Gomes-Gonçalves, E., Gzyl, H., & Mayoral, S. (2015). Maxentropic approach to decompound aggregate risk losses. *Insurance: Mathematics and Economics*, 64, pp. 326-336.

DOI: 10.1016/j.insmatheco.2015.07.003

# Maxentropic approach to decompound aggregate risk losses

Erika Gomes-Gonçalves,[1] Henryk Gzyl [2] and Silvia Mayoral[1]

[1] Dept. of Business Administration, Universidad Carlos III de Madrid

[2] Centro de Finanzas, IESA, Caracas

epgomes@emp.uc3m.es; henryk.gzyl@iesa.edu.ve; smayoral@emp.uc3m.es

Version 2015

**Abstract**

A risk manager may be faced with the following problem: she/he has obtained loss data collected during a year, but the data only contains the total number events and the total loss for that year. She/he suspects that there are different sources of risk, each occurring with a different frequency, and wants to identify the frequency with which each type of event occurs and if possible, the individual losses at each risk event.

The purpose of this methodological note is to examine a combination of disentangling and decompounding procedures, to get as close as possible to that goal. The disentangling procedure is actually a two step process: First, a preliminary analysis is carried out to determine the number of risks groups present. Once that is decided, the underlying model for the frequency of each type of risk is worked out. After that we use the maxentropic techniques in the decompounding stage to determine the distribution of individual losses that aggregated yield the observed total loss.

# 1 Introduction

An interesting directive proposed in the Basel II agreement, is the possibility for the banks to develop their own advanced measurement approach (AMA) to compute the regulatory capital based on the loss distribution approach (LDA). This approach, modeled on that developed by the insurance industry and later included within the Basel II in the advanced measurement approach (AMA) for operational risk, is used to determine the distribution of the total loss from two complementary ingredients, to wit, the frequency and the severity of the individual losses measured (or observed).

Let $N_h$ denote the number losses of certain type labeled by $h = 1, ..., H$, occurring in a given time interval, $H$ being the total number of risk sources, and $X_{h,k}, k \geq 1$ a sequence of positive real-valued random variables that represent the size of the $k-$th loss of type $h$. With this notation, the quantity of interest is the compound model given by

$$S = \sum_{h=1}^{H} S_h \text{ where } S_h = \sum_{k=1}^{N_h} X_{h,k} \quad \text{with} \quad S_k = 0 \quad \text{if} \quad N_h = 0. \tag{1}$$

Here, $h = 1, ...., H$ labels the business line/type of the institution. In (**??**), it is supposed for each $h$ that the $X_{h,k}, k \geq 1$ are independent and identically distributed and independent of the $N_h$, and we shall denote by $X_h$ a random variable having that common distribution.

An interesting problem confronting the risk manager occurs at the level of the $S_h$. It may be the case that the frequency of events of type $h$ is actually the sum of at least two different types of events, i.e., it may happen that $N_h = N_h^1 + N_h^2$ during the observation period, even though the individual losses in each case may be the same. In this case, the risk analyst may want to know how many types of risk events of each type occur, and if possible, what is the distribution of individual losses. This is of interest because it is at that level where loss prevention or mitigation may be applied. This question motivates the problem that we shall examine here by a combination of procedures: one

(called disentangling) which consists of determining how many risk sources are there and their frequencies, and the other, the decompounding, is used to determine what is the statistical nature of the individual severities. Both procedures will be explained in Section 2, in which we establish the methodological aspects of the paper. We shall see that unless we know that the distributions of individual losses is the same, all that one is able to obtain is a mixture of distributions, a kind of collective individual loss distribution, which becomes the true individual distribution only when it is known that all are the same.

The disentangling procedure was proposed in Gomes-Gonçalves and Gzyl (2014), and we briefly review it here for the sake of completeness. The decompounding problem was treated for the first time by Buchmann and Grubel (2003). There they estimate the frequency rate parameter and the individual severity distribution, in a univariate compound Poisson distribution non-parametrically through a discrete Panjer inversion. Later on, a different approach to the problem was presented in Van Es et. al. (2007), who proposed a kernel type, nonparametric density estimator to obtain the individual losses of a univariate and bivariate compound Poisson distribution. They report good results in the numerical examples presented. Bøgsted and Pitts (2010) extended the work of Buchmann and Grübel (2004), and constructed a generalization of a nonparametric estimator for any compound base process given any parametric model for the distribution of the frequencies. We shall carry out the decompounding procedure using the maximum entropy method, for which the input shall be the Laplace transform of the aggregated loss $S$.

The remainder of the paper is organized as follows: the second section is devoted to a more detailed description of the methodology, the third to a review of the maximum entropy method, and the fourth to the discussion of a few numerical examples, after which we gather some concluding remarks.

# 2 Methodological preliminaries

In this section we lay down the basic framework in which we shall be working. We refer to Figure (**??**) for graphic display of the setup. As mentioned, the number of risk events during a certain period of time is recorded as well as the total loss during the period. It is known, or suspected, that there may be more of one source of risk present in the recorded aggregated loss, thus the first issue to take care of, is to determine the number of different risk types and their statistical nature. Once this is achieved, the next order of business is to determine the distribution of the individual risks.

Figure 1: Procedure

In actual practice, before applying the disentangling procedure to determine the statistical nature of the various frequencies of events, a first step consisting of determining how many different type $H$ of events are there will be applied. Once that number is known ($H = 2$ in the diagram), we apply the disentangling procedure described below. The output of this preprocessing stage is the specification of $H = 2$ integer valued random variables. If we suppose that the individual losses of each type are the same, invoking

(**??**), it is clear for the independence assumptions that the Laplace transform $\psi(\alpha)$ of the total loss is given by

$$\psi(\alpha) = E[e^{-\alpha S}] = E[e^{-\alpha S_1}]E[e^{-\alpha S_2}] = G_{N_1}(\phi(\alpha))G_{N_2}(\phi(\alpha)) \tag{2}$$

where, of course, $\phi(\alpha) = E[e^{-\alpha X}]$ denotes the Laplace transform of a random variable having the common distribution of the individual losses in (**??**). Thus, once that we have determined the distributions of the $N_i,\ i = 1, 2$, and computed $\psi(\alpha)$ from the data, we can solve (**??**) for $\phi(\alpha)$ and use it as input in the maxentropic methods to obtain the probability density of the individual losses from it. This is the second stage of the process.

The result of the methodology summed up in (**??**) is a mixture of discrete distributions to describe the total frequency $N$ of events, that will be serve as input of our second step, the decompounding methodology to determine the distribution of individual loses present in the aggregate loss.

To exemplify, suppose that the frequencies are Poisson($\lambda_i$), for $i = 1, ..., H$, and that the individual severities are distributed according to $f_{X_i}$. In this case, a simple computation shows that in this case (**??**) becomes

$$\psi(\alpha) = e^{-\sum_{h=1}^{H} \lambda_i (E[e^{-\alpha X_h}]-1)} = e^{-\lambda(E[e^{-\alpha \hat{X}}]-1)} \tag{3}$$

where we put $\lambda = \sum_{h=1}^{H} \lambda_i$ and $\hat{X}$ is a random variable whose density is the mixture $f_{\hat{X}} = \sum_{h=1}^{H} \frac{\lambda_I}{\lambda} f_{X_h}$. That is, the aggregated risk is the result of compounding a risk produced with a Poisson intensity equal to the sum of the individual intensities and individual loss with probability density equal to a weighted average of the individual densities, with respect to weights equal to the proportion of the individual frequency relative to the total frequency.

