Chance constrained nonlinear optimization with skewed distributions and dependent rows *

Jia Liu†, Shen Peng‡, Abdel Lisser §

January 4, 2022

Abstract

This paper discusses chance constrained optimization problems where the constraints are linear to the random variables but nonlinear to the decision variables. For the individual nonlinear chance constraint, we derive tractable reformulation under finite Gaussian mixture distributions and design tight approximation under the generalized hyperbolic distribution. For the joint nonlinear chance constraint, we study several dependence structures between different rows, including independence, comonotonicity, countermonotonicity, and partial dependence.
Keywords: chance constraints, mixture distributions, generalized hyperbolic distributions, copulas, comonotonicity.

1 Introduction

In this paper, we consider a class of chance constrained nonlinear optimization problems where the chance constrained inequalities are linear to the random variables but nonlinear to the decision variables.

$$\text{(JP)} \quad \min_{x\in X} \quad g_0(x) \tag{1}$$

$$\text{s.t.} \quad \mathbb{P}\left(\sum_{i=1}^{I_k} c_i^k g_i^k(x) \le \beta_k, k = 1,\dots,K\right) \ge 1-\epsilon. \tag{2}$$

Here $1-\epsilon$ is a prespecified probability with $\epsilon < 0.5$, known as a confidence level. $\beta_k \in \mathbb{R}$, $k = 1,\dots,K$. c_i^k is a random variable and $g_i^k(x)$ is a real valued function, $i = 1,\dots,I_k$, $k = 0,\dots,K$. X is a compact set in $\mathbb{R}^n$. When $K \ge 2$, we name (2) a joint chance constraint and when $K = 1$, (2) reduces to an individual chance constraint, which is a special case of a joint chance constraint.

Chance constraints are first studied by Charnes and Cooper (1959) with wide applications in many areas, such as finance and industry. For instance, they are used to control the bankrupt

*This research was supported by the National Natural Science Foundation of China under grant numbers 11991023 and 11901449.

†School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, P. R. China (jialiu@xjtu.edu.cn, http://gr.xjtu.edu.cn/web/jialiu).

‡Optimization and Systems Theory, Department of Mathematics, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden (shenp@kth.se).

§Université Paris-Saclay, CNRS, CentraleSupelec, Laboratoire de Signaux et Systemes, Bat. Breguet, 3, rue Joliot Curie, 91190 Gif-sur-Yvette, France (abdel.lisser@l2s.centralesupelec.fr,https://l2s.centralesupelec.fr/u/lisser-abdel)

probability in portfolio selection problems [61] and to guarantee the stable operation of a power system with a large probability in power system scheduling problems [57].

The individual chance constraint is a special case of joint chance constraints. Thus, it is often the first step to study chance constraint. When the random variables and the decision variables are separable, some individual chance constraints can be reformulated as a deterministic equivalent problem by using the inverse distribution function [48]. Under the assumptions of Gaussian (normal) distribution, Student's T distribution or Q-radial distribution, a linear individual chance constraint is equivalent to a second-order cone constraint [8, 48]. It is also an important issue to study the convexity of the feasible region restricted by the chance constraint. Under the elliptical distributions, logarithmic concave distributions or r-concavity distribution based assumptions, e.g., Gaussian, Student's T, Wishart, Dirichlet, the linear individual chance constraint is convex when the confidence level is greater than 0.5 [24, 53]. For nonseparable problems, the linear individual chance constraint under elliptical distributions is convex when the confidence level is large enough [56].

Joint chance constraints are more troublesome. For example, a simple linear joint chance constraint with Gaussian distributed coefficient matrix is not convex [53]. In the simplest case when the random variables in different rows are pairwise independent, a linear joint chance constraint is equivalent to a deterministic optimization problem with bilinear terms [16]. Still for some special cases, joint chance constrained problems are convex. Prékopa et al. [49] proved that when the random variables follow logarithmic concave distribution, a joint separable chance constraint with quasi-convex left-hand side is convex. Henrion and Strugarek [25] proposed the concept of r-decreasing density, and proved that under the condition of independent rows and $(r+1)$-decreasing density function, a joint separable chance constraint with (-r)-convex left-hand-side term is convex. Liu et al. [36] showed that a joint chance constrained geometric optimization problem has an equivalent convex reformulation when the matrix rows are independent and normally distributed. There is a scarce literature on joint chance constraints. In the absence of general convexity conclusions due to the nonlinearity and non-continuity of the problem, how to address the characteristics of various special types of joint chance constraints is an important research topic.

From the computational perspective, (joint) chance constraints are essentially equivalent to constraints with combinations of (multiple) integral operators and indicator functions. The non-continuity and non-convexity are the key issues bringing intractability. Therefore, finding an effective approximation algorithm is another central research issue in this area.

One class of approximation approaches is to use samples, including sample average approximation [53] and scenario approach [45]. For example, using a discrete distribution generated from samples to replace the original distribution, the sample average approximation method provides a mixed integer programming approximation. Such approximation is asymptotically tight when the sample size goes to infinity, see [39, 40] for the approximation error analysis. Ahmed et al. [2] further relaxed nonanticipativity constraints in the mixed-integer programming formulation and proposed new Lagrangian dual problems/bounds. Kannan and Luedtke [32] used the smoothing technique to approximate the indicator function in the chance constraint, and then designed a stochastic approximation method based on the projected stochastic subgradient algorithm. Campi and Garatti [9] used the scenario approximation method which uses some deterministic constraints under finite samples to approximate chance constraints. They studied the relationship between the sample size and the approximation error. Cheng et al. [14] approximated partial random variables with samples to derive tractable reformulations when the chance constraint contains two groups of independent random variables. Peña-Ordieres et al. [47] used samples to approximate the quantile function reformulation of chance constraints and proposed an Sℓ_1QP-type trust-region method to solve joint

chance constraints. The sample-based approximation methods are often practically intractable as they require a high number of samples in order to ensure the stability and the precision of the approximation problems.

Therefore, many researchers investigates some convex approximation methods with better tractability. Nemirovski and Shapiro [46] studied Bernstein approximation and Conditional Value-at-Risk (CVaR) function to approximate independent chance constraints. Cheng and Lisser [16] used second-order cone programming to approximate a joint linear chance constraint with independent and normally distributed rows. Hong et al. [28] used the difference of convex (DC) programming to approximate joint nonlinear chance constraints, and designed a sequential convex approximation algorithm. Ahmed et al. [2] proposed the ALSO-X algorithm for chance-constrained stochastic programs based on hinge-loss approximation and relaxing nonanticipativity constraints. Liu et al. [36] proposed geometric program approximation for joint chance constrained geometric programs with independent and normally distributed rows. Other promising solution methods include Benders' decomposition approach [1], line search method with a gradient formulae [55], Boolean programming approach [34], p-efficient point method [53].

The distribution of random variables in chance constraints plays an important role. It is worth noting that there is a large number of recent studies which deal with the case where the distribution of the random variables is unknown or ambiguous, namely distributionally robust approch. Many contributions are made to study the properties and to get tractable reformulations of distributionally robust chance constraints under different kinds of uncertainty sets. For instance, the well-studied moments based uncertainty sets [8, 13, 62], uncertainty sets with mean and dispersion information [23], uncertainty sets involving uni-modal information [35], mixture distribution based uncertainty sets [11], phi-divergence based uncertainty sets [31] and the recent popular Wasserstain distance based uncertainty sets [10, 30, 59]. Recently, disributionally robust geometric chance constrained optimization with several uncertainty sets including nonnegative support and the dependence of the matrix rows was considered in [38].

To the best of our knowledge, we think that there are two important issues not fully investigated in existing chance constraint literature: the skewness and kurtosis of the underlying distribution and the dependence structure between different rows. Our objective in this paper is to fill this gap, i.e., we firstly focus on the skewness and kurtosis of the underlying distribution, and then we consider the dependance issues.

A major part of the chance constraints literature is dedicated to the tractable symmetric distributions, e.g., Gaussian or elliptical distributions family [16, 36, 55, 56]. However, symmetrical distributions are generally not suitable for many practical distributions. For instance, in power system scheduling problems, both wind power forecast errors and load forecasting errors are generally not normally distributed [27], and can be better fitted by generalized hyperbolic (GH) distributions. Moreover, many financial data are often non-Gaussian but leptokurtosis and skewed distributed. For instance, the stock return rates are often modeled by left-skewed, high kurtosis and fat-tailed distributions [52]. Birge and Chavez-Bedoya [5] studied the asset allocation problem under the generalized hyperbolic distribution of returns and exponential utility. Eberlein and Prause [19] showed that generalized hyperbolic distribution can efficiently correct the smile effect in pricing models induced by traditional Gaussian distribution assumptions.

Although continuous non-symmetric distributions, e.g., mixture distribution, skewed T distribution, or generalized hyperbolic distributions are more suitable in many applications, chance constraint problems under such non-symmetric distributions are not well considered in the literature. However, mixture distributions are mostly studied among all non-symmetric distributions. Hu et al. [29] considered the individual chance constraint under Gaussian mixture distribution and pro-

pose a branch-and-bound solution method. Chen et al. [11] studied individual chance constrained under Gaussian mixture distribution but with ambiguous mixture parameters. As for the generalized hyperbolic distribution, it has six parameters and its density function is composed by Basel function and a fractional structure. Thus, it is rather difficult to directly derive a deterministic reformulation. Cheng et al. [12] used Chebyshev inequality to approximate an individual chance constraint under the generalized hyperbolic distribution but the bounds are not tight.

In most of the aforementioned studies for joint chance constraints, the random variables in different rows are assumed to be pairwise independent [16, 36] or generated from the same random event, that is the key comonotonicity assumption of the sample average approximation approach [39].

However, in many practical joint chance constrained problems, random variables between different rows are generally nonlinearly dependent from each other. For example, in an asset allocation problem for a multinational company, the random investment returns in capital markets of different countries are often highly nonlinear, asymmetric, or tail dependent [41]. Watanabe and Ellis [58] used covariance to characterize the linear dependence structure between different rows in a joint chance constrained problem and designed a gradient-based algorithm. Xie et al. [60] studied optimized Bonferroni approximations of distributionally robust joint chance constraints. Henrion and Strugarek [26] used copulas to characterize the nonlinear dependence structure of random variables between different rows. Given the log exp-concave property of the copula function, they obtained a sufficient condition of convexity for a joint chance constrained problem with separated random vector. However, when random vectors and decision vectors are not separable, there are still several bottlenecks in the theoretical study of joint chance constrained problems. Cheng et al. [15] proposed a semidefinite programming approximation method for a quadratic joint chance constrained programming with Archimedean copula dependence structure. However, the approximation bounds might not be tight in some cases.

The rest of the paper is organized as follows: In Section 2, we study the mathematical properties of individual chance constraints under finite mixture Gaussian distribution and derive tight convex reformulations or sequence convex approximation method under different sub-cases. Section 3 extends the solution methods under finite mixture problems to the case of generalized hyperbolic distributions. Deterministic reformulations of joint dependent chance constraints are given in Section 4. Numerical results of our different approaches are presented in Section 5.

2 Individual chance constraints under finite mixture Gaussian distribution

In this section, we consider a class of individual chance constrained nonlinear programs where the nonlinear constraints are linear to the random variables but nonlinear to the decision variables.

$$\text{(IC)} \qquad \min_{x\in X} \quad g_0(x) \tag{3}$$

$$\text{s.t.} \quad \mathbb{P}\left(\sum_{i=1}^{I} c_i g_i(x) \le \beta\right) \ge 1-\epsilon. \tag{4}$$

Here $1-\epsilon$ is a prespecified probability with $\epsilon < 0.5$. X is a closed and bounded set in $\mathbb{R}^n$. $\beta \in \mathbb{R}$, c_i are random variables and $g_i(x)$ are real valued functions, $i = 1, \ldots, I$.

We denote $\boldsymbol{c} = [c_1, c_2, \ldots, c_I]^\top$, and f is the joint density function of $\boldsymbol{c}$. We assume that $\boldsymbol{c}$ follows a finite mixture Gaussian distribution, which can capture non-Gaussian features of random variables, e.g., the well-known skewness and leptokurtotic nature of financial return rates, and the varying dependence structure of multiple assets (see, for instances, [7]).

A mixture distribution is a convex combination of the components of given distributions. Let $f_1, \cdots, f_M$ denote densities of some basic Gaussian distributions and $w_1, \cdots, w_M$ denote the corresponding mixture weights. Then the density of a mixture Gaussian distribution is $f(\cdot) = \sum_{m=1}^{M} w_i f_m(\cdot)$, which implies that the random variable follows a distribution of density $f_m(\cdot)$ with probability w_m.

Assumption 2.1. *We assume that*

$$f(\boldsymbol{x}) = \sum_{m=1}^{M} w_m f_m(\boldsymbol{x}),$$

where $f_m(\boldsymbol{x})$ is the density function of an I-dimensional normal distribution with mean vector $\boldsymbol{\mu}_m = [\mu_1^m, \mu_2^m, \ldots, \mu_I^m]^\top$ and covariance matrix $\Gamma_m = \{\sigma_{i,p}^m,\ i, p = 1, \ldots, I\}$, $m = 1, \cdots, M$. Nonnegative constants $w_m \in \mathbb{R}_+$, $m = 1, \ldots, M$, are mixture weights such that $\sum_{m=1}^{M} w_m = 1$.

