Sample Approximations of Bilevel Stochastic Programming Problems with Probabilistic and Quantile Criteria

Sergey V. Ivanov[(✉)] and Aleksei N. Ignatov

Moscow Aviation Institute (National Research University),
Volokolamskoe Shosse, 4, Moscow 125993, Russia

Abstract. In this paper, bilevel stochastic programming problems with probabilistic and quantile criteria are considered. The lower level problem is assumed to be linear for fixed leader's (upper level) variables and fixed realizations of the random parameters. The objective function and the constraints of the lower level problem depend on the leader's strategy and random parameters. The objective function of the upper level problem is defined as the value of the probabilistic or quantile functional of the random losses on the upper level. We suggest conditions guaranteeing that the objective function of the upper level is a normal integrand. It is shown that these conditions are satisfied for a class of problems with positive coefficients of the lower level problem. This allows us to suggest sufficient conditions of the existence of a solution to the considered problem. We construct sample approximations of these problems. These approximations reduce to mixed integer nonlinear programming problems. We describe sufficient conditions of the convergence of the sample approximations to the original problems.

Keywords: Bilevel programming · Sample approximation · Stochastic programming · Value-at-Risk · Probabilistic criterion · Quantile criterion

1 Introduction

Bilevel programming problems describe hierarchical interaction between two subjects. The subject making decision first is called a leader. The second subject is called a follower. Their decisions are solutions to upper and lower level problems respectively. The parameters of the lower level problem depend on the leader's strategy. The leader takes into account the optimal follower's solution when the upper level strategy is selected. The theory of bilevel problems is described in monographs [1–3] and in the review [4].

In this paper, we study stochastic bilevel programming problems with probabilistic and quantile criteria. These criteria are used in stochastic models for taking into account reliability requirements [5]. The probabilistic criterion is

defined as the probability of successful work of the system modeled. The quantile criterion (also known as Value-at-Risk) is the minimal losses that cannot be exceeded with a given probability. There are a few works on using the probabilistic and quantile criteria in the stochastic bilevel optimization. The probabilistic criterion for a linear problem with fuzzy random data is studied in [6], where an algorithm to solve the problem is suggested. The quantile criterion for a stochastic bilevel problem with quantile criterion and discrete distribution of the random parameters is studied in [7]. A method to reduce this problem to a mixed integer programming problem is suggested in [7]. Another approach in stochastic programming to deal with reliability requirements is using coherent risk measures [8]. Properties of stochastic bilevel problems with criteria involving coherent risk measures and optimality conditions in these problems are studied in [9,10].

When the exact distribution of the random parameters is unknown, the objective function and the constraints of a stochastic problem can be estimated by using a sample. Thus, original problems are replaced by their approximations. The properties of the obtained sample approximations are studied in [11] for the expectation criterion and in [8,12] for problems with probabilistic constraints. In [13], the convergence of this method is studied for problems with probabilistic and quantile criteria.

In this paper, we study sample approximations of bilevel stochastic programming problems with probabilistic and quantile criteria. We reduce the sample approximations to mixed integer programming problems and give sufficient conditions of their convergence. Also, we describe conditions guaranteeing that the loss function of the problem is a normal integrand. These conditions are required to formulate results on the existence of an optimal solution and on the convergence of the sample approximations.

2 Statement

Let X be a random vector defined on a probability space $(\mathcal{X}, \mathcal{F}, \mathbf{P})$, where $\mathcal{X}$ is a closed subset of $\mathbb{R}^m$. The σ-algebra $\mathcal{F}$ is assumed to be complete, i.e., $S' \in \mathcal{F}$ if there exists a set $S \in \mathcal{F}$ such that $S' \subset S$ and $\mathbf{P}(S) = 0$. For simplicity, we assume that $X(x) = x$ for all $x \in \mathcal{X}$. This means that the sample space $\mathcal{X}$ is considered as the space of realizations of the random vector X.

Let $U \subset \mathbb{R}^r$ be a set feasible values of leader's variables. The follower's problem is defined by the linear programming problem

$$\mathcal{Y}^*(u, x) := \operatorname*{Arg\,min}_{y \in \mathbb{R}^s} \left\{ c(u, x)^\top y \mid y \in \mathcal{Y}(u, x) \right\}, \tag{1}$$

$$\mathcal{Y}(u, x) := \left\{ y \in \mathbb{R}^s \mid A(u, x) y \geq b(u, x),\ y \geq 0 \right\}. \tag{2}$$

where $u \in U$ is the leader's variable, y is the follower's variable, $A \colon U \times \mathcal{X} \to \mathbb{R}^{k \times s}$, $b \colon U \times \mathcal{X} \to \mathbb{R}^k$, $c \colon U \times \mathcal{X} \to \mathbb{R}^s$ are a matrix and vectors depending on u and x. Thus, the leader's variable u and the realization x of the random vector X define the constraints and the objective function of the follower's problem.

The set-valued mappings $\mathcal{Y}^*, \mathcal{Y} \colon U \times \mathcal{X} \to 2^{\mathbb{R}^s}$ are introduced by (1) and (2). We note that $\mathcal{Y}^*(u, x) = \emptyset$ if problem (1) is infeasible or unbounded.

Let $\Psi \colon U \times Y \times \mathcal{X} \to \mathbb{R}^*$ be a leader's loss function, where Y is the closure of the set

$$\bigcup_{u \in U, \, x \in \mathcal{X}} \mathcal{Y}(u, x),$$

$\mathbb{R}^* = \mathbb{R} \cup \{-\infty, +\infty\}$ is the extended real line.

To describe the losses of the leader when the follower chooses an optimal variable $y \in \mathcal{Y}^*(u, x)$, we introduce the function $\Phi \colon U \times \mathcal{X} \to \mathbb{R}^*$ by the rule

$$\Phi(u, x) := \inf_{y \in \mathcal{Y}^*(u,x)} \Psi(u, y, x). \tag{3}$$

The infimum in (3) means that the follower chooses the best decision for the leader among the optimal decisions $y \in \mathcal{Y}^*(u, x)$. Thus, the optimistic statement of the bilevel problem formulated below will be studied. We call Φ the optimistic leader's loss function.

