

2018 INFORMS International Conference
Taipei, Taiwan, June 17–20, 2018
© 2018 INFORMS| ISBN 978-0-9906153-1-6
https://doi.org/10.1287/infp.2018.XXXX
pp. 000–000

Chance Constrained Programs with Mixture Distributions

iZhaolin Hu
School of Economics and Management, Tongji University, Shanghai 200092, China,
huzhaolin@gmail.com

iWenjie Sun
School of Economics and Management, Tongji University, Shanghai 200092, China,
r13381638770@sina.com

Abstract Chance constrained programs (CCP) are important models in stochastic optimization. In the main conventional literature on CCPs, the underlying distribution that models the randomness of the problem is usually assumed to be given in advance. However, in practice, such a distribution needs to be specified by the modelers based on the information available. This is called input modeling. In this paper we consider input modeling in CCPs. We propose to use mixture distributions to fit the data available and to model the randomness. We demonstrate the merits of using mixture distributions and show how to handle the CCPs with mixture distributions. We also consider several scenarios and conduct numerical experiments to demonstrate our approach.

Keywords chance constrained programs; input modeling; data-driven stochastic optimization; mixture distributions; mixed integer programming

1. Introduction

Stochastic optimization (SO) has become a standard modeling philosophy and framework for decision making under uncertainty. Chance constrained programs (CCP), as a basic class of SO models, have been used in various areas. When there are random parameters in the constraints of an optimization problem, it is natural to require the constraints be satisfied with a given high probability and formulate the problem into a CCP. The literature on CCPs can be dated back to Charnes et al. (1958), who first considered a single CCP, and Miller and Wagner (1965), who first considered a joint CCP. Since then, both theories and applications of CCPs have been studied extensively. For a comprehensive literature review on the topic, readers are referred to Prekopa (2003), Nemirovski and Shapiro (2006), Hong et al. (2011) and references therein. In the framework of SO, the randomness of the system is modeled by the random parameters. In the literature, standard approaches typically assume that the distribution of the underlying random vector is given in advance. For instance, we often assume the random vector follows a multivariate normal (or t) distribution. However, in practice, there is usually no distribution that is ready for use. Decision makers often need to specify the distribution based on the information available. This important process is called input modeling. Traditionally there is a gap between the input modeling part and the optimization part. Analysis in the SO literature typically focuses on the optimization part of the SO problem. For instance, while there have been extensive studies on how to solve CCPs, the input modeling for CCPs has received relatively insufficient discussion. In this paper, we make an emphasis on the input modeling issue for SO: Since the optimal solutions of SO

models critically depend on the input distributions, input modeling is a critical step of SO, and is perhaps equally important as the solution techniques. Input modeling has more deeply been studied in the area of simulation, see, e.g., Law (2013). When a simulation model is being built, the modeler wants to specify the input distribution for the random parameters that drive the stochastic system. There have been many approaches for input modeling in simulation. When there are data, we often fit some distribution to the data. When there are no data, we often need to specify the distribution based on some subjective judgement. In simulation, the data based approaches can roughly be divided into the parametric approach and the nonparametric approach. In the parametric approach, decision makers typically select a parametric distribution family, e.g., a multivariate normal distribution, and then use statistical methods to determine (usually estimate) the parameters of the distribution, e.g., the mean vector and the covariance matrix. In the nonparametric approach, decision makers do not assume the distribution belongs to any parametric family. Rather, they use some nonparametric methods to fit the distribution. For instance, the kernel density estimation methods are often used in practice.

Nowadays we have witnessed an era of data. In many real SO problems, we can often obtain data for the random parameters. Thus it is often beneficial to mining the data to specify the input distribution for SO. Such a data driven approach has been very natural and attractive. In this paper we propose such an input modeling approach for CCPs. In contrast to the parametric approach and the nonparametric approach, we propose to use mixture distributions (or called mixture models) to model the randomness of the parameters. The use of mixture distribution can be viewed as a semiparametric approach (Bishop 2006). The idea of using mixture distribution can be dated back to Pearson (1894), who proposed to use a mixture of two normal distributions to fit a set of asymmetric biological data. Pearsons approach led to new findings in the biological area. Mixture models allow for sufficient flexibility. They provide a framework of constructing more complex distributions (Bishop 2006). For instance, the asymmetrical distributions and skewed distributions can often be approximated well by the mixture models. Among various mixture distributions, we are particularly interested in the Gaussian mixture distribution, which has been studied extensively in the statistical (machine) learning literature. Gaussian mixture distribution has been widely used in various areas. Bringing in the mixture models in CCPs, we show that this approach has many merits and demonstrate that the approach can in some sense extend the current research results for CCPs. It is well known that CCPs are in general difficult to solve. The modelers often need to consider special structures of the CCPs for handling them. In this paper, we consider several scenarios and discuss how the chance constraints can be simplified accordingly. As a straightforward but important result, we show that the CCPs with mixture distributions can be transformed as a weighted sum of probability functions with component distributions of the mixture models. Based on the result, we analyze in details two important scenarios.

