参赛队员姓名：刘云起

中学：南京外国语学校

省份：江苏省

国家/地区：中国

指导教师姓名：孙海琳、唐智逸

指导教师单位：南京师范大学、南京外国语学校

论文题目：多阶段随机二次规划的样本均值逼近方法和分裂算法

多阶段随机二次规划的样本均值逼近方法和分裂算法

刘云起

摘要： 在本文中，我们研究了多阶段随机二次规划问题的样本均值逼近方法和分裂算法. 我们首先研究了三阶段随机二次规划问题的性质，并进一步在独立同分布 (i.i.d.) 性质和阶段独立性 (stage-wise independment) 等条件不成立的情况下研究了其样本均值逼近方法的指数收敛速率, 并得到了更紧的样本量的估计. 这些结果可以容易地拓展到多阶段随机二次规划问题和一般的多阶段随机规划问题. 进一步，我们基于一类带有特殊临近项的增广拉格朗日乘子法，给出了一个非精确的对称高斯-塞德尔分裂算法求解多阶段随机二次规划问题. 数值试验展示了收敛性理论和该算法的优势.

关键词： 多阶段随机二次规划；样本均值逼近方法；收敛速率；分裂算法

Abstract: In this paper, we investigate the sample average approximation (SAA) method and decomposition method for solving multistage stochastic quadratic programming problems. Firstly, we consider the properties for three-stage stochastic quadratic programming problems, and then give the convergence rate of SAA method of three-stage stochastic quadratic programming problems. Those results can easily extend to multistage cases. Moreover, based on a class of augmented Lagrange multiplier methods with special proximal term, we give a symmetric Gaussian-Seidel splitting algorithm for solving multistage stochastic quadratic programming problems. Several numerical experiments show the correctness of the convergence analysis and the advantage of the proposed decomposition method.

Keywords: multistage stochastic quadratic programming problems; sample average approximation method; convergence rate; decomposition method

目录

1 引言

如何处理随机环境下的决策问题日益成为现代优化理论和实际应用中的核心问题之一, 而随机规划正是处理这类问题的核心方法之一 [1, 2]. 根据决策过程的不同, 随机规划可以分为单阶段随机规划, 两阶段随机规划和多阶段随机规划. 其中, 单阶段随机规划和两阶段随机规划利用随机变量来刻画随机环境, 处理只需要进行一次和两次决策的最优决策问题. 而多阶段随机规划则利用离散时间的随机过程刻画环境的随机性, 利用多阶段的未来信息, 处理需要进行多次决策的最优决策问题.

多阶段随机规划在 1955 年由线性规划之父 Dantzig 提出 [3], 自此，便一直是随机规划领域的核心问题之一, 参见专著 [2, 4]. 但是由于多阶段随机规划复杂的问题结构和规模, 在其分析和求解上一直存在着难度. 尤其当随机过程 $\{\xi_2, \cdots, \xi_T\}$ 服从连续分布时, 不仅期望值需要进行积分, 增加了计算难度, 而且也导致了一个无穷维优化问题. 将连续随机变量的多阶段随机规划近似为有限维优化问题的方法主要有两种. 一类是基于决策规则的方法 [5, 6], 这类方法用一类函数拟合最优解函数 (譬如线性决策规则用线性函数近似最优解函数), 从而将无穷维问题近似为有限维问题求解. 另一类常用的方法是基于样本的方法, 如样本均值逼近方法, 情景树方法等 [7, 8]. 样本均值逼近方法是近似连续分布的有效方法, 不仅可以用样本均值去近似期望值, 更可以将无穷维的多阶段随机规划问题降维为有限维问题, 是近似多阶段随机规划问题的有效方法 [7, 9, 10, 11, 12].

当考虑利用样本均值逼近问题求解多阶段随机规划时, 我们主要面临着两个问题:

1. 随着样本量的增加, 样本均值逼近问题的最优解是否会收敛到原问题的最优解, 收敛速率又如何? 针对该问题, 目前已经有部分文献对此进行分析: 在文献 [7, 9] 中, Shapiro 分析了当采用条件抽样时, 样本均值逼近问题的解的渐进一致收敛性以及复杂性分析, 后者可以帮助我们去刻画, 当我们需要达到给定精度时, 需要样本量的上界. 在文献 [10] 中, Reaiche 进一步针对一类多阶段随机规划问题, 给出了达到样本复杂性的下界 (即达到给定精度，需要样本量的下界). 在文献 [12] 中, Jiang 和 Li 研究了当随机变量服从重尾分布时的样本复杂度. 然而, 目前的分析主要还是在各阶段随机变量的阶段独立性和 i.i.d. 等假设成立的前提下进行的. 一个值得研究的方向是, 是否有方法进一步放宽这些假设呢?

2. 样本均值逼近问题本身也是一个服从离散分布的多阶段随机规划问题, 如何有效求解? 目前求解多阶段随机规划的方法相对还比较少, 其中基于非预测性 (Nonanticipativity) 的形式以及其相应的逐步对冲 (Progressive Hedging, PHA) 方法 [13] 和基于动态规划方程 (dynamic programming equation) 的形式以及相应的随机对偶动态规划 (Stochastic dual dynamic programming, SDDP) 方法 [14, 15] 是两种主要的形式和算法. 这两种算法都是分裂算法, 前者通过引入变量, 构造可分结构, 达到分裂并行求解的目的; 后者则是一种近似的割平面算法, 通过一个阶段一个阶段的构造近似割平面, 将问题转化为一个个可以并行计算的线性规划子问题, 从而进行分裂并行求解. 然而, SDDP 方法依赖于随机变量阶段独立性的假设; 而逐步对冲方法为了构造可分结构, 引入了大量变量, 在一定程度上增加了问题规模. 是否可以设计算法, 使得既不引进变量, 又不依赖随机变量的阶段独立性呢?

针对一类特殊的多阶段随机二次规划问题, 本文将利用问题的结构, 进一步回答以上两个问题. 本文的主要贡献如下:

1. 当各阶段随机变量不满足阶段独立性和 i.i.d. 时, 我们研究了多阶段随机二次规划优化问题解的存在性和有界性, 并进一步分析了其每个阶段最优值函数的 Lipschitz 连续性, 从而在更弱的条件下给出了更紧的样本均值逼近方法的样本复杂度分析. 该样本复杂性分析可以直接推广到一般的多阶段随机规划问题上;

2. 最近, 在随机变量服从离散分布时, Arpón 等人 [16] 利用多块的交替方向乘子法 (ADMM) 求解了两阶段随机规划. Li [17] 等人和 Chen 等人 [18] 利用带特殊临近项的增广 Lagrange 方法去考虑凸二次半定规划和凸的复合优化问题. 这启发了我们利用多阶段随机二次规划天然的部分可分结构, 给出一个非精确的对称高斯-塞德尔分裂算法来求解该问题. 该方法等价于一个带有特殊临近项的增广 Larange 乘子法. 特别的, 该方法并不依赖阶段独立性的假设，也不需要进一步引入松弛变量构造可分结构，仅依赖问题本身现有结构就可以构造有效的分裂并行算法. 但是由于不依赖阶段独立性的假设, 问题规模依然会随着阶段数的增加, 呈指数增长, 所以和 PHA 方法一样, 该方法也只能处理较少阶段的情况 ($T < 5$).

本文的结构如下: 在第二节中, 我们将以三阶段随机二次规划为重点, 考虑其解的有界性, 各阶段最优值函数的 Lipschitz 连续性, 并在较弱的条件下给出更紧的样本均值逼近方法的样本复杂度分析; 在第三节中, 我们将给出针对多阶段随机二次规划问题的一类基于对称高斯-塞德尔方法的分裂算法. 在第四节中, 数值试验将展示样本均值逼近方法的收敛性, 并将第三节提出的基于对称高斯-塞德尔方法的分裂算法与 PHA 方法比较, 以验证其有效性.

在本文中, 我们使用以下符号: 记 $B_\delta(x)$ 为点 x 半径为 δ 的领域; 对一个对称矩阵 A, 记 $\lambda_{min}(A)$ 为其最小特征值, $\lambda_{max}(A)$ 为其最大特征值; 记 $\mathcal{B}$ 为欧氏空间中的单位球, $\mathbb{S}^n$ 为 n 维对称阵组成的空间. 对于两个对称方阵 $A, B \in \mathbb{S}^n$, 记 $A \succeq B$ 为 $A-B$ 为半正定矩阵. 在文中, 我们根据上下文, 用 ξ_t 表示第 t 阶段的随机变量, 或者其样本 (实现).

2 多阶段随机二次规划的样本均值逼近方法

在本节中, 我们研究多阶段随机二次规划的样本均值逼近方法. 为了刻画一个 T 阶段的随机二次规划模型, 我们先给出如下的离散随机过程. 令 $\{\xi_1, \xi_2, \cdots, \xi_T\}$ 是一个离散时间的随机过程, 其中 ξ_1 是确定性的, $\xi_t : \Omega_t \to \Xi_t \subset \mathbb{R}^{s_t}$ 是第 t 时刻的定义在概率空间 $(\Omega, \mathcal{F}, P)$ 上的随机向量, $t = 2, \cdots, T$. 我们记 $\xi_{[t]} = (\xi_1, \cdots, \xi_t)$. 在这样的随机环境下, 我们考虑如下的多阶段随机规划问题:

$$\min_{x_1 \in X_1} c_1(x_1) + \mathbb{E}\left[\min_{x_2(\xi_2) \in X_2(x_1,\xi_2)} c_2(x_2(\xi_2), \xi_2) + \cdots + \mathbb{E}\left[\min_{x_T(\xi_{[T]}) \in X_T(x_{T-1}(\xi_{[T-1]}),\xi_T)} c_T(x_T(\xi_{[T]}), \xi_T)\right] \cdots\right] \tag{2.1}$$

其中 $x_1 \in \mathbb{R}^{n_1}$ 是第一阶段决策变量, $x_t(\xi_{[t]})$ 是第 t 阶段决策变量, 意味着在第 t 阶段，对于从初始阶段到第 t 阶段的每一种可能的实现, 人们都有相对应的决策 $x_t(\xi_{[t]})$, 我们将 $\{x_1, x_2(\xi_{[2]}), \cdots, x_T(\xi_{[T]})\}$ 称为策略;

$$\begin{aligned}
X_1 &:= \left\{x_1 \in \mathbb{R}^{n_1} : A_1^E x_1 = b_1^E, A_1^I x_1 \geq b_1^I\right\}, \\
X_t(x_{t-1}, \xi_t) &:= \{x_t \in \mathbb{R}^{n_t} : B_t^E(\xi_t)x_{t-1} + A_t^E(\xi_t)x_t = b_t^E(\xi_t), B_t^I(\xi_t)x_{t-1} + A_t^I(\xi_t)x_t \geq b_t^I(\xi_t)\},\ t = 2, \cdots, T, \\
c_1(x_1) &:= x_1^\top H_1 x_1 + h_1^\top x_1 + q_1, \\
c_t(x_t, \xi_t) &:= x_t^\top H_t(\xi_t)x_t + h_t(\xi_t)^\top x_t + q_t(\xi_t),\ t = 2, \cdots, T.
\end{aligned}$$

这里 $A_1 \in \mathbb{R}^{s_1 \times n_1}$, $b_1 \in \mathbb{R}^{s_1}$, $H_1 \in \mathbb{S}^{n_1}$, $h_1 \in \mathbb{R}^{n_1}$, $q_1 \in \mathbb{R}$, $A_t : \Xi_t \to \mathbb{R}^{s_t \times n_t}$, $B_t : \Xi_t \to \mathbb{R}^{s_t \times n_t}$, $b_t : \Xi_t \to \mathbb{R}^{s_t}$, $H_t : \Xi_t \to \in \mathbb{S}^{n_t}$, $q_t : \Xi_t \in \mathbb{R}$, $t = 2, \cdots, T$.

由于一般多阶段的情况可以很容易的由三阶段的情况推广得到, 而一般多阶段的情况只会增加符号的复杂程度, 我们将以三阶段随机二次规划为重点展开研究.

2.1 三阶段随机二次规划的性质

考虑如下的三阶段随机二次规划问题:

$$\min_{x_1 \in X_1} c_1(x_1) + \mathbb{E}\left[\min_{x_2 \in X_2(x_1,\xi_2)} c_2(x_2, \xi_2) + \mathbb{E}\left[\min_{x_3 \in X_3(x_2,\xi_3)} c_3(x_3, \xi_3)\right]\right]. \tag{2.2}$$

我们可以将问题 (2.2) 等价地写成一个如下动态规划的形式:

$$\min_{x_1 \in X_1} c(x_1) := c_1(x_1) + \mathcal{Q}_2(x_1), \tag{2.3}$$

其中 $\mathcal{Q}_2(x_1) := \mathbb{E}\left[Q_2(x_1, \xi_2)\right]$

$$Q_2(x_1, \xi_2) := \min_{x_2 \in X_2(x_1,\xi_2)} c_2(x_2, \xi_2) + \mathcal{Q}_3(x_2, \xi_2), \tag{2.4}$$

$\mathcal{Q}_3(x_2, \xi_2) := \mathbb{E}\left[Q_3(x_2, \xi_3)|\xi_2\right]$,

$$Q_3(x_2, \xi_3) := \min_{x_3 \in X_3(x_2,\xi_3)} c_3(x_3, \xi_3). \tag{2.5}$$

我们引入如下记号: $\mathcal{X}_1$ 是所有使得 $X_2(x_1,\xi_2)$ 非空的 x_1 的集合, $\mathcal{X}_2$ 是所有使得 $X_3(x_2,\xi_3)$ 非空的 x_2 的集合; 并做如下假设:

假设 2.1 *问题 (2.2) 满足相对完备补偿条件. 即对所有 $x_1\in X_1$ 和几乎所有的所有的 $\xi_2\in\Xi_2$, $X_2(x_1,\xi_2)$ 非空, 即 $X_1\subset\mathcal{X}_1$; 并且对几乎所有的 $\xi_2\in\Xi_2$ 和 $\xi_3\in\Xi_3$ 以及 $x_2\in X_2(x_1,\xi_2)$, $X_3(x_2,\xi_3)$ 非空, 即 $X_2(x_1,\xi_2)\subset\mathcal{X}_2$ 对几乎所有的 $\xi_2\in\Xi_2$ 成立.*

假设 2.2 *矩阵 H_1 是正定阵, $\Xi_2\subset\mathbb{R}^{s_2}$ 和 $\Xi_3\subset\mathbb{R}^{s_3}$ 是紧集; 对几乎所有的 $\xi_2\in\Xi_2$ 和 $\xi_3\in\Xi_3$, $H_2(\xi_2)$ 和 $H_3(\xi_3)$ 也是正定的, 并且存在 $k_1>0$ 使得 $\lambda_{\min}(H_2(\xi_2))\geq k_1$, $\lambda_{\min}(H_3(\xi_3))\geq k_1$; 矩阵值或向量值随机参变量 $H_2(\xi_2),h_2(\xi_2),A_2(\xi_2),B_2(\xi_2),b_2(\xi_2),H_3(\xi_3),h_3(\xi_3),A_3(\xi_3),B_3(\xi_3),b_3(\xi_3)$ 关于 ξ_2 或 ξ_3 连续.*

由矩阵值和向量值随机变量的连续性和 Ξ_2,Ξ_3 的紧性, 我们可以得到这些随机变量的有界性. 即存在 $k_2>0$, 使得对任意的 $\xi_2\in\Xi_2$ 和 $\xi_3\in\Xi_3$ 有

$$\max\{\|H_2(\xi_2)\|,\|h_2(\xi_2)\|,\|A_2(\xi_2)\|,\|B_2(\xi_2)\|,\|b_2(\xi_2)\|\}\leq k_2$$

以及

$$\max\{\|H_3(\xi_3)\|,\|h_3(\xi_3)\|,\|A_3(\xi_3)\|,\|B_3(\xi_3)\|,\|b_3(\xi_3)\|\}\leq k_2.$$

注 2.1 *支持集是紧集这一条件主要是用于刻画问题 (2.2) 及之后样本均值逼近问题的解的有界性. 如果可行解集本身就是有界集合, 那么这一条件也可以直接去掉, 并不影响后面的结果成立.*

此外, 虽然对于很多随机变量, 支撑集 $\Xi_2\subset\mathbb{R}^{s_2}$ 和 $\Xi_3\subset\mathbb{R}^{s_3}$ 是紧集这一条件是无法满足的, 但是我们可以考虑利用随机变量的紧 (tightness) 性质 (参见 [19]), 即随机变量在一个足够大的紧集外面的测度足够小, 从而近似地省略紧集之外测度的影响. 这将作为我们今后的工作进一步加以考虑.

