infty, \mu]$, that is decreasing on $(-\infty, x_1)$ and an increasing convex function on $(x_1, \mu]$. Consequently, there exists a unique intersection point $y_1^\star$ on $(-\infty, 0]$ that corresponds to a local maximum of m. The function $x \mapsto ax$ being positive on $(0, \mu]$, if there are two other intersection points, they are necessarily in $(\mu, +\infty)$. If

$$a \geq \sup_{x>\mu} \frac{(x - \mu)g(x)}{\sigma^2 x f(x)},$$

then for all $x > \mu$, $m'(x) < 0$ and as a result $y_1^\star$ is the unique global maximum of m. Otherwise, we have two intersection points. The point $y_2^\star$ corresponding to a local minimum of m is bound to be on (μ, x_2). Nevertheless, note that on (μ, x_2)

$$\psi(x) - ax \leq (x - \mu)\frac{a^\star}{\sigma^2} - ax = \frac{(a^\star - a\sigma^2)x - a^\star\mu}{\sigma^2}.$$

If $a^\star - a\sigma^2 > 0$, m is decreasing between μ and $a^\star\mu/(a^\star - a\sigma^2)$ and $y_2^\star \geq a^\star\mu/(a^\star - a\sigma^2)$.

Remark 4. *The results for $\mu < 0$ are obtained by symmetry of the problem. Finally the result for the general case of a signed mixture m of $\mathcal{N}(\mu^+, \sigma_+^2)$ and $\mathcal{N}(\mu^-, \sigma_-^2)$ can be derived using that for all $x \in \mathbb{R}$*

$$m(x) \propto af\left(\frac{x - \mu^+}{\sigma^+}\right) - g\left(\frac{x - \mu^+}{\sigma^+}\right), \quad \textit{with} \quad \mu = \frac{\mu^- - \mu^+}{\sigma_+} \quad \textit{and} \quad \sigma = \frac{\sigma_-}{\sigma_+}.$$

A.3.2 Example of Gamma distributions

Let $f \equiv \Gamma(\alpha^+, \beta^+)$, $g \equiv \Gamma(\alpha^-, \beta^-)$ (shape, rate parametrization). Condition (16) imposes $\alpha^+ \leq \alpha^-$ and $\beta^+ < \beta^-$, so that

$$(\alpha^- - \alpha^+)\log x + (\beta^+ - \beta^-)x \underset{x\to 0^+}{\longrightarrow} -\infty \quad \text{or} \quad 0,$$
$$(\alpha^- - \alpha^+)\log x + (\beta^+ - \beta^-)x \underset{x\to +\infty}{\longrightarrow} -\infty.$$

Assuming this, critical points are solutions of

$$\frac{\alpha^- - \alpha^+}{x} + \beta^+ - \beta^- = 0,$$

which leads to a unique global maximum at

$$x^\star = \frac{\alpha^+ - \alpha^-}{\beta^+ - \beta^-}.$$

Then

$$a^\star = \begin{cases} \dfrac{\Gamma(\alpha^+)(\beta^-)^{\alpha^-}}{\Gamma(\alpha^-)(\beta^+)^{\alpha^+}} \exp\left\{(\alpha^+ - \alpha^-)(1 - \log x^\star)\right\} & \text{if } \alpha^+ < \alpha^-, \\ \left(\dfrac{\beta^-}{\beta^+}\right)^{\alpha^+} & \text{if } \alpha^+ = \alpha^-. \end{cases}$$

Monotonicity of a two-component Gamma signed mixture The arguments for studying the monotonicity are similar to those used for the Gaussian case. For all $x > 0$,

$$m'(x) = \frac{f(x)}{(a-1)x}\left\{\psi(x) - a(\beta^+ x + 1 - \alpha^+)\right\}, \quad \text{where} \quad \psi : x \mapsto \frac{(\beta^- x - \alpha^- + 1)g(x)}{f(x)}.$$

If $\boldsymbol{\alpha^+ = \alpha^-}$, first and second derivatives of ψ write as

$$\psi'(x) = \left\{\beta^-(\beta^+ - \beta^-)x + \beta^+ + \alpha^-(\beta^- - \beta^+)\right\}\frac{g(x)}{f(x)}$$
$$\psi''(x) = \left[\beta^-(\beta^- - \beta^+)x - \{\beta^+ + \beta^- + \alpha^-(\beta^- - \beta^+)\}\right]\frac{(\beta^- - \beta^+)g(x)}{f(x)}.$$

We thus have a unique global maximum and a single change of convexity.