Let us consider the probability function

$$P_\varphi(u) := \mathbf{P}\{\Phi(u, X) \le \varphi\}, \tag{4}$$

where $\varphi \in \mathbb{R}^*$ is a fixed level of the optimistic leader's loss function Φ. The value $P_\varphi(u)$ in (4) is well defined when the function $x \mapsto \Phi(u, x)$ is measurable. Sufficient conditions for this will be suggested below.

The quantile function is defined by the equality

$$\varphi_\alpha(u) := \min\left\{\varphi \in \mathbb{R}^* \mid P_\varphi(u) \ge \alpha\right\}, \tag{5}$$

where $\alpha \in (0, 1]$ is a fixed probability level.

In this paper, we study the probability maximization problem

$$U^* := \operatorname*{Arg\,max}_{u \in U} P_\varphi(u), \quad \alpha^* := \sup_{u \in U} P_\varphi(u), \tag{6}$$

and the quantile minimization problem

$$V^* := \operatorname*{Arg\,min}_{u \in U} \varphi_\alpha(u), \quad \varphi^* := \inf_{u \in U} \varphi_\alpha(u). \tag{7}$$

Problems (6) and (7) are optimistic bilevel stochastic programming problems with probabilistic and quantile criteria respectively.

3 Existence of Optimal Solution

It is known [5,14] that problems (6) and (7) are well defined and have optimal solutions if the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand. When the σ-algebra $\mathcal{F}$ is complete, the normal integrand can be defined as a lower semicontinuous in $u \in U$ and $\mathcal{B}(U) \times \mathcal{F}$-measurable function, where $\mathcal{B}(U)$ is the Borel

σ-algebra of subsets U. In this section, we suggest conditions under which the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand.

Let us consider the follower's problem (1). According to duality theory, the follower's variable $y \in \mathbb{R}^s$ is optimal in (1) if $y \in \mathcal{Y}(u, x)$ and there exists a vector $\lambda \in \mathbb{R}^k$ such that

$$A(u, x)^\top \lambda \le c(u, x), \ \lambda \ge 0, \tag{8}$$

$$(A(u, x)y - b(u, x))^\top \lambda = 0, \tag{9}$$

$$\left(A(u, x)^\top \lambda - c(u, x)\right)^\top y = 0. \tag{10}$$

Denote by $\Lambda(u, y, x)$ the set of $\lambda \in \mathbb{R}^k$ satisfying (8)–(10). Let us introduce the function

$$\delta_\Lambda(u, y, \lambda, x) = \begin{cases} 0 & \text{if } y \in \mathcal{Y}(u, x) \text{ and } \lambda \in \Lambda(u, y, x), \\ +\infty & \text{otherwise.} \end{cases}$$

Then the optimistic loss function Φ can be represented in the form

$$\Phi(u, x) = \min_{y \in Y} \min_{\lambda \in \Lambda(u, y, x)} \left\{ \Psi(u, y, x) + \delta_\Lambda(u, y, \lambda, x) \right\}. \tag{11}$$

We use the convention $-\infty + \infty = +\infty$ in (11) and below.

Let us denote by Y^* the closure of the set $\bigcup_{u \in U x \in \mathcal{X}} \mathcal{Y}^*(u, x)$.

Theorem 1. *Let the following conditions hold:*

(i)   *the function $(u, y, x) \mapsto \Psi(u, y, x)$ is lower semicontinuous in $(u, y) \in U \times Y$ and $\mathcal{B}(U) \times \mathcal{B}(Y) \times \mathcal{F}$-measurable;*
(ii) *the functions $(u, x) \mapsto A(u, x)$, $(u, x) \mapsto b(u, x)$, $(u, x) \mapsto c(u, x)$ are continuous in $u \in U$ and measurable in x;*
(iii) *the set Y^* is bounded;*
(iv) *there exists a compact set Λ^* such that $\Lambda^* \cap \Lambda(u, y, x) \ne \emptyset$ if and only if $\Lambda(u, y, x) \ne \emptyset$.*

Then the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand.

Proof. Taking into account (iii), (iv) and (11), the optimistic loss function can be rewritten in the form

$$\Phi(u, x) = \min_{y \in Y^*, \ \lambda \in \Lambda^*} \left\{ \Psi(u, y, x) + \delta_\Lambda(u, y, \lambda, x) \right\}. \tag{12}$$

From (i) it follows that the function $((u, y), x) \mapsto \Psi(u, y, x)$ is a normal integrand. From (ii) it follows that functions $(u, x) \mapsto A(u, x)$, $(u, x) \mapsto b(u, x)$, $(u, x) \mapsto c(u, x)$ are normal integrands [15, Example 14.29]. Therefore, the set

$$\{(u, y, \lambda, x) \in U \times Y^* \times \Lambda^* \times \mathcal{X} \mid \lambda \in \Lambda(u, y, x)\}$$

is $\mathcal{B}(U) \times \mathcal{B}(Y^*) \times \mathcal{B}(\Lambda^*) \times \mathcal{F}$-measurable and its sections $\{(u, y, \lambda) \mid \lambda \in \Lambda(u, y, x)\}$ are closed for all $x \in \mathcal{X}$. Hence, the function

$$((u, y, \lambda), x) \mapsto \Psi(u, y, x) + \delta_\Lambda(u, y, \lambda, x)$$

defined on the set $U \times Y^* \times \Lambda^* \times \mathcal{X}$ is a normal integrand. Since the set $Y^* \times \Lambda^*$ is compact, the minimum in (12) can be written and the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand [15, Proposition 14.47]. Theorem 1 is proved.

Let us consider the case when the values of functions b and c are positive.

