arXiv:2303.00555v1 [math.OC] 1 Mar 2023

具有二次随机性的机会约束规划：基于高斯混合分布的统一方法

Xiaochuan Pang * Shushang Zhu † Zhaolin Hu ‡

2023年3月2日

摘要

本文研究随机变量服从多元高斯混合分布（GMD）时，对其二次型施加机会约束的随机规划。在一些温和条件下，本文证明了这类二次随机性的渐近分布是一元 GMD。这一发现有助于将机会约束规划转化为更易处理的问题，并据此提出一种有效的分支定界算法，利用问题的特殊结构搜索全局最优解。此外，研究表明，通过限制多元 GMD 各分量协方差矩阵的条件数，可以降低用其相应的渐近分布近似二次随机性所产生的误差。最后，本文还进行了数值模拟，以验证这一灵活且统一的方法的有效性。

关键词：机会约束规划、高斯混合分布、二次随机性、渐近分布、分支定界

1 引言

机会约束规划（CCP）是一种典型的随机规划，用于处理决策中的随机不确定性。自 Charnes 等人（1958）、Miller 和 Wagner（1965）以及 Pr´ekopa（1970）的早期研究以来，CCP 已广泛应用于工程、金融及其他领域。例如，在金融领域，最大化投资组合收益并满足风险价值（VaR）约束这一常见问题，实际上就是 CCP 的一个特例；其中，VaR 是一种以损失分布的分位数为特征的风险度量（例如，参见 Jorion（2007））。

*中山大学商学院，中国广州 510275。电子邮箱：pangxch@mail2.sysu.edu.cn。

†通讯作者。中山大学商学院，中国广州 510275。电子邮箱：zhuss@mail.sysu.edu.cn。

‡同济大学经济与管理学院，中国上海 200092。电子邮箱：rus-sell@tongji.edu.cn

然而，正如 Nemirovski 和 Shapiro（2006）所指出的，即使机会约束具有线性结构，该问题仍可能是非凸的。因此，CCP 通常在计算上难以处理，这阻碍了该模型在实践中的应用。

一般而言，机会约束规划研究的是关于某种随机性的概率约束，该随机性是若干随机变量和决策变量的函数。由于机会约束规划难以求解，先前文献主要考虑随机性取随机变量的线性形式。Shapiro 等人（2009）讨论了随机性的形式和随机变量分布满足何种条件时，机会约束规划是凸规划。Henrion 和 Moller（2012）给出了在可能为奇异的多元高斯分布下线性机会约束的梯度公式。还有一类文献通过不同方法定义随机变量分布的不确定集，研究线性分布鲁棒机会约束规划或其变体——风险价值优化；这些方法包括 El Ghaoui 等人（2003）提出的已知一阶矩和二阶矩信息、Calafiore 和 El Ghaoui（2006）提出的径向对称分布集，以及 Hanasusanto 等人（2017）提出的已知期望与离散函数。

本文考虑 CCP 中二次随机性的常见形式。更具体地说，引入如下二次随机性：

$$c(\boldsymbol{\xi}, \boldsymbol{x}) = \frac{1}{2}\boldsymbol{\xi}^\top A(\boldsymbol{x})\boldsymbol{\xi} + \boldsymbol{a}(\boldsymbol{x})^\top \boldsymbol{\xi} + a(\boldsymbol{x}), \tag{1}$$

其中，$\boldsymbol{\xi} \in \mathbb{R}^m$ 是服从概率分布 p^*的随机向量，$\boldsymbol{x} \in \mathbb{R}^n$ 是决策向量，$A: \mathbb{R}^n \to \mathbb{R}^{m\times m}$ 是实对称矩阵，$\boldsymbol{a}: \mathbb{R}^n \to \mathbb{R}^m$ 以及 $a: \mathbb{R}^n \to \mathbb{R}$。我们考虑如下 CCP：

$$\begin{aligned} \text{QCCP} \quad & \min_{\boldsymbol{x}\in\mathscr{X}} \quad \rho(\boldsymbol{x}) \\ & \text{s.t.} \quad \Pr_{\sim p^*}\{c(\boldsymbol{\xi}, \boldsymbol{x}) \le 0\} \ge 1-\alpha, \end{aligned}$$

其中，$\mathscr{X} \subseteq \mathbb{R}^n$ 是非空凸集，$\alpha \in \in (0,1)$，$\rho: \mathbb{R}^n \to \mathbb{R}$，$c: \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$，且 $\Pr_{\sim p^*}$ 表示关于 p^*所取的概率。

文献中已有一些关于 QCCP 特殊情形的相关研究。Zymler 等人（2013）考虑了一种最坏情形机会约束，其中 p^* 属于具有相同一阶矩和二阶矩的所有概率分布构成的集合。他们证明，如果 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 是$\boldsymbol{\xi}$的二次函数，那么最坏情形 QCCP 等价于最坏情形 CVaR 问题。Cui 等人（2013）研究了均值-VaR 框架下的期权投资组合选择问题，这是 QCCP 的一个特殊情形。在他们的模型中，$\boldsymbol{\xi}$ 被假定为服从正态分布。基于广泛使用的 Delta-Gamma-Normal 方法（可参见 Hull（2009）），他们进一步假设$c(\boldsymbol{\xi}, \boldsymbol{x})$服从正态分布，进而可将该问题重新表述为二阶锥规划（SOCP）。Zhu 等人（2020）证明，在温和条件下，如果 $\boldsymbol{\xi}$ 服从正态分布，那么 $c(\boldsymbol{\xi}, \boldsymbol{x})$的渐近分布仍为正态分布，这为 Delta-Gamma-Normal 方法提供了理论依据。Kishida 和 Nagahara（2023）研究了一个结构与 QCCP 类似的系统控制问题。他们假设 p^* 的分布——

——属于一个已知均值和方差的概率集合，并考虑最坏情形机会约束；为便于求解，他们进一步用最坏情形 CVaR 对其进行近似。显然，如何确定 ξ 的分布对求解 QCCP 至关重要。

指定基础随机因素分布的过程通常称为输入建模，它不仅会影响求解方法的选择，也会直接影响实际应用的合理性。因此，开发一种将输入建模与机会约束规划（CCP）相结合的统一方法，使其能够适用于尽可能多的情形，是有意义且必要的。此外，在许多应用领域，通常有大量关于随机因素的观测数据，因此，通过学习数据来估计真实分布十分重要。混合分布模型最早由 Pearson（1894）应用于生物学领域，是一种灵活的数据学习方法。混合分布最吸引人的特点是其拟合能力。混合分布的密度函数是有限个密度函数的加权和，能够以足够高的精度逼近许多密度函数。关于混合模型的更多性质，可参见 McLachlan 和 Peel（2000）。理论上，Wilson（2000）证明，任意单变量可积分布函数都可以在任意精度范围内由高斯混合分布（GMD）逼近。具有偏度、厚尾和多峰性等常见乃至不寻常特征的分布，都可以由 GMD 良好地建模（Marron 和 Wand（1992））。

近年来，机会约束规划与 GMD 的结合已在理论和应用方面得到研究。Chen 等人（2018）提出了一种凸近似方法，用于求解鲁棒机会约束规划，其中随机参数采用 GMD 建模，分量的权重则位于基于矩估计的不确定集内。Hu 等人（2022）研究了若干典型形式的联合机会约束规划和单一机会约束规划，其中随机参数采用 GMD 建模。他们分别提出了一阶方法和空间分支定界法，以搜索局部解和全局解。在分布式发电领域，Chen 等人（2022）假设风电出力服从 GMD，并求解机会约束规划，以确定发电机的最优电力调度。Ren 等人（2022）使用 GMD 对障碍物不确定状态的多模态行为进行建模，并通过机会约束规划保障轨迹规划的安全性。

有一些通用的情景近似方法可用于求解 QCCP。Calafiore 和 Campi（2005）以及 Campi 和 Garatti（2008）证明了能够以较高置信水平保证情景方法的最优解满足机会约束所需的样本数量下界。Campi 和 Garatti（2011）研究了样本剔除问题，并进一步考察了样本数量下界。Luedtke 和 Ahmed（2008）用经验违约概率替代机会约束中的违约概率，并将其重新表述为混合整数规划。本文不同于这些方法，基于 GMD 推导了一种求解 QCCP 的统一方法。具体而言，我们使用多变量高斯混合分布对随机变量的分布建模，并证明在某些温和条件下，这些随机变量的二次型的渐近分布服从一元 GMD。进一步地，我们利用这一结果将 QCCP 重新表述为