假设 2.3 *对任意 $(x_2,\xi_3)\in\mathcal{X}_2\times\Xi_3$, 如下条件之一成立:*

1. *$A_3^E(\xi_3)$ 是列满秩;*
2. *$A_3^E(\xi_3)$ 是行满秩, 并且对每一个使得 1) $A^I(\xi_3)x_3\geq b_3^I(\xi_3)-B_3^I(\xi_3)x_2,(A_3^I(\xi_3)x_3)_l=(b_3^I(\xi_3)-B_3^I(\xi_3)x_2)_l, A^E(\xi_3)x_3=b_3^E(\xi_3)-B_3^E(\xi_3)x_2$ 或者 2) $A^I(\xi_3)x_3\geq 0,(A_3^I(\xi_3)x_3)_l=0,A^E(\xi_3)x_3=0,x_3\neq 0$ 是可解的非空指标集 $l\subseteq\{1,\cdots,s_3^I\}$, 我们有或者 1) $(A_3^I(\xi_3)x_3)_l>0,A^E(\xi_3)x_3=0$ 是可解的, 或者 $((A_3^I(\xi_3))_l^\top,(A_3^E(\xi_3))^\top)^\top$ 是列满秩的.*

假设2.3来自于 [20, 定理 5.4]. 在该假设下, 第三阶段的可行解集 $X_3(x_2,\xi_3)$ 的条件数 (Hoffman 引理中线性系统误差界的系数) 局部有界.

下面我们考虑第三阶段问题

$$\min_{x_3\in X_3(x_2,\xi_3)} c_3(x_3,\xi_3). \tag{2.6}$$

令

$$\text{lev } c_3^\alpha(x_2,\xi_3):=\{x_3:x_3\in X_3(x_2,\xi_3),\ c_3(x_3,\xi_3)\leq\alpha,\}$$

为第三阶段问题在点 (x_2,ξ_3) 的水平集.

命题 2.1 *若假设2.2-2.3 成立, 则存在 α_1, 紧集 C_1 和足够小的领域半径 δ, 使得对所有的 $(x_2',\xi_3')\in B_\delta(x_2,\xi_3)\cap\mathcal{X}_2\times\Xi_3$, $\text{lev } c_3^\alpha(x_2',\xi_3')\subset C_1$ 并且非空.*

证明. 由于 c_3 是正定二次函数, 并且假设2.2 成立, 容易知道对任意的 α, $\text{lev } \bar{c}_3^\alpha(\xi_3)$ 和集合 $\text{lev } \underline{c}_3^\alpha(\xi_3)$ 是紧集, 并且有

$$\text{lev } \bar{c}_3^\alpha(\xi_3)\cap X_3(x_2',\xi_3')\subset\text{lev } c_3^\alpha(x_2',\xi_3')\subset\text{lev } \underline{c}_3^\alpha(\xi_3)\cap X_3(x_2',\xi_3')$$

对任意 $(x_2', \xi_3') \in B_\delta(x_2, \xi_3) \subset \mathcal{X}_2 \times \Xi_3$ 成立, 其中

$$\text{lev}\ \bar{c}_3^\alpha(\xi_3) := \{x_3 : \bar{c}_3(x_3, \xi_3) \le \alpha\},\ \text{lev}\ \underline{c}_3^\alpha(\xi_3) := \{x_3 : \underline{c}_3(x_3, \xi_3) \le \alpha\},$$

$$\bar{c}_3(x_3, \xi_3) := x_3^\top \bar{H}_3(\xi_3) x_3 + \bar{h}_3(\xi_3)\|x_3\| + \bar{q}_3(\xi_3),\ \ \underline{c}_3(x_3, \xi_3) := x_3^\top \underline{H}_3(\xi_3) x_3 - \bar{h}_3(\xi_3)\|x_3\| - \bar{q}_3(\xi_3)$$

并且

$$\bar{H}_3(\xi_3) \succeq H_3(\xi_3'),\ H_3(\xi_3') \succeq \underline{H}_3(\xi_3) \succ 0,\ \bar{h}_3(\xi_3) \ge \|h_3(\xi_3')\|_1,\ \bar{q}_3(\xi_3) \ge \|q_3(\xi_3')\|_1, \forall \xi_3' \in B_\delta(\xi_3).$$

由假设2.2, 这样的 $\bar{H}_3(\xi_3), \underline{H}_3(\xi_3), \bar{h}_3(\xi_3), \bar{q}_3(\xi_3)$ 存在. 这证明了水平集的一致紧性.

易见对任意的 $x \in \mathbb{R}^{n_3}$, 总存在足够大的 α 使得 $x \in \text{lev}\ \bar{c}_3^\alpha(\xi_3) \subset \text{lev}\ \underline{c}_3^\alpha(\xi_3)$. 由于 $X_3(x_2, \xi_3)$ 非空, 必然存在足够大的 α 使得 $\text{lev}\ \bar{c}_3^\alpha(\xi_3) \cap X_3(x_2, \xi_3) \ne \emptyset$. 进一步, 由 Hoffman 引理, 对任意 $y \in \text{lev}\ \bar{c}_3^\alpha(\xi_3) \cap X_3(x_2, \xi_3)$, 有

$$d(y, X_3(x_2', \xi_3')) \le \mathcal{K}(x_2', \xi_3')(\|B_3^E(\xi_3')x_2' + A_3^E(\xi_3')y - b_3^E(\xi_3')\| + \|(B_3^I(\xi_3')x_2' + A_3^I(\xi_3')y - b_3^I(\xi_3'))_-\|).$$

进一步, 由于 $y \in \text{lev}\ \bar{c}_3^\alpha(\xi_3) \cap X_3(x_2, \xi_3)$, 有

$$\begin{aligned}
&\|B_3^E(\xi_3')x_2' + A_3^E(\xi_3')y - b_3^E(\xi_3')\| \\
&\le \|B_3^E(\xi_3)x_2 + A_3^E(\xi_3)y - b_3^E(\xi_3)\| + \|B_3^E(\xi_3)x_2 + A_3^E(\xi_3)y - b_3^E(\xi_3) - (B_3^E(\xi_3')x_2' + A_3^E(\xi_3')y - b_3^E(\xi_3'))\| \\
&\le \|B_3^E(\xi_3)x_2 + A_3^E(\xi_3)y - b_3^E(\xi_3) - (B_3^E(\xi_3')x_2' + A_3^E(\xi_3')y - b_3^E(\xi_3'))\|
\end{aligned}$$

$$\begin{aligned}
&\|(B_3^I(\xi_3')x_2' + A_3^I(\xi_3')y - b_3^I(\xi_3'))_-\| \\
&\le \|(B_3^I(\xi_3)x_2 + A_3^I(\xi_3)y - b_3^I(\xi_3))_-\| + \|B_3^I(\xi_3)x_2 + A_3^I(\xi_3)y - b_3^I(\xi_3) - (B_3^I(\xi_3')x_2' + A_3^I(\xi_3')y - b_3^I(\xi_3'))\| \\
&\le \|B_3^I(\xi_3)x_2 + A_3^I(\xi_3)y - b_3^I(\xi_3) - (B_3^I(\xi_3')x_2' + A_3^I(\xi_3')y - b_3^I(\xi_3'))\|
\end{aligned}$$

由于 y 的有界性, 以及领域半径 δ 可以任意小, 并且 $(x_2', \xi_3') \in B_\delta(x_2, \xi_3)$, 由 [20, 定理 5.4] 对足够小的 $\epsilon > 0$, 存在 $\delta > 0$ 和 $\mathcal{K} > 0$, 使得 $d(y, X_3(x_2', \xi_3')) \le \max_{\xi_3' \in B_\delta(\xi_3)} \mathcal{K}(x_2', \xi_3')\epsilon \le \bar{\mathcal{K}}\epsilon$, 对任意 $(x_2', \xi_3') \in B_\delta(x_2, \xi_3')$.

易知, 存在足够大的 $\bar{\alpha}$ 使得 $\text{lev}\ \bar{c}_3^\alpha(\xi_3) + 2\bar{\mathcal{K}}\epsilon\mathcal{B} \subset \text{lev}\ \bar{c}_3^{\bar{\alpha}}(\xi_3)$, 从而紧集 $\text{lev}\ \bar{c}_3^{\bar{\alpha}}(\xi_3) \cap X_3(x_2', \xi_3') \ne \emptyset$, 这也意味着对任意的 $(x_2', \xi_3') \in B_\delta(x_2, \xi_3) \subset \mathcal{X}_2 \times \Xi_3$, 水平集 $c_3^{\bar{\alpha}}(x_2', \xi_3')$ 非空. ■

命题 2.2 *假设2.2-2.3 成立, $\bar{X}_2 \subset \mathcal{X}_2$ 是一个紧集. 令 $\tilde{x}_3(x_2, \xi_3)$ 是第三阶段问题(2.6) 的解, 则存在紧集 $\bar{X}_3 \subset \mathbb{R}^{n_3}$ 使得 $\tilde{x}_3(x_2, \xi_3) \in \bar{X}_3$ 对任意 $x_2 \in \bar{X}_2$ 和 $\xi_3 \in \Xi_3$ 成立.*

证明. 由假设2.2, 对任意给定的 $x_2 \in \bar{X}_2$ 和 $\xi_3 \in \Xi_3$, 第三阶段问题 (2.5) 也是一个强凸规划问题, 若解 $\tilde{x}_3(x_2, \xi_3)$ 存在, 则其唯一, 根据命题2.1和 [21, 命题 4.4], $\tilde{x}_3(x_2, \xi_3)$ 关于 (x_2, ξ_3) 连续. 又由 $\bar{X}_2$ 和 Ξ_3 有界可知, 存在紧集 $\bar{X}_3$ 使得对所有的 $x_2 \in \bar{X}_2, \xi_3 \in \Xi_3, \tilde{x}_3(x_2, \xi_3) \in \bar{X}_3$. 若解不存在, 则为空集 $\emptyset \subset \bar{X}_3$. ■

我们将条件期望值 $\mathbb{E}[Q_3(x_2, \xi_3)|\xi_2]$ 记为 $\mathbb{E}_{P_3(\xi_2)}[Q_3(x_2, \xi_3)]$, 其中 $P_3(\xi_2)$ 是在 ξ_2 条件下 ξ_3 的条件概率测度.

假设 2.4 *条件概率测度 $P_3(\xi_2)$ 关于 $\xi_2 \in \Xi_2$ 在全变差度量下 Lipschitz 连续, 即*

$$d_{TV}(P_3(\xi_2), P_3(\xi_2')) \le k_{TV}\|\xi_2 - \xi_2'\|,$$

这里 $d_{TV}(P, Q) := \sup_{g \in \mathcal{G}} |\mathbb{E}_P[g(\xi_3)] - \mathbb{E}_Q[g(\xi_3)]|$, $\mathcal{G} := \{g : \sup_{\xi_3 \in \Xi_3} |g(\xi_3)| \le 1, g\text{可测}\}$.

注 2.2 *若随机变量 ξ_2 和 ξ_3 相互独立, 则对任意 $\xi_2, \xi_2' \in \Xi_2$, $P_3(\xi_2) = P_3(\xi_2')$. 如果 ξ_2 和 ξ_3 不独立, 假设 2.4 要求随着 ξ_2 的变化, ξ_3 的条件分布的变化是连续的. 该假设放宽了问题对 ξ_2 和 ξ_3 之间的独立性的要求.*

命题 2.3 *假设 2.2-2.4成立. 则有 (a)$\mathcal{Q}(x_2,\xi_2)$ 关于 x_2 在 $\mathcal{X}_2$ 上是凸函数, 并且关于 (x_2,ξ_2) 在 $\mathcal{X}_2\times\Xi_2$ 内连续. 进一步, 假设对几乎所有的 $\xi_3\in\Xi_3$, $A_3(\xi_3)$, $A_3(\xi_3)H_3^{-1}(\xi_3)A_3(\xi_3)^\top$ 是正定的, 并且存在 $k_3>0$ 使得*

$$\lambda_{min}(A_3(\xi_3)H_3^{-1}(\xi_3)A_3(\xi_3)^\top)\geq k_3,$$

则有 (b) 对所有 $\xi_3\in\Xi_3$, 存在系数 $\kappa_2(\xi_3)$ 和 $\bar{\kappa}_2$ $Q_3(\cdot,\xi_3)$ 关于 x_2 局部 Lipschitz 连续, $\mathcal{Q}_3(x_2,\xi_2)$ 关于 (x_2,ξ_2) 在 $\mathcal{X}_2\times\Xi_2$ 内局部 Lipschitz 连续, , 并且关于 x_2 可微.

证明. (a) 部分: 假设 2.2-2.4, [21, 命题 4.4] 和 [2, 命题 2.21], 第三阶段最优值函数 $Q_3(x_2,\xi_3)$ 在 $\mathcal{X}_2$ 上关于 x_2 凸并且连续. 从而 $\mathcal{Q}(x_2,\xi_2)$ 在 $\mathcal{X}_2$ 上关于 x_2 是凸函数. 由命题2.2, 对任意的 x_2' 使得 $(x_2',\xi_2')\in(B_\delta(x_2)\cap\mathcal{X}_2)\times\Xi_2$, $|Q_3(x_2',\xi_3)|$ 有界, 记为 a_3. 则随着 $x_2'\to x_2$,

$$|\mathcal{Q}_3(x_2,\xi_2)-\mathcal{Q}_3(x_2',\xi_2)|\to 0.$$

进一步, 对任意 $x_2'\in B_\delta(x_2)\cap\mathcal{X}_2$,

$$|\mathcal{Q}_3(x_2',\xi_2)-\mathcal{Q}_3(x_2',\xi_2')|=|\mathbb{E}_{P_3(\xi_2)}[Q_3(x_2',\xi_3)]-\mathbb{E}_{P_3(\xi_2')}[Q_3(x_2',\xi_3)]|\leq a_3k_{TV}\|\xi_2-\xi_2'\|. \tag{2.7}$$

所以对任意 $(x_2',\xi_2')\in(B_\delta(x_2)\cap\mathcal{X}_2)\times\Xi_2$,

$$|\mathcal{Q}_3(x_2,\xi_2)-\mathcal{Q}_3(x_2',\xi_2')|\leq|\mathcal{Q}_3(x_2,\xi_2)-\mathcal{Q}_3(x_2',\xi_2)|+|\mathcal{Q}_3(x_2',\xi_2)-\mathcal{Q}_3(x_2',\xi_2')|\to 0 \tag{2.8}$$

随着 $(x_2',\xi_2')\to(x_2,\xi_2)$. 故 $\mathcal{Q}_3(x_2,\xi_2)$ 关于 x_2 和 ξ_2 在 $\mathcal{X}_2\times\Xi_2$ 内连续.

