Fitting Mixture Models with CVaR Constraints

Giorgi Pertaia, Stan Uryasev

Risk Management and Financial Engineering Lab
Department of Industrial and Systems Engineering
303 Weil Hall, University of Florida, Gainesville, FL 32611.
E-mails: `gpertaia@ufl.edu`, `uryasev@ufl.edu`.

Abstract

Standard methods of fitting finite mixture models mostly take into account the majority of observations in the center of the distribution. This paper considers the case where the decision maker wants to make sure that the tail of the fitted distribution is at least as heavy as the tail of an empirical distribution. For instance, in nuclear engineering, where probability of exceedance (POE) needs to be estimated, it is important to fit correctly tails of the distributions. The goal of this paper is to supplement the standard methodology and to assure an appropriate heaviness of the fitted tails. We consider a new CVaR distance between distributions, that is a convex function with respect to weights of mixture. We have conducted a case study demonstrating efficiency of the approach. Weights of normal mixture are found by minimizing CVaR distance between the mixture and the empirical distribution. We have suggested convex constraints on weights, assuring that the tail of the mixture is as fat as the tail of empirical distribution.

Keywords: CVaR norm, CVaR distance, conditional value-at-Risk, CVaR, finite mixture models, normal mixtures, distance between distributions, distribution

1. Introduction

Finite mixtures of distributions are used in various areas, such as risk management and statistics. An important feature of mixture models is the ability to approximate a heavy tail of a distribution. For instance, mixture distributions are used in finance, where the loss data frequently exhibit heavy tails. Expectation Maximization (EM) is a popular algorithm for fitting mixtures of normal distributions. In general, EM solves a nonconvex optimization problem with respect to parameters of a mixture. The original EM algorithm, as defined in [8], does not allow for additional constraints in the problem.

This article derives a new methodology for fitting mixture models with constraints on the tails of the mixture distribution. The methodology is based on the concept of Conditional Value at Risk distance between distributions. We present a case study showing that this method fits the tails as specified by the constraints. Some of the results of this paper are based on the unpublished research report [4].

2. Finite Mixture and $\mathbf{CVaR}_{\boldsymbol{\alpha}}$-distances Between Distributions

Let $F_1(x, \theta_1), \ldots, F_m(x, \theta_m)$ be a set of cumulative distribution functions (CDFs), where $x \in \mathbb{R}$ and θ_i is the parameter set of a distribution F_i. The CDF of the mixture of $F_1(x, \theta_1), \ldots, F_m(x, \theta_m)$ is defined as follows.

Definition 1. *Let $\boldsymbol{p} = (p_1, \ldots, p_m)^T$ be the column vector of weights of the mixture, $\boldsymbol{p} \geq 0$ and $\boldsymbol{p}^T\mathbf{1} = 1$, the CDF of a finite mixture is defined as*

$$F_{\boldsymbol{p},\boldsymbol{\theta}}(x) = \sum_{j=1}^{m} p_j F_j(x, \theta_j). \tag{1}$$

In this definition, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ is the vector of parameters. Further, we will omit $\boldsymbol{\theta}$ from $F_{\boldsymbol{p},\boldsymbol{\theta}}(x)$ and write the CDF of the mixture as $F_{\boldsymbol{p}}(x)$. Normal distributions are usually used for construction of finite mixtures.

2.1. $\mathbf{CVaR}_{\boldsymbol{\alpha}}$ *- norm of Random Variables*

We denote the CVaR of a random variable (r.v.) X at the the confidence level $\alpha \in [0, 1)$ by $\mathrm{CVaR}_\alpha(X)$.

$$\mathrm{CVaR}_\alpha(X) = \min_C \left(C + \frac{1}{1-\alpha} E[X - C]^+ \right) \tag{2}$$

In the last formula $[x]^+ = max(x, 0)$. It can be shown that $\mathrm{CVaR}_0(X) = E(X)$. For a comprehensive analysis of the $\mathrm{CVaR}_\alpha(X)$ risk-measure see Rockafellar and Uryasev [1], [2].