具有解析形式的确定性程序，进而可应用分支定界算法或 Hu 等人（2022）提出的一阶方法求得最优解。此外，我们还揭示了收敛速率与其他因素之间的关系，其中重点讨论了 GMD 各分量协方差矩阵的条件数。为降低渐近近似误差，我们建议使用条件数约束GMD对随机变量的分布建模。此外，我们证明，这种条件数约束GMD仍可在任意精度下近似任意可积密度函数。本文的主要贡献如下：

. 本文提出了一种基于 GMD 求解具有二次随机性的机会约束规划的统一方法。尽管二次随机性的真实分布没有解析形式，但结合渐近分布逼近与分支定界算法，可以获得近似的全局解。

. 本文揭示了影响二次随机性渐近分布收敛速率的因素。研究证明，采用条件数约束GMD能够加快二次随机性的收敛速度，同时不会损失拟合能力和灵活性。

. 所提出的方法可应用于数据丰富的众多领域。与其他常用分布不同，GMD 能够从数据中学习分布的特征，例如偏度、厚尾和多峰形状。因此，我们的模型尤其适用于难以描述数据特征的领域。

本文其余部分安排如下。第2节讨论服从 GMD 的随机变量的二次型的渐近分布，并探讨影响收敛速率的相关因素；在此基础上，研究如何在将渐近分布用作真实分布的代理时减少近似误差。第3节对 QCCP 进行重新表述，并在假设 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 是线性函数的条件下，应用分支定界算法求解该优化问题的全局解。第4节通过数值实验验证理论发现，并检验所提方法的有效性。第5节给出总结。

2 GMD 下二次随机性的渐近性质

鉴于 GMD 具有普适逼近能力，本文假设 $\boldsymbol{\xi}$ 服从含有 K 个高斯分量的 GMD。具体而言，$\boldsymbol{\xi}$ 的密度函数为

$$p^*(\boldsymbol{z}) = \sum_{i=1}^{K} \pi_i p(\boldsymbol{z}|\boldsymbol{\mu}_i, \Sigma_i), \tag{2}$$

其中，混合权重 $\pi_i > 0$, $\sum_{i=1}^{K} \pi_i = 1$ 和 $p(\boldsymbol{z}|\boldsymbol{\mu}_i, \Sigma_i)$（简称为 p_i）表示均值为 $\boldsymbol{\mu}_i$ 、协方差矩阵为 Σ_i 的高斯密度函数。在此设定下，

我们首先讨论 $\boldsymbol{\xi}$ 的二次型的渐近分布。然后分析该渐近分布的收敛速率。最后，讨论影响渐近分布与真实分布之间误差的因素，并在此基础上进一步探讨如何减少近似误差。

2.1 二次随机性的渐近分布

将证明，在一些温和条件下，服从GMD的随机向量的二次型的渐近分布仍为GMD。下文首先从服从GMD的随机向量的二次型的一阶矩和二阶矩开始分析。

命题1. 设 $\boldsymbol{\xi}$ 服从由式(2)定义的GMD。则对于固定的决策向量$\boldsymbol{x}$，由式(1)定义的 $c(\boldsymbol{\xi},\boldsymbol{x})$的均值和方差为

$$\begin{aligned} E_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \sum_{i=1}^{K}\pi_i E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) \\ V_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \sum_{i=1}^{K}\pi_i V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) + \sum_{1\le i<j\le K}\pi_i\pi_j\left(E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) - E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x}))\right)^2, \end{aligned}$$

其中， $E_{p_i}(\boldsymbol{\xi},\boldsymbol{x})$和 $V_{p_i}(\boldsymbol{\xi},\boldsymbol{x})$分别为关于概率分布 p_i的均值和方差，具体为

$$\begin{aligned} E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \frac{1}{2}\mathrm{tr}\left(A(\boldsymbol{x})\Sigma_i\right) + \frac{1}{2}\boldsymbol{\mu}_i^\top A(\boldsymbol{x})\boldsymbol{\mu}_i + \boldsymbol{a}(\boldsymbol{x})^\top\boldsymbol{\mu}_i + a(\boldsymbol{x}) \\ V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \frac{1}{2}\mathrm{tr}\left(\left(A(\boldsymbol{x})\Sigma_i\right)^2\right) + \left(A(\boldsymbol{x})\boldsymbol{\mu}_i + \boldsymbol{a}(\boldsymbol{x})\right)^\top\Sigma_i\left(A(\boldsymbol{x})\boldsymbol{\mu}_i + \boldsymbol{a}(\boldsymbol{x})\right) \end{aligned}$$

分别为。此处，tr(·)表示矩阵的迹。

证明。见附录A。 □

现在我们转向研究 $c(\boldsymbol{\xi},\boldsymbol{x})$ 的渐近分布，其中 $\boldsymbol{\xi}$ 服从 GMD。为便于重新表述，我们省略决策向量，并分别将 $A(\boldsymbol{x})$、 $\boldsymbol{a}(\boldsymbol{x})$ 和 $a(\boldsymbol{x})$ 记为 A、 $\boldsymbol{a}$ 和 a。

不失一般性，假设每个 $\Sigma_i, i\in\{1,\cdots,K\}$ 都是非奇异的。然后我们将 Σ_i 分解为 $\Sigma_i = \Sigma_i^{\frac{1}{2}}\Sigma_i^{\frac{1}{2}}$，将 $\Sigma_i^{\frac{1}{2}}A\Sigma_i^{\frac{1}{2}}$ 分解为 $\Sigma_i^{\frac{1}{2}}A\Sigma_i^{\frac{1}{2}} = D_i\Lambda_i D_i^\top$，其中 Λ_i 是对角矩阵，其对角元素为 λ_{i1}, $\cdots,\lambda_{im}$，而 D_i 是相应的正交矩阵。假设 Λ_i 仅前 h_i 个对角元素非零，即 $\lambda_{i1},\lambda_{i2}\cdots$, $\lambda_{ih_i}\neq 0\,(0<h_i\le m)$。

于是有

$$
\begin{aligned}
c(\boldsymbol{\xi},\boldsymbol{x}) &= \frac{1}{2}\hat{\boldsymbol{\xi}}_i^\top \Sigma_i^{\frac{1}{2}} A \Sigma_i^{\frac{1}{2}} \hat{\boldsymbol{\xi}}_i + \boldsymbol{a}^\top \Sigma_i^{\frac{1}{2}} \hat{\boldsymbol{\xi}}_i + a && \left(\hat{\boldsymbol{\xi}}_i = \Sigma_i^{-\frac{1}{2}} \boldsymbol{\xi}\right) \\
&= \frac{1}{2}\tilde{\boldsymbol{\xi}}_i^\top \Lambda_i \tilde{\boldsymbol{\xi}}_i + \boldsymbol{b}_i^\top \tilde{\boldsymbol{\xi}}_i + a && \left(\tilde{\boldsymbol{\xi}}_i = D_i^\top \hat{\boldsymbol{\xi}}_i, \boldsymbol{b}_i = D_i^\top \Sigma_i^{\frac{1}{2}} \boldsymbol{a}\right) \\
&= \frac{1}{2}\sum_{j=1}^{h_i} \lambda_{ij} \tilde{\xi}_{ij}^2 + \sum_{j=1}^{m} b_{ij} \tilde{\xi}_{ij} + a \\
&= \frac{1}{2}\sum_{j=1}^{h_i} \lambda_{ij} \left(\tilde{\xi}_{ij} + \frac{b_{ij}}{\lambda_{ij}}\right)^2 + \sum_{j=h_i+1}^{m} b_{ij} \tilde{\xi}_{ij} + c_i && \left(c_i = a - \frac{1}{2}\sum_{j=1}^{h_i} \frac{b_{ij}^2}{\lambda_{ij}}\right) \\
&= \frac{1}{2}\sum_{j=1}^{h_i} \lambda_{ij} \zeta_{ij}^2 + \sum_{j=h_i+1}^{m} b_{ij} \tilde{\xi}_{ij} + c_i. && \left(\zeta_{ij} = \tilde{\xi}_{ij} + \frac{b_{ij}}{\lambda_{ij}}\right) \qquad (3)
\end{aligned}
$$

