Gomes-Gonçalves, E., Gzyl, H. & Mayoral, S. (2016). Loss data analysis: Analysis of the sample dependence in density reconstruction by maxentropic methods. *Insurance: Mathematics and Economics, 71*, 145–153.

DOI: 10.1016/j.insmatheco.2016.08.007

Loss data analysis: Analysis of the sample dependence in density reconstruction by maxentropic methods

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

September 24, 2016

Abstract

The problem of determining probability densities of positive random variables from empirical data is important in many fields, in particular in insurance and risk analysis. The method of maximum entropy has proven to be a powerful tool to determine probability densities from a few values of its Laplace transform. This is so even when the amount of data to compute numerically the Laplace transform is small. But in this case, the variability of the reconstruction due to the sample variability in the available data can lead to quite different results. It is the purpose of this note to quantify as much as possible the variability of the densities reconstructed by means of two maxentropic methods: the standard maximum entropy method and its extension to incorporate data with errors.

The issues that we consider are of special interest for the advanced measurement approach in operational risk, which is based on loss data analysis to determine regulatory capital, as well as to determine the loss distribution of risks that occur with low frequency.

1 Introduction and preliminaries

One of the methodologies that banks can use to determine regulatory capital for operational risk is the advanced measurement approach, which is based on the possibility of determining the operational risk capital from the probability density of the yearly losses. Actually the problem of determining the probability density of compound losses has received a lot of attention since a long time ago in the literature devoted to insurance matters. But since the Basel Committee proposals to measure and manage operational risk, it had a revival. See Cruz (2002), Panjer (2006) or Shevchenko (2011) for a variety of aspects about the problem and for procedures to obtain the probability density of aggregate losses from the historical data. The two volumes just mentioned are part of a large body of literature devoted to the theme. To mention just a few papers rapidly cascading into a large pool of literature, consider Aue and Kalkbrenner (2006), Temnov and Warnung (2008), Brockman and Kalkbrenner (2010)

In a series of previous papers, see Gzyl et al. (2013), we explored the usefulness of the maximum entropy based (maxentropic) procedure to determine the density of aggregate losses, and compared the procedure to standard procedures like Fourier inversion techniques, direct computation of the total loss density by convolution or reconstruction from integral moments.

The power of the method that we apply here, stems from the possibility of inverting the Laplace transform of a positive random variable S from the knowledge of a few values of its Laplace transform $\psi(\alpha) = E[e^{-\alpha S}]$ at a small set of values of the parameter α by recasting the problem into a fractional moment problem on $[0, 1]$ after the transformation $x = e^{-s}$. Furthermore, an interesting feature of the methodology is that the statistical error in the estimation of $\psi(\alpha)$ can be incorporated into the procedure, as developed in Gomes-Gonçalves et al (2014), and that the procedure itself makes no assumptions about the statistical nature of the data.

That the maxentropic methods work when the amount of data is large, was the subject

matter of Gomes-Gonçalves et al. (2015a,b). In the first of these, the aim was to examine the performance of two maxentropic methods to determine the density of aggregate losses. In the second, it was supposed that the losses may be produced by different sources of risk, that is, that may have different types of events producing losses at different rates, but the available data consists of the total loss. The problem in this case is to disentangle the different sources of risk, and to determine the nature of the individual losses.

In our previous work, we have seen that maximum entropy based techniques are quite powerful to determine density distributions when the amount of data is large. The maxentropic techniques work equally well when the amount of data is small, a situation which happens when analyzing operational risk data for example. In this case one expects the resulting densities to depend on the sample used to compute the moments. Fortunately the dependence of the maxentropic density on the sample is such that its variability can be analyzed explicitly. It is our aim here to analyze the variability of the reconstructed densities, and on the other hand to examine the impact of this variability on the estimation of the regulatory capital.

To state the problems with which we shall be concerned, let us begin saying that we are interested in compound variables of the type $S = \sum_{k=0}^{N} X_k$, where N is an integer valued random variable describing the (yearly, say) frequency of events, and X_k is the individual severity of each loss. What the analyst observes each year is a collection $\{n; x_1, ..., x_n\}$, where n is the number of risk events and $\{x_1, ..., x_n\}$ are the losses occurring at each event. The aggregate loss for that year is $s = \sum_{k=0}^{n} x_k$. When $n = 0$ there were no losses,the sum is empty and $s = 0$. Suppose that the record consists of M years of data. From these, the analyst has to determine the distribution of losses, which is the intermediate step in the calculation of regulatory capital or some other measure of risk, or perhaps when some insurance premium is to be calculated. When we need to specify the year j we shall write $(n_j, x_1, ..., x_{n_j})$ and $s_j = \sum_{k=0}^{n_j} x_k$. For us, an observed sample (of losses) will be an (M) vector $\omega = (s_1, ..., s_M)$.

The Laplace transform of S is estimated by

$$\psi(\alpha) = \frac{1}{M}\sum_{j=1}^{M} e^{-\alpha S_j} \tag{1}$$

where S_j denotes the losses experienced during the $j-$th year. Later on we shall consider the moments corresponding to K values of the parameter α. Since the distribution function of S has a probability $P(N=0) = P(S=0) > 0$ at $S=0$, to determine the probability density of the losses we have to condition out this event and replace $\psi(\alpha)$ by

$$\mu(\alpha) = \frac{\psi(\alpha) - P(N=0)}{1 - P(N=0)} \tag{2}$$

where $P(N=0)$ is the estimated as the fraction of the number of years of observation in which there were no losses. Notice that if we use the change of variables $y = e^{-s}$ we can rewrite (1) as

$$\psi(\alpha) = \frac{1}{M}\sum_{j=1}^{M} y^{\alpha} \tag{3}$$

which is the empirical version of

$$\psi(\alpha) = \int_0^1 y^{\alpha} dF_Y(y) = \int_0^{\infty} e^{-\alpha x} dF_S(x).$$

With this notation our problem consists of finding a density $f_Y(y)$ on $[0,1]$ such that

$$\int_0^1 y^{\alpha} f_Y(y) dy = \mu(\alpha),$$

and once $f_Y(y)$ has been obtained, the change of variables $f_S(x) = e^{-x} f_Y(e^x)$ provides us with the desired density.