We denote by $\|X\|_\alpha$ the CVaR_α-norm of X at the confidence level $\alpha \in [0, 1)$, as defined by Mafusalov and Uryasev [3]:

$$\|X\|_\alpha = \mathrm{CVaR}_\alpha(|X|). \tag{3}$$

CVaR_α-norm satisfies the following standard properties:

(i) If $\|X\|_\alpha = 0 \Rightarrow X \equiv 0$ almost surely (a.s.),
(ii) $\|\lambda X\|_\alpha = |\lambda| \|X\|_\alpha$ for any $\lambda \in \mathbb{R}$ (positive homogeneity),
(iii) $\|X + Y\|_\alpha \leq \|X\|_\alpha + \|Y\|_\alpha$ for any r.v.s X, Y, defined on the same probability space $(\Omega, \mu, \mathcal{F})$ (triangle inequality).

2.2. $\mathbf{CVaR}_{\boldsymbol{\alpha}}$ *-distance*

This section introduces the concept of CVaR_α-distance between distributions. The CVaR_α-distance was defined by Pavlikov and Uryasev [9] in the context of discrete distributions.

A *distance* function on a set V is defined as a map $d : V \times V \mapsto \mathbb{R}$ satisfying the following conditions $\forall x, y \in V$:

1. $d(x, y) \geq 0$ (non-negativity axiom);
2. $d(x, y) = 0$ if and only if $x = y$ (identity of indiscernibles);

3. $d(x, y) = d(y, x)$ (symmetry);
4. $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Assume that there are two r.v.s Y and Z, with corresponding CDFs, $F(x)$ and $G(x)$. Assume also that there is some auxiliary r.v. H with CDF $W(x)$. We define a new r.v. X^W, representing the difference between $F(x)$ and $G(x)$, as

$$X^W(F, G) = F(H) - G(H).$$

Note, that the auxiliary r.v. H may coincide with one of the r.v.s Y and Z, i.e., $W(x)$ may be equal to $F(x)$ or $G(x)$.

Definition 2. CVaR_α *distance at some confidence level* $\alpha \in [0, 1)$*, between distributions of two r.v.s* Y *and* Z *with corresponding CDFs* F_Y *and* G_Z *is defined as*

$$d_\alpha^W(F, G) = \|X^W(F, G)\|_\alpha, \tag{4}$$

where H *is an auxiliary r.v. with CDF* W_H*.*

3. Distribution Approximation by a Finite Mixture

3.1. $\mathbf{CVaR_\alpha}$ *-distance Minimization*

We approximate CDF $F(x)$ with the mixture $F_{\boldsymbol{p}}(x)$ by finding weights $\boldsymbol{p}$ in the following minimization problem:

$$\begin{aligned} & \min_{\boldsymbol{p}} d_\alpha^W(F, F_{\boldsymbol{p}}) \\ \text{s.t.} \quad & \boldsymbol{p}^T \mathbf{1} = 1 \\ & \boldsymbol{p} \geq 0 \end{aligned} \tag{5}$$

Further we prove that, function $Q(\boldsymbol{p}) = d_\alpha^W(F, F_{\boldsymbol{p}})$ is a convex function of weights $\boldsymbol{p}$.

Proposition 3.1. $Q(\boldsymbol{p}) = d_\alpha^W(F, F_{\boldsymbol{p}})$ *is a convex function of* $\boldsymbol{p}$*.*

Proof. Let $\lambda \in [0, 1]$. From the definition of $F_{\boldsymbol{p}}(x)$ and properties of CVaR norm:

$$\begin{aligned} & Q(\lambda \boldsymbol{p} + (1-\lambda)\hat{\boldsymbol{p}}) = d_\alpha^W(F, F_{\boldsymbol{\lambda p + (1-\lambda)\hat{p}}}) = \|X^W(F, F_{\boldsymbol{\lambda p + (1-\lambda)\hat{p}}})\|_\alpha = \\ & = \|F(H) - F_{\boldsymbol{\lambda p + (1-\lambda)\hat{p}}}(H)\|_\alpha = \|F(H) - \sum_{j=1}^m (\lambda p_j + (1-\lambda)\hat{p}_j) F_j(H)\|_\alpha = \\ & = \|\lambda[F(H) - \sum_{j=1}^m p_j F_j(H)] + (1-\lambda)[F(H) - \sum_{j=1}^m \hat{p}_j F_j(H)]\|_\alpha \leq \\ & \leq \lambda\|F(H) - \sum_{j=1}^m p_j F_j(H)\|_\alpha + (1-\lambda)\|F(H) - \sum_{j=1}^m \hat{p}_j F_j(H)\|_\alpha = \\ & = \lambda Q(\boldsymbol{p}) + (1-\lambda) Q(\hat{\boldsymbol{p}}). \end{aligned} \tag{6}$$

□

3.2. $\mathbf{CVaR}_{\boldsymbol{\alpha}}$ *-constraint*

This section adds CVaR_α constraints to the problem (5). With CVaR_α constraints we ensure a specified fatness of the tail. For example, if we approximate some portfolio loss distribution by a mixture, we can guarantee that the CVaR_α of the mixture will be greater than or equal to some specified threshold.

