UNIVERSITÀ DEGLI STUDI DI TRENTO

DIPARTIMENTO DI ECONOMIA

A note on maximum likelihood estimation of a Pareto mixture

Marco Bee
Roberto Benedetti
Giuseppe Espa

Discussion Paper No. 3, 2009

The Discussion Paper series provides a means for circulating preliminary research results by staff of or visitors to the Department. Its purpose is to stimulate discussion prior to the publication of papers.

Requests for copies of Discussion Papers and address changes should be sent to:

Dott. Luciano Andreozzi
E.mail luciano.andreozzi@economia.unitn.it
Dipartimento di Economia
Università degli Studi di Trento
Via Inama 5
38100 TRENTO ITALIA

A note on maximum likelihood estimation of a Pareto mixture

Marco Bee

Department of Economics, University of Trento

Roberto Benedetti

Department of Business, Statistical, Technological and Environmental Sciences

University "G. d'Annunzio" of Chieti-Pescara

Giuseppe Espa

Department of Economics, University of Trento

Abstract. In this paper we study Maximum Likelihood Estimation of the parameters of a Pareto mixture. Application of standard techniques to a mixture of Pareto is problematic. For this reason we develop two alternative algorithms. The first one is the Simulated Annealing and the second one is based on Cross-Entropy minimization. The Pareto distribution is a commonly used model for heavy-tailed data. It is a two-parameter distribution whose shape parameter determines the degree of heaviness of the tail, so that it can be adapted to data with different features. This work is motivated by an application in the operational risk measurement field: we fit a Pareto mixture to operational losses recorded by a bank in two different business lines. Losses below an unknown threshold are discarded, so that the observed data are truncated. The thresholds used in the two business lines are unknown. Thus, under the assumption that each population follows a Pareto distribution, the appropriate model is a mixture of Pareto where all the parameters have to be estimated.

1 Introduction

Parameter estimation of the parameters of a finite mixture distribution is a well-known topic in the statistical literature. The starting point is the pioneering contribution by Pearson (1894), who tried to use the method of moments for the estimation of the parameters of a two-population normal mixture. Currently, two main approaches play a key role in this setup: the likelihood-based and the Bayesian approach. In the present paper we will focus on the first one. It is well known that Maximum Likelihood Estimation (MLE) of the parameters of a finite mixture is rather difficult, because the likelihood equations are highly nonlinear. A major step further was the introduction of

the EM algorithm (Dempster *et al.* 1977). This algorithm is mainly designed for the maximization of likelihood functions with missing data, but can often be applied in non-missing data frameworks.

The Pareto distribution was introduced by the Italian economist Pareto as a model for income distribution, and has subsequently been used mostly as a model for heavy-tailed data, in particular in hydrology, insurance and finance. In all of these fields, a small number of very large observations is of crucial importance for the computation of some quantity of interest. For example, the estimation of a large quantile (say 99% or more) depends heavily on few observations in the right tail. In this case it is extremely important to choose a probabilistic model that accounts for these observations. The Pareto distribution is a two-parameter distribution whose shape parameter determines the degree of heaviness of the tail, so that it can be adapted to data with different features.

The density of a k-population Pareto mixture is the convex combination of k Pareto densities. When $k = 2$, it is possible to build mixtures with (i) the same location and different shapes or (ii) the same shape and different locations or (iii) different shapes and locations. The first setup is easier to deal with, as estimation can be performed by means of the EM algorithm.

The mixture of two Pareto distributions obtained in case (i), namely under the restriction $a_1 = a_2$, sometimes called double Pareto distribution, has recently been employed in various fields of application. In particular, it has been proposed for the statistical analysis of the Chinese airport network (Li and Cai, 2004) and as a model for human settlements, income, and size distributions (Reed, 2002, 2003; Reed and Jorgensen, 2004). On the theoretical side, Nadarajah (2004) derived the Fisher information matrix.

As will be shown below, when $a_1 \neq a_2$ the EM algorithm breaks down. Thus, in this paper we study MLE of a general Pareto mixture, namely a Pareto mixture corresponding to (iii). From the methodological point of view, the paper originates from the remark that the EM algorithm cannot be applied if the largest of the two location parameters is not known, so that the most common numerical procedure used for MLE in a mixture setup has to be ruled out. For this reason we develop two alternative algorithms; the first one is the Simulated Annealing and the second one is based on Cross-Entropy minimization. From an applied point of view, a general Pareto mixture can be used as a model for heavy-tailed data sampled from k different populations: in particular, we will present an example based on loss data from the field of operational risk.

The rest of the paper is organized as follows. Section 2 outlines MLE of a Pareto mixture and gives some details about the failure of the EM algorithm. Sections 3 and 4 respectively show how to implement the Simulated Annealing and the Minimum Cross-Entropy algorithms. Section 5 presents a detailed simulation study of the properties of the two approaches and an example based on real data. Finally, Section 6 discusses the results and reviews the problems open for future research.

2 Maximum Likelihood Estimation of a Pareto mixture

The density of a Pareto r.v. Y is given by

$$f(y; a, c) = \frac{ca^c}{y^{c+1}} \mathbf{1}_{\{y \geq a\}}, \tag{1}$$

where $a, c \in \mathbb{R}^+$ and $\mathbf{1}_{\{y \geq a\}}$ is the indicator function of the set $\{y \geq a\}$. Given a random sample $y_1, \ldots, y_n$ from the Pareto distribution with density (1), it is well known that the MLEs of the parameters are

$$\hat{a} = \min_{1 \leq i \leq n} y_i, \qquad \hat{c} = \frac{n}{\sum_{i=1}^n y_i \log(y_i/\hat{a})}.$$

It is quite clear that the partial derivative $\partial l(a, c; \boldsymbol{y})/\partial a$ of the log-likelihood function does not exist in $\hat{a}$, where the log-likelihood is not continuous. Although this feature does not keep us from finding the MLE of the parameters of a single Pareto r.v., this is going to be a major problem in the mixture setup.

2.1 The EM algorithm in the Pareto mixture case

The EM algorithm is the preferred method for MLE of finite mixture distributions. However, the purpose of the algorithm is more general, as it is an iterative method for MLE with missing data. In order to fix the notation we summarize here the main features of the algorithm referring the interested reader to McLachlan and Krishnan (1996) for details.

Let $\boldsymbol{Y}$ be the p-dimensional random vector of the observed data, $g(\boldsymbol{y}; \boldsymbol{\theta})$ its density function and $l(\boldsymbol{\theta})$ the corresponding log-likelihood function, where $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is the parameter vector. Let $\boldsymbol{Z}$ be the (hypothetical) vector of missing data, which are unobservable, but whose knowledge would allow for a straightforward application of the maximum likelihood method. Finally, let $\boldsymbol{X} = (\boldsymbol{Y}, \boldsymbol{Z})$ be the complete-data vector, whose density and log-likelihood functions will be denoted respectively by $g_c(\boldsymbol{x}; \boldsymbol{\theta})$ and $l_c(\boldsymbol{\theta})$.

The first step of the algorithm (called E-step, where E stands for Expectation) consists in computing the conditional expectation of the complete log-likelihood $l_c(\boldsymbol{\theta})$, given the current value of $\boldsymbol{\theta}$ and the observed data $\boldsymbol{y}$. The second step (M-step, where M stands for Maximization) consists in maximizing, with respect to $\boldsymbol{\theta}$, the conditional expectation of the complete log-likelihood computed in the E-step.

Formally, the algorithm can be described as follows: let $\boldsymbol{\theta}^{(0)}$ be the initial value of the parameters vector; the E-step computes

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{(0)}) = \mathrm{E}_{\theta^{(0)}}\{l_c(\boldsymbol{\theta}) | \boldsymbol{y}\}.$$

Then one maximizes $Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(0)})$ with respect to $\boldsymbol{\theta}$, i.e. one chooses $\boldsymbol{\theta}^{(1)}$ such that

$$Q(\boldsymbol{\theta}^{(1)};\boldsymbol{\theta}^{(0)}) \geq Q(\boldsymbol{\theta};\boldsymbol{\theta}^{(0)}), \qquad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

To obtain the estimators, the two steps are iterated until some convergence criterion is met. The convergence of the algorithm is relatively slow, but typically does not depend on the choice of the starting value; the estimators enjoy the usual properties of MLEs, in particular they are consistent and asymptotically normal.

One of the reasons why the algorithm was successful is the fact that it is also suitable for the solution of problems where no data are missing. This is the case of MLE of the parameters of a finite mixture distribution. Let $\boldsymbol{Z} = (Z_1, \ldots, Z_k)'$ be a multinomial random vector, $Z_j = 0$ or 1 for all $j = 1, \ldots, k$, $\sum_{j=1}^{k} Z_j = 1$, and $\pi_j = P(Z_j = 1)$, where $\sum_{j=1}^{k} \pi_j = 1$. Let X be a random variable such that, conditionally on $Z_j = 1$, $X \sim f_j$, where f_j is a pdf. Then the joint pdf of $\boldsymbol{Z}$ and X is

$$f_{\boldsymbol{Z},X}(\boldsymbol{z}, x) = \prod_{j=1}^{k} [\pi_j f_j(x)]^{z_j}, \quad x \in \mathbb{R}, \quad \text{all } z_j = 0 \text{ or } 1, \quad \sum_{j=1}^{k} z_j = 1.$$

The marginal density of X is the mixture pdf:

$$f_X(x) = \sum_{j=1}^{k} \pi_j f_j(x). \tag{2}$$

The conditional distribution of Z_j given X, is given by

$$\tau_{jx} = P[Z_j = 1 | X = x] = \frac{\pi_j f_j(x)}{f_X(x)} = \mathrm{E}[Z_j | X = x],$$

and $P[Z_j = 0 | X = x] = 1 - \tau_{jx}$. The τ_{jx} are commonly referred to as posterior probabilities.

Let's now specialize (2) to the case where $k = 2$ and f_j is given by (1), namely the density of a Pareto distribution:

$$f(x; \pi, a_1, c_1, a_2, c_2) = \pi \frac{c_1 a_1^{c_1}}{x^{c_1+1}} \mathbf{1}_{\{x \geq a_1\}} + (1 - \pi) \frac{c_2 a_2^{c_2}}{x^{c_2+1}} \mathbf{1}_{\{x \geq a_2\}},$$

with $a_2 > a_1 > 0$, $c_1, c_2 > 0$. A graphical representation is given in figure 1.