As we shall have to emphasize the dependence of f_Y on the size M of the sample, we shall drop the Y and simply write f for it, and we shall use the notation $f_M(\omega, x)$ to denote the maxentropic density reconstructed from the collection of K moments as in (2) . We describe how to obtain f_M in Section 2 below when we explain the maximum entropy methods. Note that as (2) depends on the sample ω, then f_M depends on ω. To further specify our goals, there are three things that we want to understand, or develop intuition about. First, how much does f_M change when we change ω. Second, how much

do some basic risk measures change when we change ω, and third, what happens as M becomes very large.

As we shall recall in Section 2 below, the connection between the moments $\mu(\alpha, \omega)$ is quite non-linear, the study of the variability of f_M is not that simple, nevertheless, a few things can be said. We shall carry this out in Section 3, while in Section 4 we examine this issue by numerical simulations. The data that we use as input consists of an aggregation of risks of different nature, so it is not a simple compound model as that considered in our previous work.

We close this section mentioning that there are other methods to deal with the problem of inferring loss densities from data, some simpler and some more elaborate. Consider the well known Panjer recursion technique, or the fast Fourier transform as described, say in Shevchenko (2011), or the cubic interpolation spline proposed in den Iseger et al. (1997). One of the interesting features of the representation provided by the maxentropic approach, is that it provides an explicit analytic representation of the density which can be used as a starting point for a systematic analysis of the sample dependence

2 The maximum entropy inversion techniques

We shall describe two complementary approaches to the density reconstruction problem. First, the standard maximum entropy (SME) method and then the standard maximum entropy with error (SMEE) in the data, which is useful to cope with the issue of data uncertainty.

2.1 The standard maximum entropy method

The procedure 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:

$$\int_0^1 y^{\alpha_k} f_Y(y) dy = \mu_Y(\alpha_k) \qquad \text{for} \qquad k = 0, 1, ..., K. \tag{4}$$

seems to have been originally proposed in Jaynes (1957). We set $\alpha_0 = 0$ and $\mu_0 = 1$ to take care of the natural normalization requirement on $f_Y(y)$. The intuition is rather simple: The class of probability densities satisfying (4) is convex. One can pick up a point in that class one by maximizing (or minimizing) a concave (convex) functional (an "entropy") that achieves a maximum (minimum) in that class. That extremal point is the "maxentropic" solution to the problem. It actually takes a standard computation to see that, when the problem has a solution, it is of the type

$$f(y) = \exp\left(-\sum_{k=0}^{K} \lambda_k^* y^{\alpha_k} \right) \tag{5}$$

which depends on the α's through the λ's. It is usually customary to write $e^{-\lambda_0^*} = Z(\boldsymbol{\lambda}^*)^{-1}$, where $\boldsymbol{\lambda^*} = (\lambda_1^*, ..., \lambda_K^*)$ is a $K-$dimensional vector. Clearly, the generic form of the normalization factor is given by

$$Z(\boldsymbol{\lambda}) = \int_0^1 e^{-\sum_{k=1}^{K} \lambda_k y^{\alpha_k}} dy. \tag{6}$$

With this notation the generic form of the solution looks like

$$f^*(y) = \frac{1}{Z(\boldsymbol{\lambda}^*)} e^{-\sum_{k=1}^{K} \lambda_k^* y^{\alpha_k}} = e^{-\sum_{k=0}^{K} \lambda_k^* y^{\alpha_k}}. \tag{7}$$

To complete, it remains to specify how the vector $\boldsymbol{\lambda}^*$ can be found. For that one has to minimize the dual entropy:

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \ln Z(\boldsymbol{\lambda}) + \langle \boldsymbol{\lambda}, \boldsymbol{\mu}_Y \rangle \tag{8}$$

with respect to $\boldsymbol{\lambda}$ for each fixed $\boldsymbol{\mu}$. There $\langle \mathbf{a}, \mathbf{b} \rangle$ denotes the standard Euclidean scalar product and $\boldsymbol{\mu}$ is the $K-$vector with components μ_k, and obviously, the dependence on $\boldsymbol{\alpha}$ is through $\boldsymbol{\mu}_Y$.

We mention, that when the solution of this dual problem exists, then we have

$$H(f^*) := -\int_0^1 f^*(y)\ln f^*(y)dy = \Sigma(\boldsymbol{\lambda}^*, \boldsymbol{\mu}) = \ln Z(\boldsymbol{\lambda}^*) + \langle \boldsymbol{\lambda}^*, \boldsymbol{\mu}_Y \rangle \tag{9}$$

2.2 The standard maxentropic approach with error estimation

In this section we present an extension of the method of maximum entropy that allows us to estimate the errors in the data. In our case these come from the estimation of the moments from the empirical data, in which an integral is approximated by a sum, to begin with. These are not measurement errors in the science/engineering sense but they reflect uncertainty in the input to the maxentropic procedure. We should add that perhaps this procedure could be invoked when one suspects that there may be under-reporting or over-reporting errors in the data gathering process.