Let $X_{\boldsymbol{p}}$ be a r.v. having CDF of the mixture of distributions $F_{\boldsymbol{p}}(x)$, defined by (1).

Proposition 3.2. $\mathrm{CVaR}_\alpha(X_{\boldsymbol{p}})$ *is a concave function of* $\boldsymbol{p}$.

Proof. Using the definition of CVaR_α and X

$$\mathrm{CVaR}_\alpha(X_{\boldsymbol{\lambda p+(1-\lambda)\hat{p}}}) = \min_C \left(C + \frac{1}{1-\alpha} \int_{\mathbb{R}} \left[x-C\right]^+ \mathrm{d}F_{\boldsymbol{\lambda p+(1-\lambda)\hat{p}}}(x) \right) =$$
$$= \min_C \left(C + \frac{1}{1-\alpha} \int_{\mathbb{R}} \left[x-C\right]^+ \mathrm{d}\left(\sum_{j=1}^m (\lambda p_j + (1-\lambda)\hat{p}_j) F_j(x) \right) \right) =$$
$$= \min_C \left(C + \frac{1}{1-\alpha} \sum_{j=1}^m (\lambda p_j + (1-\lambda)\hat{p}_j) \int_{\mathbb{R}} \left[x-C\right]^+ \mathrm{d}F_j(x) \right).$$

Let

$$z_j(C) = \frac{1}{1-\alpha} \int_{\mathbb{R}} \left[x-C\right]^+ \mathrm{d}F_j(x),$$

then

$$\mathrm{CVaR}_\alpha(X_{\boldsymbol{\lambda p+(1-\lambda)\hat{p}}}) = \min_C \left(C + \sum_{j=1}^m (\lambda p_j + (1-\lambda)\hat{p}_j) z_j(C) \right) =$$
$$= \min_C \left(\lambda[C + \sum_{j=1}^m p_j z_j(C)] + (1-\lambda)[C + \sum_{j=1}^m \hat{p}_j z_j(C)] \right) \geq$$
$$\geq \lambda \min_C \left(C + \sum_{j=1}^m p_j z_j(C) \right) + (1-\lambda) \min_C \left(C + \sum_{j=1}^m \hat{p}_j z_j(C) \right) =$$
$$= \lambda \mathrm{CVaR}_\alpha(X_{\boldsymbol{p}}) + (1-\lambda)\mathrm{CVaR}_\alpha(X_{\boldsymbol{\hat{p}}}).$$

□

Further we add CVaR_α constraints to problem (5). Let u be the number of CVaR constraints, then:

$$\begin{aligned} & \min_{\boldsymbol{p}} d_\alpha^W(F, F_{\boldsymbol{p}}) \\ \text{s.t.} \quad & \mathrm{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}}) \geq \mathrm{CVaR}_{\alpha(k)}(Y), \quad k = 1, \dots, u \\ & \boldsymbol{p}^T \mathbf{1} = 1 \\ & \boldsymbol{p} \geq 0 \end{aligned} \tag{7}$$

where $\alpha(k) \in [0,1),\ k = 1,\dots,u$ are the confidence levels. The objective function in (7) is convex and the feasible region is the intersection of convex sets, thus (7) is a convex optimization problem.

