The Maximum Entropy Method for Bilevel Stochastic Programming

ZHOU Wanna1

(College of Engineering and Technology, Xi’an Fanyi University, Xi’an 710105, China)
zhangna@xupt.edu.cn

Abstract: Bi-level stochastic programming is a process of taking the optimal value of the lower level as the feedback to the upper level, and the constraint that the lower level has one single optimal solution can be relaxed in calculation. Using the maximum entropy method, the present study proposed an approximate calculation method for stochastic programming. The proposed method constructs a maximum entropy function for lower-level programming to approximately express the lower-level optimal value function, converts the bi-level stochastic programming problem into a single-level stochastic programming problem, and hence obtains a calculation method for an approximation optimal solution ($\varepsilon$-optimal solution) of the bi-level stochastic programming problem.

1 Introduction

The bi-level programming is a systematic problem with a hierarchical structure. It includes upper-level problems and lower-level problems. The upper-level and lower-level problems have their own objective functions and constraints. The bi-level programming is mainly problems of derived from actual resources allocation and price control, which was widely used in transportation, management, agriculture, power utilization, optimal design, and many other issues, and attracted more and more researchers' attention. When random variables are introduced into mathematical programming, random programming is formed, when random variables infiltrate into the bi-level programming, the corresponding bi-level stochastic programming problem is obtained. The literature [1-2] discusses the approximate calculation method of single-level stochastic programming. The literature [3] discussed the bi-level optimization problem with constraints in lower level, and proposed an approximate calculation method. Literature [4-6] discussed the stability of the approximation solution set of bi-level stochastic programming. Literature [7] discussed Lipschitz continuity of the objective function of single-level stochastic programming. Literature [8] discusses an approximate calculation method of convex programming. Inspired by the literature [3,8], this paper studied a kind of approximate calculation methods of bilevel stochastic programming.

This article studies the following bi-level stochastic programming

$$\min_{x \in X} \int_{R^P} T(x, v(x), u) \mu(du)$$

$$\left\{ \begin{array}{l}
  v(x) = \min_{y \in Y} \int_{R^P} f(x, y, u) \mu(du) \\
  s.t. g_i(x, y, u) \leq 0, i = 1, 2, \ldots, m
\end{array} \right. \quad (1)$$

In the equation, $\mu = p \circ \xi^{-1}$, $\xi$ is a continuous random variable defined on $\Omega, F, P$, $\Omega$ is a
bounded closed set, and the probability density of $\xi$ is $g(t)$. $X$ and $Y$ are the bounded closed convex sets, $T, f, g_i (i = 1, 2, \cdots, m)$ are continuous differentiable functions, which are bounded on $X \times Y \times R$. In this paper, the maximum entropy function is used to approximately replace the optimal value function of the lower layer of bilevel stochastic programming, so as to obtain the approximate optimal solution ($\varepsilon - \text{optimal solution}$) of the bi-level stochastic programming.

2. Solving the lower-level stochastic programming problem

For lower-level programming

$$v(x) = \min_{y \in Y} \int_{R^P} f(x, y, u) \mu(du)$$

s.t. $g_i(x, y, u) \leq 0, i = 1, 2, \cdots, m,$

(2)

this article assumed that $g_i (i = 1, 2, \cdots, m)$ is a differentiable convex function about $(x, y)$, and for $u$, it is a bounded and measurable continuous function.

In this paper, we divided the $\Omega$ using the division method of the set in [2], and divided $\Omega$ into $N_k$ sub-blocks $A_j (j = 1, 2, \cdots, N_k)$, which satisfies:

1. $P(A_j) \geq 0$, (where $P(A_j) = \int_{A_j} f(x, y, u) \mu(du)$);

2. $A_j \cap A_j = \emptyset, \forall j \neq j, \bigcup_{j=1}^{N_k} A_j = \Omega$

3. $P(\partial A_j) = 0, (j = 1, 2, \cdots, N_k), diam A_j = \sup_{t, r \in A_j} |t - r| \to 0, (k \to \infty)$;

4. $u_j \in A_j, (j = 1, 2, \cdots, N_k), \max \{\mu_j/P(A_j)\} \to 1 (k \to \infty)$,