Corollary 1. *Let conditions (i), (ii) of Theorem 1 hold. Suppose that*

(v) there exist $\underline{c}, \underline{b} \in \mathbb{R}$ such that

$$\inf_{u \in U, x \in \mathcal{X}} \min_{i=\overline{1,s}} c_i(u, x) > \underline{c} > 0,$$

$$\inf_{u \in U, x \in \mathcal{X}} \min_{j=\overline{1,k}} b_j(u, x) > \underline{b} > 0;$$

(vi) there exists $\bar{c} \in \mathbb{R}$ such that

$$\sup_{u \in U, x \in \mathcal{X}} \min_{y \in \mathcal{Y}(u, x)} c(u, x)^\top y < \bar{c};$$

Then the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand.

Proof. Let us notice that from (vi) it follows that $\mathcal{Y}(u, x) \neq \emptyset$ for all $u \in U$, $x \in \mathcal{X}$. From (v) and (vi) we get

$$\bigcup_{u \in U x \in \mathcal{X}} \mathcal{Y}^*(u, x) \subset \left\{ y \in \mathbb{R}^s \mid \sup_{u \in U, x \in \mathcal{X}} c(u, x)^\top y \leq \bar{c}, \ y \geq 0 \right\}$$

$$\subset \left\{ y \in \mathbb{R}^s \mid \max_{i=\overline{1,s}} y_i \leq \frac{\bar{c}}{\underline{c}}, \ y \geq 0 \right\}.$$

Thus, condition (iii) of Theorem 1 is satisfied.

From duality theory it is known that for $y \in \mathcal{Y}^*(u, x)$ there exists a vector $\lambda \in \Lambda(u, y, x)$ such that $c(u, x)^\top y = b(u, x)^\top \lambda$. Hence, the set Λ^* satisfying condition (iv) of Theorem 1 can be taken in the form

$$\Lambda^* = \left\{ \lambda \in \mathbb{R}^k \mid \max_{j=\overline{1,k}} \lambda_j \leq \frac{\bar{c}}{\underline{b}}, \ \lambda \geq 0 \right\}.$$

By Theorem 1, the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand. Corollary 1 is proved.

Example 1. Let us consider the follower's problem

$$\mathcal{Y}^*(u, x) = \operatorname*{Arg\,min}_{y \in \mathbb{R}} \left\{ \max\{0, u\} y \mid y \geqslant 0 \right\},$$

where $u \in U = [-1, 1]$. It easily seen that $\mathcal{Y}^*(u, x) = \{0\}$ if $u > 0$ and $\mathcal{Y}^*(u, x) = [0, +\infty)$ if $u \le 0$. Then for the normal integrand $\Psi(u, y, x) = e^{uy}$ the infimum-function $\Phi(u, x) = \inf_{y \in \mathcal{Y}^*(u,x)} \Psi(u, y, x)$ is not lower semicontinuous, because $\Phi(u, x) = 1$ if $u \ge 0$ and $\Phi(u, x) = 0$ if $u < 0$. This example shows that condition (v) of Corollary 1 cannot be replaced by the conditions

$$\inf_{u \in U, x \in \mathcal{X}} \min_{i=\overline{1,s}} c_i(u, x) \ge 0, \qquad \inf_{u \in U, x \in \mathcal{X}} \min_{j=\overline{1,k}} b_j(u, x) \ge 0.$$

Example 2. Let us consider a production planning model. In this model, the leader and the follower are the head office and the production division of a company. The leader can get a contract for the production of several types of products. The prices of these products manufactured according to the contract are deterministic. The leader gives the follower a task to produce the products. The leader supplies resources to the follower. The follower pays for using resources. The prices for using resources are known when the follower produces products, but the prices are considered to be random when the task for the follower is stated. The leader's variable $u \in \mathbb{R}^r$ ($u \ge 0$) consists of production volumes required by the contract. The follower's variable $y \in \mathbb{R}^s$ ($y \ge 0$) consists of volumes of required resources. Let $c(u, X) = \tilde{c}(X)$ be a random vector of prices for using resources such that $\bar{c} > c_i(u, x) = \tilde{c}_i(x) > \underline{c} > 0$ for all $x \in \mathcal{X}$. The matrix $A(u, x)$ is constant such that $A(u, x)y$ is the vector of manufactured products. Let $b(u, x) = u$. Thus, the constraint $A(u, x)y \ge b(u, x)$ means that the follower must produce the products according to the leader's decision u. The leader's loss function has the form $\Psi(u, y, x) = -\pi^\top u - \tilde{\pi}(x)^\top (A(u, x)y - b(u, x)) + f(y)$, where $f(y)$ is the cost of buying resources y, π is the vector of prices for manufactured products according to the contract, $\tilde{\pi}(x)$ is a random vector of prices for additionally manufactured products. The function f can be linear or convex (if big volumes of resources require additional costs). Notice that the conditions of Corollary 1 are satisfied for this model if $U = \{u \in \mathbb{R}^r \mid \underline{u} \le u \le \bar{u}\}$, where $0 < \underline{u} < \bar{u}$.

Let us formulate a corollary from Theorem 1 on the existence of optimal solutions to problems (6) and (7).

Corollary 2. *Let the conditions of Theorem 1 (or the conditions of Corollary 1) hold. Let the set U be compact. Then the set U^* of optimal solutions to problem (6) is nonempty. If there exists a point $u \in U$ such that $\varphi_\alpha(u) < +\infty$, then the set V^* of optimal solutions to problem (7) is nonempty.*

Proof. It is proved in [14, Theorem 6] that, if the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand, then the probability function $u \mapsto P_\varphi(u)$ is upper semicontinuous for all $\varphi \in \mathbb{R}$ and the quantile function $u \mapsto \varphi_\alpha(u)$ is lower semicontinuous for all $\alpha \in (0, 1]$ for any normal integrand $(u, x) \mapsto \Phi(u, x)$. The conditions of Theorem 1 (or the conditions of Corollary 1) guarantees that the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand. Thus, the assertion of Corollary 2 follows from the Weierstrass theorem.