(b) 部分: 我们考虑第三阶段问题(2.6)的对偶问题

$$\begin{array}{ll}\min\limits_{\lambda_3} & -(b_3(\xi_3)-B_3(\xi_3)x_2+A_3(\xi_3)H_3^{-1}(\xi_3)h_3(\xi_3))^\top\lambda_3+\frac{1}{2}\lambda_3^\top(A_3(\xi_3)H_3^{-1}(\xi_3)A_3(\xi_3)^\top)\lambda_3+q_3(\xi_3)\\ \text{s.t.} & (\lambda_3)_i\geq 0, i\in I.\end{array} \tag{2.9}$$

由条件, 问题(2.9)是一个解集非空的强凸二次规划问题, 从而有唯一解 $\lambda_3(x_2,\xi_3)$. 又由 [21, 命题 4.4], 对偶问题最优解 $\lambda_3(x_2,\xi_3)$ 关于 (x_2,ξ_3) 连续, 故而在紧集 $(B_\delta(x_2)\cap\mathcal{X}_2)\times\Xi_3$ 上有界. 由 [2, 命题 2.22 和推论 2.23], $Q_3(x_2,\xi_3)$ 可微, 并且 $\nabla_{x_2}Q_3(x_2,\xi_3)=B_3(\xi_3)^\top\lambda_3(x_2,\xi_3)$. 又由 $B_3(\xi_3)^\top\lambda_3(x_2,\xi_3)$ 在 $(B_\delta(x_2)\cap\mathcal{X}_2)\times\Xi_3$ 上有界, $Q_3(x_2,\xi_3)$ 关于 x_2 是局部 Lipschitz 连续的. 结合(2.7)-(2.8), $\mathcal{Q}_3(x_2,\xi_2)$ 关于 (x_2,ξ_2) 在 $\mathcal{X}_2\times\Xi_2$ 内局部 Lipschitz 连续, 并且关于 x_2 可微. ■

下面我们证明问题 (2.2)解的有界性. 我们需要如下符号和假设. 令

$$\text{lev } c_2^\alpha(x_1,\xi_2):=\{x_2:x_2\in X_2(x_1,\xi_2),\ c_2(x_2,\xi_2)+\mathcal{Q}_3(x_2,\xi_2)\leq\alpha,\}$$

为第二阶段问题在点 (x_1,ξ_2) 的水平集.

假设 2.5 *对任意 $(x_1,\xi_2)\in\mathcal{X}_1\times\Xi_2$, 如下条件之一成立:*

1. *$A_2^E(\xi_2)$ 是列满秩;*
2. *$A_2^E(\xi_2)$ 是行满秩, 并且对每一个使得 1) $A^I(\xi_2)x_2\geq b_2^I(\xi_2)-B_2^I(\xi_2)x_2,(A_2^I(\xi_2)x_2)_l=(b_2^I(\xi_2)-B_2^I(\xi_2)x_1)_l, A^E(\xi_2)x_2=b_2^E(\xi_2)-B_2^E(\xi_2)x_1$ 或者 2) $A^I(\xi_2)x_2\geq 0,(A_2^I(\xi_2)x_2)_l=0, A^E(\xi_2)x_2=0, x_2\neq 0$ 是可解的非空指标集 $l\subseteq\{1,\cdots,s_2^I\}$, 我们有或者 1) $(A_2^I(\xi_2)x_2)_l>0, A^E(\xi_2)x_2=0$ 是可解的, 或者 $((A_2^I(\xi_2))_l^\top,(A_2^E(\xi_2))^\top)^\top$ 是列满秩的.*

假设2.5与假设2.3类似, 在该假设下, 第二阶段的可行解集 $X_2(x_1,\xi_2)$ 的条件数 (Hoffman 引理中线性系统误差界的系数) 局部有界.

命题 2.4 *若假设2.2-2.5成立, 则存在 α_2 和紧集 C_2 和足够小的领域半径 δ_2, 使得对所有的 $(x_1',\xi_2')\in B_{\delta_2}(x_1,\xi_2)\cap\mathcal{X}_1\times\Xi_2$, lev $c_2^\alpha(x_1',\xi_2')\subset C_2$ 并且非空.*

证明. 由命题2.3, $\mathcal{Q}_3$ 关于 x_2 凸并且可微. 故对任意 $(x_2', \xi_3') \in X_2 \times B_\delta(\xi_3)$, 有

$$\mathcal{Q}_3(x_2', \xi_2') \geq \mathcal{Q}_3(\bar{x}_2, \xi_2') + \mathbb{E}[\nabla_{x_2} Q_3(\bar{x}_2, \xi_3)|\xi_2']^\top (\bar{x}_2 - x_2')$$

从而有

$$\underline{c}_2(x_2, \xi_2) := x_2^\top \underline{H}_2(\xi_2) x_2 - \underline{h}_2(\xi_2) \|x_2\|_1 - \underline{q}_2(\xi_2) \leq c_2(x_2, \xi_2') + \mathcal{Q}_3(x_2, \xi_2')$$

其中 $\bar{x}_2 \in X_2$, 并且对任意的 $\xi_2' \in B_\delta(\xi_2)$,

$$H_2(\xi_2') \succeq \underline{H}_2(\xi_2) \succ 0,\ \underline{h}_2(\xi_2) \geq \|h_2(\xi_2') + \mathbb{E}[\nabla_{x_2} Q_3(\bar{x}_2, \xi_3)|\xi_2']\|_1,\ \underline{q}_3(\xi_3) \geq \|q_3(\xi_3') + \mathcal{Q}_3(\bar{x}_2, \xi_2')\|_1.$$

进一步, 由 $\mathcal{Q}_3$ 的连续性以及 X_2 的有界性, 存在正定二次函数 $\tilde{H}_2(\xi_2)$ 使得对任意 $\xi_2' \in B_\delta(\xi_2)$,

$$\mathcal{Q}_3(x_2, \xi_2') \leq x_2^\top \tilde{H}_2(\xi_2) x_2 + \tilde{h}_2(\xi_2) x_2 + \tilde{q}_2(\xi_2).$$

令

$$\bar{c}_2(x_2, \xi_2) := x_2^\top \bar{H}_2(\xi_2) x_2 + \bar{h}_2(\xi_2) \|x_2\|_1 + \bar{q}_2(\xi_2) \geq c_2(x_2, \xi_2') + \mathcal{Q}_3(x_2, \xi_2')$$

其中对任意的 $\xi_2' \in B_\delta(\xi_2)$,

$$\bar{H}_2(\xi_2) \succeq H_2(\xi_2') + \tilde{H}(\xi_2),\ \bar{h}_3(\xi_3) \geq \|h_2(\xi_2') + \tilde{h}(\xi_2)\|_1,\ \bar{q}_2(\xi_2) \geq \|q_2(\xi_2') + \tilde{q}_2(\xi_2)\|_1.$$

由假设2.2-2.4以及命题2.3, 这样的 $\bar{H}_2(\xi_2), \underline{H}_2(\xi_2), \bar{h}_2(\xi_2), \underline{h}_2(\xi_2), \bar{q}_2(\xi_2), \underline{q}_2(\xi_2)$ 存在.

由于 c_2 是正定二次函数, 并且假设2.2 成立, 容易知道对任意的 α, lev $\bar{c}_2^\alpha(\xi_2)$ 和集合 lev $\underline{c}_2^\alpha(\xi_2)$ 是紧集, 并且有

$$\text{lev } \bar{c}_2^\alpha(\xi_2) \cap X_2(x_1', \xi_2') \subset \text{lev } c_2^\alpha(x_2', \xi_2') \subset \text{lev } \underline{c}_2^\alpha(\xi_2) \cap X_2(x_1', \xi_2')$$

对任意 $(x_1', \xi_2') \in B_\delta(x_1, \xi_2) \subset \mathcal{X}_1 \times \Xi_2$ 成立. 余下的证明与命题2.1类似, 我们略去了详细证明. ■

命题 2.5 *令假设 2.1-2.5 成立, 则问题 (2.2) 是一个解集非空的强凸优化问题, 并存在紧集 X_2, X_3 和唯一解 $(x_1^*, x_2^*(\xi_2), x_3^*(\xi_2, \xi_3))$ 使得 $x_2^*(\xi_2) \in X_2$, $x_3^*(\xi_2, \xi_3) \in X_3$ 对几乎所有的 $\xi_2 \in \Xi_2$ 和 $\xi_3 \in \Xi_3$ 成立.*

证明. 由假设 2.1-2.2, 问题 (2.2)是一个解集非空的二次强凸优化问题可由假设 2.1-2.2直接得到, 而解集非空的二次强凸问题有唯一解 ([22, 定理 12.2 和引理 12.2]).

由假设 2.1-2.4 以及 [2, 第二章], 对给定的 x_1^* 和 $\xi_2 \in \Xi_2$, 第二阶段问题 (2.4)是一个强凸优化问题, 其解 $x_2^*(\xi_2)$ 唯一. 进一步由命题2.4和 [21, 命题 4.4], $x_2^*(\xi_2)$ 关于 ξ_2 连续. 又由 Ξ_2 有界性可知, 存在紧集 X_2 使得对所有的 $\xi_2 \in \Xi_2$, $x_2^*(\xi_2) \in X_2$.

最后, 考虑第三阶段问题 (2.5). 注意到 $x_3^*(\xi_2, \xi_3)$ 实质上是与 $(x_2^*(\xi_2), \xi_3)$ 相关的, 为方便起见, 我们将其记为 $\bar{x}_3^*(x_2^*(\xi_2), \xi_3)$. 由假设 2.1-2.4, 对给定的 $x_2^*(\xi_2) \in X_2$ 和 $\xi_3 \in \Xi_3$, 第三阶段问题 (2.5) 也是一个强凸优化问题, 其解 $\bar{x}_3^*(x_2^*(\xi_2), \xi_3)$ 唯一. 同样根据 [21, 命题 4.4], $\bar{x}_3^*(x_2^*(\xi_2), \xi_3)$ 关于 $(x_2^*(\xi_2), \xi_3)$ 连续. 又由 X_2 和 Ξ_2 有界性可知, 存在紧集 X_3 使得对所有的 $x_2^*(\xi_2) \in X_2, \xi_3 \in \Xi_3$, $\bar{x}_3^*(x_2^*(\xi_2), \xi_3) \in X_3$, 即 $x_3^*(\xi_2, \xi_3) \in X_3$. ■

推论 2.1 *令假设 2.1-2.5 成立, $X_1 \subset \mathcal{X}_1$ 是一个紧集. 对任意 $x_1 \in X_1$ 和 $\xi_2 \in \Xi_2$, 令 $\hat{x}_2(x_1, \xi_2)$ 是第二阶段问题(2.4)的解, 则存在紧集 $X_2 \subset \mathbb{R}^{n_2}$ 使得 $\hat{x}_2(x_1, \xi_2) \in X_2$.*

因为推论 2.1 的证明类似于命题 2.5 中 $x_3^*(\xi_2, \xi_3)$ 的有界性证明, 所以这里我们省略了这一证明.

下面, 我们进一步考虑 $\mathcal{Q}_2(x_1, \xi_2)$ 的 Lipschitz 连续性.

命题 2.6 *假设 (a) 命题2.3的条件及假设2.1, 2.5成立, (b) 存在 $k_1 > 0$ 使得对任意 $\xi_2 \in \Xi_2$, 有*

$$\lambda_{min}((A_2(\xi_2) A_2(\xi_2)^\top)^{-1}) \geq k_1,$$

则对任意 $\xi_2 \in \Xi_2$, $Q_2(\cdot, \xi_2)$ 和 $\mathcal{Q}_2(\cdot)$ 关于 x_1 在紧集 X_1 上 Lipschitz 连续.

证明. 由命题2.3, 对给定的 $\xi_2 \in \Xi_2$, 第二阶段问题(2.4)的一阶最优性条件为

$$\begin{aligned}
&h_2(\xi_2) + H_2(\xi_2)x_2 + \nabla_{x_2}\mathcal{Q}_3(x_2,\xi_2) = A_2(\xi_2)^\top \lambda_2\\
&B_2^E(\xi_2)x_1 + A_2^E(\xi_2)x_2 = b_2^E(\xi_2)\\
&B_2^I(\xi_2)x_1 + A_2^I(\xi_2)x_2 \geq b_2^I(\xi_2)\\
&(\lambda_2^I)^\top(B_2^I(\xi_2)x_1 + A_2^I(\xi_2)x_2 - b_2^I(\xi_2)) = 0\\
&\lambda_2^I \geq 0.
\end{aligned} \tag{2.10}$$

由问题(2.4)是线性约束并且解集非空的强凸优化问题, 强对偶条件成立, 存在 KKT 对 $(x_2(x_1,\xi_2),\lambda_2(x_1,\xi_2))$ 是(2.10)的解. 进一步, 由(2.10)第一行和假设 (b),

$$\lambda_2 = (A_2(\xi_2)A_2(\xi_2)^\top)^{-1}A_2(\xi_2)(h_2(\xi_2) + H_2(\xi_2)x_2 + \nabla_{x_2}\mathcal{Q}_3(x_2,\xi_2)).$$

再由假设2.2、命题2.5和推论2.1, 对任意的 $(x_1,\xi_2) \in X_1 \times \Xi_2$, λ_2 一致有界. 故而对任意的 $(x_1,\xi_2) \in X_1 \times \Xi_2$, 第二阶段问题(2.4)的对偶问题的解唯一并且 $\lambda_2(x_1,\xi_2)$ 一致有界. 由 [2, 命题 2.22], $\nabla_{x_1}Q_2(x_1,\xi_2) = B_2^\top(\xi_2)\lambda_2(x_1,\xi_2)$ 一致有界, 从而 $\nabla\mathcal{Q}_2(x_1)$ 一致有界. 故对任意 $\xi_2 \in \Xi_2$, $Q_2(\cdot,\xi_2)$ 和 $\mathcal{Q}_2(\cdot)$ 关于 x_1 在紧集 X_1 上 Lipschitz 连续. ■

2.2 样本均值逼近问题的性质

由于问题 (2.2) 是一个无穷维的三阶段随机规划问题, 问题 (2.2) 几乎是不能直接用于数值计算的. 因此, 我们考虑如下样本均值逼近问题:

$$\min_{x_1\in X_1} c_1(x_1) + \frac{1}{N_1}\sum_{i=1}^{N_1}\left(\min_{x_2(\xi_2^i)\in X_2(x_1,\xi_2^i)} c_2(x_2(\xi_2^i),\xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2}\left(\min_{x_3(\xi_{[3]}^{ij})\in X_3(x_2(\xi_2^i),\xi_3^{ij})} c_3(x_3(\xi_{[3]}^{ij}),\xi_3^{ij})\right)\right) \tag{2.11}$$

其中 $\{\xi_2^1,\cdots,\xi_2^{N_1}\}$ 是随机变量 ξ_2 的 N_1 个随机样本. 进一步, 对每一个 ξ_2^i, $1 \leq i \leq N_1$, 产生随机变量 ξ_3 的 N_2 个条件样本 $\{\xi_3^{i1},\cdots,\xi_3^{iN_2}\}$, 并记 $\xi_{[3]}^{ij} = (\xi_2^i,\xi_3^{ij})$. 在假设 2.1-2.2 下, 问题 (2.11) 是一个有限维强凸二次规划问题. 我们记 (x_1^N,x_2^N,x_3^N) 为问题 (2.11) 的唯一最优解, 其中

$$\begin{aligned}
x_2^N &:= (x_2^N(\xi_2^1),\cdots,x_2^N(\xi_2^{N_1})),\\
x_3^N &:= (x_3^N(\xi_{[3]}^{11}),\cdots,x_3^N(\xi_{[3]}^{1N_2}),\cdots,x_3^N(\xi_{[3]}^{N_11}),\cdots,x_3^N(\xi_{[3]}^{N_1N_2})).
\end{aligned}$$

对于给定的 (x_1,ξ_2^i), 定义 $Q_2^N(x_1,\xi_2^i)$ 为第二阶段问题的样本逼近问题的最优值函数, 即

$$Q_2^N(x_1,\xi_2^i) := \min_{x_2\in X_2(x_1,\xi_2)} c_2(x_2,\xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2^i,\xi_3^{ij}), \tag{2.12}$$

则第一阶段问题 (2.3) 的样本逼近近似问题为

$$\min_{x_1\in X_1} \hat{c}_{N_1,N_2}(x_1) := c_1(x_1) + \frac{1}{N_1}\sum_{i=1}^{N_1} Q_2^{N_2}(x_1,\xi_2^i). \tag{2.13}$$

我们首先给出如下有界性结论.