^ ~若 $\boldsymbol{\xi}$ 服从正态分布 p_i，则 $\boldsymbol{\xi}_i$ 和 $\boldsymbol{\xi}_i$~的协方差矩阵均为单位矩阵。因此，$\zeta_{ij}(j = 1, \cdots, h_i)$ 和 $\xi_{ij}(j = h_i + 1, \cdots, m)$ 分别是均值为 $\delta_{ij} = d_{ij} + \frac{b_{ij}}{\lambda_{ij}}(j = 1, \cdots, h_i)$ 和 $d_{ij^1}($ $j = h_i + 1, \cdots, m)$、方差为单位方差的相互独立的正态随机变量。这里，d_{ij} 是 $D_i^\top \Sigma_i^{-\,\overline{2}} \boldsymbol{\mu}_i$的第 j 个元素。注意，如果 Σ_i 是奇异的，我们也可以将 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 分解为形式 (3)，但需要经过更复杂的重新表述过程。

根据 L´evy 连续性引理（Van der Vaart（1998）），其内容总结于附录 B 中，特征函数的收敛等价于依分布收敛。注意到上述重新表述对于每个概率分布p_i都相同，通过考察 $^{c}{}_{(\boldsymbol{\xi},}$ $_{\boldsymbol{x})\sqrt{V_{p^*}(c(\boldsymbol{\xi},\ \boldsymbol{x}))}}$的特征函数，我们得到以下关于 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 的渐近分布结果，其中 $\boldsymbol{\xi}$ 服从 GMD。

定理1. 假设 $\boldsymbol{\xi}$ 服从由式（2）定义的 GMD，且 $h_i \to +\infty$ 如 $m \to +\infty$,$i = 1, \cdots, K$。则

$$
\lim_{m \to +\infty} \left(E_{p*}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right) - \sum_{i=1}^{K} \pi_i e^{\mathrm{i}\frac{\mu_i}{\sigma}t - \frac{1}{2}\frac{\sigma_i^2}{\sigma^2}t^2} \right) = 0,
$$

其中 $\mu_i = E_{p_i}(c(\boldsymbol{\xi}, \boldsymbol{x}))$、$\sigma_i^2 = V_{p_i}(c(\boldsymbol{\xi}, \boldsymbol{x}))$和 $\sigma^2 = V_{p^*}(c(\boldsymbol{\xi}, \boldsymbol{x}))$，若

$$
\lim_{m \to +\infty} \frac{\left(\sum_{j=1}^{h_i} \left(\frac{1}{k} + \delta_{ij}^2\right) \lambda_{ij}^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h_i} \left(\frac{1}{2} + \delta_{ij}^2\right) \lambda_{ij}^2\right)^{\frac{1}{2}}} = 0 \qquad (4)
$$

对于任意 $k \geq 3$ 以及 $i = 1$、$\cdots$ 和 K，均一致成立。

证明。首先，我们证明

$$\lim_{m\to+\infty}\left(E_{p_i}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)-e^{\mathrm{i}\frac{\mu_i}{\sigma}t-\frac{1}{2}\frac{\sigma_i^2}{\sigma^2}t^2}\right)=0,\ i=1,\cdots,K.$$

由于上述方程对每个 p_i的证明均相同，不失一般性，我们考虑 $i=1$ 的情形，并将记号简化为 $\delta_{1j}=\delta_j$、$\zeta_{1j}=\zeta_j$、$\lambda_{1j}=\lambda_j$、$b_{1j}=b_j$、$d_{1j}=d_j$、$c_1=c$ 和 $h_1=h$。

记 $S_h=\prod_{j=1}^{h}E_{p_1}\left(e^{\mathrm{i}t\frac{\lambda_j\zeta_j^2}{2\sigma}}\right)$。根据(3)以及正态分布随机变量的特征函数，若 $h<m$，则有

$$E_{p_1}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)=\prod_{j=1}^{h}E_{p_1}\left(e^{\mathrm{i}t\frac{\lambda_j\zeta_j^2}{2\sigma}}\right)\prod_{j=h+1}^{m}E_{p_1}\left(e^{\mathrm{i}t\frac{b_j\tilde{\xi}_j}{\sigma}}\right)E_{p_1}\left(e^{\mathrm{i}t\frac{c}{\sigma}}\right)=S_he^{\mathrm{i}q_1t-\frac{1}{2}q_2t^2},\quad(5)$$

其中 $q_1=\frac{c}{\sigma}+\sum_{j=h+1}^{m}\frac{b_jd_j}{\sigma}$ 且 $q_2=\sum_{j=h+1}^{m}\frac{b_j^2}{\sigma^2}$。否则，对于 $h=m$，$c(\boldsymbol{\xi},\boldsymbol{x})$中不存在任何线性项，即 $E_{p_1}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)=S_he^{\mathrm{i}\frac{c}{\sigma}t}$，该证明是下述证明的一个特例。

接下来我们首先仅关注 S_h。由于 $\boldsymbol{\xi}$ 服从正态分布，其密度为 p_1，因此 ζ_j^2是一个非中心卡方随机变量，其特征函数为 $(1-\mathrm{i}2t)^{-\frac{1}{2}}e^{\mathrm{i}\frac{t}{1-\mathrm{i}2t}\delta_j^2}$。因此，我们有

$$S_h=\prod_{j=1}^{h}E_{p_1}\left(e^{\mathrm{i}\frac{\lambda_jt}{2\sigma}\zeta_j^2}\right)=\prod_{j=1}^{h}\left(1-\mathrm{i}\frac{\lambda_j}{\sigma}t\right)^{-\frac{1}{2}}e^{\frac{\mathrm{i}\frac{\lambda_j}{\sigma}t}{1-\mathrm{i}\frac{\lambda_j}{\sigma}t}\delta_j^2}.$$

记为 $\gamma_j=\frac{\lambda_j}{\sigma}$。对上述方程两边取对数，得

$$\ln(S_h)=-\frac{1}{2}\sum_{j=1}^{h}\ln(1-\mathrm{i}\gamma_jt)+\frac{1}{2}\sum_{j=1}^{h}\frac{\mathrm{i}\gamma_jt}{1-\mathrm{i}\gamma_jt}\delta_j^2.\quad(6)$$

在对方程（6）关于 $\mathrm{i}\gamma_jt$ 作泰勒展开以便进一步分析之前，我们需要证明 $\mathrm{i}\gamma_jt$ 足够小。为此，下面先证明 $\lim_{m\to+\infty}\gamma_j=0$。为简便起见，我们进一步记

$$\tilde{\sigma}_k=\left(\sum_{j=1}^{h}\left(\frac{1}{k}+\delta_j^2\right)\lambda_j^k\right)^{\frac{1}{k}},\quad k=1,\cdots,n,\cdots.$$

注意到 $\delta_j=d_j+\frac{b_j}{\lambda_j}$对于 $j=1$、$\cdots$、h 和 $\lambda_j=0$，对于 $j=h+1$、$\cdots$、m。于是根据

经过一些代数运算，我们得到

$$\begin{aligned}\tilde{\sigma}_2^2 &\leq \sum_{j=1}^{h}\left(\frac{1}{2}+\delta_j^2\right)\lambda_j^2+\sum_{j=h+1}^{m}b_j^2=\sum_{j=1}^{m}(\lambda_j d_j+b_j)^2+\frac{1}{2}\sum_{j=1}^{m}\lambda_j^2\\ &= \boldsymbol{\varsigma}_1^\top D_1^\top D_1\boldsymbol{\varsigma}_1+\frac{1}{2}\mathrm{tr}\left(\left(\Sigma_1^{\frac{1}{2}}A\Sigma_1^{\frac{1}{2}}\right)^2\right)\\ &= \left(\Sigma_1^{\frac{1}{2}}A\boldsymbol{\mu}_1+\Sigma_1^{\frac{1}{2}}\boldsymbol{a}\right)^\top\left(\Sigma_1^{\frac{1}{2}}A\boldsymbol{\mu}_1+\Sigma_1^{\frac{1}{2}}\boldsymbol{a}\right)+\frac{1}{2}\mathrm{tr}\left(\left(\Sigma_1^{\frac{1}{2}}A\Sigma_1^{\frac{1}{2}}\right)^2\right)\\ &= (A\boldsymbol{\mu}_1+\boldsymbol{a})^\top\Sigma_1(A\boldsymbol{\mu}_1+\boldsymbol{a})+\frac{1}{2}\mathrm{tr}((A\Sigma_1)^2)=\sigma_1^2,\end{aligned}$$