3.3. ***Cardinality Constraint***

Consider CVaR_α-distance minimization problem (5) or (7) with an additional constraint on the maximum number $M \le m$ of nonzero weights in $\boldsymbol{p}$. Suppose we approximate target CDF $F(x)$ by a mixture with at most M distributions. Let us denote

$$card(\boldsymbol{p}) = \sum_{i=1}^{m} g(p_i), \text{ where } g(p_i) = \begin{cases} 1 & \text{if } p_i > 0 \\ 0 & \text{if } p_i \le 0 \end{cases}.$$

Problem (5) with cardinality constraint is rewritten as

$$\begin{aligned} & \min_{\boldsymbol{p}} d_\alpha^W(F, F_{\boldsymbol{p}}) \\ \text{s.t.} \quad & card(\boldsymbol{p}) \le M \\ & \boldsymbol{p}^T \mathbf{1} = 1 \\ & \boldsymbol{p} \ge 0 \end{aligned} \tag{8}$$

Equivalently:

$$\begin{aligned} & \min_{\boldsymbol{p}} d_\alpha^W(F, F_{\boldsymbol{p}}) \\ \text{s.t.} \quad & \sum_{j=1}^{m} r_j \le M \\ & r_j \in \{0,1\},\ j = 1,\dots,m \\ & p_j \le r_j,\ j = 1,\dots,m \\ & \boldsymbol{p}^T \mathbf{1} = 1, \\ & \boldsymbol{p} \ge 0. \end{aligned} \tag{9}$$

It is possible to transform problem (9) into a mixed integer programming problem (MIP). MIP is well studied and can be solved by various commercial solvers.

4. Case Study: Fitting Mixture by minimizing CVaR_α-distance

This section solves problem (7) that fits the finite mixture to an empirical CDF. The empirical cumulative distribution for some sample $Y = \{y_1, \dots, y_n\}$ is defined as,

$$F_n(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{y \ge y_i\}}, \tag{10}$$

where n is the number of observations and $\mathbb{1}_{\{*\}}$ is the indicator function. We used data considered in the research paper [6] and the corresponding case study [7]. We used Portfolio Safeguard (PSG) version 2.3 [5] to solve the optimization problems and MATLAB for plotting and data management. The case study codes and data are posted in [10]. We used PSGs precoded CVaR function to set the constraints on the mixture. In this case study we considered the CVaR_α-distance with $\alpha = 0$ and the normal (Gaussian) mixture. Gaussian mixture is a weighted sum of normal distributions,

$$F_{\boldsymbol{p}}(x) = \sum_{j=1}^{m} p_j \Phi(x, \mu_i, \sigma_i), \tag{11}$$

where $\Phi(x, \mu_i, \sigma_i)$ is a normal CDF with mean μ_i and standard deviation σ_i. We estimated μ_i and σ_i, with EM algorithm. The estimated parameters of the mixture are in the Table 1

j	μ_j	σ_j	p_j
1	0.0020	0.0014	0.1970
2	0.0100	0.0046	0.1882
3	0.0344	0.0144	0.2382
4	0.0583	0.0206	0.2581
5	0.0957	0.0365	0.1185

Table 1: Parameters of normal distributions in the mixture fitted with EM.

For the mixture with parameters in Table 1 and the empirical distribution, we have calculated $\mathrm{CVaR}_{0.9}$, $\mathrm{CVaR}_{0.95}$, $\mathrm{CVaR}_{0.99}$ and $\mathrm{CVaR}_{0.995}$, see Table 2.

k	$\alpha(k)$	$\mathrm{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}})$	$\mathrm{CVaR}_{\alpha(k)}(Y)$	Difference
1	90%	0.1118	0.1115	0.0002
2	95%	0.1300	0.1292	0.0007
3	99%	0.1626	0.1666	-0.0040
4	99.5%	0.1735	0.1814	-0.0079

Table 2: CVaRs of empirical distribution and normal mixture fitted by the EM algorithm. $\mathrm{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}})$ is the CVaR of mixture with confidence $\alpha(k)$ and $\mathrm{CVaR}_{\alpha(k)}(Y)$ is the CVaR of empirical distribution. The entries in "Difference" column are $\mathrm{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}}) - \mathrm{CVaR}_{\alpha(k)}(Y)$.

Table 2 column "$\alpha(k)$" contains confidence levels, in column "$\mathrm{CVaR}_{\alpha(k)}(X)$" are CVaRs of the mixture, column "$\mathrm{CVaR}_{\alpha(k)}(Y)$" contains CVaRs of the empirical distribution, "Difference" column shows the difference between CVaR of mixture and CVaR of empirical distribution ($\mathrm{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}}) - \mathrm{CVaR}_{\alpha(k)}(Y)$).