And for any continuous differentiable function $h \in C(\Omega)$, we can define the convergence quadrature process [2]:

$$\sum_{j=1}^{N_k} h(u_j) \mu_j \to \int_{\Omega} h(u) \mu(du), (k \to \infty)$$

Under the division of $\Omega$, the problem (2) is transformed into the following nonlinear programming

$$\begin{align*}
v(x) &= \min_{y \in Y} \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j, \\
\text{s.t.} \ g_i(x, y, u_j) &\leq 0, i = 1, 2, \cdots, m, j = 1, 2, \cdots, N_k
\end{align*}$$

(3)

Set the feasible set be $M_k = \{y \in Y | g_i(x, y, u_j) \leq 0, i = 1, 2, \cdots, m, j = 1, 2, \cdots, N_k\}$. Suppose Slater condition of the problem (2) is satisfied, that is, there are $x \in X, y \in Y$, which can make $\forall u \in \Omega, g_i(x, y, u) < 0$.

Set

$$F_0 = \int_{R^P} f(x, y, u) \mu(du),$$

$$F_k = F_k (y) = \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j$$

Definition 1 [4] If there is $x_n \to x_0$ in $R^n$ satisfies that $f_n(x_n, y)$ converges to $f(x_0, y)$, it is denoted as $f_n \overset{epi}{\to} f$, which means

1. For any $y_n \to y_0$, there is $\lim \inf_{n \to \infty} f_n (x_n, y_n) \geq f(x_0, y_0)$,

2. There is a certain $y_n \to y_0$ satisfies $\lim \sup_{n \to \infty} f_n (x_n, y_n) \leq f(x_0, y_0)$.

Lemma 1 If the convergence condition $\sum_{j=1}^{N_k} f(x, y, u_j) \mu_j \to \int_{\Omega} f(x, y, u) \mu(du), (k \to \infty)$ is satisfied, for a fixed $x$, we can know that

$$F_k \overset{epi}{\to} F_0.$$
**Proof** For a fixed $x$, let $y_k \to y$, then
\[
\sum_{j=1}^{N_k} f(x, y_k, u_j) \mu_j - \int_{R^p} f(x, y, u) \mu(du)
\]
\[
= \sum_{j=1}^{N_k} f(x, y_k, u_j) \mu_j - \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j + \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j - \int_{R^p} f(x, y, u) \mu(du)
\]
\[
\leq \left| \sum_{j=1}^{N_k} f(x, y_k, u_j) \mu_j - \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j \right| + \left| \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j - \int_{R^p} f(x, y, u) \mu(du) \right|
\]
\[
= \sum_{j=1}^{N_k} \left| f(x, y_k, u_j) - f(x, y, u_j) \right| \mu_j + \sum_{j=1}^{N_k} \left| f(x, y, u_j) \right| \mu_j - \int_{R^p} f(x, y, u) \mu(du) \to 0 (k \to \infty),
\]
That is
\[
\limsup \sum_{j=1}^{N_k} f(x, y_k, u_j) \mu_j \leq \int_{R^p} f(x, y, u) \mu(du),
\]
And
\[
\limsup F_k (y_k) \leq F_0,
\]
In the same way
\[
\int_{R^p} f(x, y, u) \mu(du) - \sum_{j=1}^{N_k} f(x, y_k, u_j) \mu_j \to 0 (k \to \infty),
\]
That is
\[
\liminf F_k (y_k) \geq F_0,
\]
Therefore $F_k \overset{\text{epi}}{\to} F_0$.

Since the number of constraints of problem (3) increases rapidly with the increase of $k$, the problem (3) can be transformed into a smooth programming problem with only one constraint by introducing a maximum entropy function, and when $k$ is large enough, the optimal solution of problem (3) is the approximate optimal solution of problem (2). For a fixed $x$, let
\[
G_k (y) = \max_{1 \leq j \leq N_k} g_1(x, y, u_j),
\]
Then problem (3) can be equivalent to
\[
v(x) = \min_{y \in Y} F_k (y) = \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j
\]
\[s.t \quad G_k (y) \leq 0, \quad (4)
\]
The $\varepsilon$-feasible solution of problem (4) is written as $M_{k, \varepsilon}$, $M_{k, \varepsilon} = \{y \in Y \mid G_k (y) \leq \varepsilon\}$.