定理 2.1 *若命题2.6条件成立,, 则对任意的 $N_1,N_2 > 0$, 问题 (2.11) 是一个解集非空的强凸优化问题, 于是存在唯一的最优解 (x_1^N,x_2^N,x_3^N) 并且存在紧集 $X_1 \subset \mathbb{R}^{n_1}$, $X_2 \subset \mathbb{R}^{n_2}$, $X_3 \subset \mathbb{R}^{n_3}$, 使得当 N_1,N_2 足够大时, 有 $x_1^N \in X_1$, $x_2^N(\xi_2^i) \in X_2$ 和 $x_3^N(\xi_{[3]}^{ij}) \in X_3$ 对任意的 $i \in \{1,\cdots,N_1\}$ 和 $j \in \{1,\cdots,N_2\}$ 成立.*

证明. 由假设 2.1-2.4 可以直接得到问题 (2.11) 是一个解集非空的强凸优化问题, 并且存在唯一的最优解. 我们只需证明这个最优解序列 $\{(x_1^N,x_2^N,x_3^N)\}$ 有界即可. 为此, 我们分为如下三步来证明.

步 1: 首先, 我们采用反证法证明 $\{x_1^N\}$ 的有界性. 假设该序列无界, 则最优解序列必存在一个子列, 为方便起见, 仍然记为 $\{(x_1^N, x_2^N, x_3^N)\}$, 并且 $x_1^N \to \infty$. 容易看出 $(x_1^*, x_2^{*N}, x_3^{*N})$ 是问题(2.11)的一个可行解, 其中 $x_2^{*N} := (x_2^*(\xi_2^1), \cdots, x_2^*(\xi_2^{N_1}))$, $x_3^{*N} := (x_3^*(\xi_{[3]}^{11}), \cdots, x_3^*(\xi_{[3]}^{N_1 N_2}))$, 其中 $(x_1^*, x_2^*(\cdot), x_3^*(\cdot))$ 是原文题 (2.2) 的唯一最优解. 令,

$$z_1^{*N} := \frac{1}{N_1}\sum_{i=1}^{N_1}\left(c_2(x_2^*(\xi_2^i), \xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2}\left(c_3(x_3^*(\xi_{[3]}^{ij}), \xi_3^{ij})\right)\right)$$

关于 N 一致有界, 即存在 $\bar{z}_1$ 使得 $z_1^{*N} \le \bar{z}_1$ 对所有 N 成立.

进一步, 由于 c_2 和 c_3 都是强凸的二次函数，有

$$z_1^2(\xi_2) := \frac{4q_2(\xi_2) - h_2(\xi_2)^\top H_2(\xi_2)^{-1}h_2(\xi_2)}{4} = \min_{x_2\in\mathbb{R}^{n_2}} x_2^\top H_2(\xi_2)x_2 + h_2(\xi_2)^\top x_2 + q_2(\xi_2),$$

$$z_1^3(\xi_3) := \frac{4q_3(\xi_3) - h_3(\xi_3)^\top H_3(\xi_3)^{-1}h_3(\xi_3)}{4} = \min_{x_3\in\mathbb{R}^{n_2}} x_3^\top H_2(\xi_3)x_3 + h_3(\xi_3)^\top x_3 + q_3(\xi_3),$$

并且由假设 2.2, 对任意的 $\xi_2 \in \Xi_2$ 和 $\xi_3 \in \Xi_3$, $z_1^2(\xi_2)$ 和 $z_1^3(\xi_3)$ 一致有界. 令

$$z_1^N := \frac{1}{N_1}\sum_{i=1}^{N_1}\left(\min_{x_2(\xi_2^i)\in\mathbb{R}^{n_2}} c_2(x_2(\xi_2^i), \xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2}\left(\min_{x_3(\xi_{[3]}^{ij})\in\mathbb{R}^{n_3}} c_3(x_3(\xi_{[3]}^{ij}), \xi_3^{ij})\right)\right),$$

则有 $z_1^N = \frac{1}{N_1}\sum_{i=1}^{N_1}\left(z_1^2(\xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2} z_1^3(\xi_{[3]}^{ij})\right)$ 关于 N 是一致有界的.

由于 $x_1^N \to \infty$ 以及 z_1^N 的一致有界性, 随着 $N_1, N_2 \to +\infty$

$$+\infty \leftarrow c_1(x_1^N) + z_1^N \le c_1(x_1^N) + \frac{1}{N_1}\sum_{i=1}^{N_1} Q_2^N(x_1^N, \xi_2^i) \le c_1(x_1^*) + z_1^{*N} \le c_1(x_1^*) + \bar{z}_1 < \infty,$$

矛盾. 故存在 X_1 使得当 N_1, N_2 足够大时, $x_1^N \in X_1$.

步 2: 我们再考虑 $\{x_2^N\}$ 的有界性. 假设该序列无界, 则最优解序列必存在一个子列, 为方便起见, 仍记为 $\{(x_1^N, x_2^N, x_3^N)\}$, 并且 $x_1^N \in X_1$, $\sup_{i\in\{1,\cdots,N_1\}} x_2^N(\xi_2^i) \to \infty$. 也容易看出 $(x_1^N, \hat{x}_2^N, \hat{x}_3^N)$ 是问题 (2.11) 的一个可行解, 其中

$$\hat{x}_2^N := \{\hat{x}_2(x_1^N, \xi_2^1), \cdots, \hat{x}_2(x_1^N, \xi_2^{N_1}), \quad \hat{x}_3^N := \{\hat{x}_3(x_1^N, \xi_{[3]}^{11}), \cdots, \hat{x}_3(x_1^N, \xi_{[3]}^{N_1 N_2}),$$

$\hat{x}_2(x_1^N, \xi_2^1)$ 定义于推论 2.1, $\hat{x}_3(x_1^N, \xi_{[3]}^{ij})$ 是问题 (2.5) 中 x_1 取 x_1^N, (ξ_2, ξ_3) 取 (ξ_2^i, ξ_3^{ij}) 时第三阶段问题的解. 则由推论 2.1 和假设 2.1-2.4,

$$\hat{z}_2^N := \max_{x_1\in X_1, i\in\{1,\cdots,N_1\}} c_2(\hat{x}_2(x_1, \xi_2^i), \xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2}\left(c_3(\hat{x}_3(x_1, \xi_{[3]}^{ij}), \xi_3^{ij})\right)$$

关于 N 一致有界, 即存在 $\bar{z}_2$ 使得 $\hat{z}_2^N \le \bar{z}_2$ 对所有 N 成立. 令

$$z_2^N(i) := \frac{1}{N_2}\sum_{j=1}^{N_2}\left(\min_{x_3(\xi_{[3]}^{ij})\in\mathbb{R}^{n_3}} c_3(x_3(\xi_{[3]}^{ij}), \xi_3^{ij})\right) = \frac{1}{N_2}\sum_{j=1}^{N_2} z_1^3(\xi_3^{ij})$$

关于 N_2 和 $i \in \{1, \cdots, N_1\}$ 一致有界. 由于 $\sup_{i\in\{1,\cdots,N_1\}} x_2^N(\xi_2^i) \to \infty$, 随着 $N_1 \to \infty$,

$$\begin{aligned}\infty \leftarrow \max_{i\in\{1,\cdots,N_1\}} c_2(x_2^N(\xi_2^i)) + z_2^N(i) \quad &\le \quad \max_{i\in\{1,\cdots,N_1\}} c_2(x_2^N(\xi_2^i), \xi_2^i) + \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3^N(x_2^N(\xi_2^i), \xi_3^{ij}) \\ &\le \quad \hat{z}_2^N \le \bar{z}_2 < \infty\end{aligned}$$

矛盾. 由上述矛盾对 $x_1^N \in X_1$ 和 $\xi_2 \in \Xi_2$ 的一致性, 故存在 X_2 使得当 N_1, N_2 足够大时, $x_2^N(\xi_2^i) \in X_2$, $i \in \{1, \cdots, N_1\}$.

步 3: 我们进一步考虑 $\{x_3^N\}$. 记样本均值逼近问题解序列为 $\{(x_1^N, x_2^N, x_3^N)\}$, 并且 $x_1^N \in X_1$, $x_2^N(\xi_2^i) \in X_2$, $i \in \{1, \cdots, N_1\}$, 容易看出 $x_3^N(\xi_3^{ij}) = \tilde{x}_3(x_2^N(\xi_2^i), \xi_3^{ij})$, 从而由推论 2.1, 存在紧集 X_3, 使得 $x_3^N(\xi_{[3]}^{ij}) \in X_3$. ■

命题 2.7 *若命题2.6条件成立, 则对任意的 $N_1, N_2 > 0$, $Q_2^{N_2}(x_1, \xi_2^j)$ 是 Lipschitz 连续.*

利用定理2.1的有界性结果, 命题2.7的证明与命题2.6类似, 这里我们省略了证明.

2.3 样本均值逼近方法的收敛性

在我们给出多阶段随机二次规划问题的样本均值逼近的收敛性结论前, 我们先给出一些预备知识. 考虑 $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$, $\psi(x) := \mathbb{E}[\phi(x, \xi)]$ 和样本均值逼近项

$$\psi_N(x) := \frac{1}{N} \sum_{i=1}^{N} \phi(x, \xi^i),$$

其中 $\xi^1, \cdots, \xi^N$ 是 ξ 的 N 个一般样本 (可以不是独立同分布的). 记

$$M_x^N(t) := \mathbb{E}[e^{t(\psi_N(x) - \psi(x))}].$$

假设 2.6 *对任意的 $x \in X \subseteq \mathbb{R}^n$ 和 $t \in \mathbb{R}$, 下面的极限*

$$M_x(t) := \lim_{N \to \infty} M_x^N(t)$$

在扩展的实数范围内总存在, 并且 $M_x(t) < \infty$ 对任意的 t 充分靠近 0.

假设 2.7 *ϕ 在 X 上 Lipschitz 连续, 也即是,*

$$|\phi(x, \xi) - \phi(\bar{x}, \xi)| \le \kappa(\xi) \|x - \bar{x}\|, \forall x, \bar{x} \in X.$$

此外, 令 $\hat{\kappa}_N := \frac{1}{N} \sum_{i=1}^{N} \kappa(\xi^i)$. 对任意的 $\epsilon > 0$, 存在独立于 N 的 $\alpha(\epsilon) > 0$ 和 $\beta(\epsilon) > 0$, 使得

$$\mathrm{Prob}\{|\hat{\kappa}_N - \mathbb{E}[\kappa(\xi)]| \ge \epsilon\} \le \alpha(\epsilon) e^{-N\beta(\epsilon)}.$$

显然, 我们从假设 2.7 知道 $\psi(x)$ 是在 X 上连续的. 为了后文方便描述, 我们称 $(\phi(x, \xi), \psi(x))$ 满足假设 2.6-2.7, 如果如上定义的 $\phi(x, \xi)$ 和 $\psi(x)$ 满足假设 2.6-2.7.

引理 2.1 ([23, 定理 3.1]) *如果: (i) 假设 2.6 和 2.7 成立; (ii) $X \subseteq \mathbb{R}^n$ 是紧集, 则对任意的 $\epsilon > 0$ 存在独立于 N 的 $\alpha(\epsilon) > 0$ 和 $\beta(\epsilon) > 0$, 使得*

$$\mathrm{Prob}\left\{\sup_{x \in X} |\psi_N(x) - \psi(x)| \ge \epsilon\right\} \le \alpha(\epsilon) e^{-N\beta(\epsilon)}.$$

对任意 $x_1 \in X_1$, 我们引入如下中间变量:

$$\hat{c}_{N_1}(x_1) := c_1(x_1) + \frac{1}{N_1} \sum_{i=1}^{N_1} Q_2(x_1, \xi_2^i).$$

对任意的误差 $\epsilon>0$, 通过加上和减去 $\hat{c}_{N_1}(x_1)$, 我们有

$$
\begin{aligned}
&\operatorname{Prob}\left\{\sup_{x_1\in X_1}|\hat{c}_{N_1,N_2}(x_1)-c(x_1)|\geq\epsilon\right\} \qquad (2.14)\\
=&\operatorname{Prob}\left\{\sup_{x_1\in X_1}|\hat{c}_{N_1,N_2}(x_1)-\hat{c}_{N_1}(x_1)+\hat{c}_{N_1}(x_1)-c(x_1)|\geq\epsilon\right\}\\
\leq&\operatorname{Prob}\left\{\sup_{x_1\in X_1}|\hat{c}_{N_1,N_2}(x_1)-\hat{c}_{N_1}(x_1)|+\sup_{x_1\in X_1}|\hat{c}_{N_1}(x_1)-c(x_1)|\geq\epsilon\right\}\\
\leq&\operatorname{Prob}\left\{\sup_{x_1\in X_1}|\hat{c}_{N_1,N_2}(x_1)-\hat{c}_{N_1}(x_1)|\geq\frac{\epsilon}{2}\right\}+\operatorname{Prob}\left\{\sup_{x_1\in X_1}|\hat{c}_{N_1}(x_1)-c(x_1)|\geq\frac{\epsilon}{2}\right\}. \qquad (2.15)
\end{aligned}
$$

假设 2.8 如下条件成立:

(A1) 存在紧集 $X_1\subset\mathbb{R}^{n_1}$, $X_2\subset\mathbb{R}^{n_2}$, $X_3\subset\mathbb{R}^{n_3}$, 使得当 N_1,N_2 足够大时, 有 $x_1^*,x_1^N\in X_1$, $x_2^*(\xi_2),x_2^N(\xi_2^i)\in X_2$ 和 $x_3^*(\xi_{[3]}),x_3^N(\xi_{[3]}^{ij})\in X_3$ 对任意的 $i\in\{1,\cdots,N_1\}$ 和 $j\in\{1,\cdots,N_2\}$ 成立.