Further we minimized CVaR distance as given in Problem (7). We constrainted CVaRs of the mixture to be greater or equal to the empirical CVaRs.

$$\begin{aligned} &\min_{\boldsymbol{p}} d_\alpha^W(F_n, F_{\boldsymbol{p}}) \\ \text{s.t.} \quad &\text{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}}) \geq \text{CVaR}_{\alpha(k)}(Y), \quad k = 1, \dots, u \\ &\boldsymbol{p}^T \mathbf{1} = 1 \\ &\boldsymbol{p} \geq 0 \end{aligned} \tag{12}$$

Optimal weights of the mixture are given in Table 3.

j	p_j
1	0.1936
2	0.2911
3	0.1226
4	0.2071
5	0.1857
objective:	0.030791

Table 3: Weights of the mixture calculated with CVaR_α-distance minimization (7).

The CVaRs for the mixture with CVaR constraints are in Table 4.

k	$\alpha(k)$	$\text{CVaR}_{\alpha(k)}(X)$	$\text{CVaR}_{\alpha(k)}(Y)$	Difference
1	90%	0.126	0.1115	0.0145
2	95%	0.1428	0.1292	0.0136
3	99%	0.1715	0.1666	0.0049
4	99.5%	0.1814	0.1814	0.0000

Table 4: CVaRs of empirical distribution and normal mixture fitted by minimizing CVaR distance with CVaR constraints. $\text{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}})$ is the CVaR of mixture with confidence $\alpha(k)$ and $\text{CVaR}_{\alpha(k)}(Y)$ is the CVaR of empirical distribution. The entries in "Difference" column are $\text{CVaR}_{\alpha(k)}(X_{\boldsymbol{p}}) - \text{CVaR}_{\alpha(k)}(Y)$.

Table 4 shows that CVaR constraints are satisfied, i.e. $\text{CVaR}_{\alpha(k)}(X) \geq \text{CVaR}_{\alpha(k)}(Y), \; k = 1, \dots, 4$. However only the CVaR with $\alpha = 99.5\%$ is active ($\text{CVaR}_{99.5\%}(X) = \text{CVaR}_{99.5\%}(Y) = 0.1814$), for other CVaRs the inequality is strict.

Figure 1 shows the QQ plot of the mixture fitted with EM. The mixture is fitted well in the center of the distribution (the mixture quantiles and empirical quantiles form a straight line with 45 degree slope). However, the points corresponding to the quantiles of the tails are above the line, i.e. the mixture has thinner tails than the empirical distribution. Figure 2 shows the QQ plot of the mixture fitted with the CVaR constraints. In this case, the quantiles on tails are closer to the empirical, however the quantiles towards center are below the line, indicating that quantiles in the center of the mixture are larger than corresponding

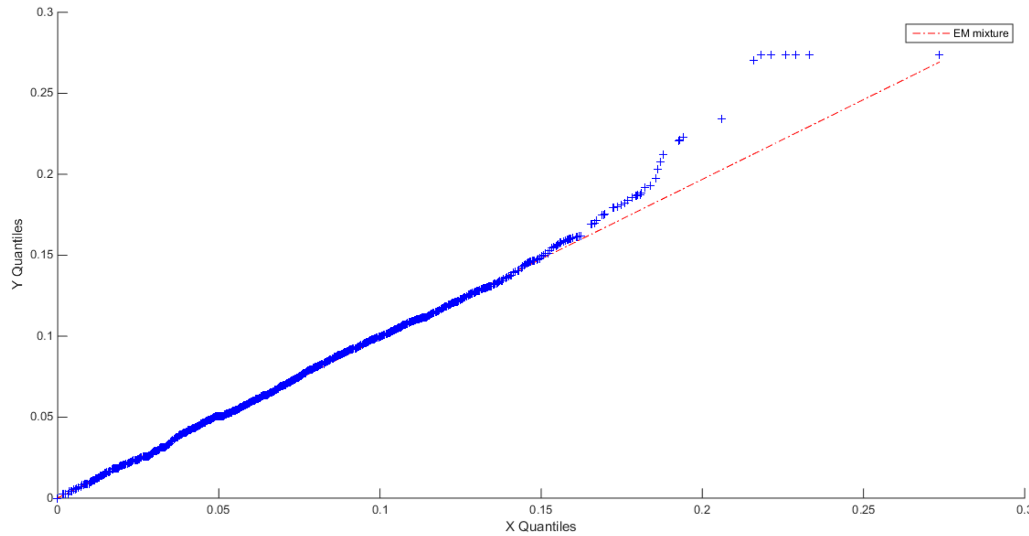

Figure 1: QQ plot of mixture with parameters calculated with EM. "X" axis shows quantiles of the mixture and "Y" axis shows quantiles of the empirical distribution.