Instead of supposing that $f_Y(y)$ satisfies (4) we present a totally different approach. To restate the problem, consider the following argument. Suppose that the measurement error in the determination of the $k-$th moment lies within the confidence interval $C_k = [a_k, b_k]$, then the estimate of the unknown measurement error can be written as $p_k a_k + (1-p_k)b_k$ for appropriate but unknown $p_k, q_k = 1 - p_k$. To determine the p_k's we extend the original problem to a problem consisting of determining a density f_Y on $[0,1]$ and parameters $0 < p_k < 1$ such that

$$\int_0^1 y^{\alpha_k} f_Y(y)dy + p_k a_k + (1-p_k)b_k = \mu_k \qquad \text{for} \qquad k = 1, ..., K. \tag{10}$$

The idea behind the proposal is clear. In order to obtain the μ_k's either experimentally or numerically, we average over a collection of observations (simulations) and the errors in each average additively. Thus, the observed value of μ_k, consists of the true (unknown) moment which is to be determined, plus an error term that has to be estimated as well. We shall suppose that all intervals $[a_k, b_k]$ are equal to a single $[-c, c]$, independently of the sample. The interval is chosen to be symmetric about 0 to allow for the error to be of any sign.

This computational simplification is possible when the support of the distribution of errors is bounded. To further simplify the notation, we write probability distributions concentrated on a two point set $\{-c, c\}$ as $q_k(dz) = p_k\epsilon_{-c}(dz) + (1 - p_k)\epsilon_c(dz)$, that is, the probability that we are after is a mixture of continuous and discrete distributions. To determine it we define, on the appropriate space of product probabilities, the entropy

$$H(f, \mathbf{p}) = -\int_0^1 f(y)\ln(f(y))dy - \sum_{k=1}^{M}\left(p_k \ln p_k + (1 - p_k)\ln(1 - p_k)\right). \tag{11}$$

Needless to say, $0 < p_k < 1$, for $k = 1, ..., K$, with this notation, our problem becomes

$$\begin{gathered}\text{Find probability density } f^*(y) \text{ and numbers } \ 0 < p_k < 1 \ \text{ satisfying:} \\ \text{Constraints (10) and the normalization constraint } \int_0^1 f_Y(y)dy = 1.\end{gathered} \tag{12}$$

The usual variational argument in Jaynes (1957), or the more rigorous proofs in Maslov and Cherny (2003), yield

$$\begin{aligned} f^*(y) &= \frac{e^{-\sum_{k=1}^{K}\lambda_k^* y^{\alpha_k}}}{Z(\boldsymbol{\lambda}^*)} \\ p_k &= \frac{e^{c\lambda_k^*}}{e^{c\lambda_k^*} + e^{-c\lambda_k^*}}. \end{aligned} \tag{13}$$

Here, the normalization factor $Z(\boldsymbol{\lambda})$ is as above. This time the vector $\boldsymbol{\lambda}^*$ of Lagrange multipliers is to be found minimizing the dual entropy

$$\Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \ln Z(\boldsymbol{\lambda}) + \sum_{k=1}^{K}\ln\left(e^{c\lambda_k} + e^{-c\lambda_k}\right) + <\boldsymbol{\lambda}, \boldsymbol{\mu}> \tag{14}$$

Once $\boldsymbol{\lambda}^*$ is found, the estimator of the measurement error is, as implicit in (10), given by

$$\epsilon_k = \frac{-ce^{c\lambda_k^*} + ce^{-c\lambda_k^*}}{e^{c\lambda_k^*} + e^{-c\lambda_k^*}}. \tag{15}$$

Notice that, although the formal expression for $f^*(y)$ is the same as that for the first method, the result is different because the $\boldsymbol{\lambda}^*$ is found minimizing a different functional. For more details about this, we direct the reader to Gomes-Gonçalves et al. (2014).

2.3 Remarks on minimization process

Let us examine some properties of $\Sigma(\boldsymbol{\lambda}, \boldsymbol{\mu})$ related to the minimization process. First, observe that in our set up, it is a strictly convex function and twice continuously differ-

entiable function, defined on all $\mathbb{R}^K$. If we denote by

$$\phi(\boldsymbol{\lambda}) = \begin{cases} -\nabla_{\boldsymbol{\lambda}} \ln Z(\boldsymbol{\lambda}) \\ -\nabla_{\boldsymbol{\lambda}} \left(\ln Z(\boldsymbol{\lambda}) + \sum_{k=1}^{K} \ln \left(e^{c\lambda_k} + e^{-c\lambda_k} \right) \right) \end{cases} \tag{16}$$

then the first order condition for $\boldsymbol{\lambda}^*$ to be a minimizer is that $\phi(\boldsymbol{\lambda}^*) = \boldsymbol{\mu}$. In practice this is not of much help and the minimization of (8) has to be carried out numerically.

Notice that $\phi(\boldsymbol{\lambda})$ is differentiable and its Jacobian (the negative of the Hessian matrix of $\ln(Z(\boldsymbol{\lambda}))$ is the negative of the covariance matrix of the fractional powers y^{α_i}, that is of the positive definite matrix with components $\mu(\alpha_i + \alpha_j) - \mu(\alpha_i)\mu(\alpha_j)$ with respect to the maximum entropy density.

Observe as well that $(\boldsymbol{\mu})$ has all of its components bounded by 1, thus the image of $[0, 1]$ by ϕ^{-1}, which is continuous due to finiteness of the Hessian matrix, is compact. Thus the following assumption, that plays a role in the analysis carried out in the next section is natural:

Assumption A *Suppose that there is a ball $B(0, R)$ in $\mathbb{R}^K$ such that all solutions to $\phi(\boldsymbol{\lambda}^*) = \boldsymbol{\mu}$ lies there for every $(\boldsymbol{\mu}) \in [0, 1]^K$. Suppose as well that the Hessian matrix*

$$\frac{\partial^2 \Sigma(\boldsymbol{\lambda}^*, \boldsymbol{\mu})}{\partial \lambda_i \partial \lambda_j} = -\frac{\partial \phi_i}{\partial \lambda_j}(\boldsymbol{\lambda}^*)$$

has its upper eigenvalue uniformly bounded above, and its minimum eigenvalues uniformly bounded away from 0 in $B(0, R)$.