The observed and complete log-likelihood functions are respectively given by

$$l(\pi, a_1, a_2, c_1, c_2; \boldsymbol{y}) = \sum_{i=1}^{n} \log \left[\pi \frac{c_1 a_1^{c_1}}{y_i^{c_1+1}} \mathbf{1}_{\{y_i \geq a_1\}} + (1 - \pi) \frac{c_2 a_2^{c_2}}{y_i^{c_2+1}} \mathbf{1}_{\{y_i \geq a_2\}} \right] \tag{3}$$

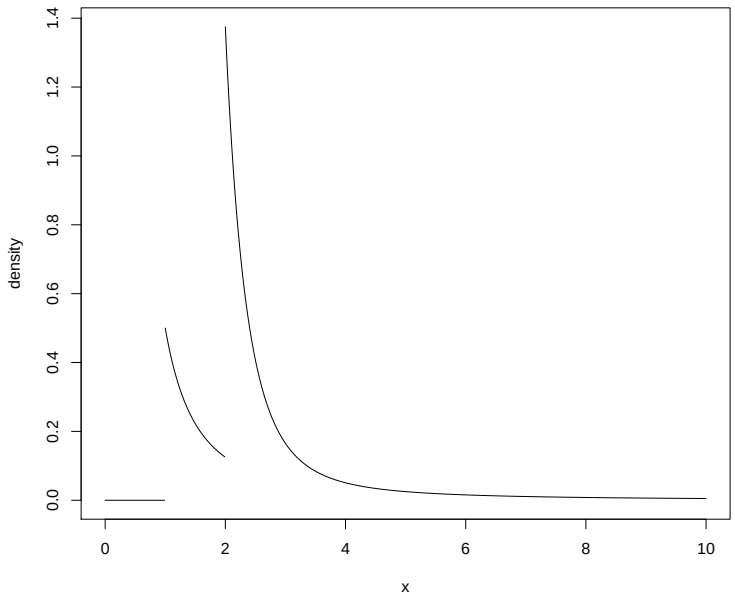

Figure 1: Density of a Pareto mixture ($\pi = 0.5,\ a_1 = 1,\ c_1 = 1,\ a_2 = 2,\ c_2 = 5$).

and

$$l_c(\pi, a_1, a_2, c_1, c_2; \boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{n} \log \left\{ \left[\pi \frac{c_1 a_1^{c_1}}{y_i^{c_1+1}} \mathbf{1}_{\{y_i \geq a_1\}} \right]^{z_{1i}} \left[(1-\pi) \frac{c_2 a_2^{c_2}}{y_i^{c_2+1}} \mathbf{1}_{\{y_i \geq a_2\}} \right]^{1-z_{1i}} \right\} =$$

$$= \sum_{i=1}^{n} z_{1i} \log(\pi) + \sum_{i=1}^{n} z_{1i} [\log(c_1) + c_1 \log(a_1) - (c_1+1)\log(y_i) + \log(\mathbf{1}_{\{y_i \geq a_1\}})] +$$

$$+ \sum_{i=1}^{n} (1 - z_{1i}) \log(1-\pi) + \sum_{i=1}^{n} (1 - z_{1i}) [\log(c_2) + c_2 \log(a_2) - (c_2+1)\log(y_i) + \log(\mathbf{1}_{\{y_i \geq a_2\}})].$$

Suppose now that c_1 and c_2 are known, so that we are only interested in the estimation of π, a_1 and a_2; then we have to maximize the function

$$\tilde{l}_c(\pi, a_1, a_2; \boldsymbol{y}, \boldsymbol{z}) = \sum_{i=1}^{n} z_{1i} \log(\pi) + \sum_{i=1}^{n} \{ z_{1i} [c_1 \log(a_1) + \log(\mathbf{1}_{\{y_i \geq a_1\}})] +$$

$$+ \sum_{i=1}^{n} (1 - z_{1i}) \log(1-\pi) + \sum_{i=1}^{n} (1 - z_{1i}) [c_2 \log(a_2) + \log(\mathbf{1}_{\{y_i \geq a_2\}})] \}.$$

As often happens when the support of the r.v. depends on the parameters, differentiation is not going to help us here, because the likelihood function is non-differentiable at the maximum. With

complete data, MLEs are given by

$$\hat{a}_1 = \min_{z_{1i}=1} y_i, \qquad \hat{a}_2 = \min_{z_{1i}=0} y_i.$$

What is the conditional expectation of the complete log-likelihood function? By linearity, we have

$$\mathrm{E}[\tilde{l}_c(\pi, a_1, a_2; \boldsymbol{y}, \boldsymbol{z})] = \sum_{i=1}^{n} \{\tau_{1i}[c_1 \log(a_1) + \log(\mathbf{1}_{\{y_i \geq a_1\}})] + \tau_{2i}[c_2 \log(a_2) + \log(\mathbf{1}_{\{y_i \geq a_2\}})]\}.$$

The problem to be solved is

$$\max_{\pi, a_1, a_2} \sum_{i=1}^{n} \{\tau_{1i}[c_1 \log(a_1) + \log(\mathbf{1}_{\{y_i \geq a_1\}})] + \tau_{2i}[c_2 \log(a_2) + \log(\mathbf{1}_{\{y_i \geq a_2\}})]\}.$$

The solution for a_2 is clearly given by $\hat{a}_2 = \min y_i$; notice that $\mathrm{E}[\tilde{l}_c(a_2; \boldsymbol{y}, \boldsymbol{z})] = -\infty$ for any $a_2 > \min y_i$. Unfortunately, this implies that $\hat{a}_2$ does not depend on the posterior probability τ_{2j}, so that it is not updated as the iterations increase; in other words, $\hat{a}_2$ is determined with probability one by the initial value $a_2^{(0)}$.

This happens because the maximum of $Q(\pi, a_1, a_2; \pi^{(t)}, a_1^{(t)}, a_2^{(t)}) = \mathrm{E}[\tilde{l}_c(\pi, a_1, a_2; \pi^{(t)}, a_1^{(t)}, a_2^{(t)}]$ does occur at a point where the first derivative with respect to a_2 is not equal to zero (more precisely, where the function is not differentiable). As a consequence, although the main convergence theorem for the EM algorithm (Little and Rubin 1987, Theorem 7.1) remains valid, Theorem 7.2 of Little and Rubin (1987, pag. 136), which would guarantee that the sequence $a_2^{(t)}$ converges to a stationary point, does not apply.

For this reason one has to resort to alternative algorithms, and in particular to techniques which do not employ derivatives. Notice indeed that standard numerical methods using derivatives, as quasi-Newton methods, also fail in this case because the maximum of the log-likelihood function occurs at a point where, as noted above, one of the partial derivatives does not exist.

3 The Simulated Annealing Algorithm

The Simulated Annealing (SA) algorithm (Metropolis *et al.* 1953, Kirkpatrick *et al.* 1983, Geman and Geman 1984; see Casella and Robert 2004, sect. 5.2.3, for a review) is a powerful algorithm for function maximization and optimization on a continuous set. Ingrassia (1992) proposed an application to MLE of the parameters of a normal mixture and performed a detailed comparison of the EM and SA algorithms.

Let $H(\boldsymbol{\theta})$ be a real-valued function defined on a compact subset $D \subset \mathbb{R}^p$. The algorithm is based on two fundamental ideas. First, the decrease of a scale parameter T, called temperature,

guarantees a faster exploration of the surface of the function to maximize. Second, the point for the next iteration is chosen by means of a Metropolis step, i.e., it may be accepted (in the sense that the acceptance probability is larger than zero) even if it corresponds to a decrease of the objective function. The latter feature allows the algorithm to escape the "traps" of local maxima. On the theoretical side, the main result states that the probability distribution of $\boldsymbol{\theta}$ converges, as the temperature tends to zero from above, to a probability measure concentrated on the set of points of global maximum of H.

The algorithm can be either homogeneous or inhomogeneous: in the first case it is described by a sequence of homogeneous Markov chains, in the second one by a single inhomogeneous Markov chain.

3.1 An anisotropic algorithm

In this section we develop, in the setup of a Pareto mixture, the homogeneous version of the algorithm illustrated by Ingrassia (1992; see also McLachlan and Krishnan 1996, sect. 6.9.2). The steps for the implementation are as follows.

(i) Initialize the algorithm;

(ii) Determine a rule for the selection of the width of the interval where the algorithm chooses the point for the next iteration;

(iii) Choose a cooling schedule, and, in particular, an initial value of the temperature;

(iv) Determine a stopping criterion.

Steps (ii) and (iii) are by far the most important ones in the definition of an SA algorithm, so that they will be given more attention. Both the choice of the initial value of the temperature and of the step distribution should be guided by the aim of maximizing the number of accepted transitions.

The algorithm is initialized randomly. After that, the transition density of the n-th Markov chain has to be determined. Notice that the n-th Markov chain, corresponding to temperature T_n, is homogeneous, namely the transition kernel only changes across chains or, in other words, when the temperature changes. Let $\boldsymbol{\theta}_l^{(n)}$ be the parameter vector at temperature T_n in the l-th position of the chain.

A possible implementation of the algorithm consists in sampling the space near the current value in an isotropic way. In other words, the direction is randomly sampled from the distribution:

$$\boldsymbol{\theta}_{l+1}^{(n)} = \boldsymbol{\theta}_l^{(n)} + \Delta r \boldsymbol{v}_n, \tag{4}$$

where $\boldsymbol{v}_n$ is a random direction vector with $\|\boldsymbol{v}_n\| = 1$ and Δr is the fixed step size.

If the objective function has a markedly different behavior in different directions, this procedure is probably not very efficient. In this case, a better solution would consist in exploring the parameter space in an anisotropic way, according to the shape of the function in a neighborhood of the current parameter value. This strategy requires a definition of the probability distribution of the next candidate value such that its support is "as similar as possible" to the shape of the function to be maximized. To this aim, Vanderbilt and Louie (1984) propose the following way of reasoning.

Consider first the artificially simple case where the Hessian $\hat{\boldsymbol{H}} = \boldsymbol{H}(\hat{\boldsymbol{\theta}})$ in the global optimum point $\hat{\boldsymbol{\theta}}$ is known and positive definite, and the current $\boldsymbol{\theta}^{(n)}$ is in a neighborhood of $\hat{\boldsymbol{\theta}}$. Let the matrix $\boldsymbol{\Sigma}$ be the inverse of $\hat{\boldsymbol{H}}$, that is $\boldsymbol{\Sigma} = \hat{\boldsymbol{H}}^{-1}$. Under these hypotheses, the Choleski decomposition allows to find a matrix $\boldsymbol{Q}$ such that $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{Q}'$. Similarly to what has been done in (4), the next candidate value can be generated as

$$\boldsymbol{\theta}_{l+1}^{(n)} = \boldsymbol{\theta}_{l}^{(n)} + \boldsymbol{Q}\boldsymbol{u},$$

where $\boldsymbol{u}$ is an m-dimensional uniform random vector with expected value equal to $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}_m$ (equivalently, it is uniformly distributed in the hypercube $[-\sqrt{3}, \sqrt{3}]^m$). This procedure produces big steps in the directions where the objective function is "less steep" (varies slowly) and small steps in the directions along which the objective function is "more steep" (varies quickly). To see why, assume $\hat{\boldsymbol{H}}$ to be diagonal, so that $\boldsymbol{Q}$ is diagonal as well; thus the diagonal elements of $\boldsymbol{S}$ and $\boldsymbol{Q}$ are large (small) when the corresponding elements of $\hat{\boldsymbol{H}}$ are small (large), so that big (small) steps take place in the direction along which the function varies slowly (quickly).

In practice, however, $\hat{\boldsymbol{H}}$ is almost invariably unknown. Vanderbilt and Louie (1984) suggest to estimate it on the basis of the information about the shape of the objective function collected during the n-th iteration of the algorithm or, in other words, in the L points visited by the n-th Markov chain.

The scheme used for updating the parameters consists of L replications of the following two steps for each value T_n of the temperature.

1. Generate a point $\boldsymbol{\theta}_{l+1}^{(n)} = \boldsymbol{\theta}_{l}^{(n)} + \boldsymbol{Q}\boldsymbol{u}$, where $\boldsymbol{Q}$ is an $m \times m$ positive definite matrix. The matrix $\boldsymbol{Q}$ plays a crucial role because it determines the width of the interval where the algorithm chooses $\boldsymbol{\theta}_{l+1}^{(n)}$, so that it deserves particular attention. According to the tests proposed by Louie and Vanderbilt (1984), it can be updated at the end of each chain on the basis of the mean and covariance matrix of the n-th path itself. More precisely, compute the mean and

covariance matrix of the n-th sample path:

$$\bar{\boldsymbol{\theta}}^{(n)} = \frac{1}{L}\sum_{l=0}^{L-1}\boldsymbol{\theta}_l^{(n)};$$
$$\boldsymbol{S}^{(n)} = \frac{1}{L}\sum_{l=0}^{L-1}(\boldsymbol{\theta}_l^{(n)} - \bar{\boldsymbol{\theta}}^{(n)})(\boldsymbol{\theta}_l^{(n)} - \bar{\boldsymbol{\theta}}^{(n)})'.$$

At the n-th iteration the matrix $\boldsymbol{\Sigma}^{(n)}$ is estimated as

$$\boldsymbol{\Sigma}^{(n)} = \frac{\chi}{\beta L}\boldsymbol{S}^{(n)},$$

where χ and β are appropriate constants: Vanderbilt and Louie (1984) show that $\beta = 1/6$ and use $\chi = 3$.

Now the matrix $\boldsymbol{Q}$ is still such that $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{Q}'$, but with $\boldsymbol{\Sigma}$ replaced by $\boldsymbol{\Sigma}^{(n)}$. Thus $\boldsymbol{Q}$ depends on n as well, so that in the following it can be denoted by $\boldsymbol{Q}^{(n)}$. The simulation of the next candidate value is based on the recursion $\boldsymbol{\theta}_{l+1}^{(n+1)} = \boldsymbol{\theta}_l^{(n)} + \boldsymbol{Q}^{(n)}\boldsymbol{u}$.