- If $\alpha^- = \alpha^+ > 1$, then $a(1 - \alpha^+) < \psi(0)$ and m admits a unique global maximum on $(0, +\infty)$.
- If $\alpha^- = \alpha^+ \leq 1$, $a(1 - \alpha^+) \geq \psi(0)$ and m admits a local minimum and local maximum on $(0, +\infty)$ solely when

  $$a < \sup_{x \geq 0} \frac{\beta^-(\beta^- x + 1 - \alpha^-)}{\beta^+(\beta^+ x + 1 - \alpha^-)} \exp\left\{(\beta^+ - \beta^-)x\right\}.$$

  Otherwise, m is decreasing on $(0, +\infty)$.

If $\boldsymbol{\alpha^+ < \alpha^-}$,

$$\psi'(x) = \Big[\beta^-(\beta^+ - \beta^-)x^2 + \{\beta^+(1 - \alpha^-) + \beta^-(2\alpha^- - \alpha^+)\}x + (\alpha^- - \alpha^+)(1 - \alpha^-)\Big]\frac{g(x)}{xf(x)}$$

The univariate polynomial has necessarily two real roots x_1 and x_2 (otherwise ψ would be a continuous decreasing function on $[0, +\infty)$ and hence constant since its limit at 0 and $+\infty$ is 0). It is straightforward to show that the smallest root is non-positive when $\alpha^- \leq 1$ and non-negative when $\alpha^- > 1$ while the largest is always positive. The convex properties are identical to the Gaussian example as

$$\psi''(x) = \frac{g(x)}{x^2 f(x)}\left\{\beta^-(\beta^- - \beta^+)^2 x^3 + Q_2(x)\right\},$$

where $Q_2(x)$ is a univariate polynomial of degree 2.

- If $\alpha^+ < 1$, then $a(1-\alpha^+) > 0$. m admits a local minimum and local maximum on $(\max\{0, (\alpha^- - 1)/\beta^-\}, +\infty)$ solely when

$$a < \sup_{x \geq 0} \frac{\{\beta^- x + 1 - \alpha^-\} g(x)}{\{\beta^+ x + 1 - \alpha^+\} f(x)}.$$

  Otherwise, m is decreasing on $(0, +\infty)$.

- If $\alpha^+ \geq 1$, then $a(1-\alpha^+) \leq 0$. The behaviour depends on the relative position of the modes of each component.

  - If $\beta^-(\alpha^+ - 1) < \beta^+(\alpha^- - 1)$, then m admits a local maximum in $[0, (\alpha^+ - 1)/\beta^+]$. It is then decreasing on $[(\alpha^+ - 1)/\beta^+, +\infty)$ when

$$a \geq \sup_{\beta^- x + 1 - \alpha^- > 0} \frac{\{\beta^- x + 1 - \alpha^-\} g(x)}{\{\beta^+ x + 1 - \alpha^+\} f(x)}.$$

    Otherwise, m admits a local minimum and a local maximum within the latter interval.

  - If $\beta^-(\alpha^+ - 1) > \beta^+(\alpha^- - 1)$, then m admits a local maximum in $[(\alpha^+ - 1)/\beta^+, +\infty)$. On $[0, (\alpha^+ - 1)/\beta^+]$, m is increasing when

$$a \geq \sup_{\beta^- x + 1 - \alpha^- < 0} \frac{\{\beta^- x + 1 - \alpha^-\} g(x)}{\{\beta^+ x + 1 - \alpha^+\} f(x)}.$$

    Otherwise, m admits a local maximum and a local minimum within the latter interval.