Following Wang (1997) we may extend the previous result as follows

**Proposition 1.** *Suppose that the $H$ compound risks to be aggregated have frequencies $N_i$ with a common mixing distributions $F(\theta)$ such that, given $\theta$ the $N_{|\theta} \sim P(\theta\lambda_i)$ and are independent. Let $S = \sum_{h=1}^{H} S_i$ be as above. Then $S \sim \sum_{n=1}^{N} \hat{X}_n$ with $N_{|\theta} \sim P(\theta\lambda)$, where $\lambda = \sum_{h=1}^{H} \lambda_i$, and as above, the $\hat{X}_n$ have common density $\sum_{h=1}^{H} \frac{\lambda_I}{\lambda} f_{X_h}$.*

*Proof.* The proof is actually easy, and hinges on the fact that $\frac{\lambda_i}{\lambda} = \frac{\theta\lambda_i}{\theta\lambda}$ is independent of the mixing parameter $\theta$, thus proceeding as if to prove (**??**) we would be led to

$$\psi(\alpha) = \int e^{-\theta\lambda(E[e^{-\alpha\hat{X}}]-1)} dF(\theta) = E[e^{-\alpha\sum_{n=1}^{N}\hat{X}_n}]$$

as claimed □

Notice that for the proposed mixing we have $\lambda_i / \sum_j \lambda_j = E[N_i] / \sum jE[N_j]$. This type of mixing appears as well in what Wang (1997) calls common shock models, which include correlated Poisson frequencies. The class of frequencies for which this is valid includes the Poisson and Negative Binomial random variables. These results were turned into a proposal to define an equivalent individual loss distribution by Wang, which we state as

**Definition 1** (Equivalent mixture)**.** *Let $f_{X_h}$ denote the common probability density of the individual losses for the compound losses of the $h-th$ type, $h = 1, ..., H$, and let $N_h$ denote the respective loss frequency. The aggregated individual loss has a density given by*

$$f_{\hat{X}} = \sum_{h=1}^{H} \frac{E[N_h]}{E[N_{agg}]} f_{X_h}, \tag{4}$$

*which we shall refer to as the "equivalent mixture", where $E[N_{agg}] = \sum_{h=1}^{H} E[N_j]$ is the mean of the aggregate loss frequency and by $\hat{X}$ we denote the individual loss which compounded according to $N_{agg}$, yields the total aggregate risk $S$.*

Thus the equivalent mixture of (individual) densities, is the density of a random variable, which compounded are the total frequency of events, yields the total observed losses. Our goal will be to determine the distributions of the $N_i$ and that of the equivalent mixture of densities, from the knowledge of the aggregate risk data.

In the general case we are within the scope of the identifiability problem considered for example in chapter 3 of Titterington et. al (1985). The importance of that work for us being, that unless we know that the individual losses in the risk sources are identically distributed, from the aggregate loss we shall not be able to obtain the individual losses of the different risk types. At the most, we shall be able to obtain the equivalent mixture

of densities given by (**??**). At this point it is interesting to advance, that the equivalent mixture coincides reasonably well with the distribution obtained by the decompounding procedure in the numerical experiments that we carried.

## 2.1 Disentangling frequencies

Consider the situation in which the frequency data $N$ is such that we can not distinguish between two or more sub-populations of risk sources. In this case it is necessary to separate those sub-populations in order to calculate the distribution of total losses for each type of risk. For that, essentially a two stage procedure is necessary. The important underlying supposition is that the frequency of events of all risks can be modeled by a distribution in the $(a, b, 0)$ Panjer class as described right below. The members of this family are characterized by a common recursion formula and two parameters that can be determined by a linear regression. The details are summed up in Table (**??**).

**Recursion formula**

$$p_k/p_{k-1} = a + b/k; \qquad k \geq 1$$

where $p_k = P(N = k)$ for $k \in \mathbb{N}$, and $p_k = N_k / \sum_k N_k$

| | **Density** | $a$ | $b$ | $p_0$ |
|---|---|---|---|---|
| Poisson | $p_k = \frac{e^{-\ell}\ell^k}{k!}$ | $0$ | $\ell$ | $e^{-\ell}$ |
| Binomial | $p_k = \binom{m}{k} q^k (1-q)^{m-k}$ | $-\frac{p}{1-p}$ | $(n+1)\frac{p}{1-p}$ | $(1-p)^n$ |
| Neg. Binomial | $p_k = \binom{k+r-1}{k} \left(\frac{1}{1+\beta}\right)^r \left(\frac{\beta}{1+\beta}\right)^k$ | $\frac{\beta}{1+\beta}$ | $(r-1)\frac{\beta}{1+\beta}$ | $(1+\beta)^{-r}$ |
| Geometric | $p_k = \left(\frac{1}{1+\beta}\right) \left(\frac{\beta}{1+\beta}\right)^k$ | $\frac{\beta}{1+\beta}$ | $0$ | $(1+\beta)^{-r}$ |

where $k, \beta, \ell, r, n \in \mathbb{N}$

Table 1: Relation between parameters $a$ and $b$ and discrete family distributions

The interesting fact about the recursion $p_k/p_{k-1} = a + b/k$ mentioned in Table (**??**) is that it can be rewritten as $kr(k) = ak + b$, where we set $r(k) = p_k/p_{k-1}$. Then, the plot $kr(k)$ vs. $k$ may suggest if there is more than one subpopulation present and may

indicate the parametric family to which it belongs. For example, if the plot consists of a line with zero slope, i.e. $a = 0$, this suggests that the underlying probability distribution is a Poisson, on the contrary if the slope is positive or negative ($a > 0$ or $a < 0$), then the underlying distribution is Negative Binomial or Binomial, respectively. The identification of these distributions will depend of the size of the data set and the degree of overlap. Additionally, identification of the models entails a visual comparison of plots with different scales, to see how steeper are the slopes of the data points.

The first stage consists of an initial analysis of the data in order to check for inconsistencies, this exploratory analysis may include a way to determine whether there are more than one sub-populations in the data. This can be done by criteria like the Akaike information criterion (AIC) proposed in Akaike (1973)-(1974), or the modified Akaike information criterion (AIC3) in Bozdogan (1987), the Bayesian integrated criterion (BIC) in Schwarz (1978), the integrated classification likelihood BIC type (ICL-BIC), see Biernacki et al (2000) or McLachlan and Peel, (2004).

In Gomes-Gonçalves (2014) we supposed that we knew how many sub-populations were present. In the present work, as part of the second stage we apply advanced techniques like the projection pursuit analyzed in Caussius and Ruiz Gazen (2009), Stone (2004), Peña and Prieto (2001 a,b) and Werner (2003), combined with the criteria mentioned above to determine the number of clusters (that is, the number of regression lines) present. Also one can apply the elbow method described in Horton and Everritt (2000) or the negentropy criterion described in Ramaswamy (1993) or Swait (2011) as confirmation tests of the number of clusters in the mixture.