(A2) 对任意的 $\xi_2\in\Xi_2$, $Q_2(\cdot,\xi_2)$ 在 X_1 上 *Lipschitz* 连续, 记 *Lipschitz* 系数为 $\kappa_1(\xi_2)$;

(A3) 对任意的 $\xi_3\in\Xi_3$, $Q_3(\cdot,\xi_3)$ 在 X_2 上 *Lipschitz* 连续, 记 *Lipschitz* 系数为 $\kappa_2(\xi_3)$;

(A4) $(Q_2(x_1,\xi_2),\mathbb{E}[Q_2(x_1,\xi_2)])$ 在 X_1 上满足假设 2.6-2.7;

(A5) 对任意 $\xi_2\in\Xi_2$, $(Q_3(x_2,\xi_3),\mathbb{E}[Q_3(x_2,\xi_3)|\xi_2])$ 在 X_2 上满足假设 2.6-2.7;

注 2.3 假设2.8是证明样本均值逼近方法指数收敛速率的主要假设. 如果命题2.6的条件成立, 那么由命题2.3-2.7及定理2.1, 我们有 *(A1)-(A3)* 成立. 进一步, 如果随机变量 $\{\xi_2\}_{i=1}^{N_1}$, $\{\xi_3^{ij}\}_{j=1}^{N_2}$, $i=1,\cdots,N_1$ 是 *i.i.d.* 样本, 由 *2.1-2.2* 节中解的有界性和 *Lipschitz* 系数的有界性及 *Cramér's* 大偏差定理, *(A4)-(A5)* 也成立.

根据假设 (A1), (A3), (A5) 和引理 2.1, 我们有: 对任意的 $\xi_2\in\Xi_2$ 和 $\epsilon>0$, 存在 $\alpha(\epsilon,\xi_2)$ 和 $\beta(\epsilon,\xi_2)$ 使得

$$
\operatorname{Prob}\left\{\sup_{x_2\in X_2}\left|\frac{1}{N_2}\sum_{j=1}^{N_2}Q_3(x_2,\xi_3^j)-\mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2\right]\right|\geq\epsilon\right\}\leq\alpha(\epsilon,\xi_2)e^{-N\beta(\epsilon,\xi_2)}. \qquad (2.16)
$$

假设 2.9 对任意的 $\epsilon>0$, 存在 $\alpha_2(\epsilon)>0$ 和 $\beta_2(\epsilon)>0$ 使得

$$
\operatorname{Prob}\left\{\sup_{x_2\in X_2,\xi_2\in\Xi_2}\left|\frac{1}{N_2}\sum_{j=1}^{N_2}Q_3(x_2,\xi_3^{ij})-\mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2^i\right]\right|\geq\epsilon\right\}\leq\alpha(\epsilon)e^{-N\beta(\epsilon)}.
$$

当 ξ_2 和 ξ_3 独立时, 由(2.16)成立可得假设 2.9 成立. 该假设放宽了问题对 ξ_2 和 ξ_3 之间的独立性要求.

为了建立目标函数和最优解集的关系, 我们给出如下引理.

引理 2.2 ([23, 引理 4.1]) 考虑一般的约束优化问题

$$
\min_{x\in\mathcal{X}}\ f(x), \qquad (2.17)
$$

其中 $f:\mathbb{R}^n\to\mathbb{R}$, $\mathcal{X}\subseteq\mathbb{R}^n$. (2.17)的扰动问题记为

$$
\min_{x\in\mathcal{X}}\ \tilde{f}(x), \qquad (2.18)
$$

其中 $\tilde{f}:\mathbb{R}^n \to \mathbb{R}$ 是 f 的一个扰动函数. 定义 $\mathcal{S}^*$ 和 $\tilde{\mathcal{S}}^*$ 分别为问题 (2.17) 和 (2.18)的最优解集. 对任意的 $\epsilon > 0$, 存在 $\delta > 0$ (依赖于 ϵ), 如果

$$\sup_{x\in\mathcal{X}} \left| f(x) - \tilde{f}(x) \right| \le \delta,$$

那么 $\mathbb{D}(\tilde{\mathcal{S}}^*, \mathcal{S}^*) \le \epsilon$.

现在我们给出本章的最主要结果.

定理 2.2 令假设 *2.8-2.9* 成立. 对任意的 $\epsilon > 0$, 存在 $\alpha_1(\epsilon) > 0$, $\beta_1(\epsilon) > 0$, $\alpha_2(\epsilon) > 0$, $\beta_2(\epsilon) > 0$, $\tilde{\alpha}_1(\epsilon) > 0$, $\tilde{\beta}_1(\epsilon) > 0$, $\tilde{\alpha}_2(\epsilon) > 0$ 和 $\tilde{\beta}_2(\epsilon) > 0$ (不依赖于 N_1 和 N_2), 使得

$$\text{Prob}\left\{ \sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - c(x_1)| \ge \epsilon \right\} \le \alpha_1(\epsilon) e^{-N_1\beta_1(\epsilon)} + \alpha_2(\epsilon) e^{-N_2\beta_2(\epsilon)}, \tag{2.19}$$

$$\text{Prob}\left\{ \mathbb{D}(\mathcal{S}_{N_1,N_2}, \mathcal{S}) \ge \epsilon \right\} \le \tilde{\alpha}_1(\epsilon) e^{-N_1\tilde{\beta}_1(\epsilon)} + \tilde{\alpha}_2(\epsilon) e^{-N_2\tilde{\beta}_2(\epsilon)}. \tag{2.20}$$

证明. 为了估计 (2.19), 我们考虑 (2.15). 根据 (A1), (A2), (A4) 和引理 2.1, 我们可得

$$\text{Prob}\left\{ \sup_{x_1\in X_1} |\hat{c}_{N_1}(x_1) - c(x_1)| \ge \frac{\epsilon}{2} \right\} \le \alpha_1(\epsilon) e^{-N_1\beta_1(\epsilon)} \tag{2.21}$$

其中 $\alpha_1(\epsilon) > 0$ 和 $\beta_1(\epsilon) > 0$.

进一步, 由根据 (A1), (A3), (A5), 我们有

$$\begin{aligned}
&\text{Prob}\left\{ \sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - \hat{c}_{N_1}(x_1)| \ge \frac{\epsilon}{2} \right\} \\
=\ &\text{Prob}\left\{ \sup_{x_1\in X_1} \left| \frac{1}{N_1}\sum_{i=1}^{N_1} Q_2^{N_2}(x_1,\xi_2^i) - \frac{1}{N_1}\sum_{i=1}^{N_1} Q_2(x_1,\xi_2^i) \right| \ge \frac{\epsilon}{2} \right\} \\
\le\ &\text{Prob}\left\{ \sup_{x_1\in X_1,\xi_2\in\Xi_2} \left| Q_2^{N_2}(x_1,\xi_2) - Q_2(x_1,\xi_2) \right| \ge \frac{\epsilon}{2} \right\}. \\
=\ &\text{Prob}\left\{ \sup_{x_1\in X_1,\xi_2\in\Xi_2} \left| \min_{x_2\in X_2(x_1,\xi_2)\cap X_2} \left(c_2(x_2,\xi_2) + \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) \right) \right.\right. \\
&\left.\left. - \min_{x_2\in X_2(x_1,\xi_2)\cap X_2} \left(c_2(x_2,\xi_2) + \mathbb{E}\left[Q_3(x_2,\xi_3) | \xi_2 \right] \right) \right| \ge \frac{\epsilon}{2} \right\} \\
\le\ &\text{Prob}\left\{ \sup_{x_1\in X_1, x_2\in X_2(x_1,\xi_2)\cap X_2, \xi_2\in\Xi_2} \left| \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) - \mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2 \right] \right| \ge \frac{\epsilon}{2} \right\} \\
\le\ &\text{Prob}\left\{ \sup_{x_2\in X_2,\xi_2\in\Xi_2} \left| \frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) - \mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2 \right] \right| \ge \frac{\epsilon}{2} \right\} \le \alpha_2(\epsilon) e^{-N_2\beta_2(\epsilon)},
\end{aligned} \tag{2.22}$$

其中 $\{\xi_3^j\}$ 是从分布 $P_3(\xi_2)$ 中的 i.i.d. 样本, 最后一个不等式来自于假设 2.9. (2.22) 意味着

$$\text{Prob}\left\{ \sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - \hat{c}_{N_1}(x_1)| \ge \frac{\epsilon}{2} \right\} \le \alpha_2(\epsilon) e^{-N_2\beta_2(\epsilon)}. \tag{2.23}$$

最后, 我们由(2.21) 和 (2.23)可得到 (2.19).

根据引理2.2, 对任意的 $\epsilon > 0$, 存在 $\delta := \delta(\epsilon) > 0$ 使得

$$\begin{aligned}
\text{Prob}\left\{ \mathbb{D}(\mathcal{S}_{N_1,N_2}, \mathcal{S}) \ge \epsilon \right\} &\le \text{Prob}\left\{ \sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - c(x_1)| \ge \delta \right\} \\
&\le \alpha_1(\delta(\epsilon)) e^{-N_1\beta_1(\delta(\epsilon))} + \alpha_2(\delta(\epsilon)) e^{-N_2\beta_2(\delta(\epsilon))},
\end{aligned}$$

其中最后一个不等式来自于 (2.19). 令 $\tilde{\alpha}_1(\epsilon) := \alpha_1(\delta(\epsilon))$, $\tilde{\beta}_1(\epsilon) := \beta_1(\delta(\epsilon))$, $\tilde{\alpha}_2(\epsilon) := \alpha_2(\delta(\epsilon))$ 以及 $\tilde{\beta}_2(\epsilon) := \beta_2(\delta(\epsilon))$, 我们便得到(2.20). ■

定理2.2给出了当阶段独立性和 i.i.d 性质不满足时, 基于条件抽样的样本均值逼近方法的指数收敛速率. 可以看出, 这一定理并不只针对三阶段随机二次规划问题, 任何满足假设2.8-2.9的三阶段随机优化问题的样本均值逼近方法的收敛速率都可以被定理2.2刻画. 下面我们介绍如何利用定理2.2估计样本大小, 并且展示定理2.2不仅放宽了 [9, 定理 2] 的条件, 也给出了更紧的样本量的估计.

注 2.4 *如果存在正数 $\rho > 0$ 使得假设2.6中的 $M_x(t) \le e^{\rho^2 t^2/2}$ 对所有 $t \in \mathbb{R}$, Q_2 和 Q_3 成立 (见 [23, 注 3.1] 中有关该条件的进一步讨论), 由 [23, 注 3.2], [24, p410] 和 [9], 对任意 $\epsilon > 0$, 有*

$$\text{Prob}\left\{\sup_{x_1\in X_1} |\hat{c}_{N_1}(x_1) - c(x_1)| \ge \frac{\epsilon}{2}\right\} \le O(1)\left(D_1\left(\frac{8\mathbb{E}[\kappa_1(\xi_2)]}{\epsilon}\right)\right)^{n_1} \exp\left\{-\frac{O(1)N_1\epsilon^2}{\rho^2}\right\}$$

$$\text{Prob}\left\{\sup_{x_2\in X_2}\left|\frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) - \mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2\right]\right| \ge \frac{\epsilon}{2}\right\} \le O(1)\left(D_2\left(\frac{8\mathbb{E}[\kappa_2(\xi_3)|\xi_2]}{\epsilon}\right)\right)^{n_2} \exp\left\{-\frac{O(1)N_2\epsilon^2}{\rho^2}\right\}.$$

若进一步假设

$$\text{Prob}\left\{\sup_{x_2\in X_2,\xi_2\in\Xi_2}\left|\frac{1}{N_2}\sum_{j=1}^{N_2} Q_3(x_2,\xi_3^j) - \mathbb{E}\left[Q_3(x_2,\xi_3)|\xi_2\right]\right| \ge \frac{\epsilon}{2}\right\} \le O(1)\left(D_2\left(\frac{8\bar{\kappa}_2}{\epsilon}\right)\right)^{n_2} \exp\left\{-\frac{O(1)N_2\epsilon^2}{\rho^2}\right\}. \tag{2.24}$$

则有

$$\begin{aligned}\text{Prob}\left\{\sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - c(x_1)| \ge \epsilon\right\} \quad &\le \quad O(1)\left(\left(D_1\left(\frac{8\mathbb{E}[\kappa_1(\xi_2)]}{\epsilon}\right)\right)^{n_1} \exp\left\{-\frac{O(1)N_1\epsilon^2}{\rho^2}\right\}\right.\\ &\quad + \left.\left(D_2\left(\frac{8\bar{\kappa}_2}{\epsilon}\right)\right)^{n_2} \exp\left\{-\frac{O(1)N_2\epsilon^2}{\rho^2}\right\}\right).\end{aligned} \tag{2.25}$$

从而可以依据(2.25)估计达到指定精度所需的样本大小. 注意到当 ξ_2 与 ξ_3 独立时, 设 $\bar{\kappa}_2 = \mathbb{E}[\kappa_2(\xi_3)]$, 在上述条件下(2.24)成立.

进一步, 我们还可以注意到, 在 [9, 定理 2] 的条件下, Shapiro 给出的估计是

$$\begin{aligned}\text{Prob}\left\{\sup_{x_1\in X_1} |\hat{c}_{N_1,N_2}(x_1) - c(x_1)| \ge \epsilon\right\} \quad &\le \quad O(1)\left(\left(D_1\left(\frac{8\mathbb{E}[\kappa_1(\xi_2)]}{\epsilon}\right)\right)^{n_1} \exp\left\{-\frac{O(1)N_1\epsilon^2}{\rho^2}\right\}\right.\\ &\quad + \left.\left(D_1\left(\frac{8\mathbb{E}[\kappa_3(\xi_2)]}{\epsilon}\right)\right)^{n_1}\left(D_2\left(\frac{8\bar{\kappa}_2}{\epsilon}\right)\right)^{n_2} \exp\left\{-\frac{O(1)N_2\epsilon^2}{\rho^2}\right\}\right).\end{aligned} \tag{2.26}$$

其中 $\kappa_3(\xi_2)$ 是 $Q_2^{N_2}(\cdot,\xi_2)$ 的 Lipschitz 系数. 比较(2.25)和(2.26), 可以发现定理2.2改进了样本量的估计, 给出了更紧的样本复杂度分析.