  - If $\beta^-(\alpha^+ - 1) = \beta^+(\alpha^- - 1)$, both components have the same mode that is the unique global maximum of m when

$$a \geq \sup_{\beta^+ x + 1 - \alpha^+ \neq 0} \frac{\{\beta^- x + 1 - \alpha^-\} g(x)}{\{\beta^+ x + 1 - \alpha^+\} f(x)} = \psi'\left(\frac{\alpha^+ - 1}{\beta^+}\right),$$

    and a local minimum otherwise. In the latter situation, m admits two local maxima, one in $(0, (\alpha^+ - 1)/\beta^+)$ and one in $((\alpha^+ - 1)/\beta^+, +\infty)$.

A.3.3 Construction of the partition

We compute $D_0 = (q_\alpha, q_{1-\alpha})^c$, where q_α and $q_{1-\alpha}$ are respectively α and $1-\alpha$-quantiles of g, with

$$\alpha = \frac{(a-1)\{1 - \delta(\varepsilon + 1)\}}{2\delta}.$$

In the specific setting of a two-component Gamma signed mixture with both shape parameters larger than 1, we consider $D_0 = [q_{1-2\alpha}, +\infty)$.

We partition D_0^c into S subsets $D_1, \ldots, D_S$ relying on the monotonic properties of the signed mixture. The aim is to decide whether, on subdivisions $[x_i, x_{i+1}[$ of D_i, we use

$$\text{(A)} \quad h_i = \sup_{[x_i, x_{i+1}[} (af - g)(x) \quad \text{or} \quad \text{(B)} \quad h_i = \sup_{[x_i, x_{i+1}[} af(x) - \inf_{[x_i, x_{i+1}[} g(x).$$

On each subset, the signed mixture has one of the following properties:

1. the signed mixture is a monotonic function. On such a subset, we use the version (A) on every subdivision $[x_i, x_{i+1}[$;

2. the signed mixture changes monotonicity only once on the subset. For all subdivisions $[x_i, x_{i+1}[$ such that $m'(x_i)m'(x_{i+1}) > 0$, we use the version (A). Otherwise, we use the version (B) but that happens solely once;

3. the signed mixture changes monotonicity more than once on the interval. On such subset, we use the version (B) on every subdivision $[x_i, x_{i+1}[$.

Note that for two-component Gamma and Normal signed mixtures, we can restrict ourselves to use only the first two types of subsets by numerically computing some of the local extrema.

For a given subset D_s, $1 \leq s \leq S$, we start with the partition $[x_1, x_2[, \ldots, [x_{n-1}, x_n[$, such that $x_{i+1} - x_i = |D_0|/100$, $1 \leq i \leq n$. The length of each partition element of D_s is divided by two until we achieve

$$\sum_{x_i \in D_s} (x_{i+1} - x_i) h_i = m(D_i) + \frac{\varepsilon}{S+1}.$$

Appendix B Pairing mechanism

B.1 Proof of Lemma 2

Proof. As mentioned in the paper, to get one sample from m, we need to propose in average C samples from π. Let now detail the number N of proposed sample from π according to the sampling strategy of Lemma 2. The probability to randomly a pick pair $(i,j) \in F$ is $(\omega_{ij}^+ - \omega_{ij}^-)/C$, while the one to pick a residual $i \in \{1, \ldots, P\}$ is r_i/C. To get one sample from a pair $(i,j) \in F$, the vanilla scheme requires proposing $\omega_{ij}^+/(\omega_{ij}^+ - \omega_{ij}^-)$ random variables, while, by assumption, the piecewise sampling scheme requires less than $1/\delta$ random variables. When sampling from a residual, we have an exact and immediate sampler that requires solely to propose one draw. Overall, we then have

$$N \leq \sum_{(i,j)\in F} \frac{\omega_{ij}^+}{C} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- \geq 0\}} + \sum_{(i,j)\in F} \frac{\omega_{ij}^+ - \omega_{ij}^-}{\delta C} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}} + \sum_{i=1}^{P} \frac{r_i}{C}.$$