Once we know the number of clusters that are there, we can proceed like in our previous work, and apply a combination of EM and k-means procedures to determine the different regressions that are in present in the data, that is, to infer the values of the coefficients $a$ and $b$ of each regression. Then, recurring to Table (**??**), to determine the parameters of the corresponding distribution. The quality of the parameters obtained is determined by a chi-squared test. As mentioned above, the value of the slope is a key indicator of which

of the $(a, b, 0)$ models should be selected.

Additionally it is also possible to compare the appropriateness of the distribution by looking at the relationship between the variance and the mean, or to apply methodologies like chi-squared goodness-of-fit test or others. These may help to improve the preliminary results obtained with the Panjer diagnostic display. We want to insist on the fact that this is not a mechanical procedure. It may involve handling some parameters by trial and error and visual comparisons with different family distribution plots, which may also help in the identification of the underlying model. But, as stressed in Panjer (2006), we must remember that whatever model we select it is only an approximation of reality, so our goal is to determine a model that is good enough to use to answer the question being considered, which depends on the particular application. Besides the decision can be influenced by the success of particular models in similar situations or the value of a particular model for its intended use.

To sum up, this methodology gives quite reasonable estimates especially for large samples and small values of $H$. Unfortunately, this procedure involves handling some parameters by either trial and error, which could be complicated in cases in which the counting distribution are too close from each other, distorting the overall picture. Even so, this will be useful as starting point for other methodologies or combinations of them. Additionally, previous knowledge about the data might be useful for the analysis. For example, in operational risk, the Poisson as well as the Negative Binomial models are adequate to describe the frequency of the losses. That is, previous experience should be brought in to determine a parsimonious model.

## 2.2 Decompounding Model

As we mentioned above, sometimes it may not be possible to record the individual severities in each risk event separately, then the only available data consists of the aggregated losses. Despite the fact that this information may be enough to infer the

distributions of the aggregated loss, interesting details and valuable information for risk management is hidden. This motivates the question: Is it possible to infer the distribution of the individual severities from that of the aggregated loss. Below, we shall see that in some cases this is possible, and that maximum entropy methods provide an efficient way to do it.

In the case in which all individual losses have a common distribution, the connection between the distribution of individual severities and the aggregate loss, is contained in the extended version of (**??**), namely

$$\psi(\alpha) = E[e^{-\alpha S}] = \prod_{i=1}^{H} E[e^{-\alpha S_i}] = \prod_{i=1}^{H} G_{N_i}(\phi(\alpha)) \tag{5}$$

Thus, whenever we can solve for $\phi(\alpha)$ in terms of $\psi(\alpha)$, then we can apply the procedure described below to determine the distribution of individual severities from it. When it makes sense to suppose that the densities of the individual losses are all equal, that is, that $\phi_1(\alpha) = ... = \phi_H(\alpha) \equiv \phi(\alpha)$, we may solve (**??**) for $\phi(\alpha)$, either analytically or numerically. Otherwise we have to resort to (**??**) to obtain the equivalent mixture of densities.

In Table (**??**), we show the relationship between $\phi(\alpha)$ and $\psi(\alpha)$ for the different models of frequency distributions in the $(a, b, 0)$ class for the case of only one source of risk. But in the general case we must use (**??**).

| **Starting point for the decompounding procedure** | |
|---|---|
| Distributions | $\phi(\alpha_k)$ |
| Poisson($\ell$) | $\frac{1}{\ell} ln(\psi(\alpha_k)) + 1$ |
| Binomial ($n$) | $\frac{\psi(\alpha_k)^{(1/n)} - (1-p)}{p}$ |
| Neg. Binomial($r$, $\beta$) | $\frac{1-\psi(\alpha_k)^{(-1/r)}}{\beta} + 1$ |
| Geometric($\beta$) | $\frac{1-\psi(\alpha_k)^{-1}}{\beta} + 1$ |

Table 2: *Laplace transform of the Severities*

We point out that the Poisson case and the case covered by Proposition (**??**) are of

special interest for us. In this case we are in the framework of the list of examples in Table (**??**). In this case we compute numerically the Laplace transform of the aggregate losses, use the disentangling procedure to identify the mixture of frequencies and then the decompounding procedure to obtain the individual severity distribution or a mixture of probability densities of individual severities. In this case, with some more information we may be able to obtain the individual severities.

The $\psi(\alpha)$ that appears in all of the identities listed above is estimated as

$$\psi(\alpha_k) = E(e^{-\alpha S}) = \frac{1}{M}\sum_{n=1}^{M} e^{-\alpha S_n} \tag{6}$$

in which $M$ is the number of data points, $S_n$ is the $n-$th loss recorded and $\alpha_k$ is the $k-$th value of the Laplace transform parameter. Actually, this computation is the first step for the determination of the probability density aggregate loss distribution, which we carry out because it is of independent interest, and useful as input for other purposes. A detail to be kept in mind, is the relationship between the $\psi(\alpha)$ and the fractional moments of the density $f_S$, contained in the following simple lemma.

**Lemma 1.** *Let $S$ be a compound random variable describing losses, and suppose that it has a density $f_S$. Then*

$$\psi(\alpha) = E[e^{-\alpha S}] = P(N=0)+(1-P(N=0))\int_0^1 y^{\alpha} f_Y(y)dy. = P(N=0)+(1-P(N=0))\mu(\alpha).$$

The proof consists of conditioning out the event $\{S=0\} = \{N=0\}$ and making the change of variables $Y = e^{-S}$. The probabilities $P(N=0)$ are obtained in the disentangling step.

# 3 Maximum entropy method

Here we rapidly recall two different approaches to the maximum entropy methods, that use as inputs the numerical approximation of the Laplace of the aggregated losses $\psi(\alpha_k)$ (as in eq. (**??**)) and a approximation to the Laplace of the severity, shown in Table (**??**) or an interval for this last value, which is calculated by bootstrapping.

The maxent methodologies that we use are the standard maximum entropy method (SME) and the maximum entropy with errors in the data (SMEE) which is a variation of the SME that allows to include the possibility of error in the data, these and additional maxent methodologies are described in Gomes-Gonçalves et al (2014).

## 3.1 Standard Maximum Entropy Method (SME)

The material presented below is standard, we recall the results that we use for the numerical computations and direct the reader to Gomes et al (2015) for the necessary details. The standard maximum entropy method (SME) is the name of a variational procedure proposed by Jaynes (1957) to solve the (inverse) problem consisting of finding a probability density $f_Y(y)$ (on $[0,1]$ in this case), satisfying the following integral constraints:

$$\int_0^1 y^{\alpha_k} f_Y(y)dy = \mu_Y(\alpha_k) \qquad \text{for} \qquad k = 0, 1, ..., M. \tag{7}$$

where the input $\mu_Y(\alpha_k)$ are the sample moments obtained from eq. (**??**) as explained in (**??**).