3 Variability of the reconstructions

The results presented next are a particular case of the sample dependence in the generalized moment problem studied in Gzyl (2015). As mentioned above, there several issues to take care of. The simplest of all is to make sure that when the sample is large, the

reconstructed f_M converges to the reconstruction obtained from the moments of the exact Laplace transform. This is rather direct if we suppose that the samples are independent. For the next statement we need the following notations.

Denote by $\mu_M(\alpha)$ the moments computed as in (2) and $\mu_e(\alpha)$ the exact moments. Denote as well $\boldsymbol{\lambda}_M^*$ and $\boldsymbol{\lambda}_e^*$ the corresponding minimizers of (8) and by $f_M(x)$ and $f_e(x)$ the maxentropic densities.

Lemma 1. *With the notations just introduced, suppose that the risks observed during M consecutive years are independent of each other.*
Then $\mu_M(\alpha) \to \mu_e(\alpha)$ and $\boldsymbol{\lambda}_M^* \to \boldsymbol{\lambda}_e^*$, and therefore $f_M(x) \to f_e(x)$ when $M \to \infty$.

Proof. The proof hinges on an application of the law of large numbers combined with an invocation to the continuity of ϕ^{-1} based on Assumption A. □

To relate the sample variability of f_M to the sample variability of the $\mu(\alpha)$, starting from $\boldsymbol{\lambda}_M^* = \phi^{-1}(\boldsymbol{\mu})$, and applying the chain rule, it follows that up to terms of $o(\delta\boldsymbol{\mu})$

$$\delta\boldsymbol{\lambda}_M^* = \mathbf{D}\delta\boldsymbol{\mu} \tag{17}$$

where $\mathbf{D}$ is the inverse matrix of the Jacobian of ϕ evaluated at $\boldsymbol{\lambda}_M^*$. Recall that the maxentropic solution to the inverse problem is

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

in which a reference to the sample size M is made explicit. It is to make the statement of the following results where the independence of the parameter c from the sample enters. Had we considered the intervals $[-c, c]$ to be sample dependent, the function ϕ defined in (16) would be sample dependent and the arguments that follow would not be true. With these notations another application of the chain rule leads to the proof of the following simple lemma.

Lemma 2. *With the notations introduces above, up to terms that are $o(\delta\boldsymbol{\mu})$, using the*

computation in (17)

$$f_M(x) - f_e(x) = \sum_{i,j=1}^{K} \left(\mu(\alpha_i) - e^{-x\alpha_i}\right) f_e(x) D_{i,j} \delta\mu_j. \tag{18}$$

Here K is the number of moments used to determine the density and $\delta\mu_j = \mu_M(\alpha_j) - \mu_e(\alpha_j)$.

This result makes explicit the deviation of f_M from its exact value f_e, up to first order in the $\delta\mu$. Notice if we integrate both sides of (18) with respect to x we get 0 on both sides. And if we multiply both sides both $e^{-x\alpha_k}$, both sides integrate to $\delta\mu_k$.

If we think of the f_M as random realizations of a density, they happen to be vectors in a convex set in $L_1(dx)$. The values $f_M(x)$ at each point can be thought as the components of those vectors, and results of the type of the central limit theorem for such values bear out in the simulations that we carried out.

But these results are of not much use for banking practice because neither the exact moments $\mu_e(\alpha)$ are known, nor is the maxentropic density f_e that they determine. But, a potentially useful result goes along the following lines. Suppose that a bank has a relatively large number of agencies, call it N_A say, and that each of them has been collecting data for the last M years. If all risks befalling these agencies can be considered identically distributed, the computations in Lemma 2, lead to an interesting consequence. To establish it, we need to introduce extra labeling. Denote by $\mu_{M,m}(\alpha_i)$ the $i-$th moment computed from the $M-$year data at branch $m = 1, ..., N_A$ of the bank. Denote as well, by $f_{M,m}(x)$ the density reconstructed from the data from branch m of the bank. We have

Lemma 3. *Set*

$$\hat{f}_M = \frac{1}{N_A} \sum_{m=1}^{N_A} f_{M,m}(x) \tag{19}$$

Then we have, $\hat{f}_M = f_e(x) + o(\delta\boldsymbol{\mu})$.

Proof. Let us rewrite (18) in the proposed notation:

$$f_{M,m}(x) - f_e(x) = \sum_{i,j=1}^{K} \left(\mu_e(\alpha_i) - e^{-x\alpha_i}\right) f_e(x) D_{i,j} \left(\mu_{M,m}(\alpha_j) - \mu_e(\alpha_j)\right).$$

Summing over m we obtain

$$\hat{f}_M(x) - f_e(x) = \sum_{i,j=1}^{K} \left(\mu_e(\alpha_i) - e^{-x\alpha_i}\right) f_e(x) D_{i,j} \left(\mu_e(\alpha_j) - \mu_e(\alpha_j)\right) = 0$$

up to terms of order $o(\delta\boldsymbol{\lambda})$, which concludes the proof of this intuitive result. □

As measure of the variability we shall estimate the L_1 difference between the approximate reconstruction f_M for each sample and the true f_e density. For this the following results come in handy. An interesting way to compare two densities f and g (on $[0,\infty)$ say) is in terms of their Kullback's divergence (or relative entropy) defined by

$$K(f,g) = \int f(x) \ln\left(\frac{f(x)}{g(x)}\right) dx. \tag{20}$$

Two properties of (20) which are important for us here are collected under

Proposition 1. *Consider two densities f and g on $[0,\infty)$ and let the Kullback divergence among them be defined by (20). Then*
(a) $K(f,g) \geq 0$ and $K(f,g) = 0 \Leftrightarrow f = g a.e$
(b) (Kullback's inequality) $\frac{1}{4}\left(\|f-g\|_1\right)^2 \leq K(f,g)$.