Expectation maximization (EM) algorithm deals with the estimation of the parameters in finite normal mixtures, see for instance [18, 3].

Given Assumption 2.1, an individual chance constraint can be reformulated as a group of SOCP constraints and linear constraints, shown in Theorem 2.1 hereafter. A similar result of Theorem 2.1 was obtained in [29]. Still, we keep the theorem and the proof for self-contained purpose.

Theorem 2.1. *Given Assumption 2.1, individual chance constrained problem (3)-(4) is equivalent to*

$$(\mathrm{IC_{mg}}) \quad \min_{x \in X, z \in \mathbb{R}^M} \quad g_0(x) \tag{5}$$

$$\text{s.t.} \quad \Phi^{-1}(z_m)\sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I}\sigma_{i,p}^m g_i(x) g_p(x)} + \sum_{i=1}^{I}\mu_i^m g_i(x) \le \beta,\ m = 1, \ldots, M, \tag{6}$$

$$\sum_{m=1}^{M} w_m z_m \ge 1 - \epsilon, \tag{7}$$

$$0 \le z_m \le 1,\ m = 1, \ldots, M, \tag{8}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of a standard normal distribution $N(0, 1)$.

Proof. Let $\boldsymbol{v} = [g_1(x), g_2(x), \ldots, g_I(x)]^\top$, we rewrite (4) as

$$\mathbb{P}_f\left(\boldsymbol{c}^\top \boldsymbol{v} \le \beta\right) \ge 1 - \epsilon. \tag{9}$$

From Assumption 2.1, we have that

$$\begin{aligned}\mathbb{P}_f\left(\boldsymbol{c}^\top \boldsymbol{v} \le \beta\right) &= \int_{\boldsymbol{x}^\top \boldsymbol{v} \le \beta} \sum_{m=1}^{M} w_m f_m(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\ &= \sum_{m=1}^{M} w_m \int_{\boldsymbol{x}^\top \boldsymbol{v} \le \beta} f_m(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.\end{aligned}$$

Let $z \in \mathbb{R}^M$ denotes the vector of auxiliary variables, (9) can be reformulated as

$$\mathbb{P}_{f_m}\left((\boldsymbol{c}^m)^\top \boldsymbol{v} \le \beta\right) \ge z_m,\ m = 1, \ldots, M, \tag{10}$$

$$\sum_{m=1}^{M} w_m z_m \ge 1 - \epsilon, \tag{11}$$

$$0 \le z_m \le 1,\ m = 1, \ldots, M, \tag{12}$$

where $\boldsymbol{c}^m$ follows an I-dimensional normal distribution with mean vector $\boldsymbol{\mu}_m$, covariance matrix Γ_m and density function f_m, $m = 1, \cdots, M$.

From the affine invariance of multi-dimensional normal distribution, we know that $(\boldsymbol{c}^m)^\top \boldsymbol{v}$ follows an 1-dimensional normal distribution with mean value $(\boldsymbol{\mu}_m)^\top \boldsymbol{v}$ and variance $(\boldsymbol{v})^\top \Gamma_m(\boldsymbol{v})$, $m = 1, \cdots, M$. Then, from [48], (10) is equivalent to

$$(\boldsymbol{\mu}_m)^\top \boldsymbol{v} + \Phi^{-1}(z_m)\sqrt{\boldsymbol{v}^\top \Gamma_m \boldsymbol{v}} \le \beta,\ m = 1, \ldots, M.$$

Taking $\boldsymbol{\mu}_m = [\mu_1^m, \mu_2^m, \ldots, \mu_I^m]^\top$ and $\Gamma_m = \{\sigma_{i,p}^m,\ i, p = 1, \ldots, I\}$ back gives the result of the theorem. □

2.1 Convexity results: preliminary lemmas

To study the mathematical property of ($\mathrm{IC_{mg}}$) , especially its convexity, core difficulties come from the product terms $g_i(x)g_p(x)$, the quantile function of a standard normal distribution $\Phi^{-1}(z_m)$, and the product structure of the functions of x and z. Firstly, we present the following lemmas about these terms.

Lemma 2.1. *a)* $\Phi^{-1}(x)$ *is monotonically increasing on* $(0, 1)$*, nonnegative and convex on* $[0.5, 1)$*, negative and concave on* $(0, 0.5)$*;*
b) $\log(\Phi^{-1}(x))$ *is monotonically increasing on* $(0.5, 1)$*, and convex on* $[\Phi(1), 1)$*.*

Proof. a) The quantile function of a standard Gaussian distribution $\Phi^{-1}(\cdot)$ is also known as the inverse CDF function or the probit function, which can be expressed as:

$$\Phi^{-1}(x) = \sum_{p=0}^{\infty} \frac{\lambda_p}{2p+1}\left(\frac{\sqrt{\pi}}{2}(2x-1)\right)^{2p+1},$$

where $\lambda_0 = 1$ and $\lambda_p = \sum_{i=0}^{p-1} \frac{\lambda_i \lambda_{p-1-i}}{(i+1)(2i+1)} > 0,\ p = 1, 2, \ldots$ Thus, $\Phi^{-1}(x) \ge 0$ for $x \ge 0.5$; $\Phi^{-1}(x) < 0$ for $x < 0.5$. Convexity or concavity follows by the property of these power function terms.

b) The first order and second order derivatives of $\log(\Phi^{-1}(x))$ are

$$\frac{d(\log(\Phi^{-1}(x)))}{dx} = \frac{1}{\varphi\phi(\varphi)},$$

and

$$\frac{d^2(\log(\Phi^{-1}(x)))}{dx^2} = -\frac{1 + \varphi\frac{\phi'(\varphi)}{\phi(\varphi)}}{(\varphi\phi(\varphi))^2} = \frac{\varphi^2 - 1}{(\varphi\phi(\varphi))^2},$$

where $\varphi = \Phi^{-1}(x)$, $\phi(\cdot)$ is the probability density function of the standard normal distribution $N(0, 1)$, and $\phi'(\cdot)$ is the first order derivative of $\phi(\cdot)$. The last equation is due to $\frac{\phi'(\varphi)}{\phi(\varphi)} = \frac{\frac{-\varphi}{\sqrt{2\pi}}e^{-\varphi^2/2}}{\frac{1}{\sqrt{2\pi}}e^{-\varphi^2/2}} = -\varphi$.

For $x > 0.5$, $\varphi > 0$ and the first order derivative is greater than 0. Hence $\log(\Phi^{-1}(x))$ is monotonically increasing on $(0.5, +\infty)$. For $x \geq \Phi(1)$, $\varphi \geq 1$ and the second order derivative is greater than or equal to 0. Hence $\log(\Phi^{-1}(x))$ is convex on $[\Phi(1), +\infty)$. □

Lemma 2.2. *Given* $\epsilon \leq (1 - \Phi(1)) \left(\min_m\{w_m\}\right)$, *any feasible solution of* ($\mathrm{IC_{mg}}$) *satisfies that* $z_m \geq \Phi(1)$. *Given* $\epsilon \leq \frac{1}{2} \left(\min_m\{w_m\}\right)$, *any feasible solution of* ($\mathrm{IC_{mg}}$) *satisfies* $z_m \geq 0.5$.

Proof. From (7), we have for any $m = 1, \ldots, M$

$$\begin{aligned} z_m &\geq \frac{1}{w_m}\Big(1 - \epsilon - \sum_{i=1,\ldots,M, i\neq m} w_i z_i\Big) \geq \frac{1}{w_m}\Big(1 - \epsilon - \sum_{i=1,\ldots,M, i\neq m} w_i\Big) \\ &= \frac{1}{w_m}(w_m - \epsilon) \geq \Phi(1) \end{aligned}$$

The sufficient condition for $z_m \geq 0.5$ can be obtained similarly. □

When the mixture distributions are equally weighted, i.e., $w_m = \frac{1}{M}$, $m = 1, \ldots, M$, the conditions in Lemma 2.2 reduce to $\epsilon \leq \frac{1-\Phi(1)}{M}$ and $\epsilon \leq \frac{1}{2M}$.

Due to the biconvex terms, it is impossible to built general convex conclusion for ($\mathrm{IC_{mg}}$) . In the following subsections, we give some convexity results of ($\mathrm{IC_{mg}}$) under some specific cases.

2.2 Case study: logarithmically convex functions

We consider the case when $g_i(x)$ is positive and logarithmically convex (log-convex for short), $g_i(x) = e^{f_i(x)}$, where $f_i(x)$ is convex. This case includes examples with exponential functions, e.g., the population growth increasing function or the wealth increasing function with random compound interests.

Lemma 2.3. *Given two nonnegative valued, convex and twice differentiable functions* $f(x)$: $\mathbb{R}^n \to \mathbb{R}_+$ *and* $g(y)$: $\mathbb{R}^n \to \mathbb{R}_{++}$, $f(x)g(y)$ *is convex if and only if*

$$f(x)\nabla^2 g(y) - \frac{1}{g(y)}\nabla^2 f(x)^{-1}\nabla f(x)\nabla g(y)^\top \nabla f(x)\nabla g(y)^\top \succeq_S 0.$$

Here, $A \succeq_S 0$ means that A is a positive semidefinite matrix.

Proof. This can be easily obverted from the Jordan decomposition of the Hessian of $f(x)g(y)$,

$$\nabla^2(f(x)g(y)) = \begin{bmatrix} g(y)\nabla^2 f(x) & \nabla f(x)\nabla g(y)^\top \\ \nabla f(x)\nabla g(y)^\top & f(x)\nabla^2 g(y) \end{bmatrix}$$

. □

Lemma 2.4. *If* $f(x)\nabla^2 f(x) - \nabla f(x)\nabla f(x)^\top \succeq_S 0$ *and* $g(y)\nabla^2 g(y) - \nabla g(y)\nabla g(y)^\top \succeq_S 0$, *then* $f(x)\nabla^2 g(y) - \frac{1}{g(y)}\nabla^2 f(x)^{-1}\nabla f(x)\nabla g(y)^\top \nabla f(x)\nabla g(y)^\top \succeq_S 0$.

Lemma 2.5 (Section 3.5.2 [6])**.** *For positive valued and twice differentiable function* $f(x)$, $f(x)\nabla^2 f(x) - \nabla f(x)\nabla f(x)^\top \succeq_S 0$ *if and only if* $f(x)$ *is log-convex.*

Theorem 2.2. *If $g_0(x)$ is convex, $g_i(x)$ is positive and logarithmically convex, $i = 1, \ldots, I$, X is a convex and compact set, $\mu_i^m \geq 0$, $i = 1, \ldots, I$, $m = 1, \ldots, M$, $\sigma_{i,p}^m \geq 0$, $i, p = 1, \ldots, I$, $m = 1, \ldots, M$, and $\epsilon \leq (1 - \Phi(1)) \left(\min_m \{w_m\}\right)$, then* ($\mathrm{IC_{mg}}$) *is a convex optimization problem.*

Proof. As $\epsilon \leq (1 - \Phi(1)) \left(\min_m \{w_m\}\right)$, by Lemma 2.2 we have that $z_m \geq \Phi(1) > 0$. Then by Lemma 2.1, we have that $\Phi^{-1}(z_m)$ is log-convex. As $g_i(x)$ is assumed to be log-convex, $g_i(x)^{\frac{1}{2}}$ is also log-convex. Then by Lemma 2.3-2.5, we have that $g_i(x)^{\frac{1}{2}} g_p(x)^{\frac{1}{2}} \Phi^{-1}(z_m)$ is a convex function, for any i, p. Thus, given $\sigma_{i,p}^m \geq 0$,

$$\Phi^{-1}(z_m) \sqrt{\sum_{i=1}^{I} \sum_{p=1}^{I} \sigma_{i,p}^m g_i(x) g_p(x)} = \sqrt{\sum_{i=1}^{I} \sum_{p=1}^{I} \left(\sqrt{\sigma_{i,p}^m} g_i(x)^{\frac{1}{2}} g_p(x)^{\frac{1}{2}} \Phi^{-1}(z_m)\right)^2},$$

is a convex function, as it is a composition of an outer-level L_2-norm function and the inner-level convex terms. □

Given the conditions in Theorem 2.2, we can reformulate ($\mathrm{IC_{mg}}$) as the following convex optimization problem

$$\begin{aligned}
\min_{x \in X, z, Y} \quad & g_0(x) \\
\text{s.t.} \quad & ||Y^m||_F + \sum_{i=1}^{I} \mu_i^m g_i(x) \leq \beta, \ m = 1, \ldots, M, \\
& \sqrt{\sigma_{i,p}^m} \exp\left\{\frac{1}{2} \log(g_i(x)) + \frac{1}{2} \log(g_p(x)) + \log(\Phi^{-1}(z_m))\right\} \leq Y_{i,p}^m, \ i, j = 1, \ldots, I, \ m = 1, \ldots, M, \\
& \sum_{m=1}^{M} w_m z_m \geq 1 - \epsilon, \\
& 0 \leq z_m \leq 1, \ m = 1, \ldots, M.
\end{aligned}$$

2.3 Case study: log-exp-revealed-convex functions

We introduce a class of positive functions which is log-convex after exponential transformation, which would have some potential use in practical applications.