其中 $\boldsymbol{\varsigma}_1=\Lambda_1 D_1^\top\Sigma_1^{-\frac{1}{2}}\boldsymbol{\mu}_1+D_1^\top\Sigma_1^{\frac{1}{2}}\boldsymbol{a}$。第二个等式成立，因为 d_j 和 b_j 分别是 $D_1^\top\Sigma_1^{-\frac{1}{2}}\boldsymbol{\mu}_1$ 和 $D_1^\top\Sigma_1^{\frac{1}{2}}\boldsymbol{a}$ 的第 j 个元素。现在，我们可以将 γ_j 改写为三个项的乘积

$$\gamma_j=\frac{\lambda_j}{\sigma}=\frac{\lambda_j}{\tilde{\sigma}_2}\frac{\tilde{\sigma}_2}{\sigma_1}\frac{\sigma_1}{\sigma},$$

其中第二项显然有界。此外，根据命题1，$\sigma^2\geq\sum_{i=1}^{K}\pi_i\sigma_i^2\geq\pi_1\sigma_1^2$，这进一步意味着 $\sigma_1/\sigma\leq 1/\sqrt{\pi_1}$ 也有界。对于第一项，由于 $\frac{\tilde{\sigma}_4}{\tilde{\sigma}_2}\geq\left(\frac{1}{4}\right)^{\frac{1}{4}}\frac{|\lambda_j|}{\tilde{\sigma}_2}$，根据假设（4），我们有

$$0=\lim_{m\to+\infty}\frac{\tilde{\sigma}_4}{\tilde{\sigma}_2}\geq\lim_{m\to+\infty}\left(\frac{1}{4}\right)^{\frac{1}{4}}\frac{|\lambda_j|}{\tilde{\sigma}_2}\geq 0,$$

这意味着 $\lim_{m\to+\infty}\frac{\lambda_j}{\tilde{\sigma}_2}=0$。因此 $\lim_{m\to+\infty}\gamma_j=0$。

回到方程（6）。对于任意给定的 t，存在充分大的 m，使得 $|\mathrm{i}\gamma_j t|<1$，因为 $\lim_{m\to+\infty}\gamma_j=0$。根据泰勒展开，对于 $|x|<1$，$\ln(1-x)=\sum_{n=1}^{\infty}\frac{-1}{n}x^n$ 且 $\frac{x}{1-x}=\sum_{n=1}^{\infty}x^n$。那么，当 m充分大时，方程（6）可改写为

$$\ln(S_h)=\frac{1}{2}\sum_{j=1}^{h}\sum_{k=1}^{\infty}\left(\delta_j^2+\frac{1}{k}\right)(\mathrm{i}\gamma_j t)^k=\frac{1}{2}\sum_{k=1}^{\infty}\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)(\mathrm{i}t)^k. \tag{7}$$

回想一下，$\tilde{\sigma}_2\leq\sigma_1$、$\sigma_1\leq\frac{\sigma}{\sqrt{\pi_1}}$ 和 $\gamma_j=\frac{\lambda_j}{\sigma}$，我们有

$$\left|\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)^{\frac{1}{k}}\right|=\left|\frac{\tilde{\sigma}_k}{\sigma}\right|\leq\left|\frac{\tilde{\sigma}_k}{\sqrt{\pi_1}\sigma_1}\right|\leq\left|\frac{\tilde{\sigma}_k}{\sqrt{\pi_1}\tilde{\sigma}_2}\right|.$$

根据假设（4），对于任意给定的 t 和足够小的 $\epsilon>0$，存在一个足够大的 m，且与 k 无关，使得对于每个 $k\geq 3$，$\left|\mathrm{i}t\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)^{\frac{1}{k}}\right|<\epsilon$。

因此,

$$\left|\sum_{k=3}^{\infty}\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)(\mathrm{i}t)^k\right| \leq \sum_{k=3}^{\infty}\left|\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)(\mathrm{i}t)^k\right| < \sum_{k=3}^{\infty}\epsilon^k = \frac{\epsilon^3}{1-\epsilon},$$

这意味着

$$\lim_{m\to+\infty}\sum_{k=3}^{\infty}\left(\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k\right)(\mathrm{i}t)^k = 0. \tag{8}$$

根据方程（5）和（7），以及 $\sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{k}\right)\gamma_j^k = \frac{\tilde{\sigma}_k^k}{\sigma^k}$ 这一事实，我们有

$$\begin{aligned}
\ln\left(E_{p_1}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{\varpi})}{\sigma}}\right)\right) &= \ln(S_h)+\mathrm{i}q_1t-\frac{1}{2}q_2t^2 \\
&= \frac{1}{2}\sum_{k=1}^{\infty}\frac{\tilde{\sigma}_k^k}{\sigma^k}(\mathrm{i}t)^k+\mathrm{i}q_1t-\frac{1}{2}q_2t^2 \\
&= \mathrm{i}\left(\frac{1}{2}\frac{\tilde{\sigma}_1}{\sigma}+q_1\right)t-\frac{1}{2}\left(\frac{\tilde{\sigma}_2^2}{\sigma^2}+q_2\right)t^2+\frac{1}{2}\sum_{k=3}^{\infty}\frac{\tilde{\sigma}_k^k}{\sigma^k}(\mathrm{i}t)^k \\
&= \mathrm{i}\frac{\mu_1}{\sigma}t-\frac{1}{2}\frac{\sigma_1^2}{\sigma^2}t^2+\frac{1}{2}\sum_{k=3}^{\infty}\frac{\tilde{\sigma}_k^k}{\sigma^k}(\mathrm{i}t)^k,
\end{aligned} \tag{9}$$

其中，最后一个等式是由于

$$\begin{aligned}
\frac{1}{2}\tilde{\sigma}_1+\sigma q_1 &= \frac{1}{2}\sum_{j=1}^{h}(\delta_j^2+1)\lambda_j+\sum_{j=h+1}^{m}b_jd_j+c \\
&= \frac{1}{2}\sum_{j=1}^{h}\left(d_j+\frac{b_j}{\lambda_j}\right)^2\lambda_j+\frac{1}{2}\sum_{j=1}^{h}\lambda_j+\sum_{j=h+1}^{m}d_jb_j+a-\frac{1}{2}\sum_{j=1}^{h}\frac{b_j^2}{\lambda_j} \\
&= \frac{1}{2}\sum_{j=1}^{m}d_j^2\lambda_j+\frac{1}{2}\sum_{j=1}^{m}\lambda_j+\sum_{j=1}^{m}d_jb_j+a \\
&= \frac{1}{2}\boldsymbol{\mu}_1^\top\Sigma_1^{-\frac{1}{2}}D_1\Lambda D_1^\top\Sigma_1^{-\frac{1}{2}}\boldsymbol{\mu}_1+\frac{1}{2}\mathrm{tr}(A\Sigma_1)+\boldsymbol{\mu}_1^\top\Sigma_1^{-\frac{1}{2}}D_1D_1^\top\Sigma_1^{\frac{1}{2}}\boldsymbol{a}+a \\
&= \frac{1}{2}\boldsymbol{\mu}_1^\top A\boldsymbol{\mu}_1+\frac{1}{2}\mathrm{tr}(A\Sigma_1)+\boldsymbol{a}^\top\boldsymbol{\mu}_1+a \\
&= \mu_1
\end{aligned}$$

以及事实 $\tilde{\sigma}_2^2+\sigma^2q_2 = \sum_{j=1}^{h}\left(\delta_j^2+\frac{1}{2}\right)\lambda_j^2+\sum_{j=h+1}^{m}b_j^2 = \sigma_1^2$。