4   Sample Approximations

In this section, we construct sample approximations of the probability maximization problem (6) and the quantile minimization problem (7) by using a sample $\left(X^1, X^2, \ldots, X^N\right)$ generated by the random vector X. The sequence of random vectors $\left(X^N\right)$, $N \in \mathbb{N}$, is defined on a complete probability space $(\Omega, \mathcal{F}', \mathbf{P}')$. The distribution functions of independent random variables X^N coincide with the distribution function of X.

Let us estimate the probability function (4) by the frequency of the event $\{\varPhi(u, X) \leq \varphi\}$:

$$P_\varphi^{(N)}(u) := \frac{1}{N} \sum_{\nu=1}^{N} \chi_{[-\infty, \varphi]}(\varPhi(u, X^\nu)), \ N \in \mathbb{N}, \tag{13}$$

where

$$\chi_S(x) := \begin{cases} 1, & x \in S; \\ 0, & x \notin S. \end{cases}$$

Replacing the probability function in (5) by the estimator (13), we obtain the sample estimator of the quantile function:

$$\varphi_\alpha^{(N)}(u) := \min \left\{ \varphi \in \mathbb{R}^* \mid P_\varphi^{(N)}(u) \geq \alpha \right\}.$$

We consider the sample approximation of the probability maximization problem in the form

$$U^{(N)} := \operatorname{Arg} \max_{u \in U} P_\varphi^{(N)}(u), \quad \alpha_N := \sup_{u \in U} P_\varphi^{(N)}(u), \tag{14}$$

and the sample approximation of the quantile minimization problem in the form

$$V^{(N)} := \operatorname{Arg} \min_{u \in U} \varphi_\alpha^{(N)}(u), \quad \varphi_N := \inf_{u \in U} \varphi_\alpha(u). \tag{15}$$

When a realization $\left(x^1, x^2, \ldots, x^N\right)$ of the sample $\left(X^1, X^2, \ldots, X^N\right)$ is fixed, problems (14) and (15) can be considered as stochastic programming problems with discrete distribution of the random parameters. This allows us to use the technique suggested in [16,17] for reducing the problems to deterministic mixed integer programming problems.

Recall that Y^* is the closure of the set $\bigcup_{u \in U, \ x \in \mathcal{X}} \mathcal{Y}^*(u, x)$. Suppose that a set $\varLambda^* \subset \mathbb{R}^k$ is chosen in such a way that, for all $u \in U$, $y \in Y^*$, $x \in \mathcal{X}$, the set $\varLambda^* \cap \varLambda(u, y, x) \neq \emptyset$ if and only if $\varLambda(u, y, x) \neq \emptyset$.

Let functions $\gamma_1 \colon U \times Y \times \mathcal{X} \times \mathbb{R} \to \mathbb{R}$, $\gamma_2 \colon U \times Y \times \mathcal{X} \to \mathbb{R}^k$, $\gamma_3 \colon U \times Y \times \mathcal{X} \to \mathbb{R}^k$, $\gamma_4 \colon \mathbb{R}^k \mapsto \mathbb{R}^k$, $\gamma_5 \colon U \times \mathbb{R}^k \times \mathcal{X} \to \mathbb{R}^s$, $\gamma_6 \colon U \times \mathbb{R}^k \times \mathcal{X} \to \mathbb{R}^s$, $\gamma_7 \colon Y \to \mathbb{R}^s$ satisfying the following conditions be known:

1. $\varPsi(u, y, x) - \varphi \leq \gamma_1(u, y, x, \varphi)$ for all $u \in U$, $y \in Y^*$, $x \in \mathcal{X}$, $\varphi \in \mathbb{R}$;
2. $-\gamma_2(u, y, x) \leq A(u, x)y - b(u, x) \leq \gamma_3(u, y, x)$ for all $u \in U$, $y \in Y^*$, $x \in \mathcal{X}$;

3. $\lambda \leq \gamma_4(\lambda)$ for all $\lambda \in \Lambda^*$.
4. $-\gamma_6(u, \lambda, x) \leq A(u, x)^\top \lambda - c(u, x) \leq \gamma_5(u, \lambda, x)$ for all $u \in U$, $\lambda \in \Lambda^*$, $x \in \mathcal{X}$;
5. $y \leq \gamma_7(y)$ for all $y \in Y^*$.

Let us introduce the variables y^ν and λ^ν corresponding to realizations x^ν, $\nu = \overline{1, N}$, and the vectors of auxilary binary variables $\delta \in \{0,1\}^N$, $\eta^\nu \in \{0,1\}^k$, $\zeta^\nu \in \{0,1\}^s$, $\nu = \overline{1, N}$. The sense of these variables follows from the proof of Theorem 2 given below. The value of δ_ν is equal to 1 if $\Phi(u, x^\nu) \leq \varphi$. If $\delta_\nu = 1$, then zero elements of η^ν correspond to nonzero elements of the dual variables for constraints $A(u, x^\nu)y^\nu \geq b(u, x^\nu)$, zero elements of ζ^ν correspond to nonzero elements of the vector y^ν. Problem (14) reduces to the problem

$$\frac{1}{N} \sum_{\nu=1}^{N} \delta_\nu \to \max_{u \in U,\, y^\nu \in Y,\, \lambda^\nu \in \mathbb{R}^k,\, \delta \in \{0,1\}^N,\, \eta^\nu \in \{0,1\}^k,\, \zeta^\nu \in \{0,1\}^s,\, \nu = \overline{1,N}} \tag{16}$$

subject to

$$\Psi(u, y^\nu, x^\nu) \leq \varphi + (1 - \delta_\nu)\gamma_1(u, y^\nu, x^\nu, \varphi), \tag{17}$$

$$A(u, x^\nu)y^\nu \geq b(u, x^\nu) - (1 - \delta_\nu)\gamma_2(u, y^\nu, x^\nu), \tag{18}$$

$$A(u, x^\nu)y^\nu \leq b(u, x^\nu) + (\eta^\nu + (1 - \delta_\nu)e_k) \circ \gamma_3(u, y^\nu, x^\nu), \tag{19}$$