Since

$$\sum_{(i,j)\in F} \omega_{ij}^+ + \sum_{i=1}^{P} r_i = \sum_{i=1}^{P} \omega_i^+,$$

we end up with

$$\begin{aligned} C \times N &\leq \sum_{i=1}^{P} \omega_i^+ - \sum_{(i,j)\in F} \omega_{ij}^+ \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}} + \sum_{(i,j)\in F} \frac{\omega_{ij}^+ - \omega_{ij}^-}{\delta} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}} \\ &\leq \sum_{i=1}^{P} \omega_i^+ + \frac{1}{\delta} \sum_{(i,j)\in F} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\} \mathbb{1}_{\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^- < 0\}}. \end{aligned}$$

□

B.2 Objective function for the simplex method

Minimizing the objective function (12) provides the optimal pairing for Lemma 2. Let $\{\tilde{\omega}_{ij}^+, \tilde{\omega}_{ij}^-\}_{(i,j)\in E}$ be a minimizer of (12), which, as a reminder, is given by

$$\sum_{(i,j)\in E} \left\{(1-\delta)\omega_{ij}^+ - \omega_{ij}^-\right\},$$

and assume there exists a pair $(k,\ell) \in E$ such that $(1-\delta)\tilde{\omega}^+_{k\ell} - \tilde{\omega}^-_{k\ell} > 0$. Then,

$$\sum_{(i,j)\in E} \left\{(1-\delta)\tilde{\omega}^+_{ij} - \tilde{\omega}^-_{ij}\right\} > \sum_{(i,j)\in E\setminus\{(k,\ell)\}} \left\{(1-\delta)\tilde{\omega}^+_{ij} - \tilde{\omega}^-_{ij}\right\} + \left\{(1-\delta)\omega^+_{k\ell} - \omega^-_{k\ell}\right\} \mathbb{1}_{\{(\omega^+_{k\ell},\omega^-_{k\ell})=(0,0)\}}.$$

This contradicts the fact that $\{\tilde{\omega}^+_{ij}, \tilde{\omega}^-_{ij}\}_{(i,j)\in E}$ is a minimizer of (12). Therefore a minimizer $\{\tilde{\omega}^+_{ij}, \tilde{\omega}^-_{ij}\}_{(i,j)\in E}$ of (12) satisfies for all $(i,j) \in E$, $(1-\delta)\tilde{\omega}^+_{ij} - \tilde{\omega}^-_{ij} \leq 0$. Consequently, (12) has the same set of minimizers than

$$\sum_{(i,j)\in E} \left\{(1-\delta)\omega^+_{ij} - \omega^-_{ij}\right\} \mathbb{1}_{\left\{(1-\delta)\omega^+_{ij}-\omega^-_{ij}<0\right\}}.$$

Appendix C Numerical inversion of the cdf

Consider n ordered points $q_1, \ldots, q_n$ in the support of f and $p_1, \ldots, p_n$ the value of the cdf associated with m at these points, that is $p_i = m((-\infty, q_i])$, $1 \leq i \leq n$. Furthermore, set a user-specified precision ε. In the paper, we used $\varepsilon = 10^{-10}$.

The set of points $q_1, \ldots, q_n$ and $p_1, \ldots, p_n$ provides a piecewise affine approximation of the inverse cdf. The aim of our numerical inversion is to refine the affine approximation so that we can find the quantile associated with a probability arbitrary close to a point $u \in [0,1]$ using one of the following steps.