3 多阶段随机二次规划的对称高斯-塞德尔分裂算法

在第 2 节里, 我们已经证明了多阶段随机二次规划问题的样本均值逼近的收敛性. 本章我们考虑如何求解样本均值近似后的一类多阶段随机二次规划问题:

$$\begin{aligned}\min_{x_1\in X_1} c_1(x_1) + \frac{1}{N_2}\sum_{j_1=1}^{N_2}&\left(\min_{x_2(\xi_2^{j_1})\in X_2(x_1,\xi_2^{j_1})} c_2(x_2(\xi_2^{j_1}),\xi_2^{j_1}) + \cdots\right.\\ &\left.+\frac{1}{N_T}\sum_{j_T=1}^{N_T}\left(\min_{x_T(\xi_{[T]}^{j_{[T]}})\in X_T(x_{T-1}(\xi_{[T-1]}^{j_{T-1}}),\xi_T^{j_{[T]}})} c_T(x_T(\xi_{[T]}^{j_{[T]}}),\xi_T^{j_T})\right)\cdots\right)\end{aligned} \tag{3.27}$$

其中

$$c_t(x_t(\xi_{[t]}^{j_{[t]}}),\xi_t^{j_t})=\frac{1}{2}x_t(\xi_{[t]}^{j_{[t]}})^\top H_t(\xi_t)x_t(\xi_{[t]}^{j_{[t]}})+h_t(\xi_t)^\top x_t(\xi_{[t]}^{j_{[t]}})$$

$$X_1=\{x\in K_1|A_1x_1=b_1\}$$

$$X_t(x_{t-1}(\xi_{[t-1]^{j_{T-1}}}),\xi_t^{j_{[t]}})=\left\{x\in K_t|B_t(\xi_{[t]}^{j_{[t]}})x_{t-1}(\xi_{[t-1]}^{j_{T-1}})+A_t(\xi_{[t]}^{j_{[t]}})x_t(\xi_{[t]}^{j_{[t]}})=b_t(\xi_{[t]}t^{j_{[t]}})\right\},$$

$x_t\in K_t:=\mathbb{R}^{n_t^I}\times\mathbb{R}_+^{n_t^I}$ 表示 x_t 的后 n_t^I 个分量大于等于 0, $j_{[t]}=j_2\cdots j_t, t=2,\cdots,T$. 注意到在本节中, 除了 $x\geq 0$ 以外, 我们不讨论不等式约束的情况.

我们先给出它的增广拉格朗日函数:

$$\begin{aligned}
L_\sigma(x_1,\cdots,x_T;z)=&c_1(x_1)+I_K(x_1)+\frac{1}{N_2}\sum_{j_2=1}^{N_2}\Big(c_2(x_2(\xi_2^{j_2}),\xi_2^{j_2})+I_K(x_2(\xi_2^{j_2}))+\cdots\\
&+\left(\frac{1}{N_T}\sum_{j_T=1}^{N_T}c_T(x_T(\xi_{[T]}^{j_{[T]}}),\xi_T^{j_T})+I_K(x_T(\xi_{[T]}^{j_T}))\right)\cdots\Bigg)+z_1^\top(A_1x_1-b_1)+\frac{\sigma}{2}\|A_1x_1-b_1\|^2\\
&+\sum_{j_2=1}^{N_2}\Big(z_2(\xi_2^{j_2})^\top\Big(B_2(\xi_2^{j_2})x_1+A_2(\xi_2^{j_2})x_2(\xi_2^{j_2})-b_2(\xi_2^{j_2})\Big)\\
&+\frac{\sigma}{2}\|B_2(\xi_2^{j_2})x_1+A_2(\xi_2^{j_2})x_2(\xi_2^{j_2})-b_2(\xi_2^{j_2})\|^2+\cdots\\
&+\sum_{j_T=1}^{N_T}\Big(z_T(\xi_{[T]}^{j_{[T]}})^\top\Big(B_T(\xi_{[T]}^{j_{[T]}})x_{T-1}(\xi_{[T-1]}^{j_{[T-1]}})+A_T(\xi_{[T]}^{j_{[T]}})x_T(\xi_{[T]}^{j_{[T]}})-b_T(\xi_{[T]}^{j_{[T]}})\Big)\\
&+\frac{\sigma}{2}\|B_T(\xi_{[T]}^{j_{[T]}})x_{T-1}(\xi_{[T-1]}^{j_{[T-1]}})+A_T(\xi_{[T]}^{j_{[T]}})x_T(\xi_{[T]}^{j_{[T]}})-b_T(\xi_{[T]}^{j_{[T]}})\|^2\Big)\cdots\Big)
\end{aligned}$$

其中 $I_k(x)$ 是 K 上的指示函数.

令 $x=(x_1,x_2(\xi_{[2]}),\cdots,x_T(\xi_{[T]})), z==(z_1,z_2(\xi_{[2]},\cdots,z_T(\xi_{[T]})))$, 那么拉格朗日函数可以写成

$$L(x;z)=\frac{1}{2}x^\top Gx+h^\top x+z^\top(Ax-b)+\frac{\sigma}{2}\|Ax-b\|^2.$$

其中

$$G=\begin{pmatrix}H_1&&&\\&\frac{1}{N_2}H_2(\tilde{\xi}_{[2]})&&\\&&\ddots&\\&&&\frac{1}{N_2\cdots N_T}H_T(\tilde{\xi}_{[T]})\end{pmatrix},h=\begin{pmatrix}h_1\\\frac{1}{N_2}h_2(\tilde{\xi}_{[2]})\\\vdots\\\frac{1}{N_2\cdots N_T}h_T(\tilde{\xi}_{[T]})\end{pmatrix}$$

$$A=\begin{pmatrix}A_1&&&&&\\B_2(\tilde{\xi}_2)&A_2(\tilde{\xi}_2)&&&&\\&B_3(\tilde{\xi}_{[3]})&A_3(\tilde{\xi}_{[3]})&&&\\&&\ddots&\ddots&&\\&&&B_{T-2}(\tilde{\xi}_{[T-1]})&A_{T-1}(\tilde{\xi}_{[T-1]})&\\&&&&B_{T-1}(\tilde{\xi}_{[T]})&A_T(\tilde{\xi}_{[T]})\end{pmatrix},b=\begin{pmatrix}b_1\\b_2(\tilde{\xi}_2)\\b_3(\tilde{\xi}_{[3]})\\\vdots\\b_{T-1}(\tilde{\xi}_{[T-1]})\\b_T(\tilde{\xi}_T)\end{pmatrix}$$

$H_t(\tilde{\xi}_{[t]}),c_t(\tilde{\xi}_{[t]})$ 的定义如下:

$$H_t(\tilde{\xi}_{[t]})=\begin{pmatrix}H_t(\xi_{[t]}^{\overbrace{1\cdots11}^{t-1}})&&&\\&H_t(\xi_{[t]}^{\overbrace{1\cdots12}^{t-1}})&&\\&&\ddots&\\&&&H_2(\xi_{[2]}^{\overbrace{N_2\cdots N_t}^{t-1}})\end{pmatrix},h_t(\tilde{\xi}_{[t]})=\begin{pmatrix}h_t(\xi_{[t]}^{\overbrace{1\cdots11}^{t-1}})\\h_t(\xi_{[t]}^{\overbrace{1\cdots12}^{t-1}})\\\vdots\\h_t(\xi_{[t]}^{\overbrace{N_2\cdots N_t}^{t-1}})\end{pmatrix}.t=2,\cdots,T.$$

$A(\tilde{\xi}), B(\tilde{\xi}), b(\tilde{\xi})$ 的定义如下：

$$A_t(\tilde{\xi}_{[t]}) = \begin{pmatrix} A_t(\xi_{[t]}^{\overbrace{1\cdots 11}^{t-1}}) & & & \\ & A_t(\xi_{[t]}^{\overbrace{11\cdots 12}^{t-1}}) & & \\ & & \ddots & \\ & & & A_t(\xi_{[t]}^{\overbrace{N_2\cdots N_t}^{t-1}}) \end{pmatrix}, b_t(\tilde{\xi}_{[t]}) = \begin{pmatrix} b_t(\xi_{[t]}^{\overbrace{1\cdots 11}^{t-1}}) \\ b_t(\xi_{[j]}^{\overbrace{1\cdots 12}^{t-1}}) \\ \vdots \\ b_t(\xi_{[j]}^{\overbrace{N_2\cdots N_t}^{t-1}}) \end{pmatrix} .t = 2, \cdots, T.$$

$$B_2(\tilde{\xi}_{[2]}) = \begin{pmatrix} B_2(\xi_{[2]}^{1}) \\ B_2(\xi_{[2]}^{2}) \\ \vdots \\ B_2(\xi_{[2]}^{N_2-1}) \\ B_2(\xi_{[2]}^{N_2}) \end{pmatrix}, B_t(\tilde{\xi}_{[t]}) = \begin{pmatrix} B_t(\xi_{[t]}^{\overbrace{1\cdots 11}^{t-1}}) & & \\ \vdots & & \\ B_t(\xi_{[t]}^{\overbrace{1\cdots 1N_t}^{t-1}}) & & \\ & \ddots & \\ & & B_t(\xi_{[t]}^{\overbrace{N_2\cdots N_{t-1}1}^{t-1}}) \\ & & \vdots \\ & & B_t(\xi_{[t]}^{\overbrace{N_2\cdots N_{t-1}N_t}^{t-1}}) \end{pmatrix}, t = 3, \cdots, T.$$

这里 $\xi_{[t]}^{\overbrace{i_2\cdots i_t}^{t-1}} := (\xi_2^{i_2}, \cdots, \xi_t^{i_t})$.

我们给出问题(3.27)的 KKT 条件:

$$Gx + h + A^{\top}z = 0, Ax - b = 0, x \in K, z \in K^*.$$

下面直接给出算法.

Algorithm 1: 用于多阶段随机二次规划问题的非精确高斯-塞德尔分裂算法

$\tau \in (0, \frac{1+\sqrt{5}}{2})$ 是步长,$\{\tilde{\varepsilon}_k\}$ 是非负序列, $\tilde{\delta} \leq \tilde{\varepsilon}, \delta \leq \tilde{\varepsilon}$ 初始点 x^0 和 z^0. 令 $k \leftarrow 0$

while 不收敛 **do**

  for $i = 1, \cdots, T-1$

$$x_i^{k+\frac{1}{2}} = \arg\min_{x_i \in K} L_\sigma((x_{\leq i-1}^{k+\frac{1}{2}}, x_i, x_{\geq i+1}^{k}); z^k) + \langle \tilde{\delta}_i^k, x_i \rangle,$$
$$\tilde{\delta}_i^k \in \partial_{x_i} L_\sigma((x_{\leq i-1}^{k+\frac{1}{2}}, x_i^{k+\frac{1}{2}}, x_{\geq i+1}^{k}); z^k), \|\tilde{\delta}_i^k\| \leq \tilde{\varepsilon}_k;$$

  for $i = T, T-1, \cdots, 1$

$$x_i^{k+1} = \arg\min_{x_i \in K} L_\sigma(x_{\leq i-1}^{k+\frac{1}{2}}, x_i, x_{\geq i+1}^{k+1}); z^k) + \langle \delta_i^k, x_i \rangle,$$
$$\delta_i^k \in \partial_{x_i} L_\sigma(x_{\leq i-1}^{k+\frac{1}{2}}, x_i^{k+1}, x_{\geq i+1}^{k+1}); z^k), \|\delta_i^k\| \leq \tilde{\varepsilon}_k;$$
$$z^{k+1} = z^k + \tau\sigma(Ax^{k+1} - b)$$

  $k \leftarrow k+1$

end while

由于 $L_\sigma(x; z)$ 有可分结构, 子问题可以进行拆分并行求解. 我们可以简单地考虑, $T = 3$, $N_2 = 2$,

$N_3 = 2$ 的情形: 先给出拉格朗日函数,

$$\begin{aligned}L_\sigma(x_1^{k+1}, x_2, x_3^k; z^k) = c_1(x_1^{k+1}) + z_1^\top(A_1x_1^{k+1} - b_1) + \frac{\sigma}{2}\|A_1x_1^{k+1} - b_1\|^2 \\ + \frac{1}{2}\sum_{i=1}^{2}\big(c_{2i}(x_{2i}) + z_{2i}^T(B_{1i}x_1^{k+1} + A_{2i}x_{2i} - b_{2i}) \\ + \frac{1}{2}\sum_{j=1}^{2}c_{3ij}(x_{3ij}) + z_{3ij}^T(B_{2ij}x_{2i} + A_{3ij}x_{3ij}^k - b_{3ij})\big) \\ + \frac{\sigma}{2}\sum_{i=1}^{2}\big(\|B_{1i}x_1^{k+1} + A_{2i}x_{2i} - b_{2i}\|^2 + \sum_{j=1}^{2}\|B_{2ij}x_{2i} + A_{3ij}x_{3ij}^k - b_{3ij}\|^2\big)\end{aligned}$$

注意到求解

$$\min_{x_2\in K} \mathcal{L}_\sigma(x_1^{k+1}, x_2, x_3^k; z^k)$$

等价于求解

$$\begin{aligned}\min_{x_{2i}\in K} \quad & c_{2i}(x_{2i}) + z_{2i}^T(B_{2i}x_1^{k+1} + A_{2i}x_{2i} - b_{2i}) + \tfrac{1}{2}\textstyle\sum_{j=1}^{2} z_{3ij}^T(B_{3ij}x_{2i} + A_{3ij}x_{3ij}^k - b_{3ij}) \\ & + \tfrac{\sigma}{2}(\|B_{2i}x_1^{k+1} + A_{2i}x_{2i}^k - b_{2i}\|^2 + \textstyle\sum_{j=1}^{2}\|B_{3ij}x_{2i} + A_{3ij}x_{3ij}\|^2 - b_{3ij})\end{aligned}$$

$i = 1, 2$. 同样地, 求解问题

$$\min_{x_3\in K} \mathcal{L}_\sigma(x_1^{k+1}, x_2^{k+1}, x_3, z^k)$$

等价于求解

$$\min_{x_{3ij}\in K} \quad c_{3ij}(x_{3ij}) + z_{3ij}^T(B_{3ij}x_{2i}^{k+1} + A_{3ij}x_{3ij} - b_{2i}) + \tfrac{\sigma}{2}\|B_{3ij}x_{2i}^{k+1} + A_{3ij}x_{3ij}\|^2$$

$i = 1, 2,\ j = 1, 2$.