$$0 \leq \lambda^\nu \leq ((2 - \delta_\nu)e_k - \eta^\nu) \circ \gamma_4(\lambda^\nu), \tag{20}$$

$$A(u, x^\nu)^\top \lambda^\nu \leq c(u, x^\nu) + (1 - \delta_\nu)\gamma_5(u, \lambda^\nu, x^\nu), \tag{21}$$

$$A(u, x^\nu)^\top \lambda^\nu \geq c(u, x^\nu) - (\zeta^\nu + (1 - \delta_\nu)e_s) \circ \gamma_6(u, y^\nu, \lambda^\nu, x^\nu), \tag{22}$$

$$0 \leq y^\nu \leq ((2 - \delta_\nu)e_s - \zeta^\nu) \circ \gamma_7(y^\nu),\ \nu = \overline{1, N}, \tag{23}$$

where e_s, e_k are vectors consisting of ones with dimension s and k respectively, $\circ$ denotes the element-wise product of two vectors.

Denote by $\left(\bar{u}, (\bar{y}^\nu), (\bar{\lambda}^\nu), \bar{\delta}, (\bar{\eta}^\nu), (\bar{\zeta}^\nu)\right)$ the optimal solution to problem (16), where $(\bar{y}^\nu) := (y^1, y^2, \ldots, y^N)$. Notation $(\bar{\lambda}^\nu)$, $(\bar{\eta}^\nu)$, $(\bar{\zeta}^\nu)$ has the same sense.

Theorem 2. *Let the conditions of Theorem 1 hold. Suppose that Λ^* satisfies condition (iv) of Theorem 1. Then,*

1. if $\bar{u}$ is an optimal value of the variable u in problem (16), then $\bar{u} \in U^{(N)}$;
2. if $\bar{\delta}$ is an optimal value of the variable δ in problem (16), then

$$\alpha_N = \frac{1}{N} \sum_{\nu=1}^{N} \bar{\delta}_\nu;$$

3. for any optimal $\tilde{u} \in U^{(N)}$ there exist values $\tilde{y}^\nu \in Y$, $\tilde{\lambda}^\nu \in \mathbb{R}^k$, $\tilde{\delta} \in \{0,1\}^N$, $\tilde{\eta}^\nu \in \{0,1\}^k$, $\tilde{\zeta}^\nu \in \{0,1\}^s$, $\nu = \overline{1, N}$ such that

$$\left(\tilde{u}, (\tilde{y}^\nu), (\tilde{\lambda}^\nu), \tilde{\delta}, (\tilde{\eta}^\nu), (\tilde{\zeta}^\nu)\right)$$

is an optimal solution to problem (16).

Proof. Let $\left(\bar{u}, (\bar{y}^\nu), (\bar{\lambda}^\nu), \bar{\delta}, (\bar{\eta}^\nu), (\bar{\zeta}^\nu)\right)$ be an optimal solution to problem (16). Let
$$K := \left\{\nu \mid \bar{\delta}_\nu = 1, \ \nu = \overline{1, N}\right\}.$$
It follows from inequalities (18)–(23) that, for all $\nu \in K$, $\bar{y}^\nu$ belongs to $\mathcal{Y}(\bar{u}, x^\nu)$, and inequalities (8)–(10) hold for $\lambda = \bar{\lambda}^\nu$, $y = \bar{y}^\nu$. Therefore, $\bar{y}^\nu \in \mathcal{Y}^*(\bar{u}, x^\nu)$. Hence, for any $\nu \in K$,
$$\Phi(\bar{u}, x^\nu) \leq \Psi(\bar{u}, \bar{y}^\nu, x^\nu) \leq \varphi.$$
Thus,
$$P_\varphi^{(N)}(\bar{u}) = \frac{1}{N}\sum_{\nu=1}^N \chi_{[-\infty,\varphi]}(\Phi(\bar{u}, x^\nu)) \geq \frac{1}{N}\sum_{\nu\in K}\bar{\delta}_\nu = \frac{1}{N}\sum_{\nu=1}^N \bar{\delta}_\nu = \bar{\alpha}^*, \qquad (24)$$
where $\bar{\alpha}^*$ is the optimal objective value of problem (16). We notice that inequality (24) is written for the fixed realization $(x^1, x^2, \ldots x^N)$ of the sample.