Definition 2.1. *A function $g : \mathbb{R}_{++}^n \to \mathbb{R}_+$ is called log-exp-revealed-convex on $[\boldsymbol{a}, \boldsymbol{b}]$ ($\boldsymbol{a}, \boldsymbol{b} > 0$) if $h(\boldsymbol{t}) = \log g(e^{\boldsymbol{t}})$ is convex to $\boldsymbol{t}$ on $[\log(\boldsymbol{a}), \log(\boldsymbol{b})]$, where $\boldsymbol{t} = (t_1, \cdots, t_n)$, $e^{\boldsymbol{t}} = \left(e^{t_1}, \cdots, e^{t_n}\right)$, $\log(\boldsymbol{a}) = (\log(a_1), \cdots, \log(a_n))$, $\log(\boldsymbol{b}) = (\log(b_1), \cdots, \log(b_n))$.*

A common example of log-exp-revealed-convex functions is $g(x) = \prod_{j=1}^{M} x_j^{a_j}$ with $x \in \mathbb{R}_{++}^M$ and $a \in \mathbb{R}^M$, known as monomial functions. Other examples of log-exp-revealed-convex functions includes many fractional functions ,e.g., $g(x) = \frac{x_1 x_2}{\log(x_2)}$, $g(x) = \frac{1}{\log(x_1) + \log(x_2)}$.

A similar definition is the log-revealed-concave function proposed by [56] which requires the concavity of $h(t) = f(e^{\boldsymbol{t}})$. Log-revealed-concavity is helpful in revealing the exponential decreasing property of the distribution function. What we are concerned with here is the log-revealed-convexity in order to characterize the structure of the decision functions.

Definition 2.2. *A function $f : \mathbb{R}^n_{++} \to \mathbb{R}$ is called log-revealed-convex on $[\boldsymbol{a}, \boldsymbol{b}]$ $(\boldsymbol{a}, \boldsymbol{b} > 0)$ if $h(\boldsymbol{t}) = f(e^{\boldsymbol{t}})$ is convex to $\boldsymbol{t}$ on $[\log(\boldsymbol{a}), \log(\boldsymbol{b})]$.*

Convex and increasing functions are naturally log-revealed-convex. Notice that Log-exp-revealed-convex functions are also log-revealed-convex. Log-revealed-convex functions contain some concave functions, e.g., $f(x) = \sum_{i=1}^{M} \log(x_i)$.

Definition 2.3. *A set $X \in \mathbb{R}^n_+$ is called a log-revealed-convex set if $\log(X)$ is a convex set, where $\log(X) = \{\log(x) | x \in X\}$.*

Theorem 2.3. *If $g_0(x)$ is log-revealed-convex, $g_i(x)$ is log-exp-revealed-convex, $i = 1, \ldots, I$, X is a log-revealed-convex and compact set, $\mu_i^m \geq 0$, $i = 1, \ldots, I$, $m = 1, \ldots, M$, $\sigma_{i,p}^m \geq 0$, $i, p = 1, \ldots, I$, $m = 1, \ldots, M$, and $\epsilon \leq (1 - \Phi(1)) \left(\min_m \{w_m\} \right)$, then the optimal value of $(\mathrm{IC_{mg}})$ is equal to the optimal value of the following convex optimization problem*

$$
\begin{aligned}
\min_{t \in \log X, z} \quad & g_0(e^t) \\
\text{s.t.} \quad & \sqrt{\sum_{i=1}^{I} \sum_{p=1}^{I} \sigma_{i,p}^m \exp(\log g_i(e^t) + \log g_p(e^t) + 2 \log \Phi^{-1}(z_m))} + \sum_{i=1}^{I} \mu_i^m g_i(e^t) \leq \beta, \ m = 1, \ldots, M, \\
& \sum_{m=1}^{M} w_m z_m \geq 1 - \epsilon, \ z_m \in [0, 1], \ m = 1, \ldots, M,
\end{aligned}
$$

and the exponential of its optimal solution is an optimal solution of $(\mathrm{IC_{mg}})$.

Proof. The reformulation techniques are similar to those in Theorem 2.2. A further variable transformation from x to $t = \log(x) = [\log(x_i), \ldots, \log(x_n)]$ can be used for the sake of the proof. □

When $g_i(x) = \prod_{j=1}^{M} t_j^{a_{ij}}$, it is usually known as a monomial, and $\sum_{i=1}^{I} c_i g_i(x)$ is a posynomial. An optimization problem with posynomial constraints is a geometric program where the coefficients c_i, $i = 1, \ldots, I_k$, $k = 0, \ldots, K$, are nonnegative real numbers. In chance constrained stochastic geometric programs, the logarithm terms in the exponentials $\log g_i(e^t) + \log g_p(e^t)$ are linear to t [36, 37, 38].

Many practical constraints have nondecreasing, convex and positive terms, such as the knapsack constraints, resource allocation constraints, and the floor/wall area constraints. Thus, we can apply Theorem 2.3 to deal with a chance constrained knapsack problem [13], a chance constrained wealth allocation problem [11], or a chance constrained shape optimization problem [36], with nonnegative dependence [4].

2.4 Disciplined convex programming: log-quantile function approximation

The reformulations in Theorem 2.2 and Theorem 2.3 might have computational difficulties from the term $\log(\Phi^{-1}(z_{k,m}))$. Although Lemma 2.2 shows its convexity at the right tail, this quantile function of a normal distribution does not have a closed-form. The piecewise linear approximations can be used here to get a disciplined convex programming reformulation more easier to solve. Similar to [36], we can choose S different increasing points $\xi_1, \xi_2, \ldots, \xi_S$, in $(1 - \epsilon, 1)$. We use the piecewise linear function

$$
G_k^L(x) = \max_{s=1,\ldots,S} \left\{ \frac{1}{\Phi^{-1}(\xi_s)\phi(\Phi^{-1}(\xi_s))}(x - \xi_s) + \log(\Phi^{-1}(\xi_s)) \right\},
$$

composed by the tangent lines at those points to approximate $\log(\Phi^{-1}(x))$, and derive a lower approximation. Moreover, we use the

$$G^U_{s,k}(x) = \max_{s=1,\dots,S} \left\{ \frac{\log(\Phi^{-1}(\xi_{s+1})) - \log(\Phi^{-1}(\xi_s))}{\xi_{s+1} - \xi_s}(x - \xi_s) + \log(\Phi^{-1}(\xi_s)) \right\},$$

composed by the segments between those points to get an upper approximation.

2.5 Solution methods under non-convex cases

When the conditions in Theorem 2.2 and Theorem 2.3 fail to hold, we may meet some non-convex cases.

We can use sequential convex approximation approaches to deal with the biconvex terms. The basic idea of the sequential convex approximation consists in decomposing the original problem into subproblems where a subset of variables is fixed alternatively. For problem ($\mathrm{IC_{mgu}}$), we first fix $z = z^n$ and update x by solving

$$\begin{aligned} (\mathrm{SQ}_1) \qquad \min_{x \in X} \quad & g_0(x) \\ \text{s.t.} \quad & \Phi^{-1}(z^n_m)\sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I}\sigma^m_{i,p} g_i(x) g_p(x)} + \sum_{i=1}^{I}\mu^m_i g_i(x) \le \beta,\ m = 1, \dots, M, \end{aligned}$$

and then fix $x = x^n$ and update z by solving

$$\begin{aligned} (\mathrm{SQ}_2) \qquad \min_{z \in \mathbb{R}^M_+} \quad & \sum_{m=1}^{M} \phi_m z_m \\ \text{s.t.} \quad & z_m \le \Phi\left(\frac{\beta - \sum_{i=1}^{I}\mu^m_i g_i(x)}{\sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I}\sigma^m_{i,p} g_i(x) g_p(x)}} \right),\ m = 1, \cdots, M, \\ & \sum_{m=1}^{M} w_m z_m \ge 1 - \epsilon, \\ & 0.5 \le z_m \le 1,\ m = 1, \dots, M \end{aligned}$$

Here, ϕ_m is a given searching direction. The sequential approximation is given by Algorithm 2.1. To guarantee the convexity at each iteration step, we require $z_m \ge 0.5$. The condition can be reached when $\epsilon \le \frac{1}{2}\left(\min_m\{w_m\}\right)$.

Algorithm 2.1. *Sequential convex approximation*

Initialization:

Choose an initial point z^0 of z feasible for (7) and (8). Set $n = 0$.

Iteration:

while *$n \le n_{max}$ and $||z^{n-1} - z^n|| \ge \epsilon$* ***do***

- *Solve problem* (SQ_1)*; let x^n, θ^n and v^n denote an optimal solution, an optimal solution of the Lagrangian dual variable θ and the optimal value of (SQ_1), respectively.*

• *Solve problem* (SQ_2) *with*

$$\phi_m = \theta_m^n \cdot (\Phi^{-1})'(z_m^n) \sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I} \sigma_{i,p}^m g_i(x^n) g_p(x^n)};$$

let $\tilde{z}$ *denote an optimal solution of* (SQ_2).

• $z^{n+1} \leftarrow z^n + \tau(\tilde{z} - z^n)$, $n \leftarrow n+1$. *Here,* $\tau \in (0,1)$ *is the step length.*

end while

Output: x^n, v^n

From Theorem 2 in [36], Algorithm 2.1 converges in a finite number of iterations and the returned value v^n is an upper bound of problem $(\mathrm{IC_{mg}})$. Algorithm 2.1 can be seen as a particular case of the alternate convex search or block-relaxation methods [21]. When these sub-problems are all convex, the objective function is continuous, the feasible set is closed, the alternate convex search algorithm converges monotonically to a partial optimum point (Theorem 4.7 [21]). When the objective function is a differentiable and biconvex function, (x, z) is a partial optimum point if and only if (x, z) is a stationary point (Corollary 4.3 [21]). Thus, if $\sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I} \sigma_{i,p}^m g_i(x) g_p(x)}$ and $\sum_{i=1}^{I} \mu_i^m g_i(x)$ are convex functions for any $m = 1, \ldots, M$, Algorithm 2.1 converges to a stationary point. For numerical experiments purposes, we can require that $\mu_i^m \geq 0$, $\sigma_{i,p}^m \geq 0$ and $g_i(x)$ is log-convex.

If $\mu_i < 0$ or $\sigma_{i,p}^k < 0$ for some i, p, while other conditions hold, $(\mathrm{IC_{mg}})$ can be generally seen as a difference of convex (DC) programming problem. Therefore, the sub-problem (SQ_1) in the sequential convex approximation can be reformulated as a DC programming problem. In this case, global optimization algorithms, e.g., successive convex approximation methods, can be used to deal with the nonconvex problems. At each iteration step, we can approximate the terms with negative means or negative correlation by a linear function at the current iteration point.

3 Individual chance constraints under generalized hyperbolic distribution

In this section, we consider the individual chance constrained problem (IC) when c_i follows the generalized hyperbolic (GH) distributions. We denote the problem under GH distribution as $(\mathrm{IC_{gh}})$.

3.1 Introduction to generalized hyperbolic distribution

Generalized hyperbolic distribution group constrains a wide spectrum of continuous distributions, which can be uniformly defined as follows.

Definition 3.1. *A random vector* $X \in \mathbb{R}^d$ *is said to follow a Generalized Hyperbolic (GH) distribution, denoted by* $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, *if its density function is*

$$f(x) = c \frac{K_{\lambda-(d/2)}(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))\,(\psi + \gamma'\Sigma^{-1}\gamma)})\mathrm{e}^{(x-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi + (x-\mu)'\Sigma^{-1}(x-\mu))\,(\psi + \gamma'\Sigma^{-1}\gamma)})^{(d/2)-\lambda}}$$

where $c = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^{\lambda}\left(\psi+\gamma'\Sigma^{-1}\gamma\right)^{(d/2)-\lambda}}{(2\pi)^{d/2}|\Sigma|^{1/2}K_{\lambda}(\sqrt{\chi\psi})}$, K_λ *is the modified Bessel function of the third kind with index* λ.

Generalized Hyperbolic (GH) distribution has six parameters. $\mu \in \mathbb{R}^d$ is the location parameter (notice that the mean value of a GH distribution is not on μ but on $\mu + \gamma(\chi/\psi)^{1/2}\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}$). The dispersion matrix $\Sigma \in \mathbb{R}^{d\times d}$ is the scale parameter. $\gamma \in \mathbb{R}^d$ is the skewness parameter, if $\gamma = 0$, then the distribution is symmetric.

$\lambda \in \mathbb{R}$ defines the subclasses of GH distribution and is related to the tail flatness, $\chi \in \mathbb{R}, \psi \in \mathbb{R}$ determine the distribution shape, specifying the weight to assign to the tails compared with the center. The larger those parameters are, the closer the distribution is to the normal distribution. It is required that $\chi > 0, \psi \geqslant 0$ if $\lambda < 0$; $\chi > 0, \psi > 0$ if $\lambda = 0$; and $\chi \geqslant 0$ $\psi > 0$ if $\lambda > 0$.

By specifying the values of λ, χ, ψ and γ, we have different sub-case distributions from the GH distribution group. We give some examples in Table 1.