接下来我们给出关于对称高斯-塞德尔技巧的一些结论 [17]:

设 $s \geq 2, \mathbb{U} = \mathbb{U}_1 \times \cdots \mathbb{U}_s, \mathbb{U}_i i$ 有限维空间, 考虑问题:

$$\min_u \theta(u) + h(u),$$

其中 $\theta : \mathbb{U}_1 \to (-\infty, \infty)$ 是给定的闭凸 proper 函数, $h(u) := \frac{1}{2}\langle u, Hu\rangle - \langle b, u\rangle, H : \mathbb{U} \to \mathbb{R}$, 是对称半正定线性算子, 且有以下分解:

$$H = \begin{pmatrix} 0 & H_{12} & \cdots & H_{1s} \\ & \ddots & \ddots & \vdots \\ & & \ddots & H_{(s-1)s} \\ & & & 0 \end{pmatrix} + \begin{pmatrix} H_{11} & & & \\ & \ddots & & \\ & & \ddots & \\ & & & H_{ss} \end{pmatrix} + \begin{pmatrix} 0 & & & \\ H_{12} & \ddots & & \\ \vdots & & \ddots & \\ H_{s1} & \cdots & H_{s(s-1)} & 0 \end{pmatrix} = H_u + H_d + H_u^*.$$

命题 3.1 (*SGS 分解 [17]*) 设 $\mathcal{H}_{ii}, i = 1, \ldots, s$ 是正定的, 有

$$\hat{H} := H + sGs(\mathcal{H}) = (\mathcal{H}_d + H_s)\,\mathcal{H}_d^{-1}\,(\mathcal{H}_d + H_u^*) > 0$$

对于 $i = s, s-1, \ldots, 2$,

$$\widetilde{u}_i := \arg\min_{u_i}\left\{\theta\left(u_1^-\right) + h\left(u_{\leq i-1}^-, u_i, \widetilde{u}_{\geq i+1}\right) - \langle\widetilde{\delta}_i, u_i\rangle\right\}.$$

对于 $i=1,2,\ldots,s$

$$\begin{cases} u_1^+ := \arg\min\limits_{u_1} \{\theta(u_1) + h(u_1, \widetilde{u}_{\geq 2}) - \langle \delta_1, u_1 \rangle\} \\ u_i^+ := \arg\min\limits_{u_i} \left\{\theta(u_1^+) + h\left(u_{\leq i-1}^+, u_i, \widetilde{u}_{\geq i+1}\right) - \langle \delta_i, u_i \rangle\right\}, \end{cases}$$

等价于

$$u^+ := \arg\min_{u_i} \left\{\theta(u_1) + h(u) + \frac{1}{2}\|u - u^-\|_{SGS(H)}^2 - \langle d\left(\tilde{\delta}, \delta\right) u_i \rangle\right\}, \quad i = 2, \ldots, s$$

其中 $SGS(H) := H_u H_d^{-1} H_u^*, d(\widetilde{\delta}, \delta) := \delta + H_u H_d^{-1}(\delta - \widetilde{\delta})$. 进一步地, $d(\widetilde{\delta}, \delta)$ 满足

$$\left\|\widehat{\mathcal{H}}^{-\frac{1}{2}} d(\widetilde{\delta}, \delta)\right\| \leq \left\|\mathcal{H}_d^{-\frac{1}{2}}(\delta - \widetilde{\delta})\right\| + \left\|\mathcal{H}_d^{\frac{1}{2}}(\mathcal{H}_d + \mathcal{H}_u)^{-1}\widetilde{\delta}\right\|.$$

由命题3.1可知算法 1 等价于:

$$\begin{aligned} x^{k+1} &= \arg\min_{x\in K} L_\sigma(x; z^k) + \frac{1}{2}\|x - x^k\|_{SGS(M)}^2 + \langle d(\tilde{\delta}^k, \delta^k), x \rangle \\ z^{k+1} &= z^k + \tau\sigma(Ax^{k+1} - b) \end{aligned}$$

其中 $M := G + \sigma A^\top A$ 为拉格朗日函数的二次项. 令 $\hat{M} = M + SGS(M), d_x^k \in \partial_x L_\sigma(x^{k+1}; z^k) + \frac{1}{2}\|x - x^k\|_S^2, d_x^k = d(\tilde{\delta}^k, \delta^k)$, 由命题3.1, 有 $\hat{M} \geq 0$,

$$\begin{aligned} \left\|\hat{M}^{-\frac{1}{2}} d_x^k\right\| &\leq \left\|M_d^{\frac{1}{2}}\right\| \left\|\delta^k - \widetilde{\delta}^k\right\| + \left\|M_d^{\frac{1}{2}}(M_d + M_u)^{-1}\right\| \left\|\tilde{\delta_k}\right\| \\ &\leq \left(2\sqrt{1 + \sum_{k=1}^T N_1 \cdots N_k} \left\|M_d^{\frac{1}{2}}\right\| + \sqrt{1 + \sum_{k=1}^T N_1 \cdots N_k} \left\|(M_d + M_u)^{-1}\right\|\right) \tilde{\varepsilon}_k \\ &= \kappa \tilde{\varepsilon}_k. \end{aligned}$$

令 $\varepsilon := \kappa \widetilde{\varepsilon}_k, S = SGS(M)$, 算法 1 可以看成一个带有特殊临近项的增广拉格朗日乘子法 (算法 2), 我们可以通过算法 2 的收敛性给出算法 1 的收敛性, 如下

Algorithm 2: 非精确的增广拉格朗日乘子法 (ALM)

令 $\tau \in (0, \frac{1+\sqrt{5}}{2})$ 是步长, $\{\varepsilon_k\}$ 是非负序列. $\hat{M} = M + S \geq 0$,
选择初始点 x^0 和 z^0. 设 $k \leftarrow 0$.
while 不收敛 **do**

$$\begin{aligned} & x^{k+1} \approx \arg\min_{x\in K} L_\sigma(x; z^k) + \frac{1}{2}\|x - x^k\|_S^2 \\ & d_x^k \in \partial_x\left(L_\sigma(x^{k+1}; z^k) + \tfrac{1}{2}\|x^{k+1} - x^k\|_S^2\right); \|\hat{M}^{\frac{1}{2}} d_x^k\| \leq \varepsilon; \\ & z^{k+1} = z^k + \tau\sigma(Ax^{k+1} - b) \end{aligned}$$

$k \leftarrow k + 1$
end while

定理 3.1 设问题(3.27)的 KKT 解集非空,$\bar{\omega} = (\bar{x}, \bar{z})$ 是解, 序列 $\{(x^k, z^k)\}$ 是由 ALM 产生的, 那么对 $\forall k \geq 0$, 有

$$\begin{aligned} \left\|(x^{k+1}; z^{k+1} - (\bar{x}; \bar{z})\right|^2 - \left\|(x^k; z^k) - (\bar{x}; \bar{z})\right\|^2 \geq -(\frac{2-\tau}{3\tau}\left\|x^{k+1} - x^k\right\|^2 + \left\|x^{k+1} - x^k\right\|^2 \left\|z^{k+1} - z^k\right\|_Q^2 \\ - 2\tau\sigma\left\|d_x^k, z^{k+1} - \bar{z}\right\|) \end{aligned}$$

其中 $Q = \tau\sigma(\frac{1}{2}G + S + \frac{(2-\tau)\sigma}{6}A^\top A)$. 进一步地, 有序列 $\{(x^k, z^k)\}$ 有界, 并收敛到问题(3.27)的 KKT 点.

证明. 算法2是非精确的 ALM 方法的一种特殊情况, 收敛性可以直接由 [18, 命题 3.1] 得到.

4 数值实验

在本节, 我们将用数值试验来验证上述的理论结果. 本节主要分为三部分, 第一和第二部分会考虑一个三阶段随机二次规划问题, 第一部分利用样本均值逼近方法将问题转化为离散的三阶段三阶段随机二次规划问题, 并展示随着样本量的增加, 最优解的收敛性. 第二部分会对样本均值逼近后的问题利用非精确的对称高斯-塞德尔分裂算法和 PHA 算法求解, 第三部分会考虑一个多阶段产品库存问题, 并用本文提出的算法求解.

例 4.1 本例考虑一个三阶段随机二次规划问题

$$\begin{array}{ll}\min\limits_{x_1\in\mathbb{R}^2} & \frac{1}{2}x_1^\top H_1x_1+h_1^\top x_1+\mathbb{E}\left[\min\limits_{x_2(\xi)\in\mathbb{R}^2}\dfrac{1}{2}x_2(\xi_2)^\top H_2(\xi_2)x_2(\xi_2)+h_2(\xi_2)^\top x_2(\xi_2)\right.\\ & \left.+\mathbb{E}\left[\min\limits_{x_3(\xi_{[3]})\in\mathbb{R}^2}\dfrac{1}{2}x_3(\xi_{[3]})^\top H_3(\xi_{[3]})x_3(\xi_{[3]})+h_3(\xi_{[3]})^\top x_3(\xi_{[3]})\right]\cdots\right]\\ s.t. & A_1x_1=b_1\\ & B_2(\xi_2)x1+A_2(\xi_2)x_2(\xi_2)=b_2(\xi_2)\\ & B_3(\xi_{[3]})x_2(\xi_2)+A_3(\xi_{[3]})x_3(\xi_{[3]})=b_3(\xi_{[3]})\end{array}$$

其中随机变量 $\xi_i:\Omega_i\to\Xi_i\subset\mathbb{R}, i=1,2$, 且 Ξ_3 与 Ξ_2 有关. 令 $C=A+I$, 其中 A 在 $[0,1]$ 上随机生成的半正定对称矩阵, 产生即确定不变, I 是单位矩阵, $H_1=2C, H_i=\xi_iC, i=1,2$. $h_1=[1,1.5], b_1=15, A_1=(1,-0.5)$, $h(\xi), B(\xi), A(\xi), b(\xi)$ 的定义如下: $\bar{h}=[0.2,0.7]$, $\bar{B}=[0.3,0.8]$, $\bar{A}=[0.5,0.2]$, $\bar{b}=2$,

$$h_2(\xi_2)=\xi_2\bar{h}, B_2(\xi_2)=\xi_2\bar{B}, A_2(\xi_2)=\xi_2\bar{A}, b_2(\xi_2)=\bar{b}\xi_2$$

$$h_3(\xi_{[3]})=\xi_{[3]}\bar{h}, B_3(\xi_{[3]})=\xi_{[3]}\bar{B}, A_3(\xi_{[3]})=\xi_{[3]}\bar{A}, b_3(\xi_{[3]})=\bar{b}\xi_{[3]}$$

可以直接验证上述问题满足定理 2.2 的假设. 因此, 我们考虑如下样本均值逼近问题:

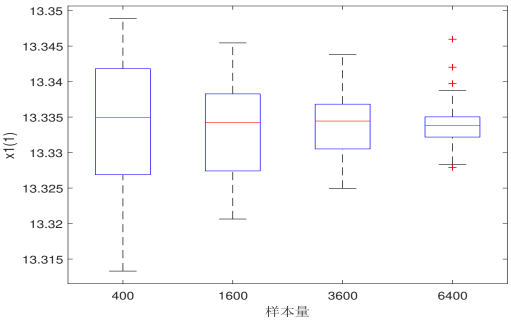
$$\begin{array}{ll}\min\limits_{x_1\in\mathbb{R}^2} & \frac{1}{2}x_1^\top H_1x_1+h_1^\top x_1+\frac{1}{N_2}\sum_{i=1}^{N_2}\left(\min\limits_{x_2(\xi_2^i)\in\mathbb{R}^2} x_2(x_1,\xi_2^i)^\top H_2(\xi_2^i)x_2(x_1,\xi_2^i)+h_2(\xi_2^i)^\top x_2(x_1,\xi_2^i)\right.\\ & \left.+\frac{1}{N_3}\sum_{j=1}^{N_3}\left(\min\limits_{x_3(\xi_{[3]}^{ij})\in\mathbb{R}^2}\dfrac{1}{2}x_3(\xi_{[3]}^{ij})^\top H_3(\xi_[3]^{ij})x_3(\xi_{[3]}^{ij})+h_3(\xi_[3]^{ij})^\top x_3(\xi_{[3]}^{ij})\right)\right)\\ s.t. & A_1x_1=b_1\\ & B_2(\xi_2^i)x_1+A_2(\xi_2^i)x_2(\xi_2^i)=b_2(\xi_2^i)\\ & B_3(\xi_{[3]}^{ij})x_2(\xi_2^i)+A_3(\xi_{[3]}^{ij})x_3(\xi_{[3]}^{ij})=b_3(\xi_{[3]}^{ij})\\ & i=1,\cdots,N_2; j=1,\cdots,N_3.\end{array}$$

令 ξ_2 服从均匀分布 $[1,50]$, ξ_3 服从均匀分布 $[\xi_2+1,\xi_2+50]$, 对于样本量 $N_2=N_1$, $N=N_1N_2=400,1600,3600,6400$, 分别取 ξ_2,ξ_3 的样本 $\{\xi_2^i\},\{\xi_3^i\}$, 随机抽样 *20* 次, 并分别求解问题. 实验结果验证了第 *2* 章的样本均值逼近方法的收敛性. 图 *1* 和图 *2* 仅列出了 x_1 的数值结果.

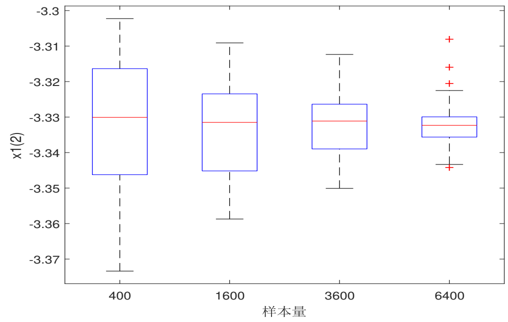
例 4.2 本例仍然考虑例 4.1 的三阶段随机二次规划问题模型:

$$\begin{array}{ll}\min\limits_{x_1\in\mathbb{R}^n} & \frac{1}{2}x_1^\top H_1x_1+h_1^\top x1+\frac{1}{N_2}\sum_{i=1}^{N_2}\left(\min\limits_{x_2(\xi_2^i)\in\mathbb{R}^n} x_2(\xi_2^i)^\top H_2(\xi_2^i)x_2(\xi_2^i)+h_2(\xi_2^i)^\top x_2(\xi_2^i)\right.\\ & \left.+\frac{1}{N_3}\sum_{j=1}^{N_3}\left(\min\limits_{x_3(\xi_{[3]}^{ij})\in\mathbb{R}^n}\dfrac{1}{2}x_3(\xi_{[3]}^{ij})^\top H_3(\xi_[3]^{ij})x_3(\xi_{[3]}^{ij})+h_3(\xi_[3]^{ij})^\top x_3(\xi_{[3]}^{ij})\right)\right)\\ s.t. & A_1x_1=b_1\\ & B_2(\xi_2^i)x1+A_2(\xi_2^i)x_2(\xi_2^i)=b_2(\xi_2^i)\\ & B_3(\xi_{[3]}^{ij})x_2(\xi_2^i)+A_3(\xi_{[3]}^{ij})x_3(\xi_{[3]}^{ij})=b_3(\xi_{[3]}^{ij})\\ & i=1,\cdots,N_1; j=1,\cdots,N_2.\end{array}$$

下面, 对于不同的样本量, 我们分别用非精确的对称高斯-塞德尔分裂算法和 *PHA* 算法进行求解, 对它们的 *CPU* 时间进行了比较.