Now, let $\tilde{u}$ be an optimal solution to problem (14). Let
$$\tilde{K} := \left\{\nu \mid \mathcal{Y}^*(\tilde{u}, x^\nu) \neq \emptyset, \ \nu = \overline{1, N}\right\}. \qquad (25)$$
For each $\nu \in \tilde{K}$ let us choose $\tilde{y}^\nu \in \mathcal{Y}^*(\tilde{u}, x^\nu)$, $\tilde{\lambda}^\nu \in \Lambda^*$ in such a way that $\Phi(\tilde{u}, x^\nu) = \Psi(\tilde{u}, \tilde{y}^\nu, x^\nu)$. The existence of such values $\tilde{y}^\nu$ follows from the representation (12), because the minimum of the lower semicontinuous function $(y, \lambda) \mapsto \Psi(\tilde{u}, y, x^\nu) + \delta_\Lambda(\tilde{u}, y, \lambda, x^\nu)$ is attained on the compact set $Y^* \times \Lambda^*$. If $\nu \notin \tilde{K}$, then we take $\tilde{y}^\nu \in Y^*$, $\tilde{\lambda}^\nu \in \Lambda^*$ arbitrarily. If $\nu \in \tilde{K}$ and $\Phi(\tilde{u}, x^\nu) \leq \varphi$, then $\tilde{\delta}_\nu = 1$; otherwise $\tilde{\delta}_\nu = 0$. Let $\tilde{\zeta}_i^\nu = 1$ if $\tilde{y}_i^\nu = 0$ and $\tilde{\zeta}_i^\nu = 0$ if $\tilde{y}_i^\nu > 0$, $i = \overline{1, s}$; $\tilde{\eta}_j^\nu = 1$ if $\tilde{\lambda}_j^\nu = 0$ and $\tilde{\eta}_j^\nu = 0$ if $\tilde{\lambda}_j^\nu > 0$, $j = \overline{1, k}$. All the constraints (17)–(23) are satisfied for the solution $\left(\tilde{u}, (\tilde{y}^\nu), (\tilde{\lambda}^\nu), \tilde{\delta}, (\tilde{\eta}^\nu), (\tilde{\zeta}^\nu)\right)$. Thus,
$$P_\varphi^{(N)}(\tilde{u}) = \frac{1}{N}\sum_{\nu=1}^N \chi_{[-\infty,\varphi]}(\Phi(\tilde{u}, x^\nu)) = \frac{1}{N}\sum_{\nu=1}^N \tilde{\delta}_\nu \leq \bar{\alpha}^*. \qquad (26)$$
Taking into account the optimality of $\tilde{u}$, we obtain from inequality (24) that
$$P_\varphi^{(N)}(\tilde{u}) \geq P_\varphi^{(N)}(\bar{u}) \geq \bar{\alpha}^*. \qquad (27)$$
Hence, $\frac{1}{N}\sum_{\nu=1}^N \tilde{\delta}_\nu = \bar{\alpha}^*$. This proves the third assertion of the theorem. Combining (26) and (27), we get
$$P_\varphi^{(N)}(\bar{u}) = P_\varphi^{(N)}(\tilde{u}) = \bar{\alpha}^*. \qquad (28)$$
This implies the first assertion of the theorem. By definition, $\bar{\alpha}^* = \frac{1}{N}\sum_{\nu=1}^N \bar{\delta}_\nu$. We conclude from (28) that
$$\alpha_N = P_\varphi^{(N)}(\tilde{u}) = \bar{\alpha}^* = \frac{1}{N}\sum_{\nu=1}^N \bar{\delta}_\nu.$$
This equality proves the second assertion. All the assertions of Theorem 2 are proved.

Remark 1. Let us consider the model from Example 2. If the sets of feasible values of leader's and follower's variables are bounded then the function γ_i can be taken constant. In this case, constraints (17)–(23) are linear or convex depending on the properties of the function f.

The quantile minimization problem (15) reduces to a mixed integer programming problem:

$$\varphi \to \min_{\varphi \in \mathbb{R}^*, \, u \in U, \, y^\nu \in Y, \, \lambda^\nu \in \mathbb{R}^k, \, \delta \in \{0,1\}^N, \, \eta^\nu \in \{0,1\}^k, \, \zeta^\nu \in \{0,1\}^s, \, \nu = \overline{1,N}} \tag{29}$$

subject to

$$\frac{1}{N} \sum_{\nu=1}^{N} \delta_\nu \geq \alpha. \tag{30}$$

and (17)–(23).

Theorem 3. *Let the conditions of Theorem 1 hold. Suppose that Λ^* satisfies condition (iv) of Theorem 1. Then,*

1. *if $\bar{u}$ is an optimal value of the variable u in problem (29), then $\bar{u} \in V^{(N)}$;*
2. *if $\bar{\varphi}$ is the optimal value of the variable φ in problem (29), then $\varphi_N = \bar{\varphi}$;*
3. *for any optimal $\tilde{u} \in V^{(N)}$ there exist values $\tilde{y}^\nu \in Y$, $\tilde{\lambda}^\nu \in \mathbb{R}^k$, $\tilde{\delta} \in \{0,1\}^N$, $\tilde{\eta}^\nu \in \{0,1\}^k$, $\tilde{\zeta}^\nu \in \{0,1\}^s$, $\nu = \overline{1,N}$ such that*

$$\left(\varphi_N, \tilde{u}, (\tilde{y}^\nu), (\tilde{\lambda}^\nu), \tilde{\delta}, (\tilde{\eta}^\nu), (\tilde{\zeta}^\nu) \right)$$

   is an optimal solution to problem (29).

Proof. Let $\left(\bar{\varphi}, \bar{u}, \bar{y}^\nu, \bar{\lambda}^\nu, \bar{\delta}, \bar{\eta}^\nu, \bar{\zeta}^\nu \right)$ be an optimal solution to problem (16). Due to constraint (30),

$$\frac{1}{N} \sum_{\nu=1}^{N} \bar{\delta}_\nu \geq \alpha. \tag{31}$$

It follows from inequalities (18)–(23) that

$$\Phi(\bar{u}, x^\nu) \leq \Psi(\bar{u}, \bar{y}^\nu, x^\nu) \leq \bar{\varphi}. \tag{32}$$

if $\bar{\delta}_\nu = 1$ (see the proof of Theorem 2). Since inequalities (31) and (32) hold,

$$\varphi_\alpha^{(N)}(\bar{u}) = \min \left\{ \varphi \mid \frac{1}{N} \sum_{\nu=1}^{N} \chi_{[-\infty,\varphi]}(\Phi(\bar{u}, x^\nu)) \geq \alpha \right\} \leq \bar{\varphi}. \tag{33}$$

Now, let $\tilde{u}$ be an optimal solution to problem (15). For each $\nu \in \tilde{K}$ let us choose $\tilde{y}^\nu \in \mathcal{Y}^*(\tilde{u}, x^\nu)$, $\tilde{\lambda}^\nu \in \Lambda^*$ in such a way that $\Phi(\tilde{u}, x^\nu) = \Psi(\tilde{u}, \tilde{y}^\nu, x^\nu)$, where $\tilde{K}$ is defined in (25). If $\nu \notin \tilde{K}$, then we take $\tilde{y}^\nu \in Y^*$, $\tilde{\lambda}^\nu \in \Lambda^*$ arbitrarily. Let $\tilde{\zeta}_i^\nu = 1$ if $\tilde{y}_i^\nu = 0$ and $\tilde{\zeta}_i^\nu = 0$ if $\tilde{y}_i^\nu > 0$, $i = \overline{1,s}$; $\tilde{\eta}_j^\nu = 1$ if $\tilde{\lambda}_j^\nu = 0$ and $\tilde{\eta}_j^\nu = 0$ if $\tilde{\lambda}_j^\nu > 0$, $j = \overline{1,k}$. If $\nu \in \tilde{K}$ and $\Phi(\tilde{u}, x^\nu) \leq \varphi_N$, then $\tilde{\delta}_\nu = 1$; otherwise