Table 1: Special cases of the Generalized Hyperbolic (GH) distribution [42]

λ	χ	ψ	μ	Σ	γ	distribution
$(d+1)/2$	χ	ψ	μ	Σ	γ	hyperbolic distribution
$-1/2$	χ	ψ	μ	Σ	γ	NIG distribution
$-v/2$	v	0	μ	Σ	γ	skewed T distribution
$v/2$	v	0	μ	Σ	0	Student's T distribution
$+\infty$	$+\infty$	0	μ	Σ	0	Gaussian distribution

The GH distribution can be viewed as a kind of Gaussian mixture distribution with infinite many components.

Proposition 3.1. *[19] If $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$, then $X \overset{\mathrm{d}}{=} \mu + W\gamma + \sqrt{W}Z$, where $Z \sim N_d(\mathbf{0}, \Sigma)$ follows a d-dimensional Gaussian distribution, $W \sim N^-(\lambda, \chi, \psi)$ follows a nonnegative, scalar-valued generalized inverse Gaussian (GIG) distribution[1], which is independent of Z.*

By Proposition 3.1, we have that $X|W = w \sim N_d(\mu + W\gamma, w\Sigma)$. For this reason, these distributions are known as mean-variance mixtures of Gaussians. In general, such distributions are not symmetric when $\gamma \neq 0$. From the mean-variance mixtures formulation in Proposition 3.1, we can easily see that the mean and the variance of the GH distribution are biased from the location and the scale parameters, i.e., $\mathbb{E}(X) = \mu + \mathbb{E}(W)\gamma$, $\mathrm{cov}(X) = \mathbb{E}(W)\Sigma + \mathrm{var}(W)\gamma\gamma'$. Furthermore, the GH distribution class is closed under linear operations.

Proposition 3.2. *(Proposition 3.13 [42]) If $X \sim GH_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $Y = BX + b$, where $B \in \mathbb{R}^{k\times d}$ and $b \in \mathbb{R}^k$, then $Y \sim GH_k(\lambda, \chi, \psi, B\mu + b, B\Sigma B', B\gamma)$.*

3.2 Tractable approximation

Hyperbolic distributions have already been widely used in industrial and finance research to represent the skewed and leptokurtotic nature of random variables. However, their applications in chance constraints ($\mathrm{IC}_{\mathrm{gh}}$), especially solving methods of the corresponding chance constrained optimization problems are not well studied. One can use classical approximation methods for chance constraints to solve chance constrained problems under hyperbolic distribution assumption, e.g., sample average approximation [39, 45] or Bernstein approximation [46], Chebyshev approximation

[1] A random variable X has a generalized inverse Gaussian (GIG) distribution, written $X \sim N^-(\lambda, \chi, \psi)$, if its density is $f(x) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1}\exp\left(-\frac{1}{2}\left(\chi x^{-1} + \psi x\right)\right),\ x > 0$.

[12] or conditional value-at-risk approximation [62]. However, all these methods do not use the particular structure of hyperbolic distributions. Therefore, they either need a high number of samples to guarantee the precision or can not provide tight bounds.

In this paper, we propose a new algorithm to solve chance constraints with generalized hyperbolic distribution. Our core idea is to approximate the generalized hyperbolic distribution by a finite Gaussian mixture distribution. Our key approach is the partial sampling technique.

By Proposition 3.1, we decompose a GH distribution into a mixture of a univariate GIG distribution and a multivariate Gaussian distribution. Thus, if c follows the I-dimensional GH distribution with a decomposition to W and Z, where $Z \sim N_I(\mathbf{0}, \Sigma)$, $W \sim N^-(\lambda, \chi, \psi)$, the individual chance constraint (4) can be written as

$$\mathbb{P}_{W,Z}\left(\sum_{i=1}^{I}(\mu_i + W\gamma_i + \sqrt{W}Z_i)g_i(x) \leq \beta\right) \geq 1-\epsilon,$$

which can be rewritten by the tower property of the expected value function

$$\mathbb{E}_W\left[\mathbb{P}\left(\sum_{i=1}^{I}(\mu_i + w\gamma_i + \sqrt{w}Z_i)g_i(x) \leq \beta \mid W = w\right)\right] \geq 1-\epsilon.$$

Then, we use a discrete distribution W' to approximate the univariate distribution of W. Here, we draw M i.i.d. samples , $w_1, w_2, \ldots, w_M$, of the univariate GIG distribution of W with Monte-Carlo method. Thanks to these samples, we get a sample average approximation of the outer level expected value function

$$\frac{1}{M}\sum_{m=1}^{M}\left[\mathbb{P}\left(\sum_{i=1}^{I}(\mu_i + w_m\gamma_i + \sqrt{w_m}Z_i)g_i(x) \leq \beta\right)\right] \geq 1-\epsilon,$$

which is equivalent to the following group of constraints

$$\sum_{m=1}^{M} y_m \geq M(1-\epsilon), \tag{13}$$

$$\mathbb{P}\left(\sum_{i=1}^{I}(\mu_i + w_m\gamma_i + \sqrt{w_m}Z_i)g_i(x) \leq \beta\right) \geq y_m,\ m = 1, \ldots, M. \tag{14}$$

with auxiliary variables $y_m \in [0,1]$, $m = 1, \ldots, M$.

As $f := \sum_{i=1}^{I}(\mu_i + w_m\gamma_i + \sqrt{w_m}Z_i)g_i(x)$ is linear to the Gaussian distributed random variable Z, it follows a univariate Gaussian distribution with mean value $\mu_f = \sum_{i=1}^{I}(\mu_i + w_m\gamma_i)g_i(x)$ and variance $\sigma_f^2 = \sum_{i=1}^{I}\sum_{p=1}^{I} w_m\sigma_{i,p}g_i(x)g_p(x)$. Thus (14) can be rewritten as

$$\mathbb{P}\left(\frac{f-\mu_f}{\sigma_f} \leq \frac{\beta-\mu_f}{\sigma_f}\right) = \Phi\left(\frac{\beta-\mu_f}{\sigma_f}\right) \geq y_m,$$

which is further equivalent to $\beta - \mu_f \geq \Phi^{-1}(y_m)\sigma_f$, $m = 1, \ldots, M$, where $\Phi^{-1}(\cdot)$ is the quantile of the standard univariate Gaussian distribution $N(0,1)$. Then, we have the following approximation

of ($\mathrm{IC_{gh}}$)under a finite Gaussian mixture approximation distribution,

$$
\begin{aligned}
(\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}) \quad \min_{x\in X, y\in\mathbb{R}^M} \quad & g_0(x) \\
\text{s.t.} \quad & \Phi^{-1}(y_m)\sqrt{\sum_{i=1}^{I}\sum_{p=1}^{I} w_m\sigma_{i,p}g_i(x)g_p(x)} + \sum_{i=1}^{I}(\mu_i + w_m\gamma_i)g_i(x) \le \beta,\ m=1,\dots,M, \\
& \sum_{m=1}^{M} y_m \ge M(1-\epsilon), \\
& 0 \le y_m \le 1,\ m=1,\dots,M.
\end{aligned}
$$

Notice that ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)and ($\mathrm{IC_{mg}}$) have the same form. Therefore, we can apply the theoretical results and corresponding solution methods investigated in Section 2, for the finite Gaussian mixture approximation problem ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$).

3.3 Convergence analysis of our approximation approach

The asymptotic tightness of the approximation problem ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)can be guaranteed if the GIG distribution W is approximated by its i.i.d. samples.

Theorem 3.1. *Suppose that, $g_i(x)$ is continuous, $i = 0, 1, \dots, I$, the samples in ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)are i.i.d., and X is a compact set. We assume some constraint qualification holds such that there exists a sequence of points x_M in the feasible set of ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)with M i.i.d. samples such that $x_M \to x$ w.p.1. Then, the optimal value of ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)converges to the optimal value of ($\mathrm{IC_{gh}}$)when $M \to +\infty$. The distance between the feasible set of ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)and the feasible set of ($\mathrm{IC_{gh}}$)converges to 0 when $M \to +\infty$.*

Proof. The key point of the proof is to view the inner-level conditional probability function as a random function and use the convergence results from sample average approximation to the outer-level expectation operator. Denote

$$
\begin{aligned}
C(x,w) &= \mathbb{P}\left(\sum_{i=1}^{I}(\mu_i + w\gamma_i + \sqrt{w}Z_i)g_i(x) > \beta\right) \\
&= \mathbb{E}\left(\mathbf{1}_{(0,\infty]}\left(\sum_{i=1}^{I}(\mu_i + w\gamma_i + \sqrt{w}Z_i)g_i(x) - \beta\right)\right).
\end{aligned}
$$

As $Z_i, i = 1, \dots, I$, is continuously distributed, the true chance constraints (4) can be written as

$$
p(x) = \mathbb{E}_W\left[C(x,W)\right] \le \epsilon,
$$

and the sample average approximation constraint can be written as

$$
\hat{p}_M(x) = \frac{1}{M}\sum_{m=1}^{M}\left[C(x,w_m)\right] \le \epsilon.
$$

As $g_i(x)$ is continuous and Z_i is continuously distributed, $i = 1, \dots, I$, $C(x,w)$ is thus random lower semi-continuous. Then by Theorem 7.48 and Theorem 7.51 in [53], $p(x)$ is continuous and $\hat{p}_M(x)$ converges to $p(x)$ w.p.1 uniformly on X. Then, the expected results follow naturally by Theorem 5.5 of [53]. □

Theorem 3.1 guarantees that ($\mathrm{IC}^{\mathrm{M}}_{\mathrm{mga}}$)is an asymptotically tight approximation to ($\mathrm{IC}_{\mathrm{gh}}$)when M goes to infinity.

Remark 3.1. *By using such a partially sampling method, we can derive a new approximation method for the chance constrained problem under the hyperbolic distribution. Compared with CVaR approximation, Chebyshev approximation, or Bernstein approximation, the new approximation method is asymptotically tight. Compared with the SAA method, we only generate samples for a sub-set of random variables which is univariate whilst SAA method would generate samples for all random variables which meet the dimensional curse problem in the sampling process when both methods are applied to a large-scale problem. Moreover, the new approximation method only need to solve a convex problem in some particular cases (see the cases in Section 2.2 and Section 2.3). Even under non-convex cases, we can apply the efficient sequential convex approximation method in Section 2.5. In our numerical experiences in Section 5.1.2, we show the high performance of our new partial sampling approximation method.*

Remark 3.2. *Approximating a generalized hyperbolic distribution (a kind of infinite Gaussian mixture distribution) by a finite Gaussian mixture distribution can be equivalently viewed as sampling a sub set (univariate) of random variables when the generalized hyperbolic distributed random variables are decomposed. Cheng et al. [14] proposed a method of partially sampling which uses samples to approximate parts of random variables while keeping others still. Their major results are derived for the cases when the two groups of random variables are separable in two sides of the chance constraints. Our approach can be viewed as an extension of the partially sampling method in [14], where we sample a part of a complex random variable. The key point of the partially sampling method is to find a tricky sub set of random variables, making sampling to which would bring large computational improvement. The univariate GIG factor commonly to the multivariate GH distribution provides a great example.*

4 Dependence between different rows

In this section, we study the joint chance constraint where we consider the dependence issue between the random variables in different rows of the matrix constraint.

$$\text{(JC)} \qquad \min_{x\in X} \quad g_0(x) \tag{15}$$

$$\text{s.t.} \quad \mathbb{P}\left(\sum_{i=1}^{I_k} c_i^k g_i^k(x) \le \beta_k, k = 1,\dots,K\right) \ge 1-\epsilon. \tag{16}$$

Here $1-\epsilon$ is a prespecified probability with $\epsilon < 0.5$. $\beta_k \in \mathbb{R}$, $k = 1,\dots,K$, c_i^k is a random variable and $g_i^k(x)$ is real valued function, $i = 1,\dots,I_k$, $k = 1,\dots,K$. We divide the random variables c_i^k, $i = 1,\dots,I_k$, $k = 1,\dots,K$, into K groups, and denote $\boldsymbol{c}^k = [c_1^k, c_2^k, \dots, c_{I_k}^k]^\top$, $k = 1,\dots,K$, the random variables in the k^{th} nonlinear inequality in the joint chance constraint. We denote $\boldsymbol{g}^k(x) = [g_1^k(x), g_2^k(x), \dots, g_{I_k}^k(x)]^\top$, and $\upsilon^k(x) = (\boldsymbol{c}^k)^\top \boldsymbol{g}^k = \sum_{i=1}^{I_k} c_i^k g_i^k(x)$, $k = 1,\dots,K$.

The dependence between rows $\boldsymbol{c}^k$, $k = 1, 2, \dots, K$, plays an important role in a joint chance constraint as the satisfaction probability of the K nonlinear inequalities is computed jointly. In this section, we investigate several dependence structures including the independent case, the copula structure, the extreme dependence, and the mixture of extreme dependence.

4.1 Independent rows

We start with the independent case.

Theorem 4.1. *Suppose $\boldsymbol{c}^k$, $k = 1, \ldots, K$, are pairwise independent, the joint chance constraint (16) is equivalent to*

$$\mathbb{P}\left(\sum_{i=1}^{I_k} c_i^k g_i^k(x) \leq \beta_k\right) \geq y_k, \ k = 1, \ldots, K, \tag{17}$$

$$\prod_{k=1}^{K} y_k \geq 1 - \epsilon, \ y_k \in [0,1], \ k = 1, \ldots, K. \tag{18}$$

We denote the feasible set of x constrained by (17) and (18) as M_{ind} and the feasible set of y constrained by (18) as Y_{ind}.