(a) 图 1



(b) 图 2

下面先给出问题模型中相关系数的定义, 随机变量 $\xi_i : \Omega_i \to \Xi_i \subset \mathbb{R}, i = 1, 2$, 且 Ξ_3 与 Ξ_2 有关. 令 $H_i = A + I$,A 在 $[0, 1]$ 上随机生成的半正定对称矩阵,I 是单位矩阵, $b_1 = 15$, h_1, A_1 是在 $[0, 1]$ 随机产生的,$h(\xi), B(\xi), A(\xi), b(\xi)$ 的定义如下: $\bar{h} \in R^n$ 是在 $[0, 2]$ 随机产生的,$\bar{B} \in R^n$, $\bar{A} \in R^n$ 均是在 $[0, 1]$ 随机产生的, 对于每次测试, 产生即确定, $\bar{b} = 2$,

$$h_2(\xi_2) = \xi_2\bar{h}, B_2(\xi_2) = \xi_2\bar{B}, A_2(\xi_2) = \xi_2\bar{A}, b_2(\xi_2) = \bar{b}\xi_2$$

$$h_3(\xi_{[3]}) = \xi_{[3]}\bar{h}, B_3(\xi_{[3]}) = \xi_{[3]}\bar{B}, A_3(\xi_{[3]}) = \xi_{[3]}\bar{A}, b_3(\xi_{[3]}) = \bar{b}\xi_{[3]}$$

令 $r = Ax - b, \eta = Hx + c + A^\top z, e = max\{r, \eta\}$, 当 $e \leq 10^{-3}$ 时, 对称高斯-塞德尔分裂算法迭代终止; *PHA* 方法在序列收敛时, 迭代终止.

实验里取 $n = 10$, 随机变量 ξ_2 服从均匀分布 $[0, 1]$, ξ_3 服从均匀分布 $[\xi_2, \xi_2 + 1]$. 步长 $\tau = 1.618$, 罚参数 $\sigma = 2$. 对于样本量 $N_1 = N_2$, $N = N_1N_2 = 100, 400, 900$, 每种样本量做 *10* 次测试取平均数. 同时, 再对求解时间进行比较, 结果如下.

Table 1: 时间/秒

N	100	400	900	1600
对称高斯-塞德尔分裂算法	5.0004	35.7911	89.8249	249.8039
PHA	8.5189	64.6510	246.5142	446.9530

由上表可知, 对称高斯-塞德尔分裂算法较 *PHA* 算法更有优势.

例 4.3 本例考虑多阶段商品库存问题. 假设零售商销售某商品 A, 商品 A 的进价为 c_1, 销售过程分为进货和售卖. 在第一天, 零售商只需要进货, 前一天未卖完的商品是第二天的库存, 会产生库存成本 h, 最后一天零售商不需要进货, 直接售卖之前的库存. 由于市场情况的不确定性, 每次进货数量, 商品需求量和库存数量都是不确定的. 求解最优决策使得利润最大化.

设第 t 阶段未来市场的不确定情景 $\xi_t : \Omega_t \to \Xi_t \subset \mathbb{R}^1$ 是随机变量, $x_t(\xi_t)$ 为进货数量,$y_t(\xi_t)$ 为商品需求, $z_t(\xi_t)$ 为库存数量, 下面先给出第 T 阶段的模型:

$$\begin{aligned} \min_{y_T(\xi_{[T]}), z_T(\xi_{[T]})} \quad & h_T z_T(\xi_{[T]}) - y_T(\xi_{[T]}) p_t(y_T(\xi_{[T]}), \xi_T) \\ s.t. \quad & z_T(\xi_{[T]}) = x_{T-1}(\xi_{[T-1]}) + z_{T-1}(\xi_{[T-1]}) - y_T(\xi_{[T]}) \\ & 0 \leq z_T(\xi_{[T]}), 0 \leq y_T(\xi_{[T]}) \end{aligned}$$

其中 $p_t(y_t(\xi_t), \xi_t) = \alpha(\xi_t - y_t(\xi_t))$ 是商品价格.

第 T 阶段由于不再进货, 目标函数里没有商品的进货成本, 下面给出第 t 阶段的模型:

$$\begin{array}{cl}\min\limits_{x_t(\xi_{[t]}),y_t(\xi_{[t]}),z_t(\xi_{[t]})} & W_t(x_t(\xi_{[t]}),y_t(\xi_{[t]}),z_t(\xi_{[t]}),\xi_t)+\sum_{j_{t+1}}^{N_{t+1}} Q_{t+1}(x_t(\xi_{[t+1]}^{j_{[t+1]}}),y_t(\xi_{[t+1]}^{j_{[t+1]}},z_t(\xi_{[t+1]}^{j_{[t+1]}}),\xi_{t+1}^{j_{t+1}}) \\ s.t. & 0\le x_t(\xi_{[t]})\le v_t \\ & z_t(\xi_{[t]})=x_{t-1}(\xi_{[t-1]})+z_{t-1}(\xi_{[t-1]})-y_t(\xi_{[t]}) \\ & 0\le z_t(\xi_{[t]}),0\le y_t(\xi_{[t]})\end{array}$$

在目标函数中,$W_t(x_t(\xi_t),y_t(\xi_t),z_t(\xi_t),\xi_t)=c_1x_t(\xi_{[t]})+h_tz_t(\xi_{[t]})-y_t(\xi_{[t]})p_t(y_t(\xi_{[t]}),\xi_t)$ 为损失函数,

$$Q_{t+1}(x_t(\xi_{[t+1]}^{j_{[t+1]}}),y_t(\xi_{[t+1]}^{j_{[t+1]}},z_t(\xi_{[t+1]}^{j_{[t+1]}}),\xi_{t+1}^{j_{t+1}})$$

是 $t+1$ 阶段的最优值. 约束 $x_t(\xi_{[t]})\le v_t$ 表示第 t 阶段的商品采购约束, 第二个等式约束表示第 t 阶段商品库存的平衡关系. 最后我们考虑第 *1* 阶段:

$$\begin{array}{cl}\min\limits_{x_1} & c_1x_1+\sum_{j_1=1}^{N_2} Q_{2i}(x_1,\xi_2^{j_1}) \\ s.t. & 0\le x_1\le v_1\end{array}$$

第一阶段零售商只需进货.

由此可以得到一个多阶段随机二次规划问题:

$$\begin{array}{cl}\min\limits_{x_1} & c_1x_1+\dfrac{1}{N_2}\sum\limits_{j_2=1}^{N_2}\left(\min\limits_{x_2(\xi_2^{j_2}),y_2(\xi_2^{j_2}),z_2(\xi_2^{j_2}),\xi_2^{j_2}} c_1x_2(\xi_2)+h_2z_2(\xi_2)-y_2(\xi_2)p_2(y_2(\xi_2),\xi_2)+\cdots\right. \\ & \left.+\dfrac{1}{N_T}\sum\limits_{j_T=1}^{N_T}\left(\min\limits_{y_T(\xi_{[T]}^{j_T}),z_T(\xi_{[T]}^{j_T}} h_Tz_T(\xi_{[T]}^{j_T})-y_T(\xi_{[T]}^{j_T})p_t(y_T(\xi_{[T]}^{j_T}),\xi_T^{j_T})\right)\cdots\right) \\ s.t. & z_t(\xi_{[t]})=x_{t-1}(\xi_{[t-1]})+z_{t-1}(\xi_{[t-1]})-y_t(\xi_{[t]}) \\ & 0\le x_{t-1}\le v_{t-1} \\ & 0\le z_t(\xi_{[t]}^{j_t}),0\le y_t(\xi_{[t]}^{j_t}), \\ & t=2,\cdots,T.\end{array}$$

本例中进货决策是在每一次随机变量实现之前决定的, 而销售量和库存量是跟进货决策和随机变量相关的, 因此进货决策是划分阶段数的关键决策. 由于一般的多阶段问题难以计算, 在数值实验里我们考虑两阶段和三阶段的情形. 两阶段是指第一天进货决策已经确定, 第二天和第三天不再进货, 只需要考虑销售和库存决策; 三阶段是指第一天进货决策确定下来, 第二天仍然需要进货, 考虑销售和库存决策, 第三天不再进货, 只需要考虑销售和库存决策.

令 $u=(x,y,z)$, 三阶段随机二次规划问题如下:

$$\begin{array}{cl}\min\limits_{u_1} & b_1^\top u_1+\frac{1}{N_2}\sum\limits_{j_2=1}^{N_2}\left(\min\limits_{u_2(\xi_2^{j_2}),\xi_2^{j_2}}\dfrac{1}{2}u_2(\xi_2^{j_2})^\top Hu_2(\xi_2^{j_2})+b_2(\xi_2^{j_2})^\top u_2(\xi_2^{j_2})\right. \\ & \left.+\frac{1}{N_3}\sum\limits_{j_3=1}^{N_3}\min\limits_{y_3(u_3(\xi_{[3]}^{j_3}}\dfrac{1}{2}u_3(\xi_{[3]}^{j_3})^\top Hu_3(\xi_{[3]}^{j_3})+b_3(\xi_{[3]}^{j_3})^\top u_3(\xi_{[3]}^{j_3})\right) \\ s.t. & z_2(\xi_2)=x_1-y_2(\xi_2) \\ & z_3(\xi_{[3]})=x_2(\xi_2)+z_2(\xi_2)-y_3(\xi_{[3]}) \\ & 0\le x_1\le v_1,\le x_2\le v_1 \\ & 0\le z_2(\xi_2),0\le z_3(\xi_{[3]}) \\ & 0\le y_2(\xi_2),0\le y_3(\xi_{[3]})\end{array}$$

其中 $b_1=(c_1,0,0),b_2=(c_1,-\alpha\xi_2,h_2),b_3=(0,-\alpha\xi_3,h_3)$, $H=\begin{pmatrix}0&0&0\\0&2\alpha&0\\0&0&0\end{pmatrix}$.

两阶段问题模型与三阶段问题模型类似, 系数和约束略有差别, 其中 $b_2 = (0, -\alpha\xi_2, h_2)$, 库存约束是 $z_3(\xi_{[3]}) = z_2(\xi_2) - y_3(\xi_{[3]})$.

在数值实验里, 随机变量 ξ_2 服从均匀分布 $\Xi_2 = [300, 350]$, ξ_3 服从均匀分布 $\Xi_3 = [\xi_2, \xi_2 + 50]$, $\alpha = 0.1, c_1 = 20$, 因为最后阶段的库存是永久库存, 所以取 $h_2 = 1, h_3 = 20$, 对样本量 $N_2 = N_3$, $N = N_2N_3 = 100, 200, 900$, 我们分别求解了两阶段模型和三阶段问题的最优值. 每个样本量均做了 *10* 次抽样做平均, 结果如表 *3*. 由表格 *3* 我们可以知道, 利润随着模型的阶段数增加. 三阶段模型的利润总是

N	10	20	30
两阶段	757.4473	781.9463	771.2165
三阶段	973.4980	982.8241	965.0519

Table 2: 最优利润

比两阶段模型的利润高. 且在市场的不确定性比较高的情况下, 更适合使用多阶段模型.

5 结论

在本文中我们以三阶段随机二次规划问题为例, 研究了多阶段随机二次规划问题的性质, 并在样本的独立同分布和阶段独立性假设不成立的情况下, 研究了其样本均值逼近方法的指数收敛速率, 该结果可以推广到一半的凸多阶段随机规划问题上. 并且在样本的独立同分布和阶段独立性假设成立的情况下, 我们的结果较 [9, 定理 2] 的结果更紧致一些 (见注2.4). 我们进一步利用多阶段随机二次规划问题的性质和一类带有特殊临近项的增广 Lagrange 方法, 给出了一个针对只含有等式约束和 $x \geq 0$ 约束的多阶段随机二次规划问题的非精确的对称高斯-塞德尔分裂算法. 最后我们用数值结果 1) 验证了样本均值逼近方法的收敛性; 2) 与 PHA 算法比较, 展示了非精确的对称高斯-塞德尔分裂算法的优势; 3) 以一个两阶段和三阶段问题为例, 展示了在随机环境下, 多阶段随机优化确实可以更好的利用未来的不确定性, 提高利润. 我们将进一步考虑针对更广泛多阶段随机规划问题的有效分裂算法.

References

[1] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer, Berlin, 2011.

[2] A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2014.

[3] G. B. Dantzig. Linear programming under uncertainty. *Manage. Sci.*, 1:197–281, 1955.

[4] G. Pflug and Pichler A. *Multistage stochastic optimization*. Springer, Switzerland, 2014.

[5] A. Georghiou, W. Wiesemann, and D. Kuhn. Generalized decision rule approximations for stochastic programming via liftings. *Math. Program.*, 152:301–338, 2015.

[6] A. Georghiou, W. Wiesemann, and D. Kuhn. The decision rule approach to optimisation under uncertainty: methodology and applications. *Comput. Manag. Sci.*, 16:545–576, 2019.

[7] A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Math. Methods Oper. Res.*, 58:57–68, 2003.

[8] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Math. Program.*, 118:371–406, 2009.

[9] A. Shapiro. On complexity of multistage stochastic programs. *Oper. Res. Lett.*, 34:1–8, 2006.

[10] M. Reaiche. A note on sample complexity of multistage stochastic programs. *Oper. Res. Lett.*, 44(4):430–435, 2016.

[11] R. P. Liu. On feasibility of sample average approximation solutions. *SIAM J. Optim.*, 30(3):2026–2052, 2020.

[12] J. Jiang and S. Li. On complexity of multistage stochastic programs under heavy tailed distributions. *Oper. Res. Lett.*, 49(2):265–269, 2021.

[13] R. T. Rockafellar and R. J-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Math. Oper. Res.*, 16:119–147, 1991.

[14] J. R. Birge. Decomposition and partitioning methods for multistage stochastic linear programs. *Math. Oper. Res.*, 33:989–1007, 1985.

[15] A. Shapiro. Decomposition and partitioning methods for multistage stochastic linear programs. *Eur. J. Oper. Res.*, 109:63–72, 2011.

[16] S. Arpón, T. Homem-de Mello, and B. K. Pagnoncelli. An admm algorithm for two-stage stochastic programming problems. *Ann. Oper. Res.*, 286:559–582, 2020.

[17] X. Li, D. Sun, and K-C Toh. QSDPNAL: A two-phase augmented lagrangian method for convex quadratic semidefinite programming. *Math. Program. Comput.*, 10(4):703–743, 2018.

[18] L. Chen, X. Li, and D. Sun. On the equivalence of inexact proximal alm and admm for a class of convex composite programming. *Math. Program*, 185:111–161, 2019.

[19] H. Sun and H. Xu. Convergence analysis for distributionally robust optimization and equilibrium problems. *Math. Oper. Res.*, 41(2):377–401, 2016.

[20] Z. Luo and P. Tseng. Pertutbation analysis of a condition number for linear systems. *SIAM J. Matrix Anal. Appl.*, 15:636–660, 1994.

[21] B. Frédéric and A. Shapiro. *Perturbation analysis of optimization problems.* Springer Science & Business Media, 2013.

[22] Nicolas Hadjisavvas, Sándor Komlósi, and Siegfried S Schaible. *Handbook of generalized convexity and generalized monotonicity*, volume 76. Springer Science & Business Media, New York, 2006.

[23] H. Xu. Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *J. Math. Anal. Appl.*, 368(2):692–710, 2010.

[24] A. Shapiro and H. Xu. Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. *Optimization*, 57:395–418, 2008.

致谢

感谢孙海琳教授对本工作的悉心指导，从论文选题、理论学习、文献研读、研究探索到论文撰写，都得到了孙老师的无私帮助和耐心讲解。感谢我的数学老师唐智逸在论文写作、参赛要求等方面给予的指导。感谢蔡邢菊教授、蒋杰博士和杨郭梅霞对本工作的建议和有帮助的讨论。本项工作与孙海琳教授 2019 年开始执行的国家自然科学基金项目《两阶段随机变分不等式的研究及其在两阶段随机非合作博弈问题中的应用》部分内容相关，在此一并致谢。