$\tilde{\delta}_\nu = 0$. Taking into account the definition of the sample quantile function, we get $\sum_{\nu=1}^{N} \tilde{\delta}_\nu \geq \alpha$. Thus, all the constraints (17)–(23) and (30) are satisfied for the solution $\left(\varphi_N, \tilde{u}, (\tilde{y}^\nu), (\tilde{\lambda}^\nu), \tilde{\delta}, (\tilde{\eta}^\nu), (\tilde{\zeta}^\nu)\right)$. Hence $\bar{\varphi} \leq \varphi_N$.

Since φ_N is the optimal objective value in (15), $\varphi_N \leq \varphi_\alpha^{(N)}(\bar{u})$. Due to (33), we obtain

$$\varphi_N \leq \varphi_\alpha^{(N)}(\bar{u}) \leq \bar{\varphi} \leq \varphi_N.$$

Thus, $\varphi_\alpha^{(N)}(\bar{u}) = \bar{\varphi} = \varphi_N$. This proves assertions 1 and 2 of Theorem 3. Assertion 3 follows from the existence of the solution $\left(\varphi_N, \tilde{u}, (\tilde{y}^\nu), (\tilde{\lambda}^\nu), \tilde{\delta}, (\tilde{\eta}^\nu), (\tilde{\zeta}^\nu)\right)$ such that $\bar{\varphi} = \varphi_N$. Theorem 3 is proved.

5   Convergence of the Sample Approximations

The convergence of sample approximations of stochastic programming problems with probabilistic criterion is studied in [13], where it was proved that $\lim_{N\to\infty} \alpha_N = \alpha^*$ almost surely (a.s.) (with respect to the probability measure $\mathbf{P}'$) if the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand and U is nonempty and compact. The sufficient conditions guaranteeing that the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand are given in Theorem 1 and Corollary 1. Let us formulate the theorem on the convergence of the sample approximations of the bilevel stochastic programming problem with probabilistic criterion. Denote by

$$D(S, T) := \sup_{s \in S} \inf_{t \in T} \|s - t\|$$

the deviation of the set $S \subset \mathbb{R}^r$ from the set $T \subset \mathbb{R}^r$.

Theorem 4. *Suppose that the function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand. Let the set U be nonempty and compact. Then $\lim_{N\to\infty} \alpha_N = \alpha^*$ a.s. and $\lim_{N\to\infty} D\left(U^{(N)}, U^*\right) = 0$ a.s.*

Proof. The convergence of α_N to α^* a.s. is proved in [13, Theorem 7]. Also, it was proved that, under the conditions of the theorem, that every limit point $\bar{u}$ of the sequence (u^N), where $u^N \in U^{(N)}$, is optimal in problem (6) a.s., i.e., $\bar{u} \in U^*$ a.s. To prove the set convergence, suppose that $\limsup_{N\to\infty} D\left(U^{(N)}, U^*\right) > 0$ with nonzero probability. This implies that there exists $\epsilon > 0$ such that $\limsup_{N\to\infty} D\left(U^{(N)}, U^*\right) > \epsilon$ with probability $\beta > 0$. Then, with probability β, we can find a sequence u^N such that

$$\limsup_{N\to\infty} \inf_{u \in U^*} \|u^N - u\| > \epsilon.$$

From the compactness of the set U and the continuity of the function $v \mapsto \inf_{u \in U^*} \|v - u\|$ it follows that there exists a limit point $\bar{u}$ of the sequence (u^N) such that $\inf_{u \in U^*} \|\bar{u} - u\| \geq \epsilon$. Therefore, with probability $\beta > 0$ there exists a limit point $\bar{u}$ (depending on the realization of the sample) such that $\bar{u} \notin U^*$. But $\bar{u} \in U^*$ a.s. This contradiction proves the theorem.

In applied problems, it is important to know the sample size guaranteeing a given accuracy of the approximation. This question is studied in [18, Theorem 1]. According to this result, if the set U is finite, then

$$\mathbf{P}'\{U^{(N)} \subset U_\epsilon\} \geq \beta$$

for

$$N \geq 2\frac{\ln|U| - \ln(1 - \beta)}{|\ln(1 - \epsilon^2)|},$$

where $\beta \in (0, 1)$, $\epsilon \in (0, 1)$, $U_\epsilon := \{u \in U \mid P_\varphi(u) \geq \alpha^* - \epsilon\}$ is the set of ϵ-optimal solutions to problem (6). This estimation was obtained for arbitrary functions $(u, x) \to \Phi(u, x)$ being normal integrands and taking values from $\mathbb{R}$. Replacing infinite values of $\Phi(u, x)$ in definition (3) by finite values that have the same sign does not change $P_\varphi(u)$ and $P_\varphi^{(N)}(u)$. Thus, the given estimation is valid for the considered bilevel problems (of course, if the conditions of Theorem 1 or Corollary 1 are satisfied).

Sufficient conditions of the convergence of problems with quantile criteria are given in [13, 14].

Theorem 5 ([14, Theorem 10]). *Suppose that*

(i) *The set U is compact and nonempty.*
(ii) *The function $(u, x) \mapsto \Phi(u, x)$ is a normal integrand, and $\Phi(u, x) > -\infty$ for all $(u, x) \in U \times \mathcal{X}$.*
(iii) *If $\varphi^* \neq +\infty$, then for all $\epsilon > 0$ there exists a pair $(\tilde{u}, \tilde{\varphi}) \in U \times \mathbb{R}$ such that $|\tilde{\varphi} - \varphi^*| \leq \epsilon$ and $P_{\tilde{\varphi}}(\tilde{u}) > \alpha$.*

Then $\lim_{N \to \infty} \varphi_N = \varphi^*$ a.s. and every limit point of the sequence (v^N), where $v^N \in V^{(N)}$, is optimal in problem (7) a.s.