We set $\alpha_0 = 0$ and $\mu_0 = 1$ to take care of the natural requirement on $f_Y(y)$. It actually takes a standard computation to see that when the problem has a solution it is of the type

$$f_M^*(y) = \exp\left(-\sum_{k=0}^{M} \lambda_k^* y^{\alpha_k}\right) \tag{8}$$

in which the number of moments $M$ appears explicitly. It is usually customary to write

$e^{-\lambda_0^*} = Z(\boldsymbol{\lambda}^*)^{-1}$, where $\boldsymbol{\lambda^*} = (\lambda_1^*, ..., \lambda_M^*)$ is an $M-$dimensional vector. Clearly, the generic form of the normalization factor is given by

$$Z(\boldsymbol{\lambda}) = \int_0^1 e^{-\sum_{k=1}^M \lambda_k y^{\alpha_k}} dy. \tag{9}$$

With this notation the generic form of the solution can be rewritten as

$$f_M^*(y) = \frac{1}{Z(\boldsymbol{\lambda}^*)} e^{-\sum_{k=1}^M \lambda_k^* y^{\alpha_k}} = e^{-\sum_{k=0}^M \lambda_k^* y^{\alpha_k}}. \tag{10}$$

The vector $\boldsymbol{\lambda}^*$ can be found minimizing the dual entropy:

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu}_Y) = \ln Z(\boldsymbol{\lambda}) + < \boldsymbol{\lambda}, \boldsymbol{\mu}_Y > \tag{11}$$

where $< \mathbf{a}, \mathbf{b} >$ denotes the standard Euclidean scalar product and $\boldsymbol{\mu}$ is the $M-$vector with components $\mu_k$, and obviously, the dependence on $\boldsymbol{\alpha}$ is through $\boldsymbol{\mu}_Y$.

## 3.2 Maximum Entropy with errors in the data (SMEE)

Here we present an extension of the method described above to solve the (inverse) problem consisting of finding a probability density $f_Y(y)$ (on $[0, 1]$ in this case), satisfying the following integral constraints:

$$\int_0^1 y^{\alpha_k} f_Y(y) dy \in C_k \qquad \text{for} \qquad k = 0, ..., M. \tag{12}$$

where the interval $C_k = [a_k, b_k]$, around the true but unknown $\mu_Y(\alpha_k)$, is estimated from the data for each of the moments. For $k = 0$ only we set $C_0 = \{1\}$ since for $\alpha_0 = 0$ we have $\mu_0 = 1$ to take care of the natural normalization requirement on $f_Y(y)$.

Let us put $\mathbf{C} = C_0 \times \{1\} \times C_1 \times ... \times C_M$ and define

$$\delta_{\mathbf{C}}^*(\boldsymbol{\lambda}) \equiv \sup\{< \boldsymbol{\lambda}, \mathbf{c} > | \mathbf{c} \in \mathbf{C}\} = \|\mathbf{D}^{-1}\boldsymbol{\lambda}\|_1 - < \mathbf{D}^{-1}\boldsymbol{\lambda}, \mathbf{h} >$$

where clearly

$$\|\mathbf{D}^{-1}\boldsymbol{\lambda}\|_1 = \sum_{i=1}^M \frac{(b_i - a_i)|\lambda_i|}{2} \quad \text{and} \quad < \mathbf{D}^{-1}\boldsymbol{\lambda}, \mathbf{h} > = -\sum_{i=1}^M \frac{(b_i + a_i)\lambda_i}{2}.$$

As a first step towards describing the solution, we have the simple

**Lemma 2.** *With the notations introduced above, set*

$$\Sigma(\boldsymbol{\lambda}) = \ln(Z(\boldsymbol{\lambda})) + \delta^*_{\mathbf{C}}(\boldsymbol{\lambda})$$

*Then* $\Sigma(\boldsymbol{\lambda})$ *is strictly convex in* $\boldsymbol{\lambda}$.

Observe that $\partial\delta^*_{\mathbf{C}}(\boldsymbol{\lambda})/\partial\lambda_i$ is defined except at $\lambda_i = 0$, where it is sub-differentiable (see [12]). Actually

$$\begin{aligned}
\partial\delta^*_{\mathbf{C}}(\boldsymbol{\lambda})/\partial\lambda_k &= b_k \ \text{ when } \ \lambda_k > 0, \\
\partial\delta^*_{\mathbf{C}}(\boldsymbol{\lambda})/\partial\lambda_k &= a_k \ \text{ when } \ \lambda_k < 0, \\
\partial\delta^*_{\mathbf{C}}(\boldsymbol{\lambda})/\partial\lambda_k &\in (a_k, b_k) \ \text{ when } \ \lambda_k = 0
\end{aligned} \tag{13}$$

The formal statement of the result that we are after is

**Theorem 1.** *Suppose that the infimum* $\boldsymbol{\lambda}^*$ *of* $\Sigma(\boldsymbol{\lambda})$ *is reached in the interior of the set* $\{\boldsymbol{\lambda} \in \mathbb{R}^M | Z(\boldsymbol{\lambda}) < \infty\}$. *Then maxentropic solution to the inverse problem (**??**) is given by*

$$g^*(y) = \frac{1}{Z(\boldsymbol{\lambda}^*)} e^{-\sum_{k=1}^{M} \lambda_k^* y^{\alpha_k}} = e^{-\sum_{k=0}^{M} \lambda_k^* y^{\alpha_k}}. \tag{14}$$

**Comment** From the computations displayed just before the statement of the theorem, it is clear that

$$\int_0^1 y^{\alpha_k} g^*(y) dy \in [a_k, b_k], \ \text{ for } \ k = 1, ..., M.$$

# 4 Numerical examples

We will illustrate the use of our methodology considering several cases, in which we include different data features and procedures. The results for each of the cases were obtained by simulating samples corresponding to different frequencies and severities from the same or different (continuous) distribution, and then aggregating the resulting compound sums. The pseudo code could be detailed as:

**Sample generation procedure**

- For each $h : h = 1, ..., H$, generate a random loss count $N_h$ from a discrete distribution $F_{N_h}$ with mean $E[N_h]$ and variance $VaR[N_h]$.
- For each $h$, generate a sequence of individual losses $X_{h,n}$, for $n = 1, ..., N_h$ from a continuous distribution $F_{X_h}$.
- Set $S_h = \sum_{n=1}^{N_h} X_{h,n}$, loss for each line $h$.
- Set $S = \sum_{h}^{H} S_h$, which is the total compound loss that we need.

To test the proposed methodology, we consider different combinations of frequencies and individual severity distributions. In several of the cases use the Poisson or the Negative Binomial as counting distributions because they are rather common in insurance and operational risk models, and more important, members of the $(a, b, 0)$ class. The title of each subsection describes the combination of frequency/individual severity used. In the cases considered we describe in more detail the disentangling and decompounding procedures, and then we just summarize the results obtained.

In all numerical examples we shall measure the quality of the reconstructions by several distances between the reconstructed density and the histogram as well as between the reconstructed density and the equivalent density mixture. For that we shall use the $L_1$ and $L_2$ distances with the caveat that such comparison depends on the bins assigned to the histograms, and two distances, the MAE and the RSME that use data points and cumulative distributions for comparison. The necessary details can be seen in Gomes-Gonçalves et al. (2015).

## 4.1 Case 1: Poisson frequencies and Lognormal individual losses

For this case we suppose that the aggregate risk has two sources, the frequencies of each of them being a Poisson distribution with parameters $\ell_1 = 2$, $\ell_2 = 8$, and that the individual severities $X_1$, $X_2$ follow a common Lognormal distributions, $X \sim LogNormal(-1, 0.25)$. All the variables are supposed independent. Besides we consider a sample of size 500 to compute $S = S_1 + S_2$, and all that we record, or consider as given data, is the total number of risk events and the total loss in each sample.