结合式 (8) 和式 (9) 可得

$$\lim_{m\to\infty}\left(\ln\left(E_{p_1}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{\varpi})}{\sigma}}\right)\right)-\left(\mathrm{i}\frac{\mu_1}{\sigma}t-\frac{1}{2}\frac{\sigma_1^2}{\sigma^2}t^2\right)\right) = 0,$$

或等价地，

$$\lim_{m\to\infty}\left(E_{p_1}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)-e^{\mathrm{i}\frac{\mu_1}{\sigma}t-\frac{1}{2}\frac{\sigma_1^2}{\sigma^2}t^2}\right)=0.$$

显然，上述结论对 $i=2$、$\cdots$、K 同样成立。

最后，注意到 $E_{p^*}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)=\sum_{i=1}^{K}\pi_i E_{p_i}\left(e^{\mathrm{i}t\frac{c}{(\boldsymbol{\xi},\boldsymbol{x}}}\right)$，我们有

$$\lim_{m\to\infty}\left(E_{p^*}\left(e^{\mathrm{i}t\frac{c(\boldsymbol{\xi},\boldsymbol{x})}{\sigma}}\right)-\sum_{i=1}^{K}\pi_i e^{\mathrm{i}\frac{\mu_i}{\sigma}t-\frac{1}{2}\frac{\sigma_i^2}{\sigma^2}t^2}\right)=0. \tag{10}$$

证明完成。 □

如下述推论所示，定理1中的条件非常宽松。

推论1. 如果对于每个 $i\in\{1,\cdots,K\}$、$h_i\to+\infty$，有 $m\to+\infty$，且 $0<\underline{\lambda}\le|\lambda_{ij}|\le\overline{\lambda}<+\infty$、$j=1$、$\cdots$、$h_i$，则方程（10）成立。

证明. 对于每个 $i=1$、$\cdots$、K，定理1中的条件相同。同样地，为简化分析，我们考虑 $i=1$ 的情形，并省略相关下标。注意到 d_j 和 b_j 仅由 GMD 的参数决定，因此它们自然有界。因此，存在有限的正常数 l 和 u，使得对于 $j=1$、$\cdots$、h，有 $l\le\delta_j^2=\left(d_j+\frac{b_j}{\lambda_j}\right)^2\le u$。于是，对于 $k\ge3$，我们有

$$\left|\frac{\left(\sum_{j=1}^{h}\left(\frac{1}{k}+\delta_j^2\right)\lambda_j^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}\left(\frac{1}{2}+\delta_j^2\right)\lambda_j^2\right)^{\frac{1}{2}}}\right|\le\frac{\left((\frac{1}{k}+u)\sum_{j=1}^{h}|\lambda_j|^k\right)^{\frac{1}{k}}}{\left((\frac{1}{2}+l)\sum_{j=1}^{h}\lambda_j^2\right)^{\frac{1}{2}}}\le\frac{(\frac{1}{k}+u)^{\frac{1}{k}}\overline{\lambda}}{(\frac{1}{2}+l)^{\frac{1}{2}}\underline{\lambda}}h^{\frac{1}{k}-\frac{1}{2}}\le Mh^{-\frac{1}{6}},$$

其中 $M>(\frac{1}{k}+u)^{\frac{1}{k}}\overline{\lambda}(\frac{1}{2}+l)^{\frac{1}{2}}\underline{\lambda}$是一个与 k无关的常数。注意，$(\frac{1}{k}+u)^{\frac{1}{k}}\overline{\lambda}(\frac{1}{2}+l)^{\frac{1}{2}}\underline{\lambda}$有界，因此 $k\ge3$所以 M 可以取为任何大于此上界的数。于是，对任意 $\epsilon>0$，存在一个与 k 无关的 $\overline{h}=\left(\frac{M}{\epsilon}\right)^6$，使得当 $h>\overline{h}$时，

$$\left|\frac{\left(\sum_{j=1}^{h}\left(\frac{1}{k}+\delta_j^2\right)\lambda_j^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}\left(\frac{1}{2}+\delta_j^2\right)\lambda_j^2\right)^{\frac{1}{2}}}\right|<\epsilon,$$

这意味着当 $h\to+\infty$趋于$m\to+\infty$时，$\lim_{m\to+\infty}(\sum_{j=1}^{h}(\frac{1}{k}+\delta_j^2)\lambda_j^k)^{\frac{1}{k}}(\sum_{j=1}^{h}(\frac{1}{2}+\delta_j^2)\lambda_j^2)^{\frac{1}{2}}=0$对 $k\ge3$一致成立。这验证了定理1中的条件。证明完毕。 □

通俗地说，定理1和推论1得出：如果对于每个 $i\in\{1,\cdots,K\}$，$\Sigma_i^{\frac{2}{2}}A\Sigma_i^{\frac{2}{2}}$的非零$_{1\ 1}$特征值数量足够大，那么$\Sigma_i^{\frac{2}{2}}A\Sigma_i^{\frac{2}{2}}$的分布

$c(\boldsymbol{\xi}, \boldsymbol{x})$可以用以下一元广义多元分布进行近似

$$p^*(z) = \sum_{i=1}^{K} \pi_i p(z|\mu_i, \sigma_i^2). \tag{11}$$

回顾一下， μ_i 和 σ_i^2 分别是相对于概率分布 p_i， $c(\boldsymbol{\xi}, \boldsymbol{x})$ 的均值和方差。

特别地，当 $K = 1$ 时，随机变量服从多元高斯分布，且 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 的渐近分布为高斯分布，这与 Zhu 等人（2020）的结果一致。若 $A = O$，则 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 是 $\boldsymbol{\xi}$ 的线性函数；Hu 等人（2022）已研究过这种情形。

2.2 渐近近似的误差分析

现在，我们探讨在有限 m 的情况下，使用渐近分布（11）近似 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 的真实分布时产生误差的因素。我们先从定理1提出的条件入手。注意，对于相对于 p_i 的每个 $i \in \{1, \cdots, K\}$，这些条件的形式都相同。因此，为简便起见，我们只分析 $K = 1$ 的情况，所得结果也可类似地推广到 $K > 1$ 的情况。

对于 $K = 1$，再次省略下标 i，方程 (3) 可简化为

$$c(\boldsymbol{\xi}, \boldsymbol{x}) = \frac{1}{2} \sum_{j=1}^{h} \lambda_j \zeta_j^2 + \sum_{j=h+1}^{m} b_j \tilde{\xi}_j + c, \tag{12}$$

它由三项组成：独立高斯随机变量的平方加权和、独立高斯随机变量的加权和，以及一个常数。此外，第一项与第二项相互独立，且 $c(\boldsymbol{\xi}, \boldsymbol{x})$ 的渐近分布为高斯分布。显然，渐近近似的误差源于第一项。因此，我们研究以下非中心卡方变量一般加权和的渐近行为

$$Z_h = \sum_{j=1}^{h} \omega_j z_j^2, \tag{13}$$

其中，对于每个 j， z_j 是均值为 α_j 、方差为 1 的独立高斯随机变量，即 $z_j \sim \mathcal{N}(\alpha_j, 1)$。我们将在以下定理中给出 Z_h 的特征函数的收敛速率。

定理2. 假设 $\{|\omega_j|\}_{j=1}^{h}$是一个非零递减序列。若 $\lim_{h \to +\infty} \frac{\omega_1}{\omega_h} \bar{\alpha}_h^{-\frac{1}{6}} h^{-\frac{1}{6}} = 0$,

其中$\bar{\alpha}_h = \frac{1}{2} + \frac{1}{h}\sum_{j=1}^{h}\alpha_j^2$，则

$$\left|\ln\left(E\left(e^{\mathrm{i}t\overline{Z}_h}\right)\right) - \ln\left(e^{-\frac{1}{2}t^2}\right)\right| = O\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}}h^{-\frac{1}{2}}\right)$$

对于足够大的 h，其中 $\overline{Z}_h = \frac{Z_h - E_{(Z_h)}}{\sqrt{V(Z_h)}}$。

证明。根据 Zhu 等人（2020）的命题 2，$\overline{Z}_h$ 的特征函数的对数为

$$\ln\left(E\left(e^{\mathrm{i}t\overline{Z}_h}\right)\right) = -\frac{1}{2}t^2 + \frac{1}{2}\sum_{k=3}^{\infty}\mathrm{i}^k f_k^k t^k, \tag{14}$$