2. As for the initial value T_0 of the temperature, consider that the number of accepted transitions is large if the quantity $\exp\{(L(\boldsymbol{\theta}^{(n+1)} - L(\boldsymbol{\theta}^{(n)})/T_n\}$ is approximately equal to 1 for almost all the proposed transitions $\boldsymbol{\theta}^{(n)} \to \boldsymbol{\theta}^{(n+1)}$. Borrowing a methodology first introduced by Johnson *et al.* (1986), generate randomly K values of the parameters $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(K)}$ and insert them into the likelihood function to get the corresponding values $L(\boldsymbol{\theta}^{(1)}; \boldsymbol{x}), \ldots, L(\boldsymbol{\theta}^{(K)}; \boldsymbol{x})$. Then compute the average of the decreasing variation of the values of the likelihood so obtained, i.e.

$$\overline{\Delta L^-} = \frac{1}{K}\sum_{i=1}^{K}\Delta L_k^-,$$

where $\Delta L_k^- = \min\{\Delta L_k, 0\}$ and $K = 200$. Then set

$$\rho_0 = \exp\left(\frac{\overline{\Delta L^-}}{T_0}\right).$$

Choosing a value of ρ_0 "close to 1", the initial value of the temperature is equal to:

$$T_0 = \frac{\overline{\Delta L^-}}{\log(\rho_0)}.$$

The cooling schedule is based on the exponential decay $T_{n+1} = c^{n+1} \cdot T_0$ (Kirkpatrick *et al.* 1983) with $c = 0.97$, which has often proved to be more efficient than the logarithmically decreasing function originally proposed by Metropolis *et al.* (1953).

Determining a stopping criterion for the SA algorithm is always a difficult problem, because the convergence of the algorithm is slow. The criterion adopted here is again based on the value of the likelihood function. Consider the maximum value of the likelihood function at the end of the n-th iteration:

$$L_{\max}^{(n)} = \arg \max_{l<L-1} L(\boldsymbol{\theta}_l^{(n)}|\boldsymbol{y}).$$

Let

$$\overline{L(\boldsymbol{\theta}^{(n)}|\boldsymbol{y})} = \frac{1}{L}\sum_{l=0}^{L-1} L(\boldsymbol{\theta}_l^{(n)}|\boldsymbol{y})$$

be the average of the values of the likelihood function at the n-th iteration. Then the algorithm stops either when

$$\frac{\overline{L(\boldsymbol{\theta}^{(n)}|\boldsymbol{y})} - L_{\max}^{(n)}}{\overline{L(\boldsymbol{\theta}^{(n)}|\boldsymbol{y})}} < \epsilon$$

or when the temperature is below a small predefined threshold T_f

3.2 A simple implementation

A general remark concerning the SA algorithm is that the actual implementation and performance are highly problem-dependent. Moreover, the large number of parameters that have to be set makes it difficult to find the "optimal" version of the algorithm: as pointed out by Brooks and Morgan (1995), a large L gives a more accurate solution, and a value of c close to 1 diminishes the probability of getting trapped into local maxima. The initial and final values of the temperature are also important. If T_0 is not sufficiently high, some regions of the parameter space have little or no chance of being explored. If T_f is not sufficiently low, the system may not yet be frozen. In the approach proposed in the preceding section, it is also necessary to specify a value for the parameter χ.

Thus, it may be useful to consider a simpler way of proceeding requiring fewer decisions on the input variables. Such a strategy is likely to be more robust to erroneous settings of the parameters. An actual implementation is given by Brooks and Morgan (1995): the main difference with respect to Ingrassia's (1992) proposal concerns the choice of the next candidate value, which is performed in two steps. First, choose at random one of the parameters; second, simulate a new value for that parameter within the bounds set by the problem at hand. This is a straightforward way of simulating a new value "close" to the old one. As for the remaining steps, namely the choice of the cooling schedule, the initial value of the temperature and the stopping criterion, the procedure is exactly as detailed in section 3.1.

4 The Cross-Entropy approach

In this section we consider another approach to the maximization of the likelihood of a Pareto mixture, namely the minimum Cross-Entropy (CE) approach. This method was first proposed by Rubinstein (1997) in a rare event simulation setup; Rubinstein (1999) introduced a simple modification that allows to use it for solving optimization problems. The approach is fully discussed in Rubinstein and Kroese (2004, chap. 4 and 5), to which the interested reader is referred. We detail here the application of the algorithm to the problem analyzed in this paper.

Recall that our aim is to maximize the log-likelihood function $l(\boldsymbol{\theta})$ given in (3) for $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. The main idea can be summarized in two steps. First, one has to randomize the problem, that is, consider the parameters as random variables: formally, we denote with $\{f\ (\cdot;\boldsymbol{u}); \boldsymbol{u} \in \mathcal{V} \subset \boldsymbol{\Theta}\}$ the family of pdfs of $\boldsymbol{\theta}$. Second, the so-called Associated Stochastic Problem (ASP) has to be linked to the actual optimization problem. The ASP is formulated as follows:

$$l(\gamma) = P_{\boldsymbol{u}}(l(\boldsymbol{\theta}) \geq \gamma) = \mathrm{E}_{\boldsymbol{u}}(\mathbf{1}_{\{l(\ )\geq\gamma\}}), \tag{5}$$

where γ is an unknown parameter. To understand how the method works, consider the problem of estimating $l(\gamma)$ for some γ close to γ^*. Usually, in this case, $\{l(\boldsymbol{\theta}) \geq \gamma\}$ is a rare event. This remark is fundamental, as it provides the link to the typical use of the CE method, namely the estimation of rare-events probabilities. In this kind of problems the CE method is based on an iterative algorithm that makes adaptive changes to the pdf of $\boldsymbol{\theta}$ according to the minimization of the Kullback-Leibler Cross-Entropy. As a result, one gets a sequence of pdfs $f\ (\cdot;\boldsymbol{u}), f\ (\cdot;\boldsymbol{v}^{(1)}), f\ (\cdot;\boldsymbol{v}^{(2)}), \ldots$ converging to the theoretically optimal density, which is characterized by the following theorem.

Theorem 1 *Let γ^* be the maximum of a real-valued function l on a finite set $\mathcal{X}$. Suppose that the corresponding maximizer $\boldsymbol{\theta}^*$ is unique and that the class of densities $\{f\ (\cdot)\}$ to be used in the CE algorithm detailed below contains the Dirac density with mass at $\boldsymbol{\theta}^*$:*

$$\delta_*(\boldsymbol{\theta}) = \begin{cases} 1 & if\ \boldsymbol{\theta} = \boldsymbol{\theta}^*; \\ 0 & \text{otherwise.} \end{cases}$$

Then the solution of the CE program for the estimation of $P_{\boldsymbol{u}}(l(\boldsymbol{\theta}) \geq \gamma^*)$ is given by δ_*.

Proof. See Rubinstein and Kroese (2004, pag. 132).

Roughly speaking, the theorem says that, at convergence, the density $f\ (\cdot)$ reduces to the Dirac delta density centered at $\boldsymbol{\theta}^*$; in other words, the sequence $\hat{\boldsymbol{\theta}}^{(n)}$ converges to $\boldsymbol{\theta}^*$ with probability 1.

For example, if $f\ (\cdot)$ is the $N(\mu, \sigma^2)$ density, it will converge to the $N(\theta^*, 0)$ density. It is worth stressing that the initial variance of the distribution of $\boldsymbol{\theta}$ must be "large enough", or the algorithm would not explore all the areas of the parameter space.

The algorithm works as follows (see Rubinstein and Kroese 2004, Algorithm 4.2.1).

1. Choose an initial parameter vector $\boldsymbol{\theta}^{(0)}$ for the parameters. Set $t = 1$.
2. Simulate a sample $\boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_N^{(t)}$ from the density $f\ (\cdot; \boldsymbol{v}^{(t-1)})$, where $\boldsymbol{v}^{(0)} = \boldsymbol{\theta}^{(0)}$. Compute the numerical values $l(\boldsymbol{\theta}_1^{(t)}), \ldots, l(\boldsymbol{\theta}_N^{(t)})$ of the log-likelihood for each simulated sample and compute the sample $(1-\rho)$-quantile $\hat{\gamma}_t$ of $l(\boldsymbol{\theta}_1^{(t)}), \ldots, l(\boldsymbol{\theta}_N^{(t)})$.
3. Use the same sample $\boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_N^{(t)}$ to solve the stochastic program (5). Call the solution $\boldsymbol{v}^{(t)}$.
4. Use the formula $\boldsymbol{v}^{(t)} = \alpha \boldsymbol{v}^{(t)} + (1-\alpha)\boldsymbol{v}^{(t-1)}$ $(0 < \alpha < 1)$ to smooth out the vector $\boldsymbol{v}^{(t)}$.
5. If some stopping criterion is satisfied, stop. Else, set $t = t + 1$ and restart from step 2.

The smoothing procedure of step 4 is introduced to rule out the possibility that some component of $\boldsymbol{v}^{(t)}$ is zero or one. Details about the stopping criterion and the numerical values of α and ρ will be given in the next section.

5 Simulation and application

5.1 Some simulation results

The first goal of this section consists in studying the properties of the estimators obtained by means of the SA and CE algorithms. To this aim, we performed the following experiment. For each of the sample sizes $n = 20,\ 40,\ 60,\ 80,\ 100,\ 150,\ 200,\ 300,\ 500$, we simulated $1,000$ samples from a two-population Pareto mixture with parameters $\pi = 0.5$, $a_1 = 1$, $a_2 = 5$, $c_1 = 2.5$ and $c_2 = 7.5$.

We tried to use both the implementations of the SA algorithm presented in section 3. However, step 1. of the first methodology (i.e., the part concerning the choice of the next candidate value) did not work well. Therefore we show the results obtained by combining the technique proposed by Brooks and Morgan (1995) for the transition probability and the Ingrassia (1992) approach for the cooling schedule. As for the numerical value of ρ_0, by trial and error we chose $\rho_0 = 0.7$; different values did not seem to change appreciably the final results. Moreover, we put $\epsilon = 0.005$ and $T_f = 0.2$.

In the CE algorithm the density to be used in the CE program is four-dimensional. As the first parameter must lie in the interval $[0, 1]$, for v_1 we used a Beta distribution, whereas the remaining components are assumed to follow normal distributions with different means and variances. The

components are assumed to be independent, so that updating of the parameters can be done independently. According to the suggestion by Rubinstein and Kroese (2004, p. 188-189), the ASP for the first parameter was solved numerically. The algorithm was implemented with $\alpha = 0.5$ and $\rho = 0.01$. Whereas different choices of the numerical values of the latter parameter did not seem to change the results appreciably, the value of α has a significant impact. In particular, with a larger value of α the algorithm sometimes does not converge to the global optimum. On the other hand, a smaller value avoids the traps of local maxima but makes convergence very slow. We ran several experiments and found that the choice of the numerical value of α is strictly related to the choice of $\boldsymbol{\theta}^{(0)}$: if $\boldsymbol{\theta}^{(0)}$ is "far away" from the true $\boldsymbol{\theta}$, a small α is called for, or the algorithm converges to a suboptimal $\boldsymbol{\theta}^*$. Finally, using the same strategy of Rubinstein and Kroese (2004, pag. 134), the algorithm was stopped when $\hat{\gamma}^{(t)} = \hat{\gamma}^{(t-1)} = \cdots = \hat{\gamma}^{(t-5)}$.

Figure 2 shows the averages $\bar{\hat{\theta}}_i^{SA} = (1/n_i) \sum_{i=1}^{n_i} \hat{\theta}_i^{SA}$ and $\bar{\hat{\theta}}_i^{CE} = (1/n_i) \sum_{i=1}^{n_i} \hat{\theta}_i^{CE}$ $(i = 1, \ldots, 9)$ of the numerical values of the estimators obtained at each replication of the procedure. Figures 3, 4, 5 and 6 display the simulated distributions of the SA estimators; figures 7, 8, 9 and 10 show the same distributions obtained with the CE algorithm.