The maximum entropy function for $G_k (y)$ is as follows
\[
G_k^{(k)} (y) = \frac{1}{m \cdot N_k} \ln \left[ \sum_{i=1}^{m} \sum_{j=1}^{N_k} \exp (m \cdot N_k g_1(x, y, u_j)) \right],
\]
$G_k^{(k)} (y)$ is a continuous differentiable function on $Y$, as the programming
\[
\begin{align*}
\left\{ v_k^{(k)} (x) &= \min_{y \in Y} F_k (y) = \sum_{j=1}^{N_k} f(x, y, u_j) \mu_j, \\
&\text{s.t. } G_k^{(k)} (y) \leq 0
\end{align*}
\]
In the feasible set of question (5), the $\varepsilon$-feasible solutions are denoted as $M_{k}^{\prime}$, $M_{k, \varepsilon}^{\prime}$, $M_{k}^{\prime} = \{y \in Y \mid G_k^{(k)} (y) \leq 0\}$, $M_{k, \varepsilon}^{\prime} = \{y \in Y \mid G_k^{(k)} (y) \leq \varepsilon\}$. 

3
Lemma 2 \( \lim_{k \to \infty} \left[ G_k^{(k)}(y) - G_k(y) \right] = 0. \)

Prove \( G_k^{(k)}(y) = \frac{1}{m \cdot N_k} \ln \left[ \sum_{i=1}^{m} \sum_{j=1}^{N_k} \exp \left( m \cdot N_k g_i(x, y, u_j) \right) \right], \)
\( G_k(y) = \frac{1}{m \cdot N_k} \ln \left[ \sum_{i=1}^{m} \sum_{j=1}^{N_k} \exp \left( m \cdot N_k g_i(y) \right) \right], \)
\( G_k^{(k)}(y) - G_k(y) = \frac{1}{m \cdot N_k} \ln \left[ \sum_{i=1}^{m} \sum_{j=1}^{N_k} \exp \left[ m \cdot N_k \left( g_i(x, y, u_j) - g_i(y) \right) \right] \right]. \)

As \( g_i(x, y, u_j) - G_k(y) \leq 0, \)
\( G_k^{(k)}(y) - G_k(y) \leq \frac{1}{m \cdot N_k} \ln m \cdot N_k, \)

Therefore \( 0 \leq G_k^{(k)}(y) - G_k(y) \leq \frac{1}{m \cdot N_k} \ln m \cdot N_k. \)

According to the division of \( \Omega, \) when \( k \to \infty, \)
\( \lim_{k \to \infty} \frac{1}{m \cdot N_k} \ln m \cdot N_k = 0, \)

According to the clamping criterion, we can know that
\( \lim_{k \to \infty} \left[ G_k^{(k)}(y) - G_k(y) \right] = 0. \)

Based on \( M_{k, \varepsilon}, M_{k}', M_{k, \varepsilon}' \) and Lemma 2, we can obtain that
(a) The feasible solution of question (5) must be the feasible solution of question (3);
(b) When \( k \) is large enough, the feasible solution of problem (3) must be the \( \varepsilon \)-feasible solution of problem (5).

As the above conditions of \( \forall \varepsilon > 0, \exists k > 0 \) satisfies \( M_k' \subset M_k \subset M_{k, \varepsilon}', \) that is, \( M_k' \) can be used to replace \( M_k, \) which means the feasible set of problem (5) can be approximately replaced by the feasible set of problem (3) and problem (3) and problem (5) have the same objective function. Therefore, the approximate optimal solution of problem (5) can be used to replace the approximate optimal solution of problem (3). Under the division of \( \Omega, \) the problem (2) is equivalent to the problem (3), so the approximate optimal solution of the problem (5) can be replaced by the approximate optimal solution of problem (2), and by the lemma, when \( k \) is large enough, the optimal solution of problem (3) can be used as the approximate optimal solution of problem (2), and the optimal solution of problem (5) can be used as the approximate optimal solution of problem (2), so the optimal value of problem (5) can be taken as the approximate optimal value of question (2).

3. Solution of upper level programming
For the upper-level stochastic programming, there is the following programming
\[ \min_{x \in X} \int_{R^n} T \left( x, v_k^{(k)}(u) \right) \mu(du), \]  \hspace{1cm} (6)

Definition 2 Set \( x_k \in X \) to be the \( \varepsilon \)-optimal solution of the problem (6), if for all \( x \in X, \) we can know that
\[ \int_{R^n} T \left( x_k, v_k^{(k)}(x_k) \right) \mu(du) < \int_{R^n} T \left( x, v_k^{(k)}(x_k) \right) \mu(du) + \varepsilon. \]

Theorem 1 (I) When problem (1) has an optimal solution \( x^* \), and when \( k \) is large enough, \( x^* \) must be the \( \varepsilon \)-optimal solution of problem (6).