Step A Assume we have $u \in [p_i, p_{i+1}]$. We compute the preimage $q^\star$ of u by the affine transformation on $[p_i, p_{i+1}]$

$$x \mapsto \frac{p_{i+1}-p_i}{q_{i+1}-q_i}x + \frac{p_i q_{i+1} - p_{i+1}q_i}{q_{i+1}-q_i},$$

that is

$$q^\star = \frac{(q_{i+1}-q_i)u - p_i q_{i+1} + p_{i+1}q_i}{p_{i+1}-p_i}.$$

Then we compute the cdf at $q^\star$ and denote $p^\star$ its value. This yields a new interval containing u that is strictly included in $[p_i, p_{i+1}]$. We now apply the same procedure on that interval. We repeat the process until we get a value $p^\star$ such that $|u - p^\star| < \varepsilon$.

Step B If we deal with a distribution that has an unbounded support, tails should be treated separately. Assume we have $u < p_1$. We use a scheme similar to the above except we take the preimage by the affine transformation based on the two first points (p_1, p_2) larger than u. Here we stop when we find a point $p^\star \leq u + \varepsilon$. If $u > p_n$, the reasoning is the same except we use the last two points smaller than u and we stop when $p^\star \geq u - \varepsilon$. Now, either this ending point satisfies $|u - p^\star| < \varepsilon$, or we apply Step A starting with the interval $[p^\star, p_1]$ for left tail or $[p_n, p^\star]$ for right tail.

Appendix D Random generator of signed mixture models

We used two different methods to generate the benchmark models. Both methods start by randomly setting an initial number K of positive weight components in the model. The number K is drawn uniformly between $k_{\min}$ and $k_{\max}$ for the following sets $\{5, \ldots, 10\}$, $\{10, \ldots, 30\}$, $\{30, \ldots, 50\}$, and $\{50, \ldots, 100\}$. Once the number of positive weight components is set, we randomly draw the associated parameter values.

- For Normal signed mixtures, the mean μ^+ is drawn uniformly in $[0, 20]$ and the standard deviation is drawn according to a Gamma distribution $\Gamma(3, 2.5)$ (shape, rate parametrization).
- For Gamma signed mixtures, the shape parameter α^+ is drawn according to a Gamma distribution $\Gamma(4, 0.5)$ and the rate parameter β^+ to a Gamma distribution $\Gamma)(2, 0.7)$.

We then randomly set the number of negative weight components (1 or 2) that are initially related to each positive weight component. The parameter value for the negative components as well as the weights are then computed to ensure a benchmark model for which the vanilla average acceptance probability p ranges in $[p_{\min}, p_{\max}]$ for the following sets $[0, 10^{-4}]$, $(10^{-4}, 0.001]$, $(0.001, 0.01]$, $(0.01, 0.05]$, $(0.05, 0.1]$, $(0.1, 0.2]$ and $(0.2, 0.3]$. For each set of values for K and p, and for each method, we generated 50 benchmarks.

D.1 First method

The first method is based on the properties of two-component signed mixtures. For a given positive weight component, we compute the parameter value of the associated negative weight component such that $a^\star \in [1/(1-p_{\max}), 1/(1-p_{\min})]$. The weights for this two-signed component are the ones associated with $a^\star$. If the positive weight component is associated with more than one negative component, we repeat this procedure for each negative component. As a result, we thus obtain a collection of two-component signed mixtures that all have the targeted acceptance probability. A convex combination with uniform rates of these mixtures yields a signed mixture with the vanilla average acceptance probability we aim at.

If the overall acceptance probability is lower than $p_{\max}$, we randomly decide to add positive weight components that can either balance some of the negative components already included or that can balance none. We can easily determine the maximal weight to assign to such single components so the acceptance probability remains lower than $p_{\max}$. Indeed, assume we add $\tilde{K}$ positive weight components. The new normalized signed mixture writes as

$$\frac{\sum_{i=1}^K \omega_i^+ f_i - \sum_{j=1}^N \omega_j^- g_j + \sum_{k=1}^{\tilde{K}} r_k f_{K+k}}{1 + \sum_{k=1}^{\tilde{K}} r_k}.$$

The latter is associated with a vanilla acceptance probability lower than $p_{\max}$ as long as

$$\sum_{k=1}^{\tilde{K}} r_k \leq \frac{p_{\max} \sum_{i=1}^K \omega_i^+ - 1}{1 - p_{\max}}.$$

In a given benchmark, a negative weight component is hence not naturally paired with a single positive weight component. This method aims at providing benchmarks such that the number of acceptable pairs in the model is quite important. They constitute a good basis to challenge the performances of the simplex method as it has to narrow down the pairs involved in a pairing from a large number of initial acceptable pairs.