### 4.1.1 Disentangling Procedure

This is the first step in our methodology and we shall describe it in much detail. In Figure (**??**) we show a histogram of the frequency of losses. We need to determine from the simulated data the number of sub-populations and their distributions. If we suppose that they belong to the $(a, b, 0)$ class and we make use of the plot based on the Panjer recursion formula, we observe that patterns appear as shown in Figure (**??**). There we observe a group of values around the value two, and another group of points that is more dispersed. Also, towards the right end of the abscissa axis we observe some larger values. Additionally, the groups observed in Figure (**??**) show little slope, and the only significant increase in the values occurs at the end. This seems indicate that the underlying distribution could be a mixture of Poisson distributions. On the other hand, by rescaling the vertical axis of the Figure (**??**) the slope of the points would look steeper, and a univariate Negative Binomial distribution can also be a possible candidate. But notice that the shape of the histogram of Figure (**??**) does not seem to suggest that.

To get an idea about the number of possible groups that are present in the data sample, it is customary to utilize a variety of information criteria measures like the AIC, AIC3, BIC and ICL-BIC values (see the Appendix and the notes in Table (**??**)). These measures address the goodness-of-fit of the clustering method (in our case, the EM method), and they are defined so that bigger is better. Additionally, we estimate the negentropy measure

(a) Frequencies of the losses (b) Panjer recursion formula

Figure 2: Histogram of frequencies and Panjer Recursion Formula (Case 1)

which indicates how well discriminated or separated the classes seem to be, based on posterior probabilities (as before the higher the measure, the better). The results of these estimations are listed in Table (**??**), where we consider a number of components that varies between one and six. This range was selected according to the result obtained with the sum of squared errors (SSE) for a broad number of clusters. This methodology is known in the clustering literature as the Elbow Method (see the Appendix) and it consists of calculating the sum of the squared distance between each member of a cluster to its cluster centroid, in order to give us an idea of the possible number of groups.

Table (**??**) shows that the information criteria is not very conclusive, these values indicate that the number of groups are between three and five. On the basis of the negentropy measure we decided to begin supposing that there are four sub-populations present.

Additionally, more advanced methodologies may be brought into the analysis. For example, the projection pursuit method of Peña and Prieto (2001a,b) gives us a possible clustering without the need of introduce in advance the number of groups. This algorithm detects four groups in our data, two of them being clusters and the other two are isolated points or outliers (the largest values). This result is equivalent to the one obtained with

| N. of components | AIC | AIC3 | BIC | ICL-BIC | Negentropy |
|---|---|---|---|---|---|
| 1 | 255.2 | 259.2 | 259.0 | 259.0 | 1 |
| 2 | 253.6 | 260.6 | 260.2 | 266.1 | 0.7763 |
| 3 | 231.5 | 241.5* | 240.9* | 244.9 | 0.9043 |
| 4 | 229.4 | 242.4 | 241.7 | 243.7* | 0.9616* |
| 5 | 228.4* | 244.4 | 243.5 | 247.8 | 0.9292 |
| 6 | 229.8 | 248.8 | 247.7 | 251.4 | 0.9458 |

* Value of g given by criterion.

$AIC = -2loglik + 2H$, $AIC3 = -2loglik + 3H$

$BIC = -2loglik + Hlog(n)$, $ICL = BIC + 2 * ENT$

$Negentropty = 1 - \frac{\sum_{i=1}^{n}\sum_{l=1}^{H} -p_{il} \cdot ln(p_{il})}{n \cdot ln(H)}$

$p_{il}$: is the posterior probability that the element $i$ being in group $l$

$H$: number of groups, $n$: number of elements in the data

Table 3: *AIC, AIC3, BIC, ICL-BIC & Negentropy values (Case 1)*

the EM algorithm when the input is the discrete data and the number of groups is $H = 4$. The results are displayed in Figure (**??**).

In Figure (**??**) we see that the two largest clusters have almost zero slope, this indicates that the two groups follow a Poisson distribution. Combining the results with the k-means the Poisson parameters are the centers with values of 2.08 and 7.64 for each group. Rounding these values, the density obtained is showed in Figure (**??**) as a mixture of Poisson distributions with different parameters.

Looking at Figure (**??**) we conclude that this density mixture provides a satisfactory fit to the data. Considering the $\chi^2$ goodness-of-fit test statistic we find a satisfactory fit ($\chi^2_o$=9.512, $\rho$=0.5747, 11 degrees of freedom ). Additionally, comparing this result with a simple Poisson and a simple Negative Binomial model, the mixture model always gives superior fit.

Now that we have identified the frequency distributions present in the data, we proceed to calculate the severity distribution of the individual losses through maxentropic methodologies. The input is the numerical computation of the Laplace transform (as in eq. **??**).

(a) Disentangling Result (b) Mixture of Poisson distributions

Figure 3: Results (Case 1)

Thus, by (**??**) we model the combined frequency $N$ by a Poisson distribution with $E[N] = E[N_1] + E[N_2] = \ell_1 + \ell_2 \approx 10$ and $Var[N] = Var[N_1] + Var[N_2] = \ell_1 + \ell_2 \approx 10$, because we are supposing that they are independent. Then, the combined individual severity distribution may be calculated by

$$f_{\hat{X}} = \frac{E(N_1)}{E(N_1) + E(N_2)} \times f_{X_1} + \frac{E(N_1)}{E(N_1) + E(N_2)} \times f_{X_2}$$

where $f_{X_1}$ and $f_{X_2}$ are the probability densities of the individual severities that are unknown. As this is usually the case, by the use of the maxentropic methodology we can decompound the aggregate losses to obtain the distribution of the individual severities that produces the observed aggregate loss. We take up this task next.

### 4.1.2 Decompounding procedure

Once we have determined a model for the frequency distribution, we implement the second stage of the procedure. We consider two versions on the maxentropic methodology: reconstruction with and without error in the data.

The inputs for the maxentropic methods are shown in Table (**??**). Here we observe the numerical calculation of the Laplace transform $\psi(\alpha)$ of the aggregate losses along with

the calculation of the Laplace transform $\phi(\alpha)$ of the individual severity using the fact that the combined frequency is Poisson and the computations displayed in Table (**??**) of Section 2.

| $k$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\psi(\alpha_k)$ | 0.0064 | 0.0515 | 0.1229 | 0.1978 | 0.2671 | 0.3284 | 0.3818 | 0.4283 |
| $\phi(\alpha_k)$ | 0.4959 | 0.7034 | 0.7903 | 0.8380 | 0.8680 | 0.8886 | 0.9037 | 0.9152 |

Table 4: Laplace transform of the aggregate losses and individual losses (Case 1)

To implement the maxentropic methodology for data with error we need the confidence for the numerical determination of the Laplace transform of the individual losses, which are shown in Table (**??**), these intervals are small and should contain the values of Table (**??**), they can be determined at the 1% or at the 5% confidence level using a bootstrapping procedure.