Figure 2 allows to draw some interesting conclusions. First, both estimators are rather precise even for moderate sample sizes. There is, however, a significant difference for the smallest sample size considered in the experiment (i.e. $N = 20$): in this case the CE estimator has a considerably worse performance than the SA estimator, and cannot be considered reliable. Moreover, the estimation of the shape parameters, and in particular of c_2, seems to be the most difficult one. For moderate to large sample size the CE estimator is more precise.

A thorough analysis of the results must take into account also the variability of the estimators. For these reasons we display the simulated distribution of the estimators and compute the Mean Squared Errors (MSEs).

Figures 3 to 10 reinforce the remarks arising from figure 2 above. In particular, for $N = 20$, the performance of the CE estimator is disappointing, and the numerical value of $\hat{c}_2$ is sometimes meaningless. To complete the analysis with a measure that encompasses all the results displayed so far, we computed the MSE of the estimators. The outcomes are shown in Table 1.

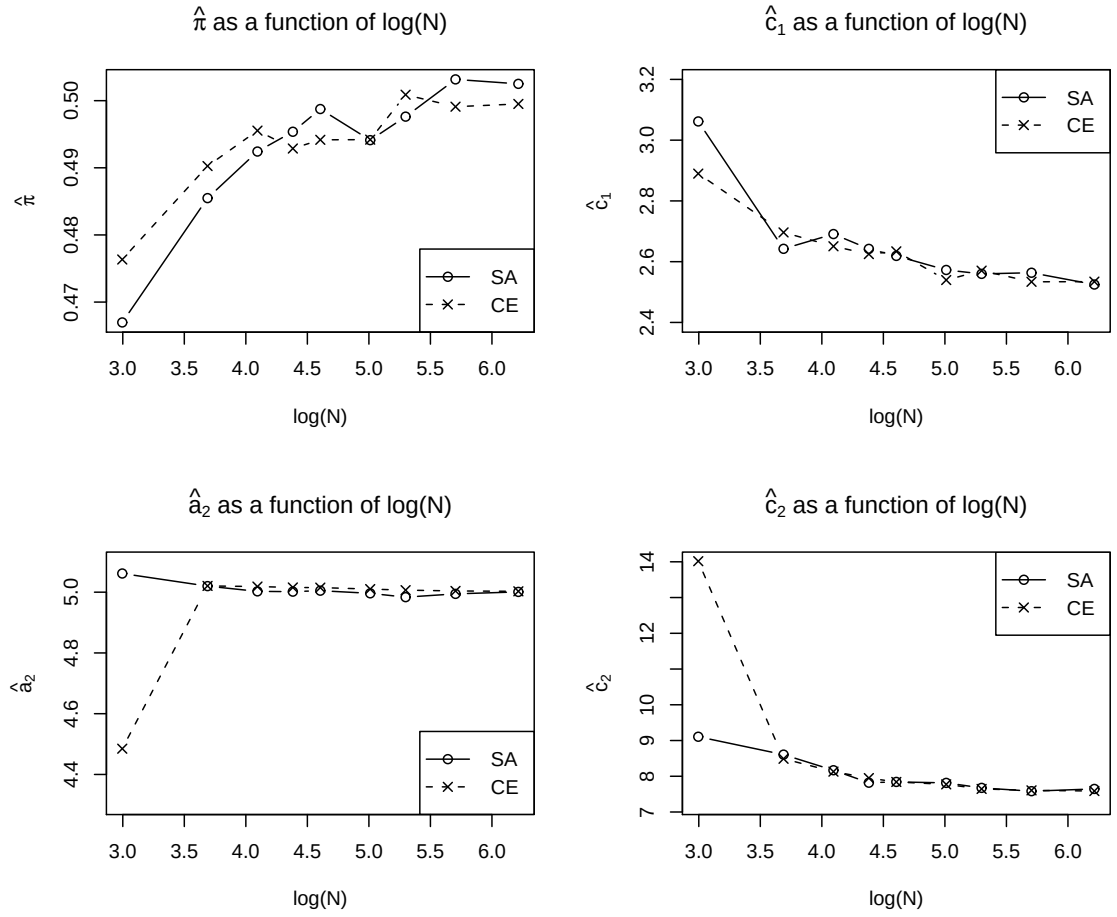

Figure 2: The values of $\bar{\hat{\theta}}_i^{SA} = (1/N_i)\sum_{i=1}^{N_i}\hat{\theta}_i^{SA}$ and $\bar{\hat{\theta}}_i^{CE} = (1/N_i)\sum_{i=1}^{N_i}\hat{\theta}_i^{CE}$ $(i = 1, \ldots, 4)$.

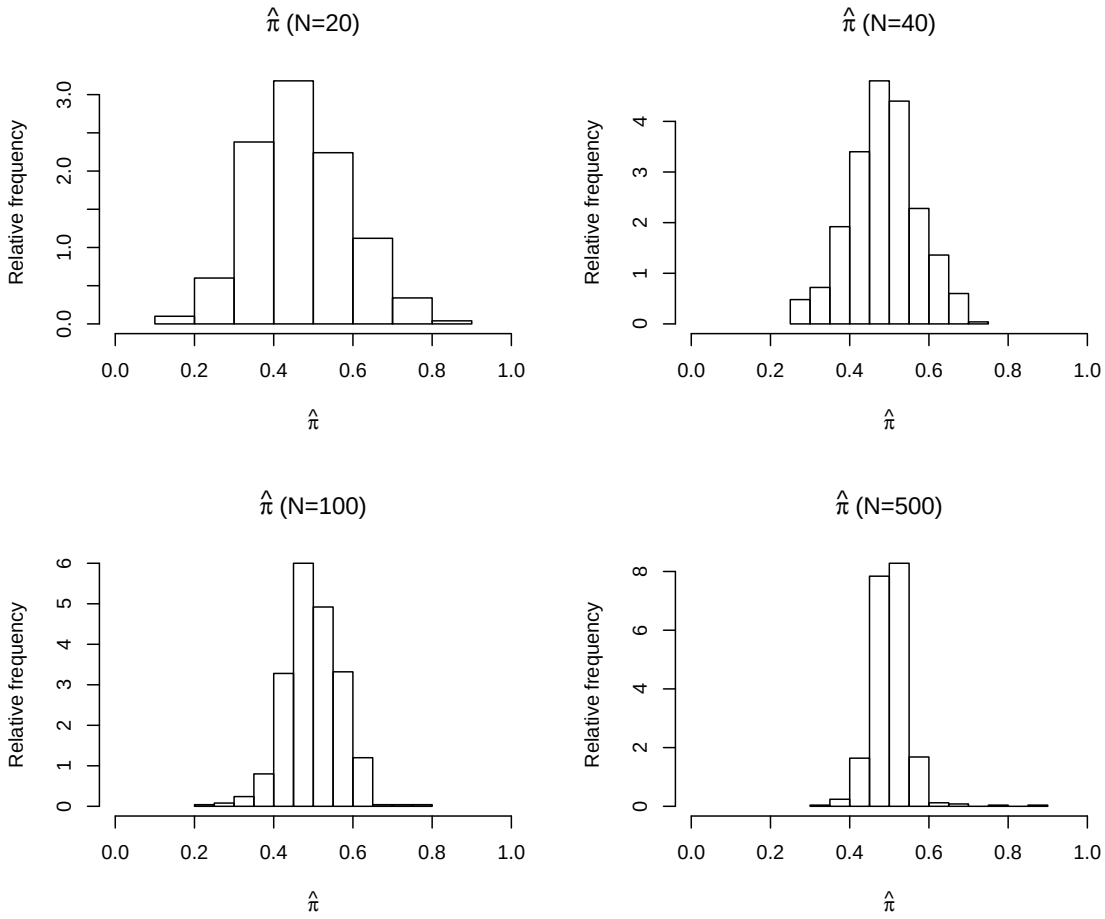

Figure 3: Simulated distributions of the SA estimator of π for $N = 20, 40, 100, 500$.

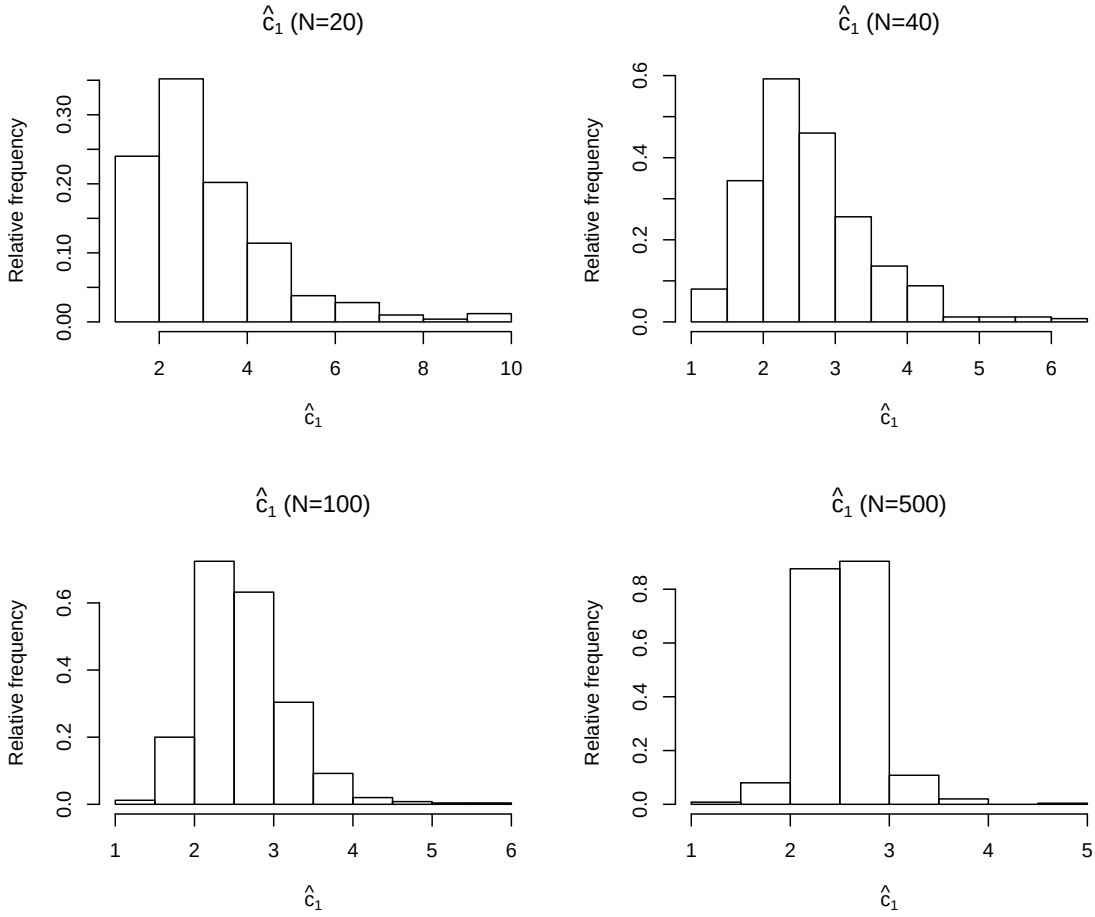

Figure 4: Simulated distributions of the SA estimator of c_1 for $N = 20, 40, 100, 500$.