其中 $f_k = \frac{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{k})\omega_j^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{2})\omega_j^2\right)^{\frac{1}{2}}}$。我们有

$$\begin{aligned}
|f_k| = \left|\frac{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{k})\omega_j^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{2})\omega_j^2\right)^{\frac{1}{2}}}\right| &\leq \frac{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{k})|\omega_j|^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{2})|\omega_j|^2\right)^{\frac{1}{2}}} \leq \frac{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{k})|\omega_1|^k\right)^{\frac{1}{k}}}{\left(\sum_{j=1}^{h}(\alpha_j^2+\frac{1}{2})|\omega_h|^2\right)^{\frac{1}{2}}} \\
&\leq \left|\frac{\omega_1}{\omega_h}\right|\frac{(\bar{\alpha}_h+\frac{1}{k})^{\frac{1}{k}}}{(\bar{\alpha}_h+\frac{1}{2})^{\frac{1}{2}}}h^{\frac{1}{k}-\frac{1}{2}} \leq \left|\frac{\omega_1}{\omega_h}\right|\bar{\alpha}_h^{\frac{1}{k}-\frac{1}{2}}h^{\frac{1}{k}-\frac{1}{2}}.
\end{aligned} \tag{15}$$

记 $g_k = \frac{f_k}{\left|\frac{\omega_1}{\omega_h}\right|\bar{\alpha}_h^{-\frac{1}{6}}h^{-\frac{1}{6}}}$。进一步有

$$|g_k| \leq \bar{\alpha}_h^{\frac{1}{k}-\frac{1}{3}}h^{\frac{1}{k}-\frac{1}{3}}. \tag{16}$$

注意到，当 $h > 2$ 时，$\left|\frac{\omega_1}{\omega_h}\right| \geq 1$ 和 $\bar{\alpha}_h h \geq \frac{1}{2}h \geq 1$。于是，对任意 $k \geq 4$ 和 $h > 2$，我们有 $\bar{\alpha}_h^{\frac{1}{k}-\frac{1}{3}}h^{\frac{1}{k}-\frac{1}{3}} \leq \bar{\alpha}_h^{-\frac{1}{12}}h^{-\frac{1}{12}}$，这意味着当 $k \geq 4$ 时，g_k 一致收敛于 0。

此外，假设 $\lim_{h\to+\infty}\frac{\omega_1}{\omega_h}\bar{\alpha}_h^{-\frac{1}{6}}h^{-\frac{1}{6}} = 0$ 意味着，对于足够大的 h，$\left|\frac{\omega_1}{\omega_h}\right|\bar{\alpha}_h^{-\frac{1}{6}}h^{-\frac{1}{6}} < 1$。因此，给定 t，对于任意 $k \geq 4$ 以及任意足够小的 $\epsilon > 0$，都存在一个与 k 无关的充分大的 h，使得

$$\frac{|\mathrm{i}f_k t|}{\left|\frac{\omega_1}{\omega_h}\right|\bar{\alpha}_h^{-\frac{1}{6}}h^{-\frac{1}{6}}} < \epsilon \text{ and } \frac{\left|\mathrm{i}^k f_k^k t^k\right|}{\left|\frac{\omega_1}{\omega_h}\right|^3\bar{\alpha}_h^{-\frac{1}{2}}h^{-\frac{1}{2}}} < \frac{\left|\mathrm{i}^k f_k^k t^k\right|}{\left|\frac{\omega_1}{\omega_h}\right|^k\bar{\alpha}_h^{-\frac{k}{6}}h^{-\frac{k}{6}}} < \epsilon^k.$$

因此，

$$\frac{\left|\sum_{k=4}^{\infty}\mathrm{i}^k f_k^k t^k\right|}{\left|\frac{\omega_1}{\omega_h}\right|^3\bar{\alpha}_h^{-\frac{1}{2}}h^{-\frac{1}{2}}} < \frac{\sum_{k=4}^{\infty}\left|\mathrm{i}^k f_k^k t^k\right|}{\left|\frac{\omega_1}{\omega_h}\right|^3\bar{\alpha}_h^{-\frac{1}{2}}h^{-\frac{1}{2}}} < \sum_{k=4}^{\infty}\epsilon^k = \frac{\epsilon^4}{1-\epsilon},$$

这意味着，当 h趋于无穷大时，$\frac{\left|\sum_{k=4}^{\infty}{}^{k}f_k^k t^k\right|}{\left|\frac{\omega_1}{\omega_h}\right|^3 \bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}}$i趋于零，即，

$$\left|\sum_{k=4}^{\infty} \mathrm{i}^k f_k^k t^k\right| = o\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right). \tag{17}$$

对于 $k=3$，根据（16），$|g_3| \leq 1$。由（15）不难看出，在某些特殊情况下可以达到等式$|g_3| = 1$。因此，我们有

$$|\mathrm{i}^3 f_3^3 t^3| = O\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right). \tag{18}$$

结合（14）、（17）和（18）可得

$$\begin{aligned}\left|\ln\left(E\left(e^{it\overline{Z}_h}\right)\right) - \ln\left(e^{-\frac{1}{2}t^2}\right)\right| = \left|\frac{1}{2}\sum_{k=3}^{\infty}\mathrm{i}^k f_k^k t^k\right| &\leq \left|\frac{1}{2}\mathrm{i}^3 f_3^3 t^3\right| + \left|\frac{1}{2}\sum_{k=4}^{\infty}\mathrm{i}^k f_k^k t^k\right| \\ &= O\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right) + o\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right) \\ &= O\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right).\end{aligned}$$

类似地，我们有

$$\left|\ln\left(E\left(e^{it\overline{Z}_h}\right)\right) - \ln\left(e^{-\frac{1}{2}t^2}\right)\right| \geq \left|\frac{1}{2}\mathrm{i}^3 f_3^3 t^3\right| - \left|\frac{1}{2}\sum_{k=4}^{\infty}\mathrm{i}^k f_k^k t^k\right| = O\left(\frac{\omega_1^3}{\omega_h^3}\bar{\alpha}_h^{-\frac{1}{2}} h^{-\frac{1}{2}}\right).$$

证明完毕。 □

定理2表明，对于给定的 $\boldsymbol{\omega}$ 和 $\boldsymbol{\alpha}$，非中心卡方随机变量加权和的分布会随着随机变量数量趋于无穷而收敛于高斯分布，且收敛速率为 $O(1/\sqrt{h})$。这还表明，$\overline{Z}_h$ 与其渐近分布的特征函数之间的差异也取决于权重向量 $\boldsymbol{\omega}$ 和均值向量 $\boldsymbol{\alpha}$。权重 ω_i'的绝对值之间的差异越小，均值 α_i'的绝对值越大，收敛速度越快。由于该结论对每个概率分布 p_i均成立，因此这些结果对于具有 $K>1$的GMD情形同样成立。

2.3 渐近近似的误差降低

渐近近似会产生两类误差：基础适应度误差和渐近近似误差。具体而言，基础适应度误差是指用 GMD 近似 $\boldsymbol{\xi}$ 的分布所产生的误差，而渐近近似误差则是指用基于基础 GMD 的渐近分布近似 $c(\boldsymbol{\xi}, \boldsymbol{x})$的分布所造成的误差。

及其基于基础 GMD 得到的渐近分布。

比较方程（12）和（13）可得

$$\alpha_j \sim d_j + \frac{b_j}{\lambda_j} \text{ and } \omega_j \sim \lambda_j,\ j = 1, \cdots, h,$$

其中， d_j和 b_j分别是 $D^\top \Sigma^{-\frac{1}{2}} \boldsymbol{\mu}$和 $D^\top \Sigma^{\frac{1}{2}} \boldsymbol{a}$的第 j个元素。根据定理2， $\left(d_j + \frac{b_j}{\lambda_j}\right)$的绝对值越大，矩阵 $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$的非零特征值中最大绝对值与最小绝对值之比越小，渐近近似误差就越小。从本质上说，误差由 $\boldsymbol{\xi}$的分布参数 $\boldsymbol{\mu}$和 Σ 、 A 以及$\boldsymbol{a}$决定。给定决策向量 $\boldsymbol{x}$后， A 和 $\boldsymbol{a}$ 是两个常数，误差仅由参数 $\boldsymbol{\mu}$ 和 Σ决定。因此，接下来我们主要讨论如何通过适当设置参数 $\boldsymbol{\mu}$ 和 Σ 来降低渐近近似误差。