The rest of this paper is organized as follows. In Section 2 we introduce the mixture models and analyze the CCPs with the mixture models. In Section 3 we consider two scenarios and demonstrate the potential applications of our approach. We conduct some preliminary numerical experiments in Section 4. Section 5 concludes the paper.

2. Mixture Distributions and Model Structures

2.1. Probability Function with Mixture Distribution

We start from the well known model chance constrained program (CCP):

$$\underset{x\in X}{\text{minimize}} \quad h(x) \tag{1}$$

$$\text{subject to } \Pr_{\sim P_*}\{A(x,\xi)\} \geq 1-\alpha. \tag{2}$$

In Problem (1), x is the decision vector, which belongs to a set $X \subset \Re^d$, $h : \Re^d \to \Re$ is a real-valued function which models the objective, ξ is a k-dimensional vector of uncertain parameters and the support of ξ, denoted as Ξ, is a closed subset of $\Re^k$, $A(x, \xi)$ is a random event (driven by ξ) that depends on x, and the notation $\Pr_{\sim P_*}$ denotes that the probability is taken with respect to (w.r.t.) P_*, where P_* is usually called the underlying distribution. Constraint (2) is called a chance constraint. A typical instance of $A(x, \xi)$ is the following

$$A(x, \xi) := c_1(x, \xi) \leq 0, \cdots, c_m(x, \xi) \leq 0,$$

where $c_i : \Re^d \times \Xi \to \Re, i = 1, \cdots, m$ are real-valued functions. In this setting, (1) is called a single CCP if $m = 1$ and a joint CCP if $m > 1$. In such a scenario, (2) requires that the m constraints are satisfied at least with a probability $1 - \alpha$. This is a very natural requirement in mathematical programs when random parameters are involved in the constraints. Throughout this paper we assume that X is a convex and compact set, which may be defined by some deterministic constraints, and the function h can be handled by deterministic optimization techniques. Thus, the major difficulty of solving (1) comes from the chance constraint (2).

When the CCP (1) is being studied, the underlying distribution P_* is typically assumed to be given. However, as discussed in Section 1, in practice, such a distribution can only be inferred. In this paper, we model that P_* is a mixture distribution that takes the following expression

$$p_*(z) = \sum_{j=1}^{K} \pi_j p_j(z), \tag{3}$$

where $p_j(z), j = 1, \cdots, K$ are probability distributions called component distributions, and $\pi_j > 0, j = 1, \cdots, K$ are mixing coefficients satisfying $\sum_{j=1}^{K} \pi_j = 1$. We also call (3) a mixture model. Note that throughout the analysis we do not differentiate the notations P_* and p_* when there is no confusion. Both notations refer to the same distribution. In (3), P_* and p_* could be probability density functions or probability mass functions. As noted, we also call (3) a mixture model. Mixture models have been studied for long in the area of statistics. For instance, McLachlan and Peel (2000) studied finite mixture models and built various interesting results for the models.

2.2. Gaussian Mixture Distribution

A most well known instance of the mixture models is the Gaussian mixture distribution. The Gaussian mixture model takes the following form