The proof of the theorem 1 is trivial, it follows from joint distribution for independent variables. Such a decomposition approach is widely adopted in the computation of a joint chance constrained programming problems with independent rows [43].

There is another equivalent reformulation of constraint (17)-(18) where the auxiliary variables are set to $y_k = e^{x_k}$,

$$\mathbb{P}\left(\sum_{i=1}^{I_k} c_i^k g_i^k(x) \leq \beta_k\right) \geq e^{x_k}, \ k = 1, \ldots, K, \tag{19}$$

$$\sum_{k=1}^{K} x_k \geq \log(1-\epsilon), \ x_k \leq 0, \ k = 1, \ldots, K. \tag{20}$$

This reformulation keeps (20) linear which leads to the design of efficient solution methods. Based on this reformulation, Cheng and Lisser [16] derived SOCP approximation for joint chance constrained linear programs, and Liu et al. [36] derived convex reformulations for joint chance constrained geometric programs.

4.2 Dependent rows: copulas

We denote $F_{\boldsymbol{c}^k}$ as the cumulative distribution function of $\boldsymbol{c}^k$, $k = 1, \ldots, K$. $F_{\boldsymbol{c}}$ denotes the joint cumulative distribution function of $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$. $F_{\upsilon^k(x)}$ denotes the cumulative distribution function of $\upsilon^k(x)$, $k = 1, \ldots, K$. $F_{\upsilon(x)}$ denotes the joint cumulative distribution function of $\upsilon^1(x), \upsilon^2(x), \ldots, \upsilon^K(x)$.

We use a copula function to describe the nonlinear dependence structure between $\boldsymbol{c}^k$, $k = 1, \ldots, K$.

Assumption 4.1. *We assume that the joint distribution of $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$ is a function of the marginal distributions,*

$$F_{\boldsymbol{c}}(\cdot) = C(F_{\boldsymbol{c}^1}(\cdot), F_{\boldsymbol{c}^2}(\cdot), \ldots, F_{\boldsymbol{c}^K}(\cdot)),$$

where $C : [0,1]^d \to [0,1]$ is a copulas function such that C is zero if one of the arguments is zero, C equals to u if one argument is u and all others are 1, and C is K-nondecreasing.

There is a large number of copulas functions as any distribution function can form a copula. For instance, the Gaussian copula is $C_{\Sigma}^{\text{Gauss}}(\boldsymbol{u}) = \Phi_{\Sigma}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$, where Φ^{-1} is the inverse cumulative distribution (quantile) function of a univariate standard Gaussian distribution $N(0,1)$ and Φ_{Σ} is the joint cumulative distribution function of a multivariate Gaussian distribution $N(\mathbf{0}, \Sigma)$.

The Archimedean copula is $C^{\psi}(\boldsymbol{u}) = \psi^{(-1)}(\sum_{k=1}^{K} \psi(u_k))$ where $\psi(u)$ is the generator of an Archimedean copula. $\psi(u)$ is a convex, nonnegative and strictly decreasing function on $u \in [0,1]$. By choosing different generators, we have different kinds of Archimedean copulas. Nelsen [44] concluded 22 classic generators, including Gumbel-Hougaard generator, $\psi(u) = (-\ln u)^{\theta}$ with $\theta \geq 1$; the Frank generator, $\psi(u) = -\ln(\frac{e^{-\theta u}-1}{e^{-\theta}-1})$ with $\theta > 0$; Clayton generator, $\psi(u) = \frac{1}{\theta}(u^{-\theta}-1)$, with $\theta > 0$; and Joe generator, $\psi(u) = -\ln[1-(1-u)^{\theta}]$ with $\theta \geq 1$.

However, the copula relationship C between $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$ can not imply the same copula relationship C between $\upsilon^1(x), \upsilon^2(x), \ldots, \upsilon^K(x)$, after the linear transformation with respect to $\boldsymbol{g}^k(x)$, $k = 1, \ldots, K$. The reason is that the copula after the linear transformation relies on x. From Sklar's theorem, there always exists a copula C_x such that $F_{\boldsymbol{\upsilon}}(\cdot) = C_x(F_{\upsilon^1(x)}(\cdot), F_{\upsilon^2(x)}(\cdot), \ldots, F_{\upsilon^K(x)}(\cdot))$. However, the copula C_x after the linear transformation may not be C or even not be uniform to x. Therefore, the copula dependence structure is not always closed to linear transformation. van Ackooij and de Oliveira [54] gave a counter-example on the Gaussian copula with 2×2 variate Gaussian random variables in 2 rows to show this point.

Considering this issue, we do not simply apply a classical copula such as Gaussian or Archimedean in the joint chance constraint. Instead, we investigate the extreme dependence cases in a joint chance constraint which corresponds to the extreme positive dependent and extreme negative dependent structures. Then, we consider a sub-class of copula by mixing the extreme dependent cases.

4.3 Extreme dependence: Hoeffding-Fréchet bounds

As revealed in the previous subsection, the joint copula C_x relies on the decision variable. Therefore, it is difficult to derive a tractable reformulation for dependent chance constraints with the dependence copula. However, we can estimate the upper and lower bounds of a joint chance constraint based on the following Hoeffding-Fréchet inequality.

Lemma 4.1. *[44] The joint distribution F_X of a random vector X with marginal distributions $F_1, \ldots, F_K$ satisfies*

$$\underline{F}(x) \leq F_X(x) \leq \overline{F}(x), \tag{21}$$

where $\underline{F}(x) = [\sum_{k=1}^{K} F_k(x) - K + 1]_+$ and $\overline{F}(x) = \min_{1 \leq k \leq K} F_k(x)$. They are called lower and upper Hoeffding-Fréchet bounds.

By using Lemma 4.1, we can obtain two bounds for $M(x)$.

Theorem 4.2. *Denote*

$$Y_{in} = \left\{ y \mid \sum_{k=1}^{K} y_k \geq K - \epsilon \right\}$$

and

$$Y_{out} = \{ y \mid y_k \geq 1 - \epsilon,\ k = 1, \ldots, K \}.$$

Constraint (17) *together with $y \in Y_{in}$ provide an inner approximation of constraint (16). Constraint* (17) *together with $y \in Y_{out}$ provide an outer approximation of constraint (16).*

Proof. Constraint (16) can be reformulated by

$$F_{\boldsymbol{v}(x)}(\boldsymbol{\beta}) \geq 1-\epsilon,$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^\top$. By Lemma 4.1, $[\sum_{k=1}^K F_{v^k(x)}(\beta_k) - K + 1]_+$ is an uniformly lower bound of $F_{\boldsymbol{v}(x)}(\boldsymbol{\beta})$, and $\min_{1\leq k\leq K} F_{v^k(x)}(\beta_k)$ is an uniformly upper bound of $F_{\boldsymbol{v}(x)}(\boldsymbol{\beta})$.

Hence, $F_{v^k(x)}(\beta_k) \geq y_k$, $k = 1, \ldots, K$, which are equivalent to (17) together with $[\sum_{k=1}^K y_k - K + 1]_+ \geq 1-\epsilon$, provides an inner approximation of (16). Furthermore, $[\sum_{k=1}^K y_k - K + 1]_+ \geq 1-\epsilon$ is equivalent to $y \in Y_{in}$ as $1-\epsilon > 0$.

Similarly, (17) together with $\min_{1\leq k\leq K} y_k \geq 1-\epsilon$ provides an outer approximation of (16). □

It is worth noting that the upper Hoeffding-Fréchet bound is always a copula which corresponds to the dependence structure of comonotonicity. While the lower Hoeffding-Fréchet bound is a copula only in two dimensions, in which case it corresponds to countermonotonic random variables.

To define the comonotonicity and countermonotonicity between random variables, we call a subset C of $\mathbb{R}^n$ comonotonic (respectively, countermonotonic, restricted to $\mathbb{R}^2$) if $(s_i - t_i)(s_j - t_j) \geq 0$ (respectively ≤ 0) for all $i \neq j$ whenever $(s_1, \ldots, s_n)$ and $(t_1, \ldots, t_n)$ belong to C. A random vector $(X_1, \ldots, X_n)$ is called comonotonic (respectively countermonotonic) if there is a comonotonic (respectively countermonotonic) set $C \subseteq \mathbb{R}^n$ such that $\mathbb{P}((X_1, \ldots, X_n) \in C) = 1$, i.e., its support is comonotonic (respectively countermonotonic). Notice that countermonotonicity is generally defined for only two dimensional random vectors. From [50], we know that the upper Hoeffding-Fréchet bound Y_{out} we obtain in Theorem 4.2 corresponds to the case where $(v^1(x), v^2(x), \ldots, v^K(x))$ is comonotonic which relies on x. Correspondingly, only when $K = 2$, the lower Hoeffding-Fréchet bound Y_{in} corresponds the case that $(v^1(x), v^2(x))$ are countermonotonic. When $K \geq 3$, the lower Hoeffding-Fréchet bound corresponds to the pairwise countermonotonicity if it exists. A random vector $(X_1, \ldots, X_n)$ is said to be pairwise countermonotonic if all its bivariate projections $(X_i, X_j), i \neq j$, are countermonotonic random vectors. When $K \geq 3$, the pairwise countermonotonicity always exists when the marginals are discrete distributions, but does not exists when the marginals are continuous distributions [51]. To show this point, we consider the following example with a preliminary lemma.

Lemma 4.2. *(Proposition 4. [20]) Let X, Y and Z be random variables with joint distribution F and continuous marginals F_1, F_2 and F_3.*

1. *If (X, Y) and (Y, Z) are both comonotonic, then (X, Z) is also comonotonic and $F(x, y, z) = \min\{F_1(x), F_2(y), F_3(z)\}$,*
2. *If (X, Y) is comonotonic and (Y, Z) is countermonotonic, then (X, Z) is countermonotonic and $F(x, y, z) = \max\{0, \min\{F_1(x), F_2(y)\} + F_3(z) - 1\}$,*
3. *If (X, Y) and (Y, Z) are both countermonotonic, then (X, Z) is comonotonic and*

$$F(x, y, z) = \max\{0, \min\{F_1(x), F_3(z)\} + F_2(y) - 1\}.$$

Consider an example with $K = 3$, i.e., there are three continuous marginal distributions F_1, F_2 and F_3. If F_1 and F_2 are countermonotonic, F_1 and F_3 are countermonotonic, then F_2 and F_3 are comonotonic (Lemma 4.2). Thus, the pairwise countermonotonicty does not hold for more than three marginal distributions. Then, in the cases where the marginals are the mixture Gaussian distributions studied in Section 2 or the generalized hyperbolic distributions studied in Section 3,

the lower Hoeffding-Fréchet bound can not be reached by a copula and Y_{in} provides a lower bound. Otherwise when the marginals are discrete distributions, one can always construct a rearrangement matrix to represent the discrete bivariate distributions with a pairwise countermonotonic dependence (see Figure 6 in [51] for an instance).

Theorem 4.2 essentially studies the extreme dependence structure between decision dependent random variables $v^1(x), v^2(x), \ldots, v^K(x)$ which are linear scalarizations of $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$. To study the extreme dependence structures between $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$, multivariate extensions of comonotonicity and countermonotonicity should be defined. Multivariate comonotonicity between several random vectors with different strength are studied in [50]. However, to the best of our knowledge, multivariate countermonotonicity between several random vectors was not studied in the literature. Due to the scalarization process, the dependence structures between $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$ which corresponds to the comonotonicity or countermonotonicity between $v^1(x), v^2(x), \ldots, v^K(x)$ might not be unique. We introduce the following concept to characterize some kinds of extreme dependence structures between $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$.

Definition 4.1. *For a group of random vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$, where $\boldsymbol{X}_k$ is I_k-dimensional, we call $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$ group-comonotonic if for any nonnegative scalarization vectors $\boldsymbol{w}_k \in [0,1]^{I_k}$, $k = 1, \ldots, K$, $(\boldsymbol{w}_1^\top \boldsymbol{X}_1, \ldots, \boldsymbol{w}_K^\top \boldsymbol{X}_K)$ is comonotonic. For the unique case where $K = 2$, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are group-countermonotonic if for any nonnegative vectors $\boldsymbol{w}_1 \in [0,1]^{I_1}$ and $\boldsymbol{w}_2 \in [0,1]^{I_2}$, $(\boldsymbol{w}_1^\top \boldsymbol{X}_1, \boldsymbol{w}_2^\top \boldsymbol{X}_2)$ is countermonotonic.*

In order to study the relationship between the group-comonotonicity (respectively group-countermonotonicity) property of $\boldsymbol{c}^1, \boldsymbol{c}^2, \ldots, \boldsymbol{c}^K$, and the classical comonotonicity (respectively countermonotonicity) property of $v^1(x), v^2(x), \ldots, v^K(x)$, we first introduce a lemma. As we study continuous skewed distributions in the previous sections, we only investigate the continuous marginal cases.