| $\phi(\alpha_k)$ | Confidence Interval |
|---|---|
| $\phi(\alpha_1)$ | [0.4948, 0.4966] |
| $\phi(\alpha_2)$ | [0.7032, 0.7039] |
| $\phi(\alpha_3)$ | [0.7903, 0.7907] |
| $\phi(\alpha_4)$ | [0.8380, 0.8383] |
| $\phi(\alpha_5)$ | [0.8680, 0.8682] |
| $\phi(\alpha_6)$ | [0.8886, 0.8888] |
| $\phi(\alpha_7)$ | [0.9037, 0.9038] |
| $\phi(\alpha_8)$ | [0.9151, 0.9153] |

Table 5: Intervals for the Laplace of the simulated individual losses (Case 1)

The results of the maxentropic methodology for the inputs of the Tables (**??**) and (**??**) are shown in Figure (**??**). In the left panel of Fig. (**??**) we display the results of the SME approach when the entries Table (**??**) are used as inputs. All curves seem to be close to the histogram of simulated data and to the equivalent mixture. In Table (**??**) we show the MAE and RMSE distances between them, and in Table (**??**) we show the $L_1$ and $L_2$ distances between them. Keep in mind that this comparison is possible because the

| Approach | Hist. vs. Real density | | Hist. vs. Maxent | | Real density vs. Maxent | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| SME | 0.0024 | 0.0029 | 0.0129 | 0.0143 | 0.0129 | 0.0143 |
| SMEE | 0.0024 | 0.0029 | 0.0143 | 0.0162 | 0.0142 | 0.0159 |

Table 6: MAE and RMSE distances between reconstructed densities, original histogram and true densities(Case 1)

simulated data and the equivalent mixture density are known to us, but in actual practice this may be unavailable.

(a) Results with SME (b) Results with SMEE

Figure 4: Maxentropic densities (Case 1)

The right panel of Fig. (**??**) displays the results the application of the SMEE to the inputs of Table (**??**). All the curves seem to be close to the histogram and the equivalent mixture density (the mixture of Lognormals). The distances between the maxentropic reconstructions and the data are shown in Tables (**??**) and (**??**).

| Approach | Hist. vs. Real density | | Hist. vs. Maxent | | Real density vs. Maxent | |
|---|---|---|---|---|---|---|
| | L1-norm | L2-norm | L1-norm | L2-norm | L1-norm | L2-norm |
| SME | 0.1120 | 0.2148 | 0.1443 | 0.2573 | 0.1116 | 0.1740 |
| SMEE | 0.1120 | 0.2148 | 0.1469 | 0.2633 | 0.1152 | 0.1822 |

Table 7: L1 and L2 distances between reconstructed densities, original histogram and densities (Case 1).

### 4.1.3 Identifiability

We know from (**??**) that the severities follow a mixture of distribution of the form

$$f_{\hat{X}} = \frac{E(N_1)}{E(N_1)+E(N_2)} \times f_{X_1} + \frac{E(N_1)}{E(N_1)+E(N_2)} \times f_{X_2}$$

the maxentropic methods give us the value of $f_{\hat{X}}$ and the disentangling methodology give us the weights of the distribution by assumption, and we can say that $f_{X_1}$ and $f_{X_2}$ are identifiable. If the two severities were different then the mixture is said to be unidentifiable, because there will be only one equation but two unknowns, in this case $f_{X_1}$ and $f_{X_2}$.

## 4.2 Case 2: Poisson frequencies with Beta and Frechet for the individual losses

This example is similar Case 1, except that now the densities of the individual losses are quite different among themselves. They are specified as follows: For the frequencies we consider $N_1 \sim Poisson(1)$ and $N_2 \sim Poisson(2)$, and the individual loss distributions are, respectively, $X_1 \sim Frechet(2,4)$ and $X_2 \sim Beta(1,5)$. This time the sample size was 1000. Simulating a sample from these populations we obtain the input for the problem, and we apply the disentangling/decompounding procedure once more.

### 4.2.1 Disentangling Procedure

The results of applying the Panjer recursion are shown in Figure (**??**). There we can observe an increasing pattern, but the increase does not seem to be large. If we truncate the values resulting of the Panjer recursion, we can see two horizontal lines (without slope) with y-values equal to 1 and 2, which suggests that we have mixture of Poisson distributions with different values of the parameters. The result of the EM algorithm is presented in Figure (**??**) and the mixture distribution is displayed in Figure (**??**).

(a) Frequencies of the losses (b) Panjer recursion formula

Figure 5: Histogram of frequencies and Panjer Recursion Formula (Case 3)

### 4.2.2 Decompounding procedure

The inputs for the maxent methods are shown in Table (**??**). In Figures (**??**) and (**??**) respectively, we present the result of the decompounding procedure using the SME and the SMEE methods, along with the plot of the equivalent mixture of densities.

| $M$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\psi(\alpha_k)$ | 0.2117 | 0.3164 | 0.3998 | 0.4673 | 0.5221 | 0.5672 | 0.6048 | 0.6365 |
| $\phi(\alpha_k)$ | 0.4824 | 0.6163 | 0.6944 | 0.7463 | 0.7833 | 0.8109 | 0.8323 | 0.8494 |

Table 8: Laplace transform of the aggregate losses and individual losses (Case 3).

(a) Disentangling Result (b) Mixture of Poisson distributions

Figure 6: Results (Case 3)

Even though the reconstruction in the middle region is not that satisfactory, the tail behavior is quite good. But keep in mind that we are using only 8 fractional moments to carry out the reconstruction. Perhaps the incorporation of more information may yield better reconstruction results, but the given 8 moments are those used for all the reconstructions that we carried out.

(a) Decompounding result. (b) Results with SMEE

Figure 7: Maxentropic Results for Case 3

| Approach | Hist. vs. Equi. Mix. | | Hist. vs. Maxent | | Equiv. Mix. vs. Maxent | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| SME | 0.0193 | 0.0233 | 0.0595 | 0.0709 | 0.0776 | 0.0908 |
| SMEE | 0.0193 | 0.0233 | 0.0515 | 0.0643 | 0.0658 | 0.08024 |

Table 9: MAE and RMSE distances between reconstructed densities, original histogram and densities (Case 3).

| Approach | Hist. vs. Equi. Mix | | Hist. vs. Maxent | | Equiv. Mix. vs. Maxent | |
|---|---|---|---|---|---|---|
| | L1-norm | L2-norm | L1-norm | L2-norm | L1-norm | L2-norm |
| SME | 0.4328 | 0.5999 | 0.4179 | 0.3954 | 0.4551 | 0.3913 |
| SMEE | 0.4328 | 0.5999 | 0.4978 | 0.4763 | 0.4554 | 0.3472 |

Table 10: L1 and L2 distances between reconstructed densities, original histogram and densities (Case 3).

## 4.3 Case 3: Binomial frequencies and Beta, Gamma, Weibull, Frechet individual losses

We consider now a more elaborate example. It consists of a larger collection of risks, each having a different individual loss density. We shall generate the data by aggregating four sources of risk in a sample of size 1000. We suppose that the individual frequencies are of the binomial type: $N_1 \sim Binomial(15, 0.50)$, $N_2 \sim Binomial(30, 0.65)$, $N_3 \sim Binomial(30, 0.90)$ and $N_4 \sim Binomial(40, 0.8)$. The individual losses to be aggregated have (in the same order) distributions given by $X_1 \sim Beta(1, 25)$, $X_2 \sim Weibull(1, 0.1)$, $X_3 \sim Frechet(0.01, 2)$ and $X_4 \sim Gamma(0.1, 3)$.