Part (a) is an easy consequence of Jensen's inequality and part (b) is a guided exercise in Kullback(1959).

Now, with the notations introduced at the beginning of this section consider to begin with $K(f_e, f_M)$. Using the representation provided in (7), definition (20) and (9), it takes a simple computation to verify that,

$$K(f_e, f_M) = \ln Z(\boldsymbol{\lambda}_M^*) + \langle \boldsymbol{\lambda}_M^*, \boldsymbol{\mu}_e \rangle - \ln Z(\boldsymbol{\lambda}_e^*) - \langle \boldsymbol{\lambda}_e^*, \boldsymbol{\mu}_e \rangle \tag{21}$$

Combining this, Proposition 1 and Lemma 1, we obtain

Proposition 2. *With the notations introduced above we have*

$$\|f_M - f_e\|_1 \leq (4K(f_e, f_M))^{1/2} \to 0 \ \ as \ \ M \to \infty. \tag{22}$$

Observe that a similar result would have been obtained had we considered $K(f_M, f_e)$. Of interest to us is that (22) allows us to estimate the sample size variability of the reconstruction for a given sample.

Actually, we can combine Lemma 1, (21) and (2) to further relate the deviations $\|f_M - f_e\|_1$ to the deviations of $\boldsymbol{\mu}_M$ from $\boldsymbol{\mu}_e$.

Proposition 3. *With the notations introduced above we have*

$$\|f_M - f_e\|_1 \leq (2\langle \delta\boldsymbol{\mu}, \mathbf{D}\delta\boldsymbol{\mu}\rangle)^{1/2} \to 0 \ \ as \ \ M \to \infty. \tag{23}$$

Here $\delta\boldsymbol{\mu}$ *and* $\mathbf{D}$ *are as above.*

The proof is simple, and drops out from the fact that up to terms of order $o((\delta\boldsymbol{\mu})^2)$, $K(f_e, f_M) = \langle \delta\boldsymbol{\mu}, \mathbf{D}\delta\boldsymbol{\mu}\rangle$.

4 Numerical Results

This section is organized as follows. We begin by describing the sample generation process. To the large sample we apply the standard maximum entropy procedure and reconstruct a density from $K = 8$ values of its Laplace transform. Since the full sample is very large, the effect of the sample variability is small. We call that density the *true* density and use it as reference to test the behavior of the densities reconstructed from smaller samples. After obtaining the "true" density, to analyze the sample variability we first plot the densities obtained for the smaller samples along with the true density, and then we present the different measures of sample variability. After that we carry out an analysis of the sample variability of the regulatory capital, computed as the VaR and TVaR of each reconstructed density.

4.1 The sample generation process

As data set we considered a loss obtained as aggregation of 8 different compound variables to model the risk coming from different risk sources. The frequencies of the compound variables were chosen as follows: Two Poisson random variables with parameters $\ell_1 = 80$ and $\ell_2 = 60$, four binomial frequencies of parameters $(n_3 = 70, p_3 = 0.5)$, $(n_4 = 62, p_4 = 0.5)$, $(n_5 = 50, p_5 = 0.5)$ and $(n_6 = 76, p_6 = 0.5)$. We considered as well two negative binomial frequencies with parameters $(n_7 = 80, p_7 = 0.3)$ and $(n_8 = 90, p_8 = 0.8)$.

The corresponding individual losses were chosen as independent identically distributed random variables listed as follows. By X_k we denote the individual loss to be compounded according to the $k-$th frequency in the previous list. X_1 and X_4 Champerowne densities with parameters $(\alpha_1 = 20, M_1 = 85, c_1 = 15)$ and $(\alpha_4 = 10, M_4 = 125, c_4 = 45)$. X_2 was chosen as a Lognormal with parameters $(-0.01, 2)$. Then X_3 and X_8 were chosen to have fat tails and to follow a Pareto distribution with (shape, scale) parameters given by $(10, 85)$ and $(5.5, 5550)$. Next in the list are X_5 and X_6, chosen to be Gamma distributed with (shape, scale) parameters given respectively by $(4500, 15)$ and $(900, 35)$. And finally, X_7 was chosen to be a Weibull of type $(200, 50)$.

The size of the total sample was $M = 5000$, from which different subsamples were chosen to study the sample variability. We considered $K = 8$ moments determined by $\alpha_k = 1.5/k,\ k = 1, ..., 8$.

As the size of the resulting aggregate losses have order of magnitude 10^4, and since the maxentropic methods use the Laplace transforms of the aggregate loss as starting point, we scale the data prior to the application of the procedure, and reverse the scaling to plot the resulting densities. The simplest scaling consists of dividing by 10^4 as we did. Another possibility is to introduce a scaled variable defined by $s - \min(s))/\max(s) - \min(s)$, where of course, $\max(s)$ and $\min(s)$ stand for the maximum and minimum values realized in the sample.

4.2 The true maxentropic density

In Figure (1) we present the result of applying both maxentropic methods, the SME and the SMEE to the full data set. For comparison, the histogram of the data used to compute the moments is also plotted. As intuitively expected, since the data set is very large, the maxentropic reconstructions agree quite well with the simulated (empirical) data.