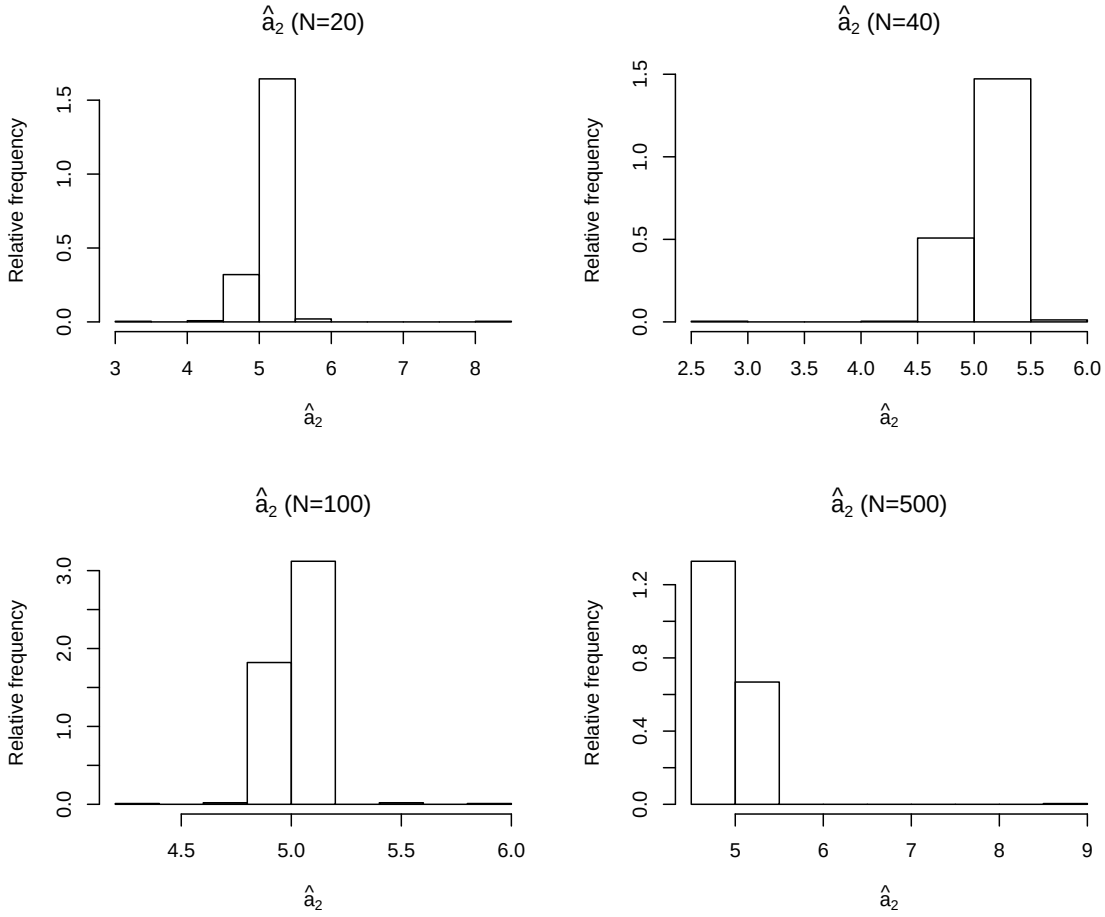

Figure 5: Simulated distributions of the SA estimator of a_2 for $N = 20, 40, 100, 500$.

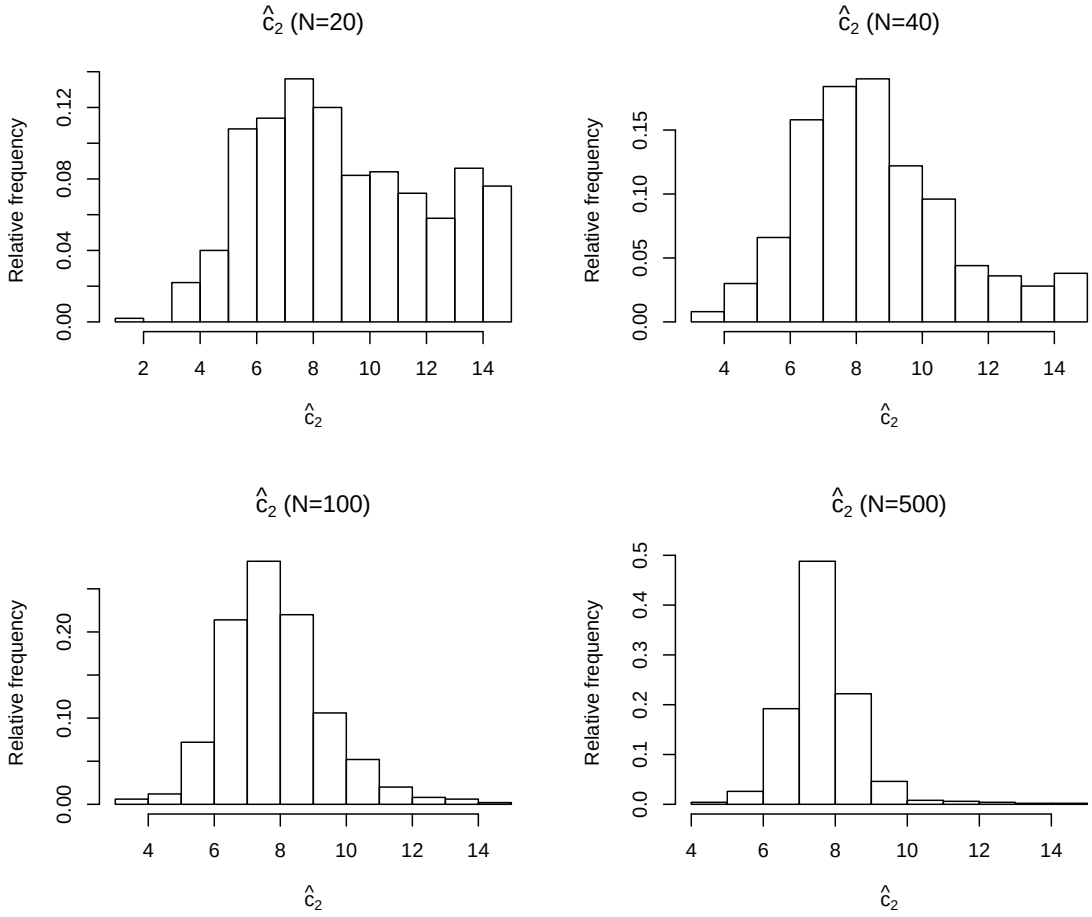

Figure 6: Simulated distributions of the SA estimator of c_2 for $N = 20, 40, 100, 500$.

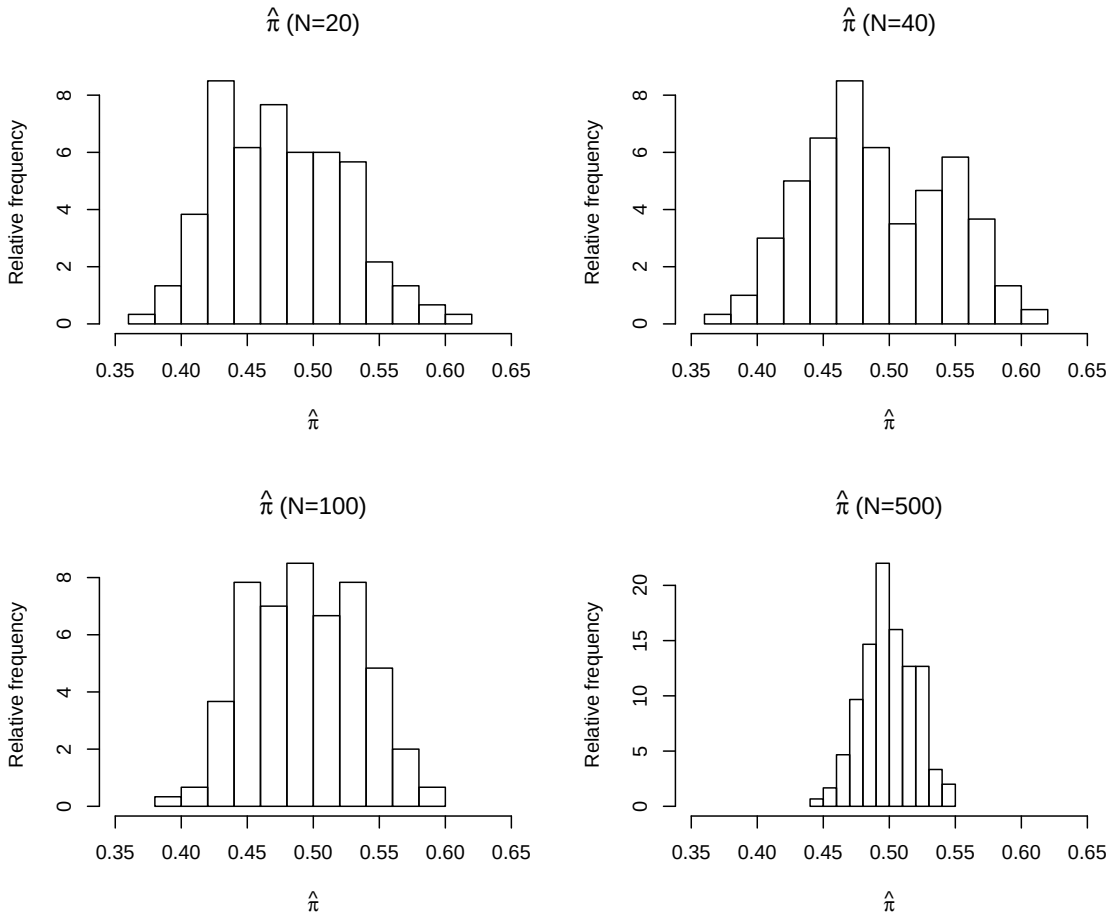

Figure 7: Simulated distributions of the CE estimator of π for $N = 20, 40, 100, 500$.

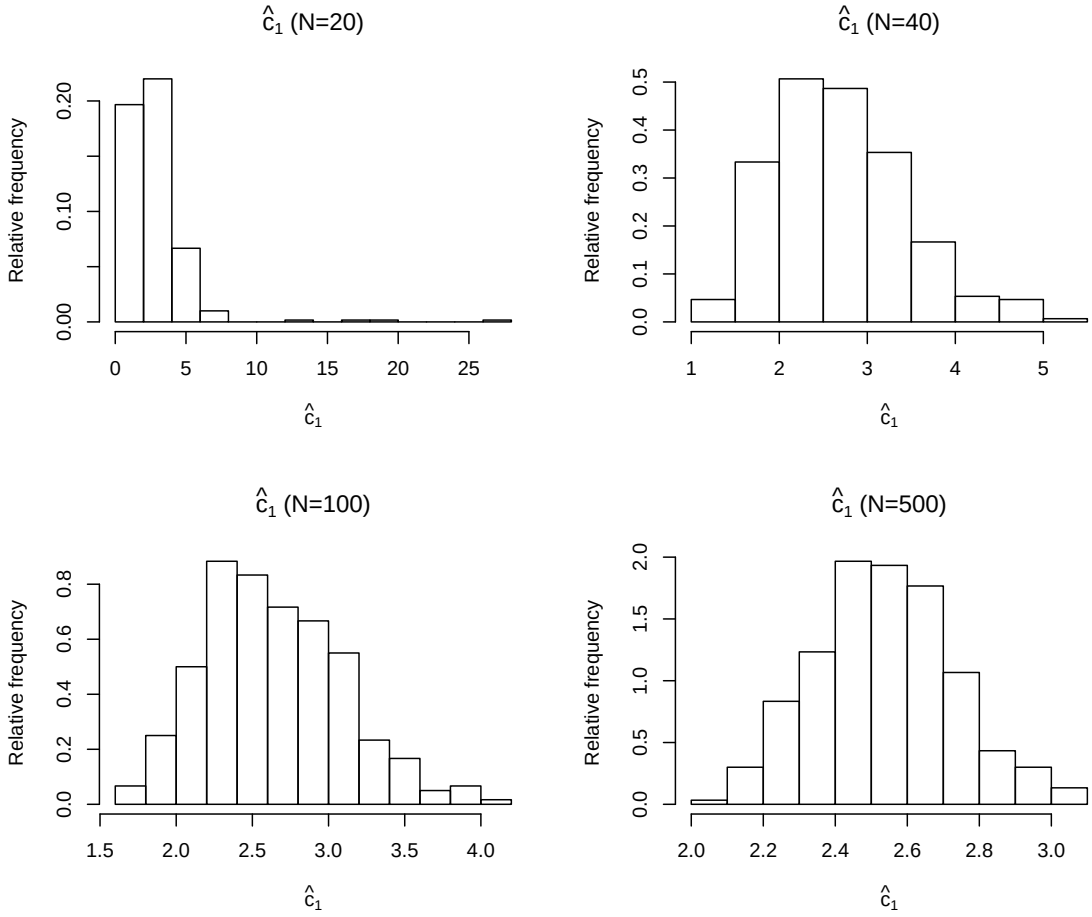

Figure 8: Simulated distributions of the CE estimator of c_1 for $N = 20, 40, 100, 500$.