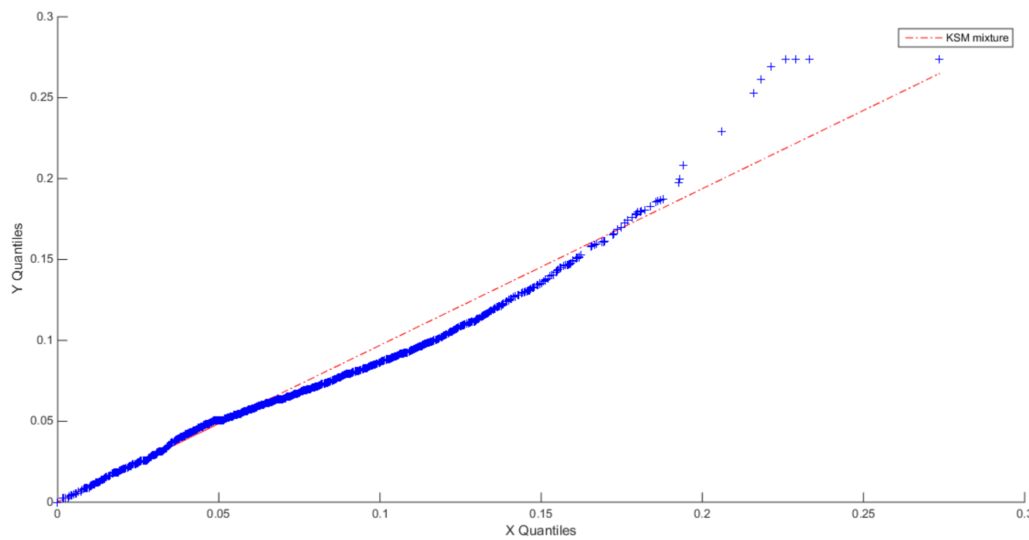

Figure 2: QQ plot of the mixture with parameters calculated by minimizing CVaR_0-distance as defined in (12). "X" axis shows quantiles of the mixture and "Y" axis shows quantiles of empirical distribution.