D.2 Second method

After generating all the positive weight components, accounting for multiplicity when more than one negative weight component is associated, for a given positive weight component, we randomly draw the parameter value of the associated negative weight component such that $a^\star \leq 10$. This constraint ensures the two-component signed mixture does not have a vanilla acceptance probability larger than 0.90 in the worst case scenario, making the next step easier. As opposed to the previous method, we now consider the linear combination

$af - g$ with a drawn uniformly in $[0, a^\star]$. The resulting function takes negative values on the support of f and, hence, does not define a distribution anymore.

We use all the positive components f_i, $1 \leq i \leq K-1$ generated, except f, to balance the negative part of that function. First, we make sure that all positive components together have enough mass over the set of negative values, that is the function is not negative in the tails of all the possible positive components. When necessary we add one or more positive weight components (we still denote K the overall number of positive weight components). We then compute the weights $\tilde{\omega}_i^+$ such that

$$af - g + \sum_{i=1}^{K-1} \tilde{\omega}_i^+ f_i \geq 0.$$

That yields a collection of K signed mixtures, each one having solely one negative component and associated with a vanilla acceptance probability p_i. We consider a convex combination of these signed mixtures to control the acceptance probability associated with the vanilla method and set it to $\min_i p_i$. This is usually not enough to ensure that p ranges in $[p_{\min}, p_{\max}]$. However, we can easily modify the model to satisfy this constraint by adding a two-signed component mixture to the model. We select at random a positive weight component included in the model and we build from it a two-signed mixture $a^\star f - g$ that fulfills the constraint on p. The new normalized signed mixture writes as

$$\frac{\sum_{i=1}^K \omega_i^+ f_i - \sum_{j=1}^N \omega_j^- g_j + \lambda(a^\star f - g)}{1 + \lambda(a^\star - 1)}.$$

The latter is associated with a vanilla acceptance probability lower than $p_{\max}$ as long as

$$\lambda \leq \frac{p_{\max} \sum_{i=1}^K \omega_i^+ - 1}{a^\star(1 - p_{\max}) - 1}.$$

This method aims at providing benchmarks that exhibit negative weight residuals. Such benchmarks allow to study the performances of the stratified method when the residual mixture obtained after the pairing step degrades the acceptance probability of the procedure (see Figure 7).