假设 $\boldsymbol{\mu} = \mathbf{0}$和 $\boldsymbol{a} = \mathbf{0}$。那么 $d_j + \frac{b_j}{\lambda_j} = 0 (i = 1, \cdots, h)$，根据定理2，这意味着与 $\boldsymbol{\alpha}$ 相关的收敛速率最低。因此，均值向量 $\boldsymbol{\mu}$ 与原点之间的偏差越大，越有可能带来更快的收敛。然而，给定 A 和 $\boldsymbol{a}$的不同取值， $\boldsymbol{\mu}$ 对渐近近似误差的影响也不同。因此，调整均值向量 $\boldsymbol{\mu}$未必能降低误差，所以下文将重点讨论 Σ 。

由于影响我们分析的是矩阵 $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ 的非零特征值个数，而非其奇异性，为简单起见，在以下分析中我们假设 $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ 是非奇异的，或者等价地， A 和 Σ 均为非奇异矩阵。关于矩阵 $\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}}$ 的非零特征值绝对值的最大值与最小值之比，我们有以下命题。为简单起见，记 $|\lambda_j(\cdot)|$ 为一个 m 维非奇异矩阵的特征值绝对值按从大到小排列后的第 j 个值。

命题 2. 假设 A 和 Σ 均为非奇异矩阵。那么

$$\left| \frac{\lambda_1(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})}{\lambda_m(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})} \right| \leq \left| \frac{\lambda_1(A)}{\lambda_m(A)} \right| \frac{\lambda_1(\Sigma)}{\lambda_m(\Sigma)}.$$

证明. 注意， $\Sigma^{\frac{1}{2}}$ 是一个满足 $\Sigma = \Sigma^{\frac{1}{2}} \Sigma^{\frac{1}{2}}$ 的定矩阵。根据 Golub 和 Van Loan（2013）的定理8.1.17，我们有

$$0 < |\lambda_i(A)\lambda_m(\Sigma)| \leq |\lambda_i(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})| \leq |\lambda_i(A)\lambda_1(\Sigma)|$$

对于 $i = 1$、 $\cdots$、 m，这表明

$$\left| \frac{\lambda_1(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})}{\lambda_m(\Sigma^{\frac{1}{2}} A \Sigma^{\frac{1}{2}})} \right| \leq \left| \frac{\lambda_1(A)\lambda_1(\Sigma)}{\lambda_m(A)\lambda_m(\Sigma)} \right| = \left| \frac{\lambda_1(A)}{\lambda_m(A)} \right| \frac{\lambda_1(\Sigma)}{\lambda_m(\Sigma)}.$$

证明完成。 □

比较方程(13)与方程(12)的第一项可知，根据定理2和命题2，当矩阵 A 和/或 Σ 的特征值绝对值最大值与最小值之比相对较小时， $c(\boldsymbol{\xi}, \boldsymbol{x})$的分布可以较容易地收敛到相应的渐近分布。

从上述分析中得到的一个直观认识是，我们可以约束 $^{\lambda_1(\Sigma_i)}\,\lambda_m(\Sigma_i)$，即构成 GMD 的高斯分布的协方差矩阵的条件数，以减小渐近近似误差。需要指出的是，我们最初选择 GMD 的动机在于，它对 $\boldsymbol{\xi}$ 的真实分布具有很强的拟合能力，从而能够降低基础适应度误差。下文将表明，即使限制协方差矩阵的条件数，也能够实现这一目的。

定理 3. 假设 $f(\boldsymbol{z}): \mathbb{R}^m \to \mathbb{R}$ 是一个满足 $\int_{\mathbb{R}^m} f(\boldsymbol{z})d\boldsymbol{z} = 1$ 的非负黎曼可积函数。那么，对于任意给定的 $\delta > 0$ 和 $q \geq 1$，都存在一个高斯混合密度函数

$$p^*(\boldsymbol{z}) = \sum_{i=1}^{K} \pi_i p(\boldsymbol{z}|\boldsymbol{\mu}_i, \Sigma_i), \tag{19}$$

其中 $\pi_i > 0$、 $i = 1$、$\cdots$、K 和 $^{\lambda_1(\Sigma_i)}\,\lambda_m(\Sigma_i) \leq q$、 $i = 1$、$\cdots$、K，使得

$$\int_{\mathbb{R}^m} |f(\boldsymbol{z}) - p^*(\boldsymbol{z})|d\boldsymbol{z} < \delta.$$

证明。根据假设 $\int_{\mathbb{R}^m} f(\boldsymbol{z})d\boldsymbol{z} = 1$, $f(\boldsymbol{z})$ 可由若干简单函数的有限和近似:

$$f_N(\boldsymbol{z}) = \sum_{i=1}^{N} \underline{f}_i \mathbf{1}_{\mathscr{I}_i},$$

其中 $\mathscr{I}_i = \prod_{j=1}^{m}(a_{ij}, b_{ij}]\,(b_{ij} > a_{ij})$, $i = 1, \cdots, N$ 是互不相交的立方体， $\underline{f}_i = \inf\{f(\boldsymbol{z}): \boldsymbol{z} \in \mathscr{I}_i\}$ ，且 $\mathbf{1}_{\mathscr{I}_i}$ 是示性函数。不考虑那些包含函数值为零的点的小立方体，我们进一步假设 $\underline{f}_i > 0$。这里， N 选取得足够大，使得

$$\int_{\mathbb{R}^m} |f(\boldsymbol{z}) - f_N(\boldsymbol{z})|d\boldsymbol{z} < \frac{\delta}{2}. \tag{20}$$

下面我们将证明：给定立方体的数量 N 和划分 $\{\mathscr{I}_i\}_{i=1}^{N}$， $f_N(\boldsymbol{z})$ 可以用高斯混合密度函数近似，且误差小于 $\frac{\delta}{2}$。具体而言，我们通过用卷积近似每个简单函数 $\underline{f}_i \mathbf{1}_{\mathscr{I}_i}$，再用高斯混合密度函数近似该卷积来实现这一点。

为简便起见，我们将该简单函数记作

$$\chi_{\mathscr{I}_i}(\boldsymbol{z}) = \underline{f}_i \mathbf{1}_{\mathscr{I}_i}(\boldsymbol{z}) = \begin{cases} \underline{f}_i & \boldsymbol{z} \in \mathscr{I}_i \\ 0 & \text{otherwise} \end{cases}$$

以及一个均值向量为零、协方差矩阵为 Σ_i 的高斯密度函数 $p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) = \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{z}^\top \Sigma_i^{-1}\boldsymbol{z}}$。

考虑 $\chi_{\mathscr{I}_i}(\cdot)$ 与 $p(\cdot|\mathbf{0}, \Sigma_i)$ 的卷积，其定义为

$$\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) = \int_{\mathscr{I}_i} \chi_{\mathscr{I}_i}(\boldsymbol{y}) p(\boldsymbol{z} - \boldsymbol{y}|\mathbf{0}, \Sigma_i) d\boldsymbol{y} = \underline{f}_i \int_{\mathscr{I}_i} p(\boldsymbol{z} - \boldsymbol{y}|\mathbf{0}, \Sigma_i) d\boldsymbol{y}. \tag{21}$$

该卷积具有以下两个性质。首先,

$$\int_{\mathbb{R}^m} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} = \int_{\mathscr{I}_i} \int_{\mathbb{R}^m} \underline{f}_i p(\boldsymbol{z} - \boldsymbol{y}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} d\boldsymbol{y} = \int_{\mathscr{I}_i} \underline{f}_i d\boldsymbol{y} = |\mathscr{I}_i| \underline{f}_i, \tag{22}$$