$$p_*(z) = \sum_{j=1}^{K} \pi_j \mathcal{N}(z | \mu_j, \Sigma_j), \tag{4}$$

where each component $\mathcal{N}(z | \mu_j, \Sigma_j)$ is a (multivariate) normal distribution with mean vector μ_j and covariance matrix Σ_j. Gaussian mixture distribution is widely used in machine learning literature, see, e.g., Bishop (2009). Using Gaussian mixture models to approximate a distribution has been very popular in the literature. The approximation theory has also been studied, see, for instance, Zeevi and Meir (1997). In practice, the random parameters in CCP sometimes follow a continuous skewed distribution, an asymmetrical distribution, or a multimodal distribution. These distribution structures typically hide in the data that are collected. Consequently, conventional parametric families, e.g., a multivariate normal distribution or an exponential family may not fit the data well. In such situations, it may be better to use some mixture distribution to model the randomness. As indicated in the literature, there are several issues for the use of mixture models. One essential issue is how to fit the mixture models to the data. There are a number of approaches that have been developed for such fitting. A widely used method is the expectation maximization (EM)

method, see, for instance, Bishop (2006). It is well known that such EM algorithm will converge to some local maxima of the log likelihood function. The EM algorithm has been imbedded in some softwares. For instance, in Matlab, we can directly call the EM algorithm package to conduct the estimation for the Gaussian mixture model.

2.3. Model Structures

While the idea of CCP is conceptually very natural and fits decision making very well, solving a CCP is often a nontrivial task. In general, CCPs are challenging problems. There are two main streams of methodologies for tackling CCPs. One stream is to use a deterministic optimization approach, which has been explored extensively in the optimization literature, see, e.g., Nemirovski and Shapiro (2006). In this approach, some deterministic optimization model is often constructed to somehow approximate the CCP. The deterministic model is then solved and the solution is used to approximate that of the original problem. The other stream is to use a simulation approach, see, e.g., Pagnoncelli et al. (2009) and Hong et al. (2011).

The simulation approach is typically general and may be used to handle various problem structures and distributions. Since our mixture model can be viewed as a general distribution, the simulation approach may be directly used to handle the problem. However, the detailed procedures of this approach often significantly depend on the probability function structures and the underlying distribution. Therefore, in the literature, the CCPs are usually assumed to have certain structures. In this paper, we explore certain structures and properties of Constraint (2), and use them to design efficient procedures to solve Problem (1) under different scenarios. As a simple but important result, we have the following theorem.

Theorem 1. *Suppose the distribution p_* takes the form of (3). Then*

$$\Pr_{\sim P_*}\{A(x,\xi)\} = \sum_{j=1}^{K} \pi_j \Pr_{\sim P_j}\{A(x,\xi)\}.$$

Proof: Let $\mathbb{1}_{\{A(x,\xi)\}}$ denote the indicator function, which equals 1 if $A(x,\xi)$ happens and 0 otherwise. Then

$$\Pr_{\sim P_*}\{A(x,\xi)\} = \mathrm{E}_{P_*}\left[\mathbb{1}_{\{A(\xi)\}}\right] = \sum_{j=1}^{K} \pi_j \mathrm{E}_{P_j}\left[\mathbb{1}_{\{A(\xi)\}}\right] = \sum_{j=1}^{K} \pi_j \Pr_{\sim P_j}\{A(x,\xi)\}.$$

This concludes the proof of the theorem.

Theorem 1 shows that the probability function with a mixture distribution can be expressed as the summation of probability functions with the underlying distributions being the component distributions. This result, from a crude simulation perspective, does not simply solve the problem. However, it has interesting applications when each term in the summation can be simplified or handled efficiently. For instance, when the distributions $p_j, j = 1, \cdots, K$ are normal, sometimes more efficient methods, than a crude simulation method, may be used to compute (or estimate) the probability function values.

Problem (1) is a nonlinear optimization problem. A natural approach of solving the problem is to use some nonlinear optimization techniques. A general nonlinear optimization algorithm typically requires the function values and the function gradients. For CCP, these values are often difficult to be derived analytically and thus need to be estimated. In the literature, various methods have been proposed for estimating the gradient of the probability function. The difficulties of the estimation procedures of course depend on the structures of the probability function. But in general, estimating the gradient of a probability function is a challenging problem. In this paper, we consider the gradient estimation of the chance constraint with mixture distribution. Since the probability function can be separated, as stated in Theorem 1, the gradient can also be separated directly. From Theorem 1 we immediately have the following theorem.

Theorem 2. *Suppose that the distribution P_* takes the form of (3). Then*

$$\nabla_x \Pr_{\sim P_*}\{A(x,\xi)\} = \sum_{j=1}^{K} \pi_j \nabla_x \Pr_{\sim P_j}\{A(x,\xi)\}.$$

Theorem 2 shows that the gradient of the probability function with a mixture distribution is simply the weighted sum of the gradients of the probability functions with the component distributions. Suppose the gradient of the probability function with each component distribution can be handled efficiently. Then Theorem 2 guarantees that we can handle the gradient for the mixture model efficiently. In next section we will give concrete examples that support the statement.