Theorem 5 was proved for arbitrary functions $(u, x) \to \Phi(u, x)$ being normal integrands and taking values from $(-\infty, +\infty]$. Thus, it holds for the considered functions Φ in bilevel problems. Due to condition (ii), Theorem 5 is not applied to functions Φ taking value $-\infty$.

In the same manner as in the proof of Theorem 4, it can be proved that the assertion on the optimality of limit points in Theorem 5 can be replaced by

$$\lim_{N \to \infty} D\left(V^{(N)}, V^*\right) = 0 \text{ a.s.}$$

The most difficult point in applying Theorem 5 is to check assumption (iii). It is hard to describe sufficient conditions for this, because the dependence $(u, x) \mapsto \Phi(u, x)$ must be known. However, in some cases (for example, in the case of linear follower's problem [19]) this dependence can be found. It is easy to check that assumption (iii) of Theorem 5 holds if the function $x \mapsto \Phi(u, x)$ is strictly increasing and X has a positive on $\mathbb{R}$ density.

6   Conclusion

In this paper, sample approximations of the stochastic optimistic bilevel programming problems with probabilistic and quantile criteria were studied. The sample approximations reduced to deterministic optimization problems. These problems can be solved by using special software for nonlinear optimization. Conditions ensuring the convergence of the sample approximations were given. Since these conditions require that the leader's loss function is a normal integrand, some classes of the considered problems with such leader's loss functions were described. Although the convergence was proved, the sufficient sample size for the infinite set of the leader's variables (and for the quantile minimization problem even when the set of the variable is finite) is still unknown. This question can be studied in future research.

References

1. Bard, J.F.: Practical Bilevel Optimization: Algorithms and Applications. Kluwer Academie Publishers, Dordrecht (1998)
2. Dempe, S.: Foundations of Bilevel Programming. Kluwer Academie Publishers, Dordrecht (2002)
3. Dempe, S., Kalashnikov, V., Pérez-Valdés, G.A., Kalashnykova, N.: Bilevel Programming Problems – Theory, Algorithms and Applications to Energy Network. Springer Verlag, Heidelberg (2015). https://doi.org/10.1007/978-3-662-45827-3
4. Dempe, S.: Bilevel optimization: theory, algorithms, applications and a bibliography. In: Dempe, S., Zemkoho, A. (eds.) Bilevel Optimization. SOIA, vol. 161, pp. 581–672. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52119-6_20
5. Kibzun, A.I., Kan, Y.S.: Stochastic Programming Problems with Probability and Quantile Functions. John Wiley & Sons, Chichester (1996)
6. Sakawa, M., Katagiri, H., Matsui, T.: Stackelberg solutions for fuzzy random bilevel linear programming through level sets and probability maximization. Oper. Res. Int. J. **12**(3), 271–286 (2012). https://doi.org/10.1007/s12351-010-0090-2
7. Dempe, S., Ivanov, S., Naumov, A.: Reduction of the bilevel stochastic optimization problem with quantile objective function to a mixed-integer problem. Appl. Stoch. Models Bus. Ind. **33**(5), 544–554 (2017). https://doi.org/10.1002/asmb.2254
8. Shapiro, A., Dentcheva, D., Ruszczyński, A.: Lectures on Stochastic Programming. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Modeling and Theory (2014)
9. Burtscheidt, J., Claus, M., Dempe, S.: Risk-Averse models in bilevel stochastic linear programming. SIAM J. Optim. **30**(1), 377–406 (2020). https://doi.org/10.1137/19M1242240
10. Burtscheidt, J., Claus, M.: Bilevel linear optimization under uncertainty. In: Dempe, S., Zemkoho, A. (eds.) Bilevel Optimization. SOIA, vol. 161, pp. 485–511. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-52119-6_17
11. Artstein, Z., Wets, R.J.-B.: Consistency of minimizers and the SLLN for stochastic programs. J. Convex Anal. **2**, 1–17 (1996)

12. Pagnoncelli, B.K., Ahmed, S., Shapiro, A.: Sample average approximation method for chance constrained programming: theory and applications. J. Optim. Theory Appl. **142**, 399–416 (2009). https://doi.org/10.1007/s10957-009-9523-6
13. Ivanov, S.V., Kibzun, A.I.: On the convergence of sample approximations for stochastic programming problems with probabilistic criteria. Autom. Remote Control **79**(2), 216–228 (2018). https://doi.org/10.1134/S0005117918020029
14. Ivanov, S.V., Kibzun, A.I.: General properties of two-stage stochastic programming problems with probabilistic criteria. Autom. Remote Control **80**(6), 1041—1057 (2019). https://doi.org/10.1134/S0005117919060043
15. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02431-3
16. Norkin, V.I., Kibzun, A.I., Naumov, A.V.: Reducing two-stage probabilistic optimization problems with discrete distribution of random data to mixed-integer programming problems *. Cybern. Syst. Anal. **50**(5), 679–692 (2014). https://doi.org/10.1007/s10559-014-9658-9
17. Kibzun, A.I., Naumov, A.V., Norkin, V.I.: On reducing a quantile optimization problem with discrete distribution to a mixed integer programming problem. Autom. Remote Control **74**(6), 951–967 (2013). https://doi.org/10.1134/S0005117913060064
18. Ivanov, S.V., Zhenevskaya, I.D.: Estimation of the necessary sample size for approximation of stochastic optimization problems with probabilistic criteria. In: Khachay, M., Kochetov, Y., Pardalos, P. (eds.) MOTOR 2019. LNCS, vol. 11548, pp. 552–564. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22629-9_39
19. Ivanov, S.V.: A bilevel stochastic programming problem with random parameters in the follower's objective function. J. Appl. Ind. Math. **12**(4), 658–667 (2018). https://doi.org/10.1134/S1990478918040063