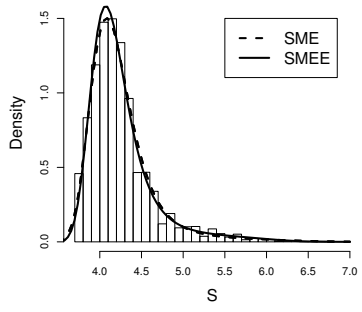


Figure 1: Histogram of the (scaled) total losses and maxentropic density

To determine how well the densities obtained approximate the data, we consider two measures that are sample dependent, which make comparisons with empirical data simple because they are bin independent, whereas standard distances like L_1 and L_2 are bin dependent. They measures that we consider are the mean absolute error (MAE) and the square root mean error (RMSE). See Hyndman and Koehler (2006) for more about such measures. These are computed as follows: If $\hat{F}_s$ denotes the empirical distribution function and F_e denote the "exact" (reconstructed) distribution function, the measures of reconstruction error is to be estimated by

$$MAE = \frac{1}{N}\sum_{n=1}^{N} |\hat{F}(x_n) - F_e(x_n)|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N} \left(\hat{F}(x_n) - F_e(x_n)\right)^2}$$

where here N is generic and stands for the size of the sample and the $\{x_n,\ n = 1, ..., N\}$ are the sample points.

Approach	**MAE**	**RMSE**
SMEE	0.005928	0.006836
SME	0.006395	0.009399

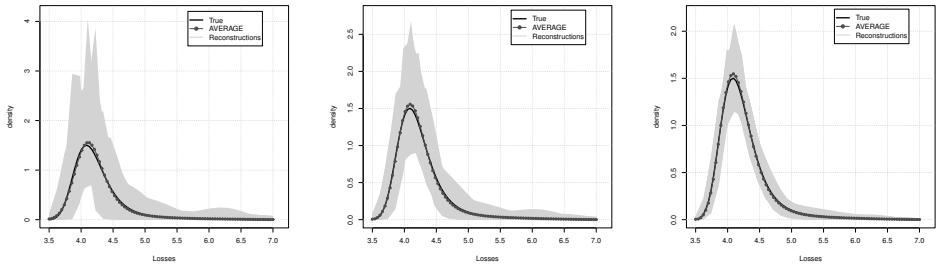
Table 1: MAE and RMSE for a sample size of 5000

In Table (1) we show the results of the computation of MAE and RSME to the full data set and how the well the maxentropic densities fit the empirical data. Clearly, both maxentropic procedures yield quite good reconstructions.
At this point we add that the reconstruction error in the maxentropic procedure, that is, the norm of the gradient of the dual entropy (8), was less than 10^{-6} in all reconstructions. This value measures the quality of the reconstructions, and is used as criterion for stopping the iterations in the process of minimization of the dual entropy.

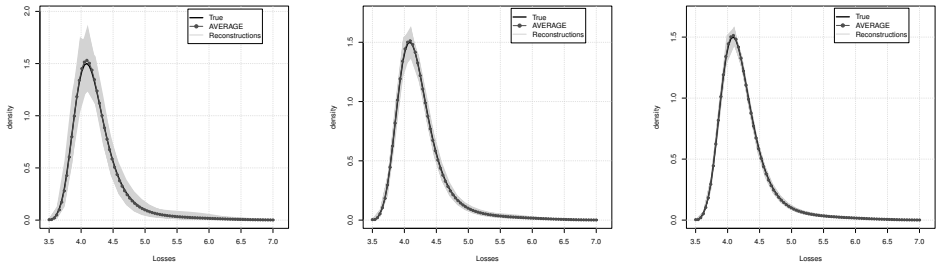
4.3 The sample variability of the densities

In this section we study how does the maxentropic density approximate the true density. To each sample, of each size, we applied both the SME and the SMEE, to obtain a density. Let us describe separately the results obtained by each method.

The results displayed in the panels of Figure (2) should be understood as follows. To study the sample variability, we applied the maxentropic procedures to 200 samples of different sizes. For example in the first panel, the gray shadow consists of the plot of the 200 densities produced as output of the maxentropic procedure estimated form the moments produced by 200 samples of size $M = 10$. In each panel we plot the average density $\hat{f}_M$ as well as the true density.
Also, for each panel we average the moments and use them as input for the maximum entropy method, and we display the density obtained from such average moments along with the density obtained from the exact moments. That the last two mentioned coincide is no surprise according to the results in the previous section. As we move from panel to



(a) M=10 (b) M=20 (c) M=50



(d) M=100 (e) M=500 (f) M=1000

Figure 2: SME reconstructions with different sample sizes

panel, the amount of data used to compute the moments increases and they are closer to their true values. This is the law of large numbers in action, thus the spread in the indeterminacy of the true density decreases, as clearly seen by the shrinking size of the gray shadow.