3. Solution Techniques for Different Scenarios

We have introduced CCPs with mixture distributions and have analyzed the model structures. In this section, we consider several scenarios in which the structured CCPs are used to model various practical decisions. We mainly analyze how the chance constraint can be handled to encompass various optimization techniques.

3.1. Linear Separable Randomness and Decision Vector

We first consider a class of CCPs that have very simple structures. The chance constraint in the models has the following expression

$$\Pr_{\sim P_*}\{B\xi \le x\} \ge 1-\alpha \tag{5}$$

where B is a matrix of appropriate dimension. Note that the constraint above is indeed a joint chance constraint. In the chance constraint the random vector ξ and the decision vector x are separated. This class of CCPs have been studied widely in the literature, see, e.g., Van Ackooij et al. (2010), Henrion and Möller (2012), and Van Ackooij et al. (2014). Henrion and Moller (2012) proposed that calculations of the gradient can be reduced to calculations of the probability function values. This result is derived for the Gaussian distribution case. For a general distribution, it is not known how to derive similar results. However, based on Theorem 2 we can obtain a similar gradient formula for the probability function with the Gaussian mixture distribution. To ease the analysis, we define

$$F^j(x) = \Pr_{\sim P_j}\{B\xi \le x\}, \;\; F^*(x) = \Pr_{\sim P_*}\{B\xi \le x\}.$$

Theorem 3.2 (or Theorem 4.1) of Henrion and Möller (2012) built an expression for the gradient of $F^j(x)$. We thus take this gradient expression $\nabla_x F^j(x)$ as given. Based on Proposition 1, we can derive an expression for the gradient of $F^*(x)$. We summarize the result in the following proposition.

Proposition 1. *Suppose the distribution P_* is a Gaussian mixture distribution of the form (4) with all Σ_j being positive definite. Then*

$$\nabla_x F^*(x) = \sum_{j=1}^{K} \pi_j \nabla_x F^j(x).$$

Proposition 1 can be viewed as a generalization of Theorem 3.2 of Henrion and Möller (2012). Given that, theoretically, Gaussian mixture distribution can approximate any distributions, such a generalization is very useful. For instance, in solving Problem (1) with constraint (5) by some nonlinear optimization methods, we can implement Proposition 1 to estimate the gradient for the chance constraint whose randomness can be approximated by Gaussian mixture distribution.

3.2. Mixed Integer Linear Product Form

We next consider a class of CCPs where part of the decision variables are integer-valued. This structured model is called a mixed integer CCP. Many real problems may be formulated as mixed integer CCPs. However, the problems are typically very difficult to solve. One rare tractability case is when the underlying distribution of the chance constraint follows a multivariate normal distribution. Let $f(x) = (f_1(x), \cdots, f_d(x))^\top$. Consider the following CCP

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to } \Pr_{\sim P_*}\left\{f_0(x) + f(x)^\top \xi \le 0\right\} \ge 1-\alpha, \\ &\qquad\qquad x \in Z^{d_1} \times \Re^{d_2}, \end{aligned} \tag{6}$$

where Z is the set of integers and $d = d_1 + d_2$. The model requires that the first d_1 elements of x are integer-valued and the rest d_2 elements are continuous. Of course, one can replace Z^{d_1} with $\{0,1\}^{d_1}$, i.e., the first d_1 elements of x are $0-1$ valued. Suppose that the distribution P_* follows a multivariate normal distribution $\mathcal{N}(z|\,\mu, \Sigma)$. Then the chance constraint is equivalent to the following constraint

$$\Phi^{-1}(1-\alpha)\sqrt{f(x)^\top \Sigma f(x)} + \left(f_0(x) + f(x)^\top \mu\right) \le 0,$$

where Φ^{-1} is the inverse of the standard normal distribution function. Suppose further that $f_0(x)$ and $f(x)$ are linear in x and $\alpha \le 0.5$. Then the constraint is known as a second order cone constraint. In this scenario, we can transform (6) into a so called mixed integer conic program (MICP). MICP has been studied in recent years. It can be solved by using the branch-and-bound type procedures, usually imbedded in softwares. To the best of our knowledge, there is relatively scarce study for Problem (6) for other distributions.