Note that now they are of quite different nature. For the disentangling procedure it is convenient to have Table (21), in which we collect a summary of the parameters of the binomial distributions used for the simulation.

| - | $\mu$ | $\sigma^2$ | $p$ | $n$ |
|---|---|---|---|---|
| $Group_1$ | 7.5 | 3.75 | 0.5 | 15 |
| $Group_2$ | 19.5 | 6.825 | 0.65 | 30 |
| $Group_3$ | 27 | 2.7 | 0.90 | 30 |
| $Group_4$ | 32 | 6.4 | 0.8 | 40 |

Table 11: Binomial parameters of each group (Original values)

### 4.3.1 Disentangling procedure

We proceed as in the previous examples. In Figure (**??**) we can see the histogram of the frequency of events and next to it, in (**??**), the result of the plot according to the Panjer recursion. Here we can observe different groups of points that seem to have a negative slope. The result obtained applying the EM algorithm is presented in Figure (**??**), and the mixture distribution may be seen in Figure (**??**).

In Table (**??**) we present a summary of the estimated parameters obtained with the regression formulation discussed in Section 2, along with a trial and error search in order to improve the results.

We considered five groups according to the result of the Negentropy measure, as the other four criteria does not seem to be consistent as observed in Table (**??**). In the left panel of Figure (**??**) we observe the five groups obtained by the EM procedure. The first set of points is excluded and for the calculations the other four sets of data are used.

### 4.3.2 Decompounding procedure

To obtain the severity of the individual losses we make use of both the SME and the SMEE procedures. In Table (**??**) we list the values of $\psi(\alpha)$ obtained by the simulation procedure and the corresponding values of $\phi(\alpha)$ to be used as input to the decompounding procedure.

(a) Frequencies of the losses (b) Panjer recursion formula

Figure 8: Histogram of frequencies and Panjer Recursion Formula (Case 5)

| **N. of components** | **AIC** | **AIC3** | **BIC** | **ICL-BIC** | **Negentropy** |
|---|---|---|---|---|---|
| 1 | 573.8 | 577.8 | 580.4 | 580.4 | 1 |
| 2 | 564.2 | 572.2* | 577.5* | 586.0* | 0.8330 |
| 3 | 561.1 | 573.1 | 581.1 | 593.6 | 0.8457 |
| 4 | 560.5 | 576.5 | 587.1 | 597.2 | 0.9011 |
| 5 | 557.8* | 577.8 | 591.1 | 597.8 | 0.9434* |
| 6 | 558.7 | 582.7 | 598.7 | 608.0 | 0.9291 |

* Value of g given by criterion.

$AIC = -2loglik + 2H$, $AIC3 = -2loglik + 3H$

$BIC = -2loglik + Hlog(n)$, $ICL = BIC + 2 * ENT$

$Negentropy = 1 - \frac{\sum_{i=1}^{n}\sum_{l=1}^{H} -p_{il}\cdot ln(p_{il})}{n\cdot ln(H)}$

$p_{il}$: is the posterior probability that the element $i$ being in group $l$

$H$: number of groups, $n$: number of elements in the data

Table 12: *AIC, AIC3, BIC, ICL-BIC & Negentropy values (Case 5)*

For the SMEE procedure we consider as input a interval which include the values of $\phi(\alpha)$ of Table (**??**). This interval is usually small. We obtained it by bootstrapping with replacement.

The results of the maxentropic methodology for the inputs of Tables (**??**) and (**??**)

(a) Disentangling Result (b) Mixture of Binomial distributions

Figure 9: Results (Case 5))

| - | $\mu$ | $\sigma^2$ | $p$ | $n$ |
|---|---|---|---|---|
| $Group_1$ | 7.67 | 4.94 | 0.36 | 22 |
| $Group_2$ | 19.2 | 7.68 | 0.60 | 32 |
| $Group_3$ | 27.3 | 2.457 | 0.91 | 30 |
| $Group_4$ | 32.37 | 5.50 | 0.83 | 39 |

Table 13: Results: Binomial parameters of each group

are shown in Figure (**??**).

In Table (**??**) we show the MAE and RMSE distances between the reconstructed densities, original histogram and the mixture of densities computed from the input data. In Table (**??**) we show the distances between the densities in the L1 and L2 norms.

In the Tables [(**??**) and (**??**)] we can observe a decrease in the value of the errors when we use as input the interval of $\phi(\alpha_k)$ in Table (**??**).

| $m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| $\psi(\alpha_k)$ | 0.0062 | 0.0688 | 0.1624 | 0.2524 | 0.3302 | 0.3957 | 0.4506 | 0.4971 |
| $\phi(\alpha_k)$ | 0.9425 | 0.9694 | 0.9791 | 0.9841 | 0.9872 | 0.9893 | 0.9908 | 0.9919 |

Table 14: Laplace transform of the aggregate losses and individual losses (Case 5).

| $\phi(\alpha_k)$ | **Confid. Interval** |
|---|---|
| $\phi(\alpha_1)$ | [0.9425, 0.9426] |
| $\phi(\alpha_2)$ | [0.9694, 0.9695] |
| $\phi(\alpha_3)$ | [0.9791, 0.9792] |
| $\phi(\alpha_4)$ | [0.9841, 0.9842] |
| $\phi(\alpha_5)$ | [0.9871, 0.9872] |
| $\phi(\alpha_6)$ | [0.9892, 0.9893] |
| $\phi(\alpha_7)$ | [0.9908, 0.9909] |
| $\phi(\alpha_8)$ | [0.9919, 0.9920] |

Table 15: Confidence Intervals for the Laplace of the simulated individual losses (Case 5)

(a) Results with SME (b) Results with SMEE

Figure 10: Maxentropic densities (Case 5)

# 5 Concluding comments

We presented a two stage procedure to determine the probability density of individual losses corresponding to an observed aggregate loss. The first consisting of determining

| Approach | Hist. vs. Equiv. mix | | Hist. vs. Maxent | | Equiv. mix vs. Maxent | |
|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| SME | 0.2120 | 0.2474 | 0.1270 | 0.1355 | 0.1052 | 0.1285 |
| SMEE | 0.2120 | 0.2474 | 0.1772 | 0.1854 | 0.0894 | 0.1009 |

Table 16: MAE and RMSE distances between reconstructed densities, original histogram and densities (Case 5).

| Approach | Hist. vs. Equiv. mix | | Hist. vs. Maxent | | Equiv. mix vs. Maxent | |
|---|---|---|---|---|---|---|
| | L1-norm | L2-norm | L1-norm | L2-norm | L1-norm | L2-norm |
| SME | 0.3943 | 1.9240 | 0.3163 | 1.1964 | 0.3147 | 1.1085 |
| SMEE | 0.3943 | 1.9240 | 0.3490 | 1.2403 | 0.3353 | 1.1201 |

Table 17: L1 and L2 distances between reconstructed densities, original histogram and densities (Case 5).

how many risk sources are present and their distributions, and the second to determine the distribution of individual losses when all risk sources have a common severity, or an equivalent individual severity when not.