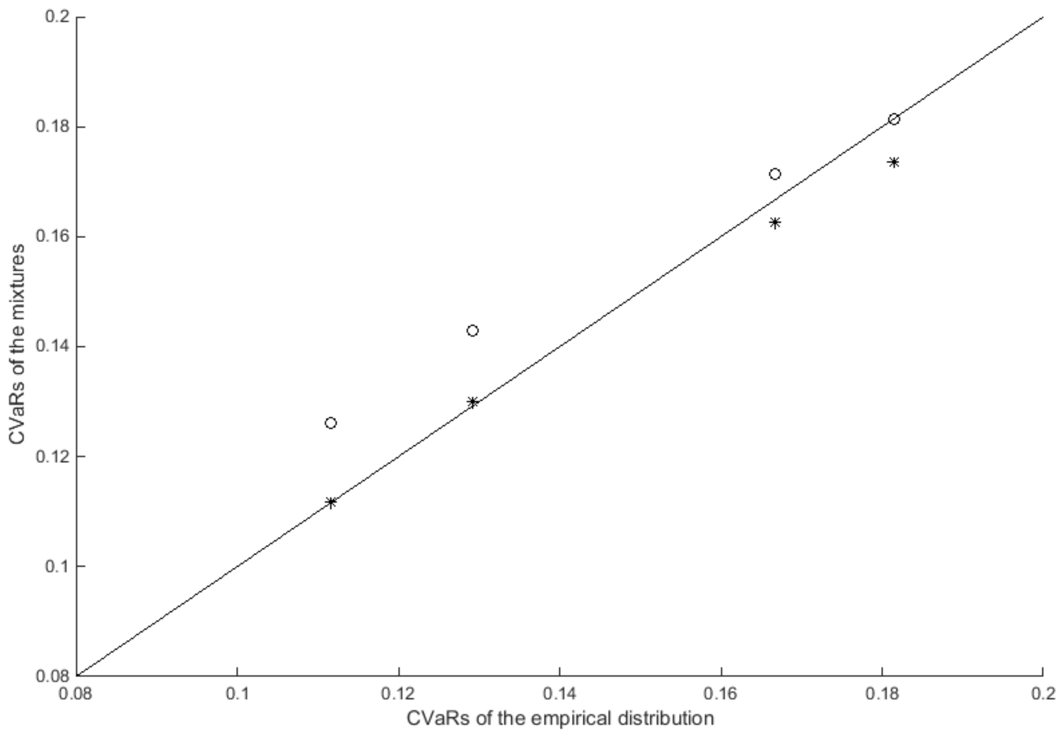

Figure 3: This is the analog of QQ plots, but CVaRs are plotted instead of the quantiles. Horizontal axis shows CVaRs of the empirical distribution and the vertical axis shows CVaRs of the mixtures. The star symbols (*) shows CVaRs of the original mixture fitted with EM. The empty circle symbols (o) shows CVaRs of mixture fitted by minimizing CVaR distance with CVaR constraints.

quantiles in the empirical distribution. Figure 3 shows CVaR to CVaR plot of the two mixtures. The CVaRs of the fitted mixture are fatter or equal to the empirical CVaRs. In this figure the points corresponding to the CVaR_αs are above the line, except for the last point, that is on the line. This indicates that only the last CVaR_α constraint ($\alpha(4) = 99.5\%$) is active and other CVaRs are fatter (larger) than specified in the right hand side of the constraints.

5. Conclusion

We presented a new method for fitting mixture distributions using CVaR distance. To assure that tails of the mixture distribution are as fat as tails of empirical distribution, we used CVaR constraints on the mixture distribution. We also considered a cardinality constraint specifying that the number of distributions with nonzero weights in the mixture is bounded by some constant. We proved that the CVaR of the mixture is a concave function with respect to the weights of mixture. The case study illustrated fitting of the mixture with CVaR constraints of 90%, 95%, 99%, 99.5% confidence levels. The case study demonstrated that the suggested procedure ensures that the tails of the fitted mixture are as fat as specified by the constraints.

6. Acknowledgments

This research was supported by the DARPA EQUiPS program, grant SNL 014150709, Risk-Averse Optimization of Large-Scale Multiphysics Systems.

References

References

[1] Rockafellar, R.T. and S. Uryasev. *Conditional Value-at-Risk for General Loss Distributions*. Journal of Banking and Finance, 26/7, 2002, 1443-1471.

[2] Rockafellar, R.T. and S. Uryasev. *Optimization of Conditional Value-At-Risk*. The Journal of Risk, Vol. 2, No. 3, 2000, 21-41.

[3] Mafusalov, A. and S. Uryasev. *CVaR (Superquantile) Norm: Stochastic Case*. European Journal of Operational Research, 249, 2016, 200-208.

[4] Zdanovskaya, V. and S. Uryasev. *Unpublished Case Study on Mixtures and CVaR Distance*. University of Florida

[5] Portfolio Safeguard, American Optimal Decisions. http://www.aorda.com/

[6] Veremyev, A., Tsyurmasto, P., and S. Uryasev *Optimal Structuring of CDO contracts: Optimization Approach*. Journal of Credit Risk, 8(4), Winter 2012/13.

[7] *Case Study: Structuring step up CDO*. Data and codes: http://www.ise.ufl.edu/uryasev/research/testproblems/financial_engineering/structuring-step-up-cdo.

[8] Dempster, A.P., Laird, N.M., Rubin, D.B (1977) *Maximum Likelihood from Incomplete Data via the EM Algorithm*. Journal of the Royal Statistical Society, Series B. 39 (1). JSTOR 2984875. MR 0501537.

[9] Pavlikov, K. and S. Uryasev. *CVaR distance between univariate probability distributions and approximation problems*. Annals of Operations Research, 2018, DOI 10.1007/s10479-017-2732-8, 1-22

[10] *Case Study: Fitting mixture models with CVaR constraints*. Data and codes: http://www.ise.ufl.edu/uryasev/research/testproblems/advanced-statistics/fitting-mixture-models-with-cvar/