Appendix E Supplementary material on methods comparison

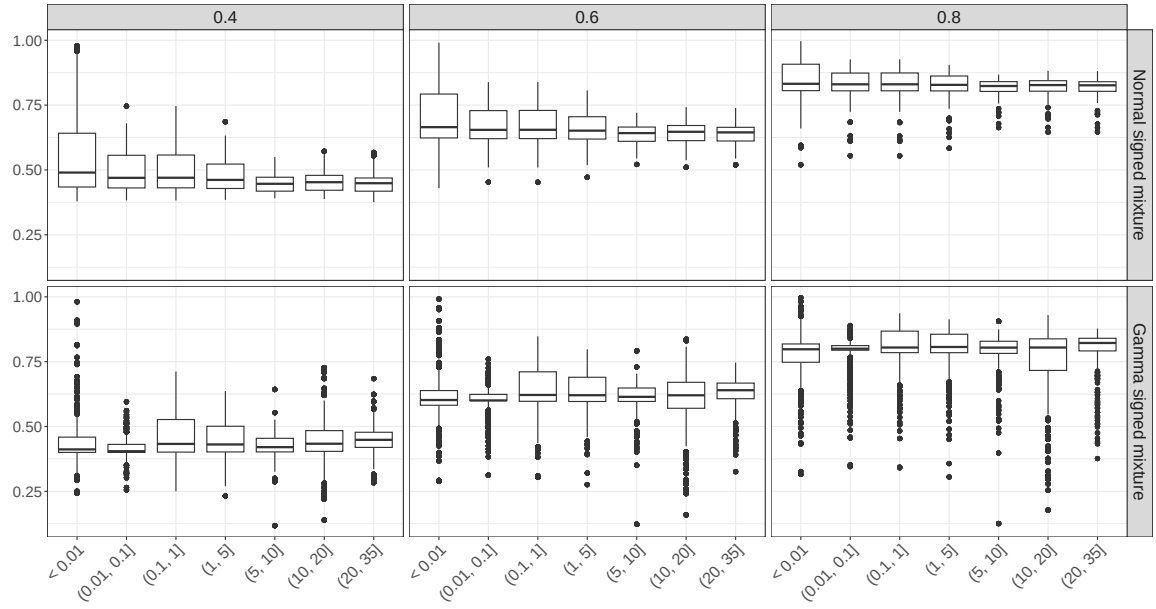

Figure 7: Theoretical acceptance probability of the stratified sampling scheme with respect to the vanilla average acceptance probability categories (x-axis in %) and user-specified acceptance probability δ for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (top row) and Gamma distributions (bottom row). An acceptance probability lower than δ signals the presence of negative weight residuals.

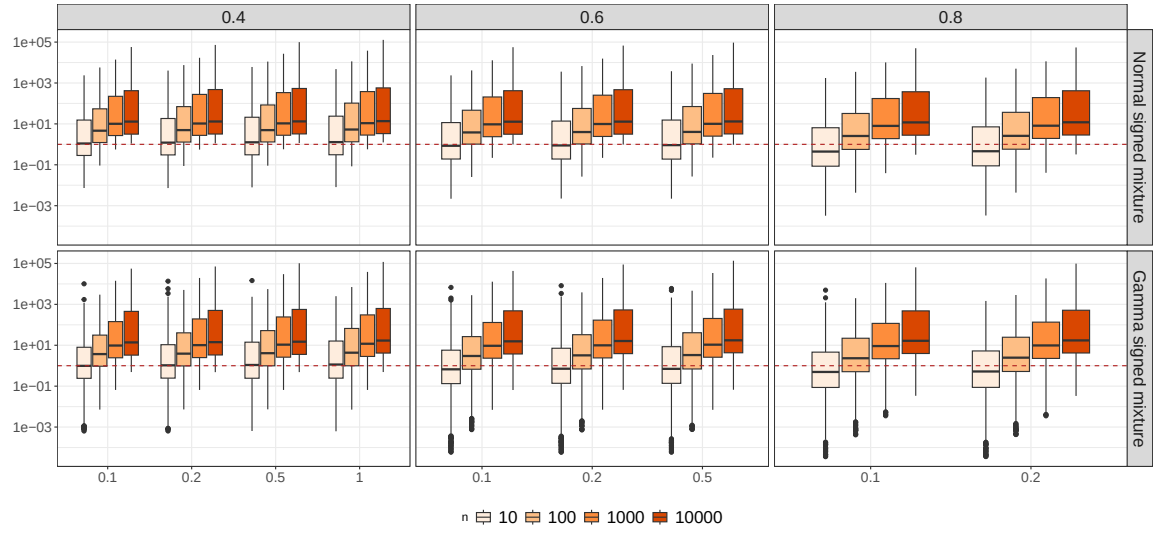

Figure 8: Relative efficiency of the vanilla method compared to the stratified method with respect to user-specified acceptance probability δ, tolerance level ϵ (x-axis) and number of draws n, for the 2,800 randomly generated signed mixtures of, respectively, Normal distributions (top row) and Gamma distributions (bottom row).