Lemma 4.3. *Let X, Y and Z be random variables with joint distribution F and continuous marginals F_1, F_2 and F_3.*

1. *If (X, Y) and (Y, Z) are both comonotonic, then $(aX + bY + c, Z)$ is comonotonic for any $a, b \in \mathbb{R}_+$, $c \in \mathbb{R}$,*

2. *If (X, Y) is comonotonic and (Y, Z) is countermonotonic, then $(aX + bY + c, Z)$ is countermonotonic for any $a, b \in \mathbb{R}_+$, $c \in \mathbb{R}$.*

Proof. 1. By Lemma 4.2, (X, Y, Z) is comonotonic. Thus, there exists a random variable $Z \sim U[0,1]$ and nondecreasing functions f_1, f_2, f_3 such that $(X, Y, Z) \overset{\text{dist}}{=} (f_1(Z), f_2(Z), f_3(Z))$ by Proposition 2.1 in [17]. Therefore,

$$(aX + bY + c, Z) \overset{\text{dist}}{=} ((af_1 + bf_2 + c)(Z), f_3(Z)),$$

which implies the comonotonicity.

2. Due to the fact that (Y, Z) is countermonotonic $\Leftrightarrow$ $(Y, -Z)$ is comonotonic, the statement 2 follows from statement 1. □

By Lemma 4.2 and Lemma 4.3, the group-comonotonicity in definition 4.1 is equivalent to the s-comonotonicity proposed by [50] given continuous marginals, but weaker than s-comonotonicity in general cases.

Proposition 4.1. *For random vectors, $\boldsymbol{X}_k = (\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{1,I_k})$, $k = 1, \ldots, K$. If they all have continuous marginal distributions, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$ are group-comonotonic if and only if $(\boldsymbol{X}_{1,1}, \ldots, \boldsymbol{X}_{1,I_1}, \boldsymbol{X}_{2,1}, \ldots, \boldsymbol{X}_{2,I_2}, \ldots, \boldsymbol{X}_{K,1}, \ldots, \boldsymbol{X}_{K,I_K})$ is comonotonic.*

Proof. Necessary conditions follow from the first statement of Lemma 4.3. For sufficient conditions, we assume that $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K$ are group-comonotonic. For any $k, l = 1, \ldots, K$, $k \neq l$, for any $i = 1, \ldots, I_k$, $j = 1, \ldots, I_l$, we set $w_{k,i} = 1$, $w_{k,\hat{i}} = 0$ for all $\hat{i} \neq i$, and $w_{l,j} = 1$, $w_{l,\hat{j}} = 0$ for $\hat{j} \neq j$. Then by the definition of group-comonotonicity, $(w_k^\top \boldsymbol{X}_k, w_l^\top \boldsymbol{X}_l) = (\boldsymbol{X}_{k,i}, \boldsymbol{X}_{l,j})$ is comonotonic. □

While, the group-countermonotonicity leads to a combination of inner comonotonicity and cross countermonotonicity.

Proposition 4.2. *Continuous marginals $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are group-countermonotonic if and only if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are comonotonic vectors and bivariate projections components $(\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j})$ is countermonotonic for any $i = 1, \ldots, I_1$, $j = 1, \ldots, I_2$.*

Proof. If $\boldsymbol{X}_1, \boldsymbol{X}_2$ are group-countermonotonic, $(\boldsymbol{X}_{1,i}, \boldsymbol{X}_{2,j})$ is countermonotonic for any $i = 1, \ldots, I_1$, $j = 1, \ldots, I_2$ by definition. Then by the third statement of Lemma 4.2, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are comonotonic vectors. Necessary conditions follow from Lemma 4.3. □

Based on the above propositions, it is easy to see that the group-comonotonicity or group-countermonotonicity are related to the Hoeffding-Fréchet bounds.

Theorem 4.3. *Suppose that $\boldsymbol{c}^1, \boldsymbol{c}^2 \ldots, \boldsymbol{c}^K$ are group-comonotonic, then constraint (16) is equivalent to constraints* (17) *together with $y \in Y_{in}$ defined in Theorem 4.2. For the two rows case, suppose that $\boldsymbol{c}^1, \boldsymbol{c}^2$ are group-countermonotonic, then constraint (16) is equivalent to constraints* (17) *together with $y \in Y_{out}$.*

Proof. This theorem shows that the upper and lower Hoeffding-Fréchet bounds are tight under the group-comonotonicity and group-countermonotonicity cases, respectively.

By Definition 4.1, we have that $(\upsilon^1(x), \upsilon^2(x), \ldots, \upsilon^K(x))$ is comonotonic or $(\upsilon^1(x), \upsilon^2(x))$ is countermonotonic. Then, by Proposition 2.1 in [17] and Theorem 2 in [33], we have that the Hoeffding-Fréchet bounds are tight when $(\upsilon^1(x), \upsilon^2(x), \ldots, \upsilon^K(x))$ is comonotonic or $(\upsilon^1(x), \upsilon^2(x))$ is countermonotonicity, which gives the desired results. □

Remark 4.1. *It is worth noting that, the multivariate extreme dependence structures between $\boldsymbol{c}^1, \boldsymbol{c}^2 \ldots, \boldsymbol{c}^K$ may not be unique Hoeffding-Fréchet bounds. In this paper, we construct two particular extreme dependence cases. Generally speaking, our definition of group-comonotonicity is weaker than the s-comonotonicity in [50] whilst they are equivalent for continuous marginals. The inner comonotonicity of $\boldsymbol{c}^1(\boldsymbol{c}^2)$ is generally unnecessary, while our group-countermonotonicity is closed to linear transformation thus is decision independent.*

In the two next subsections, we study the dependence structure between the two extreme dependence structures.We consider the joint chance constraint where the marginal distribution in each row is known and the dependence between different rows is described as a linear combination of group-comonotonicity and group-countermonotonicity.

4.4 Mixture dependence under two rows case

As the definition of countermonotonicity is meaningful only for the case $K = 2$, we first investigate the joint chance constraint with only two rows.

Assumption 4.2. *Assume $K = 2$ and the dependence between $\boldsymbol{c}^1$ and $\boldsymbol{c}^2$ is a linear combination of group-comonotonicity and group-countermonotonicity, i.e., $F_{\boldsymbol{c}}(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2) = \lambda \underline{F}_{\boldsymbol{c}}(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2) + (1 - \lambda)\overline{F}_{\boldsymbol{c}}(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2)$, where $\underline{F}_{\boldsymbol{c}}\left(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2\right) = [F_1(\boldsymbol{\xi}^1) + F_2(\boldsymbol{\xi}^2) - 1]_+$ is the copula corresponding to group-countermonotonicity and $\overline{F}\left(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2\right) = \min\{F_1(\boldsymbol{\xi}^1), F_2(\boldsymbol{\xi}^2)\}$ is the copula corresponding to group-comonotonicity.*

If $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are univariate random variables, say, for the separable chance constraints case, we can choose

$$\lambda = (\rho_{\max} - \rho) / (\rho_{\max} - \rho_{\min}) \in [\rho_{\min}, \rho_{\max}],$$

where $\rho_{\min}$ and $\rho_{\max}$ are the corresponding minimal and maximal linear correlations between $\boldsymbol{c}^1$ and $\boldsymbol{c}^2$, such that the bivariate mixture distribution given by $F_{\boldsymbol{c}}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) = \lambda \underline{F}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) + (1 - \lambda)\overline{F}(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$, has marginals F_1 and F_2 and linear correlation ρ.

Theorem 4.4. *Suppose that there are only two rows in* (JC) *with $\epsilon \leq 0.5$ and $g_i^k(x)$ is nonnegative, $i = 1, \ldots, I^k$, $k = 1, 2$. Denote*

$$Y_{mix} = \left\{ y \in \mathbb{R}^2_+ \mid \text{there exists } z \in \mathbb{R}_+ \text{ such that } \left\{ \begin{array}{l} \lambda(y_1 + y_2 - 1) + (1 - \lambda)z \geq 1 - \epsilon \\ y_1 \geq z, y_2 \geq z \end{array} \right\} \right. . \tag{22}$$

Given Assumption 4.2, constraints (17) *together with $y \in Y_{mix}$ is equivalent to the joint chance constraint (16).*

Proof. Given Assumption 4.2, by Proposition 4.3 and introducing auxiliary variables y_1, y_2, constraint (16) is equivalent to (17) together with

$$\lambda \max\{y_1 + y_2 - 1, 0\} + (1 - \lambda) \min\{y_1, y_2\} \geq 1 - \epsilon. \tag{23}$$

There are two cases for the max operator.

For the first case when $y_1 + y_2 - 1 < 0$, we have that (23) is equivalently to

$$(1 - \lambda) \min\{y_1, y_2\} \geq 1 - \epsilon,$$

which implies

$$1 - \epsilon \leq (1 - \lambda) \min\{y_1, y_2\} \leq (1 - \lambda)\frac{1}{2}(y_1 + y_2) < \frac{1 - \lambda}{2} \leq \frac{1}{2}.$$

given the condition $y_1 + y_2 - 1 < 0$. As ϵ is less than 0.5, this case in excluded from the feasible region.

For the second case when $y_1 + y_2 - 1 \geq 0$, we have that (23) is equivalent to $y \in Y_{mix}$ by introducing the auxiliary variables $z \in \mathbb{R}_+$, which gives the desired results. □

If we consider z as an auxiliary decision variable, $y \in Y_{mix}$ can be formulated as a polyhedral set jointly of y and z. Given the conditions in Assumptions 2.2 and 2.3, the joint chance constrained programming problem with two rows under the mixture dependence is convex and can be equivalently formulated as deterministic convex programming problems.

4.5 Mixture dependence under multi-rows case

We next investigate the case with multi-rows in a joint chance constraint.

In the multi-row case, we consider a linear combination between the most negative dependence and the most positive dependence to represent the dependence structure. Group-comonotonicity can still be used to represent most positive dependence structure. However, the discussion for the most negative dependence is more complex. Classical countermonotonicity is is defined only for the two-row case. To extend to multi-row case, we propose a new mixture of group-comonotonicity and group-countermonotonicity to represent the dependence structure between different multi-rows.

The idea is to divide the rows into two groups, each group belongs to a group-comonotonicity. Two rows in different groups are group-countermonotonic. If we consider all group combinations, we have $S(K,2)+1$ possible partitions. $S(K,n)$ is the Stirling numbers of the second kind, which can be computed explicitly by $S(K,n)=\frac{1}{n!}\sum_{k=0}^{K}(-1)^k\binom{n}{k}(K-k)^K$, with $S(K,2)=2^{K-1}-1$. We denote the index set of the first group in the sth case by I_s. The index set of the second group is denoted I_s^c, also known as the complementary set of I_s. We assign to each case a weight parameter which corresponds to the likelihood of occurrence.

Assumption 4.3. *Assume that the dependence between $\boldsymbol{c}^1,\boldsymbol{c}^2,\ldots,\boldsymbol{c}^K$ in K rows is a linear combination of all possible group-countermonotonicity and group-comonotonicity cases,*

$$F_{\boldsymbol{c}}(\xi_1,\xi_2,\ldots,\xi_K)=\sum_{s=1}^{S(K,2)}\lambda_s F^s(\xi_1,\xi_2,\ldots,\xi_K)+\lambda_0\overline{F}(\xi_1,\xi_2,\ldots,\xi_K),$$

where F^s is the distribution function such that $\{\boldsymbol{c}^k|k\in I_s\}$ are group-comonotonic, $\{\boldsymbol{c}_k|k\in I_s^c\}$ are group-comonotonic, and $\boldsymbol{c}_i$, $\boldsymbol{c}_j$ are group-countermonotonic for any $i\in I_s$ and $j\in I_s^c$. $\overline{F}\left(\xi_1,\xi_2,\ldots,\xi_K\right)=\min_{k=1,\ldots,K}\left\{F_k(\xi_k)\right\}$. $\lambda_s, s=0,1,\ldots,S(K,2)$, is the likelihood of occurrence which measures the probability that $\overline{F}$ or F^s appears such that $\sum_{s=0}^{S(K,2)}\lambda_s=1$.

Theorem 4.5. *Suppose that there are K ($K>2$) rows in* (JC) *with $\epsilon\le 0.5$ and $g_i^k(x)$ is nonnegative valued, $i=1,\ldots,I_k$, $k=1,\ldots,K$. Denote*

$$Y_{mmix}=\left\{y\in[1-\epsilon,1]^K\;\middle|\;\begin{array}{l}\text{there exists } z^1\in[1-\epsilon,1]^{S(K,2)},z^2\in[1-\epsilon,1]^{S(K,2)},z^0\in[1-\epsilon,1]\\ \quad\text{such that }\left\{\begin{array}{l}\sum_{s=1}^{S(K,2)}\lambda_s\left(z_s^1+z_s^2-1\right)+\lambda_0 z_0\ge 1-\epsilon,\\ y_k\ge z_s^1,\ k\in I_s, s=1,\ldots,S(K,2),\\ y_k\ge z_s^2,\ k\in I_s^c, s=1,\ldots,S(K,2),\\ y_k\ge z^0,\ k=1,\ldots,K.\end{array}\right.\end{array}\right\}. \tag{24}$$

Given Assumption 4.3, the joint chance constraint (16) is equivalent to a series of individual chance constraints (17) *together with $y\in Y_{mmix}$.*

Proof. Given Assumption 4.3, the joint probability distribution of $\boldsymbol{c}^1,\ldots,\boldsymbol{c}^K$ has a finite mixture distribution similar to Assumption 2.1. Moreover, $\upsilon(x)$ which is linear to $\boldsymbol{c}^1,\ldots,\boldsymbol{c}^K$ is also finitely mixed, thus

$$F_{\upsilon(x)}(\beta)=\sum_{s=1}^{S(K,2)}\lambda_s F_{\upsilon(x)}^s(\beta)+\overline{F}_{\upsilon(x)}(\beta). \tag{25}$$

By Lemma 4.3, $\{\upsilon^k(x),k\in I_s\}$ and $\{\upsilon^k(x),k\in I_s^c\}$ are comonotonic, and $\{\upsilon^k(x),\upsilon^j(x)|k\in I_s,j\in I_s^c\}$ is countermonotonic.