In this paper, we consider the Gaussian mixture distribution and conduct some analysis for such a scenario. Suppose that the distribution P_* takes the form of (4). Then we have the constraint

$$\sum_{j=1}^{K} \pi_j \Pr_{\sim P_j}\left\{f_0(x) + f(x)^\top \xi \le 0\right\} \ge 1-\alpha. \tag{7}$$

For the general case where $K > 1$, we can no longer derive the second order cone structure. In the rest of this subsection we introduce some heuristic procedures based on the structure of Gaussian mixture distribution to handle the difficulty. Note that by introducing an auxiliary decision vector $y = (y_1, \cdots, y_K)^\top$, we can transform the chance constraint (7) equivalently to

$$\Pr_{\sim P_j}\left\{f_0(x) + f(x)^\top \xi \le 0\right\} \ge y_j, \ \sum_{j=1}^{K} \pi_j y_j \ge 1-\alpha.$$

Then based on the analysis above, we can obtain the following equivalent formulation of Problem (6):

$$\begin{aligned} &\underset{x\in X}{\text{minimize}} \quad h(x) \\ &\text{subject to } \Phi^{-1}(y_j)\sqrt{f(x)^\top \Sigma_j f(x)} + \left(f_0(x) + f(x)^\top \mu_j\right) \le 0, \\ &\qquad \sum_{j=1}^{K} \pi_j y_j \ge 1-\alpha \\ &\qquad x \in Z^{d_1} \times \Re^{d_2}. \end{aligned} \tag{8}$$

Problem (8) becomes a deterministic optimization problem with some second order cone structures. This provides a potential to use mixed integer techniques to solve the problem.

To formally discuss the solution procedure, we define the following optimization problem parametrized by y, with decision vector being x:

$$\begin{aligned}(P(y)) \quad & \underset{x\in X}{\text{minimize}} \quad h(x) \\ & \text{subject to } \Phi^{-1}(y_j^k)\sqrt{f(x)^\top \Sigma_j f(x)} + \left(f_0(x) + f(x)^\top \mu_j\right) \le 0, \\ & \qquad x \in Z^{d_1} \times \Re^{d_2}.\end{aligned}$$

We focus on the linear case, e.g., $h(x)$, $f_0(x)$ and $f(x)$ are all linear in x. We propose the following heuristic procedure to handle Problem (8).

Sequential MICP

Step 1. Set $k = 1$. Select y^k such that $\sum_{j=1}^K \pi_j y_j^k \ge 1-\alpha$.
Step 2. Solve Problem $(P(y^k))$ and obtain an optimal solution x^k. Let J_1 be the set of j for which the inequality is binding, and J_2 be the set of j for which the inequality strictly holds.
Step 3. Set $k = k+1$. Decrease those y_j for $j \in J_1$ and increase those y_j for $j \in J_2$. Obtain a reallocation y^k such that the constraints in Problem $(P(y))$ are satisfied for y_k and x^{k-1}.
Step 4. Stop if some terminating criterion is satisfied. Otherwise, go back to Step 2.

In Sequential MICP Algorithm, we need to solve an integer program in each iteration. Suppose all $y_j^k \ge 1/2$. Then Problem is a mixed integer SOCP and we can use the conventional softwares such as CPLEX to directly solve it. However, when some of y_j^k is less than $1/2$, the problem may become non-convex. In such a scenario, we need to design procedures to handle it. Moreover, the search trajectory of the algorithm will also depend on how to update y^k in Step 3.

4. Numerical Illustrations

In this section, we conduct some preliminary numerical experiments for the scenario in which constraint (2) is of continuous linear product form

$$\Pr_{\sim P_*}\left\{f_0(x) + f(x)^\top \xi \le 0\right\} \ge 1-\alpha \tag{9}$$

Thus, with constraint (9) and P^* given by (4), Problem (1) can be transformed as

$$\begin{aligned}& \underset{x\in X\subseteq \Re^d}{\text{minimize}} \quad h(x) \\ & \text{subject to } \sum_{j=1}^K \pi_j \Phi\left(\frac{-[f_0(x) + f(x)^\top \mu_j]}{\sqrt{f(x)^\top \Sigma_j f(x)}}\right) \ge 1-\alpha,\end{aligned} \tag{10}$$