(II) Problem (6) must have an optimal solution \( x_k^* \), and when \( k \) is large enough, the optimal solution \( x_k^* \) of problem (6) must be the \( \varepsilon \)-optimal solution of problem (1).

Proof
(1) Bounded by \( g_i(x, y, u) \), for \( \exists M > 0, \) there is \( |g_i(x, y, u)| \leq M, i = 1, 2, \cdots m, \) as \( T(x, y, u) \) are consistent and continuous, \( \forall \varepsilon > 0, \exists \delta > 0, \) for \( \forall x \in X, y_i \in Y, \) \( |y_1 - y_2| < \delta, \) we can know that
\[ |T(x, y_1, u) - T(x, y_2, u)| < \frac{\varepsilon}{2}, \]  \hspace{1cm} (1)

When \( k \) is large enough, the optimal value of problem (5) can be used as the approximate optimal
value of problem (2), then \(\lim_{k \to \infty} v^{(k)}_k(x) = v(x)\), that is, when \(k\) is large enough, for \(\forall x\), we can know that

\[
|v^{(k)}_k(x) - v(x)| < \delta, \quad \textcircled{2}
\]

It can be seen from \(\textcircled{1}\), \(\textcircled{2}\) that when \(k\) is large enough, for \(\forall x \in X\), we can know that

\[
\left| T(x, v^{(k)}_k(x), u) - T(x, v(x), u) \right| < \frac{\varepsilon}{2}, \quad \textcircled{3}
\]

Obtained from \(\textcircled{3}\), when there is an optimal solution \(x^*\) in problem (1), and \(k\) is large enough, for \(\forall x \in X\), we can know that

\[
\left| T(x^*, v^{(k)}_k(x^*), u) - T(x^*, v(x^*), u) \right| < \frac{\varepsilon}{2}, \quad \textcircled{4}
\]

Since \(T(x, y, u)\) is consistent and continuous, we can know from the proposition 2.3 in [7] that

\[
ET(x^*, v^{(k)}_k(x^*), u) < ET(x, v^{(k)}_k(x), u) + \varepsilon,
\]

So

\[
\int_{\mathcal{R}^p} T(x^*, v^{(k)}_k(x^*), u) \mu(du) < \int_{\mathcal{R}^p} T(x, v^{(k)}_k(x), u) \mu(du) + \varepsilon.
\]

Therefore, \(x^*\) must be the \(\varepsilon\) –optimal solution of problem (6).

(2) Similarly, when \(k\) is large enough, the optimal solution \(x^{(k)}_k\) of problem (6) must be the \(\varepsilon\) –optimal solution of problem (1).

Acknowledgment

Foundation item: Foundation project of Shaanxi Provincial Department of Education in 2020 (20JK0641);

References:
[1] Wan Zhongping. An approximate exact penalty function method for single-stage stochastic programming[J]. Journal of Engineering Mathematics, 1996, 13(4): 37-42.
[2] LEPP R. Approximations to Stochastic Programs with Complete Recourse[J]. SIAM J control and optimization, 1990,28(2):382-394.
[3] Wu Zhiyou, Peng Jianwen, Yu Hui. Maximum entropy method for a class of bi-level optimization problems with linear constraints[J]. Journal of Operations Research, 2000, 25(2): 134-137.
[4] Zhou Wanna, Huo Yongliang. The upper half convergence of the optimal solution set of bi-level stochastic programming[J]. Pure and Applied Mathematics, 2014, 30(2): 207-215.
[5] Zhou Wanna, Huo Yongliang. A sufficient condition for approximating the upper semiconvergence of the solution set of bi-level stochastic programming[J]. Journal of Southwest Normal University, 2015,40(9):17-22.
[6] Zhou Wanna, Huo Yongliang. Hausdorff convergence of approximating \(\varepsilon\)-optimal solution set for bi-level stochastic programming[J]. Journal of Hunan Normal University, 2016,39(5):80-83.
[7] J-B Wets. Stochastic Programming, in Handbook for Operation Research and Management Sciences, Vol. 1(des)G. Nemhauser and A. Rinnooy Kan, Elsevier Science Publishers, 1989, 573-629.
[8] Li Xingsi. The Maximum Entropy Method of Convex Programming[J]. Science Bulletin, 1994,39(8):682-684.
[9] Huo Yongliang. Stability theory of stochastic programming[M]. Chengdu: Southwest Jiaotong University Press, 2010.