Then by Lemma 4.2, we have

$$F^s_{\upsilon(x)}(\beta) = \max\left\{0, (\min_{k\in I_s} F_{\upsilon^k(x)}(\beta_k)) + (\min_{k\in k_s^c} F_{\upsilon^k(x)}(\beta_k)) - 1\right\},$$

and $\overline{F}_{\upsilon(x)}(\beta) = \min_{k=1,\dots,K} F_{\upsilon^k(x)}(\beta_k)$.

Moreover, the sub-probability $F^s_{\upsilon(x)}(\beta)$ should be always larger than or equal to the total probability $F_{\upsilon(x)}(\beta)$. Thus, $F_{\upsilon^k(x)}(\beta_k) \geq 1-\epsilon$ is a necessary condition for $F_{\upsilon(x)}(\beta) \geq 1-\epsilon$, $k = 1, \dots, K$. By assumption $\epsilon \leq 0.5$, we have that $(\min_{k\in I_s} F_{\upsilon^k(x)}(\beta_k)) + (\min_{k\in k_s^c} F_{\upsilon^k(x)}(\beta_k)) - 1 \geq (1-\epsilon) + (1-\epsilon) - 1 \geq 0$. Then, by Lemma 4.2, we have

$$F^s_{\upsilon(x)}(\beta) = (\min_{k\in I_s} F_{\upsilon^k(x)}(\beta_k)) + (\min_{k\in k_s^c} F_{\upsilon^k(x)}(\beta_k)) - 1,$$

By taking $F^s_{\upsilon(x)}(\beta)$ back to (25), we have that $F_{\upsilon(x)}(\beta) \geq 1-\epsilon$ is equivalent to $F_{\upsilon^k(x)}(\beta_k) \geq y$ together with $y \in Y_{mmix}$, where

$$Y_{mmix} = \left\{ y \in [1-\epsilon, 1]^K \;\middle|\; \sum_{s=1}^{S(K,2)} \lambda_s \left((\min_{k\in I_s} y_k) + (\min_{k\in I_s^c} y_k) - 1 \right) + \lambda_0 (\min_{k=1,\dots,K} y_i) \geq 1-\epsilon \right\}.$$

Introducing a series of auxiliary variables $z_s^1 \in [1-\epsilon, 1]$, $z_s^2 \in [1-\epsilon, 1]$, $s = 1, \dots, S(K,2)$, and z_0 leads to the desired results. □

Similarly to Theorem 4.4, we can obtain a convex deterministic reformulation of (JC), given the conditions in Assumptions 2.2 and 2.3. It is worth noting that, the group-comonotonicity and group-countermonotonicity are invariant to linear transformation with only positive slopes. Thus, the deterministic reformulations in Theorem 4.4 and 4.5 must rely on the positiveness of g_i^k.

5 Numerical tests

In this section, we carry out three groups of numerical tests. Section 5.1.1 considers the convergence of the piecewise linear approximation under finite mixture Gaussian distribution used in Section 2.4. Section 5.1.2 gives the numerical performance of the proposed partial sampling method for the individual chance constrained problem under GH distribution. Section 5.2 studies different dependence structures in a joint chance constraint. All the reformulation problems are solved by MOSEK solver in CVX package [22] with Matlab 2018b, on a Laptop with an Intel Core i7-8550U CPU and 16.0 GB RAM.

5.1 Individual case

In this experiment, we first consider the individual chance constrained problem (IC) with $I = n$, $g_0(x) = \prod_{i=1}^n x_i^{-\frac{1}{n}}$, $g_i(x) = x_i, i = 1, \cdots, n$, $X = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^n x_i \leq 1\}$, β is generated uniformly on $[1, 2]$ and $\epsilon = 0.01$.

5.1.1 Piecewise linear approximation results under mixture Gaussian distribution

In this subsection, we investigate the convergence of piecewise linear approximation under finite mixture Gaussian distribution with $M = 3$, $n = 10$. The weights $w_m, m = 1, \dots, M$, are generated

randomly with $w_m \geq 0, \sum_{m=1}^{M} w_m = 1$. For $m = 1, \ldots, M$, the mean vector $\boldsymbol{\mu}_m$ is generated randomly with command "abs(randn(n,M))" and the each element in covariance matrix $\boldsymbol{\Gamma}_m$ is random generated uniformly in $[0, 1]$. Such a setting satisfies the conditions in Theorem 2.3 and thus provides a convex reformulation of the individual chance constrained problem.

We set $S = 3, 10, 50, 100, 200$, then we solve the individual chance constrained problem with two kinds of piecewise linear approximations and obtain five groups of lower bounds and upper bounds. The first column in Table 2 gives the values of S. The second and third columns give the upper bound value and the corresponding CPU time in seconds. The fourth and fifth columns give the lower bound value and the CPU time in seconds. The last column shows the relative difference between the lower bound and the upper bound.

Table 2: Bounds and CPU times of upper and lower piecewise linear approximations for mixture distribution with $M = 3$, $n = 10$.

S	Upper Bound		Lower Bound		Gap(%)
	Value	CPU Time(s)	Value	CPU Time(s)	
3	11.4734	3.5625	11.1634	3.3438	2.78
10	11.2779	11.4219	11.2730	12.1406	0.04
50	11.2739	73.7344	11.2732	74.5313	0.01
100	11.2737	205.7656	11.2737	217.5469	0.00
200	11.2737	687.9531	11.2737	670.3438	0.00

From Table 2, we can see that as S increases, the difference between the lower bound and the upper bound monotonically decreases. When the number of tangent points reaches 100, these bounds are tight. Meanwhile, the solution times in the third and fifth columns valid the tractability of the piecewise linear approximation problems. These results illustrate that, with a proper size of linear approximations, the piecewise linear approximations can provide a good approximation solution.

5.1.2 Tests of partial sampling method under GH distribution

In this subsection, we test the numerical performance of the proposed partial sampling method for the individual chance constrained problem under GH distribution.

In this subsection, we set $n = 5$. For GH distribution, we set $\lambda = -0.5$, $\chi = 1$, $\psi = 2$, γ is generated uniformly in $[0.5, 1.5]^n$, μ is generated with the command "abs(randn(n,1))" and the semidefinite matrix Σ is generated randomly with all the elements in $[0, 1]$.

To better illustrate the efficiency of partial sampling method, we compare the performances between partial sampling method and SAA method. For partial sampling method, we set the sample size $M = 2, 3, 4, 6, 8, 10$. For SAA method, the sample size N is set to $70, 100, 200, 500, 1000, 2000$. Then, we solve the the individual chance constrained problem under GH distribution by partial sampling method and SAA method. We reformulate the SAA problem as a mixed-integer programming problem [40], and solve the MIP by MOSEK solver in the same testing environment. To reduce the randomness from the sampling, we re-run each group of problems 100 times. The statistics of the results and CPU times are summarized in Table 3, where V_M and T_M are the mean of of the optimal values and the CPU times, respectively. V_V and T_V are the variances of the optimal values and the CPU times, respectively.

Table 3: Average values (V_M and T_M) and variances (V_V and T_V) of the optimal values and the CPU times from 100 re-run instances of the partial sampling method and SAA method, $n = 5$

M	partial sampling method				N	SAA			
	Value		CPU Time(s)			Value		CPU Time(s)	
	V_M	V_V	T_M	T_V		V_M	V_V	T_M	T_V
2	11.8310	6.7383	14.7469	0.5786	70	12.7405	5.6599	24.7717	0.0510
3	11.6259	2.6169	21.2169	9.5440	100	10.7351	1.7902	27.1777	0.0236
4	12.3248	2.1039	27.4059	3.9605	200	11.7070	2.3306	40.1897	0.3015
6	12.0949	1.3427	34.9122	4.4945	500	12.1059	0.9107	59.3041	67.723
8	12.0041	0.3974	50.9703	5.3930	1000	12.5771	0.4391	80.4680	1860.6
10	12.1293	0.3017	76.5669	11.8546	2000	12.5817	0.3812	180.4492	29228

From Table 3, we observe that for each row, both the average value and variance of optimal values for partial sampling method and SAA method are comparable. However, the average CPU time of partial sampling method is smaller than that the CPU time of SAA. Moreover, the variance of the CPU time of partial sampling method is smaller than that the CPU time of SAA method when the samples size increases. In addition, the variance of the CPU time of SAA method is smaller than that of the partial sampling method for only small sample sizes.

Figure 1 graphically illustrates the results given in Table 3. Figure (a) shows the box plot of optimal values of the partial sampling method and SAA method when M and N increase. As shown in Figure 1, both partial sampling method and SAA can convergence as M and N increase. Partial sampling method shows the same performance as SAA with only 10 samples while SAA requires 2000 samples.

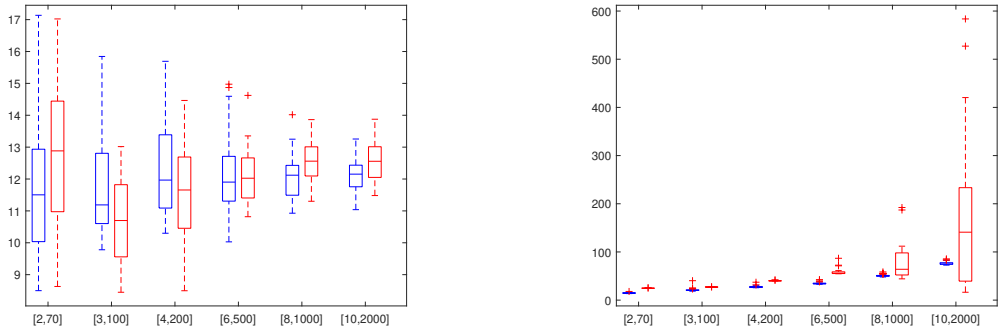
The right box plot in Figure 1 shows the corresponding CPU times of the two methods. We observe that the average CPU time of partial sampling method is similar to SAA method, while the variance of partial sampling method is smaller, especially when the sample size is large. The reason is that, the sample size M needed for partial sampling method is smaller than for SAA method, thus it has less randomness in the sampling process. Meanwhile, the partial sampling method provides a convex programming approximation problem under our setting, while the SAA method leads to a mixed-integer programming approximation problem.

From Table 3 and Figure 1, we find that the proposed partial sampling method can achieve almost the same performances as the SAA method with far smaller number of samples and less CPU time.

5.2 Joint case

In this section, we consider the joint chance constrained optimization problem (JP) with $K = 2$, $I_k = n, k = 1, \cdots, K$, $g_0(x) = \prod_{i=1}^{n} x_i^{-\frac{1}{n}}$, $g_i^k(x) = x_i, k = 1, \ldots, K, i = 1, \ldots, n$, $X = \{x \in \mathbb{R}^n : x \geq 0, \sum_{i=1}^{n} x_i \leq 1\}$. β_k is generated uniformly in $[1, 2]$, $k = 1, \ldots, K$, and $\epsilon = 0.01$. We assume that for $k = 1, \cdots, K$, $\boldsymbol{c}^k$ follows a multivariate Gaussian distribution with mean vector μ_k and covariance matrix $\Sigma_k \succeq 0$, where μ_k are generated uniformly in $[0, 1]^{I_k}$ and each element of Σ_k is generated uniformly in $[0, 1]$.

In this experiment, we investigate different dependence structures between the two rows numerically. We show the optimal values and CPU times in seconds of the joint chance constrained



a) box plot of optimal values

b) box plot of CPU times

Figure 1: Box plot of optimal values and CPU times of the partial sampling method (left blue) and the SAA method (right red) with different sample sizes $[M, N]$. For each group, the left blue box is partial sampling method; the right red box is SAA method. On each box, the central mark indicates the median, and the bottom and top edges of the box indicate the 25% and 75% percentiles from the 100 instances, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers are plotted individually using the $'+'$ marker symbol

problem with different dependence structures and different sizes of n in Table 4. The second column shows the optimal value and CPU time under the group-countermonotonicity case. The third column shows the results under the independent row case. The forth to sixth columns present the results under the mixture dependence structure, with different weight $\lambda = 0.7, 0.5, 0.3$, respectively. The last column gives the results under the group-comonotonicity case.

From Table 4, we can observe that there are slight difference between the CPU times in different dependence cases. The optimal value under the group-comonotonicity (extreme positive dependence) is always the smallest, which can be viewed as the best-case when facing ambiguity of the dependence structure. While the optimal value under the group-countermonotonicity (extreme negative dependence) is always the largest, it can be viewed as the worst-case, which makes a robust constraint on the ambiguity of the dependence structure. The mixture dependence case combines the group-countermonotonicity and group-comonotonicity and thus provides results between the best and the worst cases. Table 4 also shows numerically that the independent row case lies between the worst case and the best case, and more particularly closer to the worst-case in these groups of numerical instances.