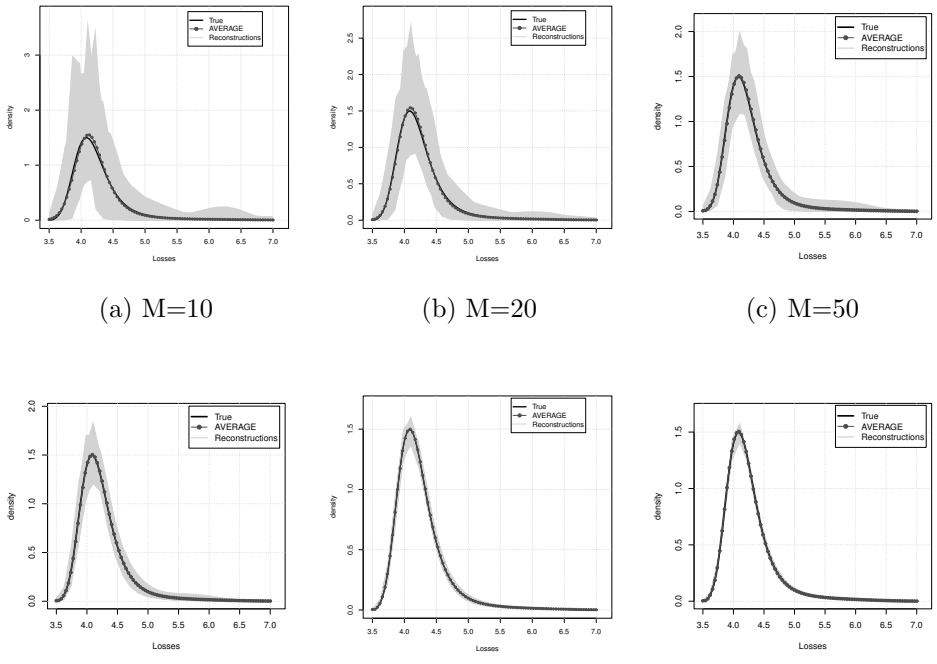
The entries of Table (2) are to be understood as follows. The first two columns describe the size of the sample and the error measure being listed. For the next three columns, a list of the MAE and RSME errors computed as described in the previous section for the 200 samples of each size, and the percentiles mentioned were computed. In column number six we list the area of the gray shadow, computed the obvious way. To obtain the last column, we averaged the moments over the 200 samples and obtained the corresponding maxentropic density, and computed the average discrepancies between that density and the densities in the sample.

To obtain the next figure, we repeat the same procedure as before, except that this time the densities were obtained using the SMEE method. This is to test whether specifying a

Size(M)	**Error Meas.**	1st Qu.	Mean	3rd Qu.	Area	AVERAGE
10	MAE	0.0500	0.0880	0.1150	2.625	0.0092
	RMSE	0.0590	0.1010	0.1310		0.0120
20	MAE	0.0359	0.0619	0.0811	1.523	0.0089
	RMSE	0.0400	0.0702	0.0915		0.0116
50	MAE	0.0203	0.0377	0.0510	0.9545	0.0082
	RMSE	0.0237	0.0429	0.0575		0.0106
100	MAE	0.0144	0.0266	0.0343	0.696	0.0053
	RMSE	0.0169	0.0304	0.0393		0.0066
200	MAE	0.0126	0.0194	0.0247	0.5375	0.0053
	RMSE	0.0148	0.0225	0.0285		0.0067
500	MAE	0.0081	0.0128	0.0163	0.3258	0.0055
	RMSE	0.0101	0.0153	0.0194		0.0076
1000	MAE	0.0067	0.0093	0.0108	0.2033	0.0054
	RMSE	0.0087	0.0115	0.0132		0.0078

Table 2: MAE & RMSE of SME for different sample sizes

measurement error improves the reconstructions. In the six panels of Figure (3) we did as described above.



(a) M=10 (b) M=20 (c) M=50

(d) M=100 (e) M=5000 (f) M=10000

Figure 3: SMEE reconstructions with different sample sizes

Similarly to the previous case, as we increased the amount of data, the reconstructions improved, but as shown in Table (3) the improvement (relative to the results obtained using (SME)) as the sample size increased is small. The entries in Table (3) were produced and have the same meaning as those in Table (2).

In Table (4) we summarize the variability of the reconstructions depending on the size of the sample in a slightly different way. There we show how the mean MAE, and the area of the gray shadow, vary with the sample size. The improvement of the reconstructions as the sample size increases is apparent.

Size(M)	**Error Meas.**	1st Qu.	Mean	3rd Qu.	Area	AVERAGE
10	MAE	0.0419	0.0690	0.0898	2.619	0.0069
	RMSE	0.0514	0.0784	0.1030		0.0110
20	MAE	0.0360	0.0620	0.0820	1.759	0.0066
	RMSE	0.0420	0.0705	0.0918		0.0109
50	MAE	0.0198	0.0378	0.0500	1.044	0.0065
	RMSE	0.0240	0.0430	0.0582		0.0102
100	MAE	0.0142	0.0267	0.0353	0.690	0.0060
	RMSE	0.0168	0.0306	0.0398		0.0082
200	MAE	0.0125	0.0196	0.0247	0.552	0.0063
	RMSE	0.0147	0.0229	0.0270		0.0072
500	MAE	0.0083	0.0131	0.0165	0.294	0.0058
	RMSE	0.0101	0.0156	0.0199		0.0083
1000	MAE	0.0068	0.0093	0.0109	0.200	0.0057
	RMSE	0.0082	0.0114	0.0133		0.0082

Table 3: MAE & RMSE of SMEE results for different sample sizes

4.4 Computation of the regulatory capital

This section is devoted to the computation of the two most used risk measures, namely, the VaR and the $TVaR$, which are used to determine the regulatory capital. We explained how to compute these risk measures as in Gomes-Gonçalves et al. (2015). We present several possible tables. The idea is that they should provide us with intuition about the possible variability of quantities essential in risk measurement.

In Table (5) we present a comparison of the values of the VaR and the $TVaR$ computed from an empirical sample of size 5000, and the VaR and the $TVaR$ computed using the SME and SMEE densities at the 95% and the 99% confidence levels. In the table, γ stands for the confidence level.

The sample considered to build Table (5) is large for operational risk purposes. Data sets corresponding to large disasters may comprise even smaller data sets. Therefore computations

	10		100		1000	
	Mean (MAE)	Area	Mean (MAE)	Area	Mean (MAE)	Area
SME	0.0880	2.625	0.0266	0.696	0.0092	0.2033
SMEE	0.0690	2.619	0.0267	0.690	0.0093	0.2000

Table 4: Summary of results

	γ	**Empirical**	**SME**	**SMEE**
VaR	0.950	5.05	4.935	5.004
	0.990	5.72	5.755	5.772
TVaR	0.950	5.45	5.443	5.461
	0.990	6.05	6.0207	6.014

Table 5: Comparison of VaR and TVaR at 95% and 99% for a unique sample of size 5000

like that leading to this table will have to take into account sample variability.