In the several numerical experiments that we performed, when the individual severities of the different sources are different among themselves but the frequencies are of the Poisson or Negative Binomial type, the method works quite well. We considered a case in which the frequencies were Poisson with different parameters and the individual distributions had a common Gamma distribution. The results were as expected and the individual density was perfectly reconstructed. Another example considered consisted of a family of different negative binomial frequencies of events and the individual losses were two different Pareto distributed losses. The disentangling procedure yielded the parameters of the frequencies quite well and the individual losses turned out to be distributed according to the equivalent mixture.

But when the frequencies are of different types, the method does not necessarily work. There may be instances in which, even though the disentangling of the frequencies might

be possible, the nature of the underlying individual losses can be such that no equivalent loss might be determined. This happened when we considered the case in which the data comes from the following combination: The frequencies were distributed according to $N_1 \sim Poisson(2)$, $N_2 \sim NBin(10, 0.5)$ and $N_3 \sim Bin(25, 0.8)$, and the corresponding loss densities according to $X_1 \sim Weibull(2, 4)$, $X_2 \sim Frechet(2, 4)$ and $X_3 \sim Beta(1, 5)$.

Notice that if we knew that the individual frequencies were the same, we could at least attempt to solve (**??**) numerically to obtain $\phi(\alpha)$, but this may not be always possible.

Also, we considered small number of risk sources because when that number is large, the disentangling procedure is hard to implement when the amount of data is small. Actually, this is an important detail, for in actual practice, this is the situation in operational risk analysis. This, combined with the fact that not all of the individual severities have the same distribution is a handicap of the method. And to finish, the performance of the two procedures depends on the availability of enough data, something that may not be always the case.

We finish by addressing several issues. One relates to an important methodological aspect: How does the estimator perform on new data? In terms of the numerical examples, how do the estimators fitted on one half of the data set perform on predicting the other half of a data set? These two are related and have a direct answer: The procedures are the same but have to be applied afresh to each new data set. As with operational risk: data sets can be augmented as time goes on. More data would mean better estimates of the loss distributions.

Another interesting issue is: Real data sets for operational risk will contain information not just on the risks but on how the data was collected. How to deal with this? Note that the problems and the methods to solve them that we deal with here, use very specific data, namely total number of events during a given period and total loss during that period. If the information on how the data is collected affects those numbers, and if that can be quantified, we could perhaps treat the resulting available data as data with errors

in the measurement process.

An issue that we do not deal with here, but will be the matter of a forthcoming study, is the relationship between the size of the sample data and the variability of the maxentropic density and regulatory capital estimation. At this point we also mention that once the Laplace transform $\psi(\alpha)$ of a collection of aggregate losses is estimated, we may be interested in using it to determine the density of the total loss distribution $f_S$. We also leave this problem to a forthcoming note.

**Acknowledgements** We thank the reviewers for their comments. They helped us to improve the presentation.

# References

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov and F. Caski (Eds.), *Proceedings of the Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.

[3] Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Trans. Pattern Analysis and Machine Intelligence*, **22**, 719-725.

[4] Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions.*Psychometrika*, **52**, 345-370.

[5] B$\phi$gsted, M., and Pitts, S. M. (2010). Decompounding random sums: a nonparametric approach. *Annals of the Institute of Statistical Mathematics*, **62**, 855-872.

[6] Buchmann, B., and Grübel, R. (2004). Decompounding Poisson random sums: recursively truncated estimates in the discrete case. *Annals of the Institute of Statistical Mathematics*, **56**, 743-756.

[7] Caussinus, H. and Ruiz-Gazen, A. (2009) Exploratory Projection Pursuit, in Data Analysis (ed G. Govaert), *ISTE, London, UK*. doi: 10.1002/9780470611777.

[8] Cruz, M. G. (2002). Modeling, measuring and hedging operational risk. *John Wiley & Sons*, New York.

[9] Hothorn, T., and Everitt, B. S. (2009). A handbook of statistical analyses using R (Vol. 12). *CRC Press*.

[10] Gomes-Gonçalves , E., and Gzyl, H. (2014). Disentangling Frequency Models, *Journal of Operational Risk*, **9**(2), 3-21.

[11] Gomes-Gonçalves , E. Gzyl, H., Mayoral, S. (2015). *Two maxentropic approaches to determine the probability density of compound risk losses*. Insurance: Mathematics and Economics, **62**, 42-53.

[12] Gomes-Gonçalves , E, Gzyl, H., Mayoral, S.(2014). Density Reconstructions with Errors in the Data. *Entropy* **16**(6), 3257-3272.

[13] Gzyl, H., Novi-Inverardi, P.L. and Tagliani, A (2013). A comparison of numerical approaches to determine the severity of losses. *J. of Oper. Risk*, *8*, 3-15.

[14] Jaynes,E.T. (1957) Information theory and statistical physics. *The Phys. Rev.*, **106**, 620-630.

[15] Mosteller, D.C., and Tukey, J.W. (Eds.) (2011). Exploring data tables, trends, and shapes (Vol. 101). *John Wiley & Sons*.

[16] Lin, G.D. (1992). Characterizations of distributions via moments, *Sankhya: The Indian Journal of Statistics*, **54**, 128132.

[17] McLachlan, G., and Peel, D. (2004). Finite mixture models. *John Wiley & Sons*.

[18] Panjer, H. (2006). Operational Risk: Modeling Analytics, John Wiley & Sons, Inc., Hoboken, NJ.

[19] Peña D., Prieto F. J.(2001a). Cluster Identification Using Projections, *Journal of the American Statistical Association*, **96**, 1433-1445.

[20] Peña D. and Prieto F. J.(2001b). Cluster identification routines for the paper Cluster Identification Using Projections, J. Amer. Stat. Assoc., **96**, 1433-1445.

[21] Peña, D., Prieto, F.J. and Rocke, D.M., (2001). Discussion of Multivariate Outlier Detection and Robust Covariance Matrix Estimation *Technometrics*, **43**, 300-303.

[22] Ramaswamy, V., Wayne Desarbo, D. Reibstein and W. Robinson (1993). An Empirical Pooling Approach for Estimating Marketing Mix Elasticities with PIMS Data, *Marketing Sciences*, **12**,103- 124

[23] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.

[24] Stone, J. V. (2004). Independent component analysis: a tutorial introduction. *Massachusetts Institute of Technology*.

[25] Swait, J. D. (2011). Discrete choice theory and modeling. The Oxford Handbook of the Economics of Food Consumption and Policy, 119.

[26] Titterington, D. M., Smith, A. F. M. and Makov, U. E. (1985). Statistical analysis of finite mixture models. *Wiley*, Chichester

[27] Van Es, B., Gugushvili, S., and Spreij, P. (2007). A kernel type nonparametric density estimator for decompounding. *Bernoulli*, **13**, 672-694.

[28] Wang, A.S. (1997) Aggregation of correlated risk portfolios: Models and algorithms.*C*asualty Actuarial Society, pp. 848-939.

[29] Werner, M. (2003). Identification of multivariate outliers in large data sets (Doctoral dissertation, University of Colorado at Denver).