6 Conclusion

In this paper, we study chance constraints with skewed distributions and joint dependent structures. We study the mathematical properties of individual chance constraints under finite mixture Gaussian distribution and derive tractable reformulations. Then, we extends the solution methods under finite mixture problems to the case of generalized hyperbolic distributions by using a partial sampling approach. We study different deterministic reformulations of joint dependent chance constraints under different kinds of dependence structures. Our numerical results show the performances of our different approaches on randomly generated instances. Our partial sampling method outperforms the standard SAA method.

Table 4: Optimal values and CPU times of (JC) under group-countermonotonicity case ($\underline{F}$), independence rows case, mixture dependence cases with weights $\lambda = 0.7, 0.5, 0.3$ and group-comonotonicity case ($\underline{F}$).

n		$\underline{F}$	Independent	Mixture Dependence			$\underline{F}$
				$\lambda = 0.7$	$\lambda = 0.5$	$\lambda = 0.3$	
5	Value	8.2754	8.2742	8.1935	8.1217	8.0291	7.8262
	CPU Time(s)	4.6094	4.8750	4.6406	4.3594	4.4531	4.7031
10	Value	15.8244	15.8225	15.6908	15.5736	15.4225	15.0911
	CPU Time(s)	15.8750	15.4531	15.5313	15.7188	15.4219	15.1875
15	Value	18.2898	18.2875	18.1344	17.9980	17.8221	17.4360
	CPU Time(s)	20.0781	19.8594	19.3281	18.6719	20.5313	19.4219
20	Value	21.8191	21.8163	21.6247	21.4541	21.2341	20.7515
	CPU Time(s)	28.9375	26.7031	27.1406	27.5156	26.5938	27.7188
30	Value	31.2285	31.2254	31.0253	30.8664	30.6861	30.3772
	CPU Time(s)	45.0625	44.8906	43.6563	43.2500	45.9688	45.8125
50	Value	77.7786	77.7698	77.1631	76.6227	75.9256	74.3956
	CPU Time(s)	98.9531	97.6250	98.6406	99.6719	98.7344	97.7500
80	Value	95.7127	95.7016	94.9361	94.2544	93.3749	91.4447
	CPU Time(s)	225.1875	245.9531	234.6719	245.6250	239.1406	240.9063

References

[1] L. Adam, M. Branda, H. Heitsch, and R. Henrion, *Solving joint chance constrained problems using regularization and Benders' decomposition*, Annals of Operations Research, 292 (2020), pp. 683–709.

[2] S. Ahmed, J. Luedtke, Y. Song, and W. Xie, *Nonanticipative duality, relaxations, and formulations for chance-constrained stochastic programs*, Mathematical Programming, 162 (2017), pp. 51–81.

[3] P. Arcidiacono and J. B. Jones, *Finite mixture distributions, sequential likelihood and the EM algorithm*, Econometrica, 71 (2003), pp. 933–946.

[4] V. S. Bawa, *On Chance Constrained Programming Problems with Joint Constraints*, Management Science, 19 (1973), pp. 1326–1331.

[5] J. R. Birge and L. Chavez-Bedoya, *Portfolio optimization under the generalized hyperbolic distribution: optimal allocation, performance and tail behavior*, Quantitative Finance, 21 (2021), pp. 199–219.

[6] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[7] I. Buckley, D. Saunders, and L. Seco, *Portfolio optimization when asset returns have the Gaussian mixture distribution*, European Journal of Operational Research, 185 (2008), pp. 1434–1461.

[8] G. C. Calafiore and L. E. Ghaoui, *On distributionally robust chance-constrained linear programs*, Journal of Optimization Theory and Applications, 130 (2006), pp. 1–22.

[9] M. C. Campi and S. Garatti, *A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality*, Journal of Optimization Theory and Applications, 148 (2011), pp. 257–280.

[10] Z. Chen, D. Kuhn, and W. Wiesemann, *Data-driven chance constrained programs over Wasserstein balls*, arXiv:1809.00210 [math], (2018). arXiv: 1809.00210.

[11] Z. Chen, S. Peng, and J. Liu, *Data-driven robust chance constrained problems: A mixture model approach*, Journal of Optimization Theory and Applications, 179 (2018), pp. 1065–1085.

[12] B. Cheng, Z. Chen, and J. Liu, *Multi-stage financial index tracking model under GH distribution*, Pakistan Journal of Statistics, 29 (2013), pp. 795–810.

[13] J. Cheng, E. Delage, and A. Lisser, *Distributionally robust stochastic knapsack problem*, SIAM Journal on Optimization, 24 (2014), pp. 1485–1506.

[14] J. Cheng, C. Gicquel, and A. Lisser, *Partial sample average approximation method for chance constrained problems*, Optimization Letters, 13 (2019), pp. 657–672.

[15] J. Cheng, M. Houda, and A. Lisser, *Chance constrained 0–1 quadratic programs using copulas*, Optimization Letters, 9 (2015), pp. 1283–1295.

[16] J. Cheng and A. Lisser, *A second-order cone programming approach for linear programs with joint probabilistic constraints*, Operations Research Letters, 40 (2012), pp. 325–328.

[17] J. Cuestaalbertos, L. Ruschendorf, and A. Tuerodiaz, *Optimal coupling of multivariate distributions and stochastic processes*, Journal of Multivariate Analysis, 46 (1993), pp. 335–361.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society: Series B (Methodological), 39 (1977), pp. 1–22.

[19] E. Eberlein and U. Keller, *Hyperbolic distributions in finance*, Bernoulli, 1 (1995), pp. 281–299.

[20] P. Embrechts, A. Mcneil, and D. Straumann, *Correlation and dependence in risk management: properties and pitfalls*, in Risk management: value at risk and beyond, Cambridge University Press, edited by m. a. h. dempster ed., 2002, pp. 176–223.

[21] J. Gorski, F. Pfeuffer, and K. Klamroth, *Biconvex sets and optimization with biconvex functions: a survey and extensions*, Mathematical Methods of Operations Research, 66 (2007), pp. 373–407.

[22] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming, version 2.1.* `http://cvxr.com/cvx`, 2014.

[23] G. A. Hanasusanto, V. Roitch, D. Kuhn, and W. Wiesemann, *Ambiguous joint chance constraints under mean and dispersion information*, Operations Research, 65 (2017), pp. 751–767.

[24] R. Henrion, *Structural properties of linear probabilistic constraints*, Optimization, 56 (2007), pp. 425–440.

[25] R. Henrion and C. Strugarek, *Convexity of chance constraints with independent random variables*, Computational Optimization and Applications, 41 (2008), pp. 263–276.

[26] R. Henrion and C. Strugarek, *Convexity of chance constraints with dependent random variables: the use of copulae*, in Stochastic Optimization Methods in Finance and Energy, M. Bertocchi, G. Consigli, and M. A. H. Dempster, eds., vol. 163, Springer New York, New York, NY, 2011, pp. 427–439. Series Title: International Series in Operations Research & Management Science.

[27] B. M. Hodge, A. Florita, K. Orwig, D. Lew, and M. Milligan, *A comparison of wind power and load forecasting error distributions: preprint*, (2012), p. 10.

[28] L. J. Hong, Y. Yang, and L. Zhang, *Sequential convex approximations to joint chance constrained programs: a Monte Carlo approach*, Operations Research, 59 (2011), pp. 617–630.

[29] Z. Hu, W. Sun, and S. Zhu, *Chance constrained programs with gaussian mixture models*, IISE Transactions, (2021), pp. 1–37.

[30] R. Ji and M. Lejeune, *Data-driven distributionally robust chance-constrained optimization With Wasserstein metric*, SSRN Electronic Journal, (2018).

[31] R. Jiang and Y. Guan, *Data-driven chance constrained stochastic program*, Mathematical Programming, 158 (2016), pp. 291–327.

[32] R. Kannan and J. R. Luedtke, *A stochastic approximation method for approximating the efficient frontier of chance-constrained nonlinear programs*, Mathematical Programming Computation, (2021).

[33] W. Lee and J. Y. Ahn, *On the multidimensional extension of countermonotonicity and its applications*, Insurance: Mathematics and Economics, 56 (2014), pp. 68–79.

[34] M. A. Lejeune and F. Margot, *Solving chance-constrained optimization problems with stochastic quadratic inequalities*, Operations Research, 64 (2016), pp. 939–957.

[35] B. Li, R. Jiang, and J. L. Mathieu, *Ambiguous risk constraints with moment and unimodality information*, Mathematical Programming, 173 (2019), pp. 151–192.

[36] J. Liu, A. Lisser, and Z. Chen, *Stochastic geometric optimization with joint probabilistic constraints*, Operations Research Letters, 44 (2016), pp. 687–691.

[37] J. Liu, A. Lisser, and Z. Chen, *Rectangular chance constrained geometric optimization*, Optimization & Engineering, 21 (2020), pp. 537–566.

[38] J. Liu, A. Lisser, and Z. Chen, *Distributionally robust chance constrained geometric optimization*, To appear in Mathematics of Operations Research, (2021).

[39] J. Luedtke and S. Ahmed, *A sample approximation approach for optimization with probabilistic constraints*, SIAM Journal on Optimization, 19 (2008), pp. 674–699.

[40] J. Luedtke, S. Ahmed, and G. L. Nemhauser, *An integer programming approach for linear programs with probabilistic constraints*, Mathematical Programming, 122 (2010), pp. 247–272.

[41] Y. Malevergne and D. Sornette, *Extreme financial risks: from dependence to risk management*, Springer, Berlin, 2006.

[42] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton series in finance, Princeton University Press, Princeton, N.J, 2005.

[43] B. L. Miller and H. M. Wagnert, *Chance constrained programming with joint constraints*, Operations Research, 13 (1965), pp. 930–945.

[44] R. B. Nelsen, *An Introduction to Copulas*, Springer series in statistics, Springer, New York, 2nd ed ed., 2006.

[45] A. Nemirovski and A. Shapiro, *Scenario approximations of chance constraints*, in Probabilistic and Randomized Methods for Design under Uncertainty, G. Calafiore and F. Dabbene, eds., Springer-Verlag, London, 2006, pp. 3–47.

[46] A. Nemirovski and A. Shapiro, *Convex approximations of chance constrained programs*, SIAM Journal on Optimization, 17 (2007), pp. 969–996.

[47] A. Peña-Ordieres, J. Luedtke, and A. Wächter, *Solving chance-constrained problems via a smooth sample-based nonlinear approximation*, SIAM Journal on Optimization, 30 (2020), pp. 2221–2250.

[48] A. Prékopa, *Probabilistic Programming*, in Handbooks in operations research and management science, 10, Elsevier, 2003, pp. 267–351.

[49] A. Prékopa, K. Yoda, and M. M. Subasi, *Uniform quasi-concavity in probabilistic constrained stochastic programming*, Operations Research Letters, 39 (2011), pp. 188–192.

[50] G. Puccetti and M. Scarsini, *Multivariate comonotonicity*, Journal of Multivariate Analysis, 101 (2010), pp. 291–304.

[51] G. Puccetti and R. Wang, *Extremal Dependence Concepts*, Statistical Science, 30 (2015).

[52] S. T. Rachev, C. Menn, and F. J. Fabozzi, *Fat-tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*, Wiley, 2005.

[53] A. Shapiro, D. Dentcheva, and A. Ruszczyński, eds., *Lectures on Stochastic Programming Modelling and Theory*, MOS-SIAM series on Optimization 9, SIAM, Philadelphia, 2009.

[54] W. van Ackooij and W. de Oliveira, *Convexity and optimization with copulæ structured probabilistic constraints*, Optimization, 65 (2016), pp. 1349–1376.

[55] W. van Ackooij and R. Henrion, *Gradient formulae for nonlinear probabilistic constraints with Gaussian and Gaussian-like distributions*, SIAM Journal on Optimization, 24 (2014), pp. 1864–1889.

[56] W. van Ackooij and J. Malick, *Eventual convexity of probability constraints with elliptical distributions*, Mathematical Programming, 175 (2019), pp. 1–27.

[57] Q. Wang, Y. Guan, and J. Wang, *A chance-constrained two-stage stochastic program for unit commitment with uncertain wind power output*, IEEE Transactions on Power Systems, 27 (2012), pp. 206–215.

[58] T. Watanabe and H. Ellis, *A joint chance-constrained programming model with row dependence*, European Journal of Operational Research, 77 (1994), pp. 325–343.

[59] W. Xie, *On distributionally robust chance constrained programs with Wasserstein distance*, Mathematical Programming, 186 (2021), pp. 115–155.

[60] W. Xie, S. Ahmed, and R. Jiang, *Optimized Bonferroni approximations of distributionally robust joint chance constraints*, Mathematical Programming, (2019).

[61] S.-S. Zhu, D. Li, and S.-Y. Wang, *Risk control over bankruptcy in dynamic portfolio selection: a generalized mean-variance formulation*, IEEE Transactions on Automatic Control, 49 (2004), pp. 447–457.

[62] S. Zymler, D. Kuhn, and B. Rustem, *Distributionally robust joint chance constraints with second-order moment information*, Mathematical Programming, 137 (2013), pp. 167–198.