In this specific example, we set $h(x)$ to be a linear function $c\top x$, $f_0(x)$ to be constant and $f(x) = (f_1(x), \ldots, f_k(x))^\top$ to be linear function (vector) $a + B^\top x$ where a is a constant vector and B is a $k \times d$ matrix where d is the dimension of decision variable x. For ease of analysis, we let $g(x) = \Pr_{\sim P_*}\left\{f_0(x) + f(x)^\top \xi \le 0\right\}$. To solve Problem (10), we make full use of some nonlinear optimization methods, which requires the function values and the gradients of the chance constraint. It is easy to verify that $\nabla h(x)$ has the following closed form:

$$\nabla h(x) = \sum_{j=1}^K \pi_j \varphi\left(\frac{-[f_0(x) + f(x)^\top \mu_j]}{\sqrt{f(x)^\top \Sigma_j f(x)}}\right)\left(\frac{f(x)^\top \mu_j B^\top \Sigma_j f(x) - B^\top \mu_j f(x)^\top \Sigma_j f(x)}{\sqrt[3]{f(x)^\top \Sigma_j f(x)}}\right)$$

where $\varphi(x)$ refers to the density function of the standard normal distribution. Then, we can embed the $\nabla h(x)$ in the nonlinear algorithm using Matlab optimization function(fmincon).

We construct two multivariate normal distributions with different mean vectors and covariance matrices. We then use the two distributions to construct three different Gaussian mixture distributions whose mixing coefficients vary from each other. We call the fmincon function in Matlab to solve Problem (10). The optimal values derived for the three mixture models are listed in Table 1.

Table 1. The optimal value of model (10) over different component proportions

component proportions	[0.3 0.7]	[0.5 0.5]	[0.7 0.3]
optimal value	-535.5741	-510.2027	-1112

Because Problem (10) is typically a nonconvex optimization problem, the optimal values reported in Table 1 may be local optimal values. Therefore, in practice, we often consider multiple starting points and run the fmincon function to find the best local optimal values.

5. Conclusions

In this paper, we studied using mixture distributions to model the randomness in CCPs. We analyzed the structures and properties of CCPs with mixture pdistributions. Based on the results we studied several scenarios and discussed how to handle the CCPs in these scenarios efficiently. We also conducted a simple numerical experiment to demonstrate the implementation of our approach. We expect that the approach proposed in this paper could generalize the scope and applicability of CCPs.

References

Bishop CM (2006) *Pattern Recognition and Machine Learning* (New York, NY: Springer).

Charnes A, Cooper WW, Symonds GH (1958) Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil. *Management Science* 4:235-263.

Henrion R, Möller A (2012) A gradient formula for linear chance constraints under Gaussian distribution. *Mathematics of Operations Research* 37(3):475-488.

Hong LJ, Yang Y, Zhang L (2011) Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach. *Operations Research* 59:617-630.

Law AM (2013) *Simulation Modeling and Analysis* (New York, NY: McGraw-Hill).

McLachlan GJ, Peel D (2000) *Finite Mixture Models* (New York, NY: Wiley).

Miller LB, Wagner H (1965) Chance-constrained programming with joint constraints. *Operations Research* 13:930-945.

Nemirovski A, Shapiro A (2006) Convex approximations of chance constrained programs. *SIAM Journal on Optimization* 17:969-996.

Pagnoncelli BK, Ahmed S, Shapiro A (2009) Sample average approximation method for chance constrained programming: Theory and applications. *Journal of Optimization Theory and Applications* 142:399-416.

Pearson K (1894) Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London, A* 185:71-110.

Prékopa A (2003) Chapter 5: Probabilistic programming. Ruszczynski, A. Shapiro, eds., *Stochastic Programming*, Handbooks in OR&MS (Amsterdam: Elsevier).

Van Ackooij, Henrion WR, Möller A, Zorgati R (2010) On probabilistic constraints induced by rectangular sets and multivariate normal distributions. *Mathematical Methods of Operations Research* 71(3):535-549.

Van Ackooij, Henrion WR, Möller A, Zorgati R (2014) Joint chance constrained programming for hydro reservoir management *Optimization and Engineering* 15(2):509-531.

Zeevi AJ, Meir R (1997) Density estimation through convex combinations of densities: approximation and estimation bounds. *Neural Networks* 10(1):99-109.