In Table (6) we consider two measures of variability of the VaR and the $TVaR$ computed from the maxentropic densities obtained for 200 samples of the indicated sizes. In each cell we present the mean and the variance (within parentheses) of each risk measure, for each sample size. And we do it at the 95% and the 99% confidence levels.

To finish, consider Table (7) in which we compute the "true" VaR and the 'true" $TVaR$ of small samples. Recall that the "true" density of a small sample of size M was obtained averaging the densities of the 200 samples of size M, and was shown to become closer to the true density of the total loss as M increased. We see that the same happens to the VaR and the $TVaR$, as described in Table (7).

5 Concluding remarks

We have seen that the maxentropic techniques are quite suitable to determine a density from a few values of its Laplace transform by recasting the problem as a fractional moment problem.

Size (M)	VaR(95%)		TVaR(95%)		VaR(99%)		TVaR(99%)	
	SME	SMEE	SME	SMEE	SME	SMEE	SME	SMEE
10	4.96	4.87	5.30	5.156	4.331	5.283	4.328	5.634
	(0.4530)	(0.6740)	(0.5678)	(0.7984)	(1.083)	(0.8597)	(0.995)	(1.3138)
20	4.96	4.91	5.30	5.282	4.518	5.502	4.633	5.818
	(0.4536)	(0.5200)	(0.5678)	(0.6515)	(1.004)	(0.7537)	(1.004)	(0.8053)
50	4.97	4.925	5.39	5.386	5.003	5.688	5.162	6.058
	(0.2902)	(0.3254)	(0.382)	(0.4286)	(0.988)	(0.5463)	(1.069)	(0.5617)
100	4.96	4.931	5.44	5.43	5.457	5.779	5.694	6.016
	(0.1994)	(0.2258)	(0.251)	(0.2794)	(0.705)	(0.3626)	(0.768)	(0.3476)
200	4.97	5.013	5.45	5.46	5.624	5.766	5.871	6.016
	(0.1537)	(0.1828)	(0.1902)	(0.1995)	(0.395)	(0.2343)	(0.417)	(0.2498)
500	4.95	4.93	5.45	5.45	5.708	5.822	5.972	6.017
	(0.08963)	(0.1038)	(0.1136)	(0.1249)	(0.153)	(0.1539)	(0.159)	(0.1385)
1000	4.95	4.95	5.45	5.45	5.729	5.828	5.977	6.064
	(0.05815)	(0.07033)	(0.07088)	(0.07765)	(0.109)	(0.1065)	(0.107)	(0.09838)

Table 6: Mean and standard deviation of the VaR and TVaR for 200 samples of different sizes

Not only that, the maxentropic procedures provide for more freedom of action: on the one hand the can help us to cope with a case in which the traditional methods are not that good when there is not enough data.

Maxentropic methods are useful to estimate loss densities when we are aggregating total losses form different sources. This is because the input to the maxentropic procedure are the empirical moments computed as in (2), which depend only on the total aggregate data and not on its statistical nature. Clearly, the equality of the reconstruction will depend on the size of the data set.

The maxentropic method, combined with a simulation or re-sampling procedure provides us with reasonable measures of the variability of quantities of interest in risk measurement. For example, if we had 200 data points, we could re-sample a large number of times, and the average over the densities determined by each sub-sample to obtain a rather accurate estimate of relevant risk measures.

Acknowledgment We want to thank the reviewers for the care with which they read our manuscript and for the suggestions that they made for the improvement of the presentation.

Size(M)	VaR(95%)		TVaR(95%)		VaR(99%)		TVaR(99%)	
	SME	SMEE	SME	SMEE	SME	SMEE	SME	SMEE
10	4.81	4.82	5.32	5.35	4.54	5.64	5.97	5.97
100	4.95	4.92	5.42	5.38	5.67	5.72	5.99	6.05
500	4.95	4.95	5.49	5.45	5.78	5.79	6.09	6.09
1000	5.01	4.95	5.50	5.45	5.78	5.79	6.05	6.05

Table 7: VaR and TVaR for the average of the Maxentropic densities for different sample sizes.

References

[1] Aue, F. and Kalkbrenner, M. (2006) *LDA at work: Deutsche Bank's approach to quantifying operational risk*, Jour. Operat. Risk, **1**, 49-93.

[2] Brockmann, M. and Kalkbrenner, M. (2010) *On the aggregation of risk*, Jour. Operat. Risk, **12**, 45-68.

[3] Cherny, A. and Maslov, V. (2003) *On minimization and maximization of entropy functionals in various disciplines*, Theory of Probab. and its Applic, **2003**, **17**, 447-464.

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

[5] Embrechts, P. and Frei, M. (2007) *Panjer recursion versus FFT for compound distributions*, Math. Meth. of Operational Res., **69**, 497-508.

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

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

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

[9] 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.

[10] Gzyl, H. (2015) *Sample dependence in the maximum entropy solution to the generalized moment problem*, Preprint, arXiv:1510.04062 [math.ST].

[11] Hyndman, R.J. and Koehler, A.B. (2006) *Another look at measures of forecast accuracy*, Int. Jour. Forecasting, **22**, 679-688.

[12] den Iseger, P.W, Smith, W.A.J. and Dekker, R. (1997) *Computing compound distributions faster*, Insurance: Mathematics and Economics, **20**, 23-34.

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

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

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

[16] Shevchenko, P.V. *Modeling Operational Risk*, Springer Verlag, Berlin, (2011).

[17] Temnov, G. and Warnung, R. (2008) A comparison of risk aggregation methods for operational risk, Jour. Operat. Risk, **3**, 3-23.