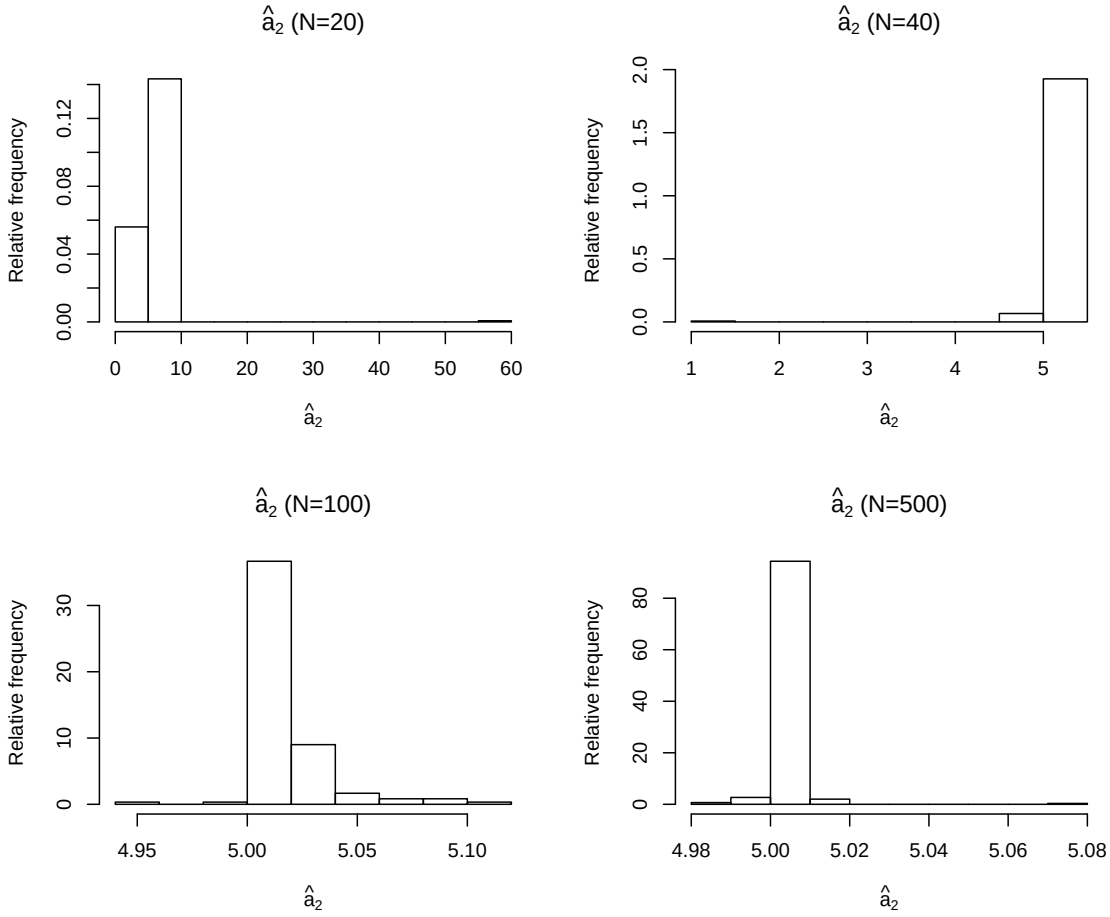

Figure 9: Simulated distributions of the CE estimator of a_2 for $N = 20, 40, 100, 500$.

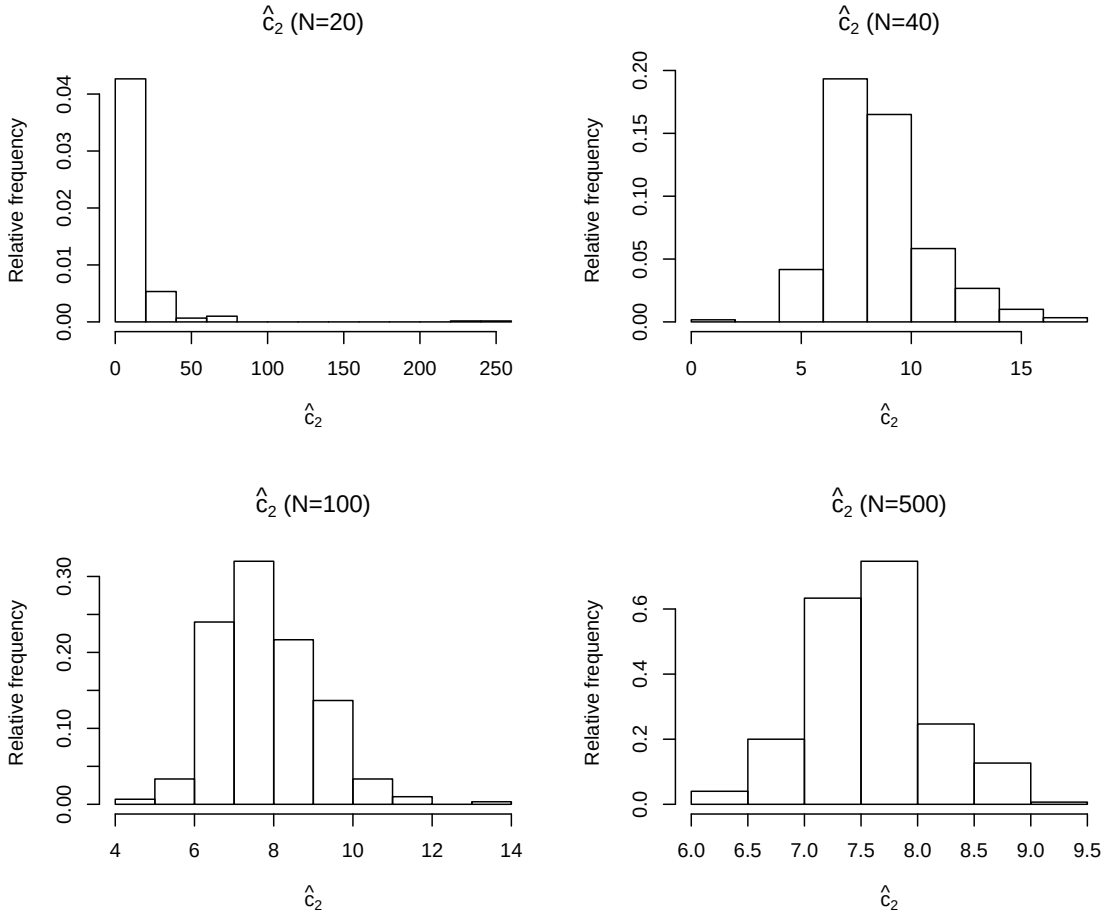

Figure 10: Simulated distributions of the CE estimator of c_2 for $N = 20, 40, 100, 500$.

Table 1: Mean Squared Errors of the SA and CE estimators.

	$\hat{\pi}^{SA}$	$\hat{\pi}^{CE}$	$\hat{c}_1^{SA}$	$\hat{c}_1^{CE}$	$\hat{a}_2^{SA}$	$\hat{a}_2^{CE}$	$\hat{c}_2^{SA}$	$\hat{c}_2^{CE}$
$N = 20$	0.016	0.003	2.644	5.772	0.043	11.367	12.043	530.230
$N = 40$	0.008	0.003	0.674	0.612	0.019	0.056	6.788	6.259
$N = 60$	0.007	0.002	0.703	0.362	0.038	0.001	4.323	3.894
$N = 80$	0.006	0.002	0.485	0.288	0.014	0.001	3.223	2.177
$N = 100$	0.004	0.002	0.330	0.223	0.005	0.001	2.579	1.706
$N = 150$	0.003	0.001	0.247	0.110	0.003	0.001	1.825	1.003
$N = 200$	0.003	0.001	0.170	0.086	0.027	0.001	1.722	0.762
$N = 300$	0.002	0.001	0.273	0.056	0.004	0.001	1.342	0.433
$N = 500$	0.002	< 0.001	0.119	0.039	0.025	0.001	1.073	0.294

Table 1 gives some more insight about the two estimators. First of all, for $N = 20$ the SA approach is definitely preferable. Second, for all the remaining sample sizes CE performs better than SA: in particular, the difference in the precision of estimation of the shape parameters c_1 and c_2 increases as the sample size gets larger, so that according to Table 1 the CE algorithm should be chosen for large N.

Finally, a remark about execution times is in order: on average, in the setup of the simulation experiments performed above, with $N = 500$, the CE approach took approximately 111 seconds and the SA algorithm 254 seconds on a 1.60GHz Pentium processor.

In conclusion, for $N = 20$ the SA algorithm is clearly superior in terms of MSE. For the remaining sample sizes, and more evidently as N increases, the CE approach has a better performance. The fact that SA is preferable when $N = 20$ is not surprising: in this setup the maximization is clearly not straightforward, and it is well-known that SA often performs quite well (in the sense that it converges at the global optimum when other algorithms get stuck at suboptimal points) in "difficult" problems. On the other hand, SA is typically slow, dependent on many parameters and less accurate in identifying the maximum, thus it may not be the best solution in more standard frameworks. The problem treated here seems to confirm these remarks, so that, considering also the larger convergence time and the difficulties in setting the parameters of SA, the CE approach may be regarded as preferable for all but very small sample sizes. However, from the practitioner's point ov view, it should also be noted that when sampling from a mixture distribution a sample size as small as 20 is hardly ever encountered, in particular in applications such as the one considered in this paper.

5.2 A real-data application

Figure 11a shows the distribution of the amounts (in thousands of Euros) of 1102 operational losses corresponding to two different business lines. The losses were measured in Banca Intesa (Milano) in a recent year and have been rescaled for confidentiality reasons. The business line each observation belongs to is unknown, so that the data are sampled from the mixture. The tail is quite heavy. Fitting a two-population Pareto mixture is a reasonbly good solution, although the frequency of big losses (approximately > 8, according to a visual inspection of the histogram) seems larger than expected under the Pareto model.

Table 2 shows the estimated parameters and the values of the log-likelihood obtained with the two methodologies presented above. Notice that the maximum of the log-likelihood found by the CE approach is larger than the value obtained with SA: given the large sample size, this is in agreement with the results of the simulation exercise of the preceding section. As for the execution times, the CE algorithm converged in 138 seconds, whereas the SA algorithm converged in 246 seconds (320 iterations).

Table 2: Estimated parameters

	$\hat{\pi}$	$\hat{a}_1$	$\hat{c}_1$	$\hat{a}_2$	$\hat{c}_2$	log-lik
SA	0.257	1.004	3.343	2.013	2.444	-1736.20
CE	0.259	1.004	2.537	2.001	2.005	-1678.72

Figure 11b shows the frequency distribution of the observed data and the two estimated densities, namely the Pareto mixtures corresponding to the parameters estimated by means of SA and CE. The two densities are very similar to each other; the fact that c_1 is the parameter for which we observe the largest difference between the estimates obtained with the two approaches is not surprising if we consider that it is related to the tail behavior of the first distribution, and the tail of this distribution is strongly contaminated by the body of the second one.

The ultimate measure of interest in risk management appplications like the present one is the so-called Value at Risk (VaR). The VaR at level α is the α quantile of the loss distribution and represents the threshold monetary amount such that the probability that the loss over the given time horizon exceeds this value is equal to α. Having estimated the parameters, we can compute the VaR at level *alpha* by simulating B observations from the estimated mixture and computing the *alpha* quantile of the simulated distribution. With α equal respectively to 0.95 and 0.99, the results are $VaR_{SA,0.95} = 6.13$, $VaR_{CE,0.95} = 7.81$, $VaR_{SA,0.99} = 11.91$ and $VaR_{CE,0.99} = 17.06$. Thus, the differences are non-negligible, in particular for large α.

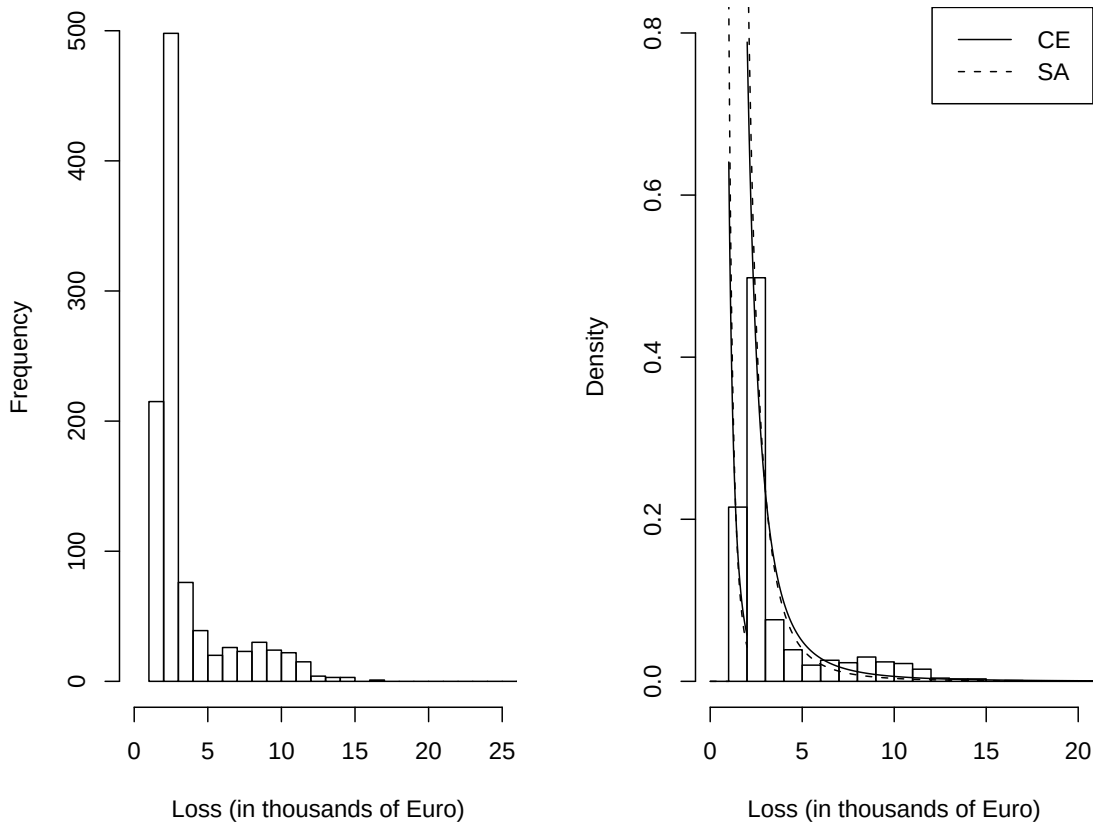

Figure 11: (a) Loss distribution; (b) Estimated densities.

6 Conclusions

In this paper we have studied Maximum Likelihood Estimation of a mixture of Pareto distributions. When the location parameters of the component distributions are different, the EM algorithm breaks down, so that different optimization techniques are required. We showed how the problem can be solved by means of the Simulated Annealing and the minimum Cross-Entropy algorithms. The results of a simulation experiment suggest that the CE algorithm has a slight advantage for moderate and large N, mainly due to a more straightforward implementation and a faster convergence. On the other hand, for very small sample sizes, SA is preferable. An application in the field of operational risk confirmed the importance of an accurate estimate of the parameters.

Some problems are open for future research. First, the algorithms used in this paper are very general and can be extended to mixtures of different distributions (just to mention an example, the Laplace mixtures studied by Naradajah 2006): their precise formulation and properties depend on the probabilistic model and need further investigation. Second, likelihood maximization in this setup may probably be performed by means of other algorithms: given the features of the log-likelihood function, direct search methods as the Nelder-Mead simplex method are likely to be a good solution.

References

Brooks, S.P. and Morgan, B.J.P. (1995) "Optimization using simulated annealing". *The Statistician*, 44, 241-257.

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

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), "Maximum likelihood from incomplete data via the EM algorithm (with discussion)". *Journal of the Royal Statistical Society B*, 39, 1-38.

Geman S. and Geman D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, **6**, 721-741.

Ingrassia, S. (1992), "A comparison between the simulated annealing and the EM algorithms in normal mixture decompositions". *Statistics and Computing*, 2, 203-211.

Johnson D.S., Aragon, C.R., McGeoch, L.A. and Schevon, C. (1986), "Optimization by simulated annealing: an experimental evaluation". *Workshop on Statistical Physics in Engineering and Biology.* Yorktown Heights, NY.

Kirkpatrick S., Gelatt, C.D. and Vecchi, M.P. (1983), "Optimization by simulated annealing". *Sci-*

ence, 220, 671-680.

Li, W. and Cai, X. (2004), "Statistical analysis of airport network of China", *Phys. Rev. E*, 69, Art. No. 046106.

Little, R.J.A. and Rubin, D.B. (2002), *Statistical Analysis with Missing Data*, second edition. Wiley, New York.

McLachlan, G.J. and Krishnan, T. (1996), *The EM algorithm and Extensions*. Wiley, New York.

McNeil, A.J., Frey, R. and Embrechts, P. (2005), *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton, Princeton University Press.

Metropolis N., Rosenbluth A.W., Rosenbluth N.M., Teller A.H. and Teller E. (1953) Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1092.

Nadarajah, S. (2006), "Information matrices for Laplace and Pareto mixtures", *Computational Statistics and Data Analysis*, 50, 950-966.

Pearson, K. (1894), "Contribution to the Mathematical Theory of Evolution", *Phil. Trans. Roy. Soc. A*, 185, 71-110.

Reed,W.J. (2002), "On the rank-size distribution for human settlements", *J. Reg. Sci.*, 42, 1-17.

Reed,W.J. (2003), "The Pareto law of incomes - an explanation and an extension", *Phys. A - Statist. Mech. Appl.*, 319, 469-486.

Reed, W.J. and Jorgensen, M. (2004), "The double pareto-lognormal distributiona new parametric model for size distributions", *Comm. Statist. Theory Methods*, 33, 1733-1753.

Rubinstein, R.Y. (1997). "Optimization of computer simulation models with rare events". *European Journal of Operational Research*, 99, 89-112.

Rubinstein, R.Y. (1999). "The Cross-Entropy method for combinatorial and continuous optimization", *Methodology and Computing in Applied Probability*, 2, 127-190.

Rubinstein, R.Y. Kroese, D.P. (2004). *The Cross-Entropy Method*. New York: Springer.

Vanderbilt, D. and Louie, S.G. (1984), "A Monte Carlo simulated annealing approach to optimization over continuous variables", *Journal of Computational Physics*, 56, 259-271.

Elenco dei papers del Dipartimento di Economia

2000.1 *A two-sector model of the effects of wage compression on unemployment and industry distribution of employment*, by Luigi Bonatti

2000.2 *From Kuwait to Kosovo: What have we learned? Reflections on globalization and peace,* by Roberto Tamborini

2000.3 *Metodo e valutazione in economia. Dall'apriorismo a Friedman* , by Matteo Motterlini

2000.4 *Under tertiarisation and unemployment.* by Maurizio Pugno

2001.1 *Growth and Monetary Rules in a Model with Competitive Labor Markets,* by Luigi Bonatti.

2001.2 *Profit Versus Non-Profit Firms in the Service Sector: an Analysis of the Employment and Welfare Implications,* by Luigi Bonatti, Carlo Borzaga and Luigi Mittone.

2001.3 *Statistical Economic Approach to Mixed Stock-Flows Dynamic Models in Macroeconomics,* by Bernardo Maggi and Giuseppe Espa.

2001.4 *The monetary transmission mechanism in Italy: The credit channel and a missing ring,* by Riccardo Fiorentini and Roberto Tamborini.

2001.5 *Vat evasion: an experimental approach,* by Luigi Mittone

2001.6 *Decomposability and Modularity of Economic Interactions,* by Luigi Marengo, Corrado Pasquali and Marco Valente.

2001.7 *Unbalanced Growth and Women's Homework,* by Maurizio Pugno

2002.1 *The Underground Economy and the Underdevelopment Trap,* by Maria Rosaria Carillo and Maurizio Pugno.

2002.2 *Interregional Income Redistribution and Convergence in a Model with Perfect Capital Mobility and Unionized Labor Markets,* by Luigi Bonatti.

2002.3 *Firms' bankruptcy and turnover in a macroeconomy,* by Marco Bee, Giuseppe Espa and Roberto Tamborini.

2002.4 *One "monetary giant" with many "fiscal dwarfs": the efficiency of macroeconomic stabilization policies in the European Monetary Union,* by Roberto Tamborini.

2002.5 *The Boom that never was? Latin American Loans in London 1822-1825,* by Giorgio Fodor.

2002.6 *L'economia senza banditore di Axel Leijonhufvud: le 'forze oscure del tempo e dell'ignoranza' e la complessità del coordinamento,* by Elisabetta De Antoni.

2002.7 *Why is Trade between the European Union and the Transition Economies Vertical?,* by Hubert Gabrisch and Maria Luigia Segnana.

2003.1 *The service paradox and endogenous economic gorwth,* by Maurizio Pugno.

2003.2 *Mappe di probabilità di sito archeologico: un passo avanti,* di Giuseppe Espa, Roberto Benedetti, Anna De Meo e Salvatore Espa.
(*Probability maps of archaeological site location: one step beyond,* by Giuseppe Espa, Roberto Benedetti, Anna De Meo and Salvatore Espa).

2003.3 *The Long Swings in Economic Understianding,* by Axel Leijonhufvud.

2003.4 *Dinamica strutturale e occupazione nei servizi,* di Giulia Felice.

2003.5 *The Desirable Organizational Structure for Evolutionary Firms in Static Landscapes,* by Nicolás Garrido.

2003.6 *The Financial Markets and Wealth Effects on Consumption* An Experimental Analysis, by Matteo Ploner.

2003.7 *Essays on Computable Economics, Methodology and the Philosophy of Science,* by Kumaraswamy Velupillai.

2003.8 *Economics and the Complexity Vision: Chimerical Partners or Elysian Adventurers?*, by Kumaraswamy Velupillai.

2003.9 *Contratto d'area cooperativo contro il rischio sistemico di produzione in agricoltura,* di Luciano Pilati e Vasco Boatto.

2003.10 *Il contratto della docenza universitaria. Un problema multi-tasking*, di Roberto Tamborini.

2004.1 *Razionalità e motivazioni affettive: nuove idee dalla neurobiologia e psichiatria per la teoria economica?* di Maurizio Pugno.
(*Rationality and affective motivations: new ideas from neurobiology and psychiatry for economic theory?* by Maurizio Pugno.

2004.2 *The economic consequences of Mr. G. W. Bush's foreign policy. Can th US afford it?* by Roberto Tamborini

2004.3 *Fighting Poverty as a Worldwide Goal* by Rubens Ricupero

2004.4 *Commodity Prices and Debt Sustainability* by Christopher L. Gilbert and Alexandra Tabova

2004.5 *A Primer on the Tools and Concepts of Computable Economics* by K. Vela Velupillai

2004.6 *The Unreasonable Ineffectiveness of Mathematics in Economics* by Vela K. Velupillai

2004.7 *Hicksian Visions and Vignettes on (Non-Linear) Trade Cycle Theories* by Vela K. Velupillai

2004.8 *Trade, inequality and pro-poor growth: Two perspectives, one message?* By Gabriella Berloffa and Maria Luigia Segnana

2004.9 *Worker involvement in entrepreneurial nonprofit organizations. Toward a new assessment of workers? Perceived satisfaction and fairness* by Carlo Borzaga and Ermanno Tortia.

2004.10 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part I): Rational Bargaining and Justification* by Lorenzo Sacconi

2004.11 *A Social Contract Account for CSR as Extended Model of Corporate Governance (Part II): Compliance, Reputation and Reciprocity* by Lorenzo Sacconi

2004.12 *A Fuzzy Logic and Default Reasoning Model of Social Norm and Equilibrium Selection in Games under Unforeseen Contingencies* by Lorenzo Sacconi and Stefano Moretti

2004.13 *The Constitution of the Not-For-Profit Organisation: Reciprocal Conformity to Morality* by Gianluca Grimalda and Lorenzo Sacconi

2005.1 *The happiness paradox: a formal explanation from psycho-economics* by Maurizio Pugno

2005.2 *Euro Bonds: in Search of Financial Spillovers* by Stefano Schiavo

2005.3 *On Maximum Likelihood Estimation of Operational Loss Distributions* by Marco Bee

2005.4 *An enclave-led model growth: the structural problem of informality persistence in Latin America* by Mario Cimoli, Annalisa Primi and Maurizio Pugno

2005.5 *A tree-based approach to forming strata in multipurpose business surveys,* Roberto Benedetti, Giuseppe Espa and Giovanni Lafratta.

2005.6 *Price Discovery in the Aluminium Market* by Isabel Figuerola-Ferretti and Christopher L. Gilbert.

2005.7 *How is Futures Trading Affected by the Move to a Computerized Trading System? Lessons from the LIFFE FTSE 100 Contract* by Christopher L. Gilbert and Herbert A. Rijken.

2005.8 *Can We Link Concessional Debt Service to Commodity Prices?* By Christopher L. Gilbert and Alexandra Tabova

2005.9 *On the feasibility and desirability of GDP-indexed concessional lending* by Alexandra Tabova.

2005.10 *Un modello finanziario di breve periodo per il settore statale italiano: l'analisi relativa al contesto pre-unione monetaria* by Bernardo Maggi e Giuseppe Espa.

2005.11 *Why does money matter? A structural analysis of monetary policy, credit and aggregate supply effects in Italy*, Giuliana Passamani and Roberto Tamborini.

2005.12 *Conformity and Reciprocity in the "Exclusion Game": an Experimental Investigation* by Lorenzo Sacconi and Marco Faillo.

2005.13 *The Foundations of Computable General Equilibrium Theory*, by K. Vela Velupillai.

2005.14 *The Impossibility of an Effective Theory of Policy in a Complex Economy*, by K. Vela Velupillai.

2005.15 *Morishima's Nonlinear Model of the Cycle: Simplifications and Generalizations,* by K. Vela Velupillai.

2005.16 Using and Producing *Ideas* in Computable Endogenous Growth, by K. Vela Velupillai.

2005.17 *From Planning to Mature: on the Determinants of Open Source Take Off* by Stefano Comino, Fabio M. Manenti and Maria Laura Parisi.

2005.18 *Capabilities, the self, and well-being: a research in psycho-economics,* by Maurizio Pugno.

2005.19 Fiscal and monetary policy, unfortunate events, and the SGP arithmetics. Evidence from a *growth-gap* model, by Edoardo Gaffeo, Giuliana Passamani and Roberto Tamborini

2005.20 *Semiparametric Evidence on the Long-Run Effects of Inflation on Growth,* by Andrea Vaona and Stefano Schiavo.

2006.1 *On the role of public policies supporting Free/Open Source Software. An European perspective,* by Stefano Comino, Fabio M. Manenti and Alessandro Rossi.

2006.2 *Back to Wicksell? In search of the foundations of practical monetary policy,* by Roberto Tamborini

2006.3 *The uses of the past,* by Axel Leijonhufvud

2006.4 *Worker Satisfaction and Perceived Fairness: Result of a Survey in Public, and Non-profit Organizations,* by Ermanno Tortia

2006.5 *Value Chain Analysis and Market Power in Commodity Processing with Application to the Cocoa and Coffee Sectors,* by Christopher L. Gilbert

2006.6 *Macroeconomic Fluctuations and the Firms' Rate of Growth Distribution: Evidence from UK and US Quoted Companies,* by Emiliano Santoro

2006.7 *Heterogeneity and Learning in Inflation Expectation Formation: An Empirical Assessment,* by Damjan Pfajfar and Emiliano Santoro

2006.8 Good *Law & Economics* needs suitable microeconomic models: the case against the application of standard agency models: the case against the application of standard agency models to the professions, by Lorenzo Sacconi

2006.9 *Monetary policy through the "credit-cost channel". Italy and Germany,* by Giuliana Passamani and Roberto Tamborini

2007.1 *The Asymptotic Loss Distribution in a Fat-Tailed Factor Model of Portfolio Credit Risk,* by Marco Bee

2007.2 *Sraffa's Mathematical Economics – A Constructive Interpretation,* by Kumaraswamy Velupillai

2007.3 *Variations on the Theme of Conning in Mathematical Economics,* by Kumaraswamy Velupillai

2007.4 *Norm Compliance: the Contribution of Behavioral Economics Models,* by Marco Faillo and Lorenzo Sacconi

2007.5 *A class of spatial econometric methods in the empirical analysis of clusters of firms in the space,* by Giuseppe Arbia, Giuseppe Espa e Danny Quah.

2007.6 *Rescuing the LM (and the money market) in a modern Macro course,* by Roberto Tamborini.

2007.7 *Family, Partnerships, and Network: Reflections on the Strategies of the Salvadori Firm of Trento,* by Cinzia Lorandini.

2007.8 *I Verleger serici trentino-tirolesi nei rapporti tra Nord e Sud: un approccio prosopografico,* by Cinzia Lorandini.

2007.9 *A Framework for Cut-off Sampling in Business Survey Design,* by Marco Bee, Roberto Benedetti e Giuseppe Espa

2007.10 *Spatial Models for Flood Risk Assessment,* by Marco Bee, Roberto Benedetti e Giuseppe Espa

2007.11 *Inequality across cohorts of households:evidence from Italy,* by Gabriella Berloffa and Paola Villa

2007.12 *Cultural Relativism and Ideological Policy Makers in a Dynamic Model with Endogenous Preferences,* by Luigi Bonatti

2007.13 *Optimal Public Policy and Endogenous Preferences: an Application to an Economy with For-Profit and Non-Profit,* by Luigi Bonatti

2007.14 *Breaking the Stability Pact: Was it Predictable?,* by Luigi Bonatti and Annalisa Cristini.

2007.15 *Home Production, Labor Taxation and Trade Account,* by Luigi Bonatti.

2007.16 *The Interaction Between the Central Bank and a Monopoly Union Revisited: Does Greater Uncertainty about Monetary Policy Reduce Average Inflation?,* by Luigi Bonatti.

2007.17 *Complementary Research Strategies, First-Mover Advantage and the Inefficiency of Patents,* by Luigi Bonatti.

2007.18 *DualLicensing in Open Source Markets,* by Stefano Comino and Fabio M. Manenti.

2007.19 *Evolution of Preferences and Cross-Country Differences in Time Devoted to Market Work,* by Luigi Bonatti.

2007.20 *Aggregation of Regional Economic Time Series with Different Spatial Correlation Structures,* by Giuseppe Arbia, Marco Bee and Giuseppe Espa.

2007.21 *The Sustainable Enterprise. The multi-fiduciary perspective to the EU Sustainability Strategy,* by Giuseppe Danese.

2007.22 *Taming the Incomputable, Reconstructing the Nonconstructive and Deciding the Undecidable in Mathematical Economics,* by K. Vela Velupillai.

2007.23 *A Computable Economist's Perspective on Computational Complexity,* by K. Vela Velupillai.

2007.24 *Models for Non-Exclusive Multinomial Choice, with Application to Indonesian Rural Households,* by Christopher L. Gilbert and Francesca Modena.

2007.25 *Have we been Mugged? Market Power in the World Coffee Industry,* by Christopher L. Gilbert.

2007.26 *A Stochastic Complexity Perspective of Induction in Economics and Inference in Dynamics,* by K. Vela Velupillai.

2007.27 *Local Credit ad Territorial Development: General Aspects and the Italian Experience,* by Silvio Goglio.

2007.28 *Importance Sampling for Sums of Lognormal Distributions, with Applications to Operational Risk,* by Marco Bee.

2007.29 *Re-reading Jevons's Principles of Science. Induction Redux,* by K. Vela Velupillai.

2007.30 *Taking stock: global imbalances. Where do we stand and where are we aiming to?* by Andrea Fracasso.

2007.31 *Rediscovering Fiscal Policy Through Minskyan Eyes,* by Philip Arestis and Elisabetta De Antoni.

2008.1 *A Monte Carlo EM Algorithm for the Estimation of a Logistic Auto-logistic Model with Missing Data,* by Marco Bee and Giuseppe Espa.

2008.2 *Adaptive microfoundations for emergent macroeconomics*, Edoardo Gaffeo, Domenico Delli Gatti, Saul Desiderio, Mauro Gallegati.

2008.3 *A look at the relationship between industrial dynamics and aggregate fluctuations*, Domenico Delli Gatti, Edoardo Gaffeo, Mauro Gallegati.

2008.4 *Demand Distribution Dynamics in Creative Industries: the Market for Books in Italy*, Edoardo Gaffeo, Antonello E. Scorcu, Laura Vici.

2008.5 *On the mean/variance relationship of the firm size distribution: evidence and some theory*, Edoardo Gaffeo, Corrado di Guilmi, Mauro Gallegati, Alberto Russo.

2008.6 *Uncomputability and Undecidability in Economic Theory*, K. Vela Velupillai.

2008.7 *The Mathematization of Macroeconomics*: *A Recursive Revolution*, K. Vela Velupillai.

2008.8 *Natural disturbances and natural hazards in mountain forests: a framework for the economic valuation*, Sandra Notaro, Alessandro Paletto

2008.9 *Does forest damage have an economic impact? A case study from the Italian Alps,* Sandra Notaro, Alessandro Paletto, Roberta Raffaelli.

2008.10 *Compliance by believing: an experimental exploration on social norms and impartial agreements,* Marco Faillo, Stefania Ottone, Lorenzo Sacconi.

2008.11 *You Won the Battle. What about the War? A Model of Competition between Proprietary and Open Source Software*, Riccardo Leoncini, Francesco Rentocchini, Giuseppe Vittucci Marzetti.

2008.12 *Minsky's Upward Instability: the Not-Too-Keynesian Optimism of a Financial Cassandra,* Elisabetta De Antoni.

2008.13 *A theoretical analysis of the relationship between social capital and corporate social responsibility: concepts and definitions,* Lorenzo Sacconi, Giacomo Degli Antoni.

2008.14 *Conformity, Reciprocity and the Sense of Justice. How Social Contract-based Preferences and Beliefs Explain Norm Compliance: the Experimental Evidence,* Lorenzo Sacconi, Marco Faillo.

2008.15 *The macroeconomics of imperfect capital markets. Whither saving-investment imbalances?* Roberto Tamborini

2008.16 *Financial Constraints and Firm Export Behavior*, Flora Bellone, Patrick Musso, Lionel Nesta and Stefano Schiavo

2008.17 *Why do frms invest abroad? An analysis of the motives underlying Foreign Direct Investments Financial Constraints and Firm Export Behavior*, Chiara Franco, Francesco Rentocchini and Giuseppe Vittucci Marzetti

2008.18 *CSR as Contractarian Model of Multi-Stakeholder Corporate Governance and the Game-Theory of its Implementation*, Lorenzo Sacconi

2008.19 *Managing Agricultural Price Risk in Developing Countries*, Julie Dana and Christopher L. Gilbert

2008.20 *Commodity Speculation and Commodity Investment*, Christopher L. Gilbert

2008.21 *Inspection games with long-run inspectors*, Luciano Andreozzi

2008.22 *Property Rights and Investments: An Evolutionary Approach*, Luciano Andreozzi

2008.23 *How to Understand High Food Price*, Christopher L. Gilbert

2008.24 *Trade-imbalances networks and exchange rate adjustments: The paradox of a new Plaza*, Andrea Fracasso and Stefano Schiavo

2008.25 *The process of Convergence Towards the Euro for the VISEGRAD – 4 Countries*, Giuliana Passamani

2009.1 *Income Shocks, Coping Strategies, and Consumption Smoothing. An Application to Indonesian Data*, Gabriella Berloffa and Francesca Modena

2009.2 *Clusters of firms in space and time*, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani e Andrea Mazzitelli

2009.3 *A note on maximum likelihood estimation of a Pareto mixture*, Marco Bee, Roberto Benedetti e Giuseppe Espa

PUBBLICAZIONE REGISTRATA PRESSO IL TRIBUNALE DI TRENTO