其中，$|\mathscr{I}_i|$ 是 $\mathscr{I}_i$ 的体积。由于 $\boldsymbol{z}$ 和 $\boldsymbol{y}$ 的积分域无关，因此交换了式（22）中的积分顺序。其次，对于任意 $\boldsymbol{z} \in \mathscr{I}_i$,

$$\chi_{\mathscr{I}_i}(\boldsymbol{z}) = \underline{f}_i \geq \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i). \tag{23}$$

结合式（22）—（23），我们可以推导出简单函数与相应卷积之间的积分绝对误差

$$\begin{aligned}
&\int_{\mathbb{R}^m} |\chi_{\mathscr{I}_i}(\boldsymbol{z}) - \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i)| d\boldsymbol{z} \\
&= \int_{\mathscr{I}_i} \left(\chi_{\mathscr{I}_i}(\boldsymbol{z}) - \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i)\right) d\boldsymbol{z} + \int_{\mathbb{R}^m \setminus \mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} \\
&= \int_{\mathscr{I}_i} \chi_{\mathscr{I}_i}(\boldsymbol{z}) d\boldsymbol{z} + \int_{\mathbb{R}^m} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} - 2\int_{\mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} \\
&= 2\underline{f}_i |\mathscr{I}_i| - 2\int_{\mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} = 2\underline{f}_i |\mathscr{I}_i| - 2\int_{\text{int}\mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z},
\end{aligned} \tag{24}$$

其中，$\text{int}\mathscr{I}_i = \prod_{j=1}^m (a_{ij}, b_{ij})$ 是相对于 $\mathscr{I}_i$ 的内点集合。该绝对误差取决于卷积在 $\mathscr{I}_i$ 上的积分。下面我们将说明积分绝对误差收敛于零的条件，或者等价地，卷积的积分收敛于 $\underline{f}_i|\mathscr{I}_i|$。

回顾式（21），卷积可以改写为

$$\begin{aligned}
\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) &= \underline{f}_i \int_{\mathscr{I}_i} \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{z})^\top \Sigma_i^{-1}(\boldsymbol{y}-\boldsymbol{z})} d\boldsymbol{y} \\
&= \underline{f}_i \int_{\mathscr{D}_{\boldsymbol{z}}} \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{x}^\top \Sigma_i^{-1}\boldsymbol{x}} |J_{\boldsymbol{y},\boldsymbol{x}}| d\boldsymbol{x} \\
&= \underline{f}_i \int_{\mathscr{D}_{\boldsymbol{z}}} \frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{x}^\top \Sigma_i^{-1}\boldsymbol{x}} d\boldsymbol{x},
\end{aligned}$$

其中， $\boldsymbol{x} = \boldsymbol{y} - \boldsymbol{z}$, $J_{\boldsymbol{y},\ \boldsymbol{x}}$ 是相应的雅可比矩阵，即 $\boldsymbol{y}$关于 $\boldsymbol{x}$的导数矩阵；此处它等于单位矩阵，且 $\mathscr{D}_{\boldsymbol{z}} = \{\boldsymbol{x}:\ \boldsymbol{x} + \boldsymbol{z} \in \mathscr{I}_i\}$。将 Σ_i分解为 $\Sigma_i = Q\Lambda Q^\top$，其中 Λ 是一个对角矩阵，其元素为 $\lambda_1, \cdots, \lambda_m$，而 Q 是一个正交矩阵；进一步记 $\mathscr{F} = \{\boldsymbol{u}:\ \boldsymbol{u} = Q^\top \boldsymbol{x}, \boldsymbol{x} \in \mathscr{D}_z\}$。对于任意 $\boldsymbol{z} \in \text{int}\mathscr{I}_i$，有 $\mathbf{0} \in \text{int}\mathscr{D}_{\boldsymbol{z}}$。注意， $\mathscr{F}$ 是 $\mathscr{D}_{\boldsymbol{z}}$的正交变换。因此， $\mathbf{0} \in \text{int}\mathscr{F}$。所以，存在一个子集 $\mathscr{F}_1 \subseteq \mathscr{F}$ 以及 $\mathscr{F}_1 = \prod_{i=1}^m (-\tau_i, \tau_i]$ ，且 $\tau_i > 0, i = 1, \cdots, m$足够小。于是，对于任意 $\boldsymbol{z} \in \text{int}\mathscr{I}_i$，有

$$\begin{aligned}
\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma) &= \underline{f}_i \int_{\mathscr{D}_{\boldsymbol{z}}} \frac{1}{(2\pi)^{\frac{m}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{x}^\top \Sigma_i^{-1} \boldsymbol{x}} d\boldsymbol{x} \\
&= \underline{f}_i \int_{\mathscr{F}} \frac{1}{(2\pi)^{\frac{m}{2}} |\Lambda|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{u}^\top \Lambda^{-1} \boldsymbol{u}} |J_{\boldsymbol{x},\boldsymbol{u}}| d\boldsymbol{u} \\
&\geq \underline{f}_i \int_{\mathscr{F}_1} \frac{1}{(2\pi)^{\frac{m}{2}} |\Lambda|^{\frac{1}{2}}} e^{-\frac{1}{2}\boldsymbol{u}^\top \Lambda^{-1} \boldsymbol{u}} d\boldsymbol{u} \\
&= \underline{f}_i \prod_{i=1}^m \int_{-\tau_i}^{\tau_i} \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{1}{2\lambda_i} u_i^2} du_i \\
&= \underline{f}_i \prod_{i=1}^m \int_{-\frac{\tau_i}{\sqrt{\lambda_i}}}^{\frac{\tau_i}{\sqrt{\lambda_i}}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u_i^2} du_i,
\end{aligned}$$

其中 $\boldsymbol{u} = Q^\top \boldsymbol{x}$, $J_{\boldsymbol{x},\boldsymbol{u}} = Q$ 且 $|J_{\boldsymbol{x},\boldsymbol{u}}| = |Q| = 1$。给定 τ_i, $i = 1, \cdots, m$，如果所有特征值λ_i, $i = 1, \cdots, m$都足够小，那么 $\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma)$ 收敛于 $\underline{f}_i$。因此，存在一个足够小的 $\lambda > 0$ 以及一个协方差矩阵 Σ_i ，使得如果

$$\bar{\lambda} > \lambda_i > \frac{\bar{\lambda}}{q},\ i = 1, \cdots, m,$$

then

$$\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) \geq \underline{f}_i \left(1 - \frac{\delta}{8N|\mathscr{I}_i|\underline{f}_i}\right).$$

注意，上述不等式对 $\boldsymbol{z} \in \text{int}\mathscr{I}_i$ 成立。对上述不等式两边在 $\text{int}\mathscr{I}_i$ 上积分，并结合式 (22)，可得

$$|\mathscr{I}_i|\underline{f}_i \geq \int_{\text{int}\mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} \geq |\mathscr{I}_i|\underline{f}_i \left(1 - \frac{\delta}{8N|\mathscr{I}_i|\underline{f}_i}\right).$$

将上述不等式与式 (24) 结合，可进一步推出

$$\int_{\mathbb{R}^m} |\chi_{\mathscr{I}_i}(\boldsymbol{z}) - \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i)| d\boldsymbol{z} = 2\underline{f}_i|\mathscr{I}_i| - 2\int_{\text{int}\mathscr{I}_i} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} < \frac{\delta}{4N}. \quad (25)$$

我们已经证明，任意简单函数都可以由相应的卷积函数逼近。下面我们将证明，卷积函数可以由

高斯混合密度函数逼近。

根据 (22)，卷积 $\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i)$ 关于 $\mathbb{R}^m$ 可积。因此，存在一个体积足够大的紧集 $\mathscr{B}$，使得

$$\int_{\mathbb{R}^m \setminus \mathscr{B}} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} < \frac{\delta}{16N}.$$

根据附录 C 中的引理 4，给定 $\mathscr{B}$，存在一个划分 $\mathscr{I}_i = \bigcup_{j=1}^{n_i} \mathscr{I}_{ij}$ 和一个函数

$$\chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z}) = \sum_{j=1}^{n_i} |\mathscr{I}_{ij}| \chi_{\mathscr{I}_i}(\boldsymbol{y}_{ij}) p(\boldsymbol{z} - \boldsymbol{y}_{ij}|\mathbf{0}, \Sigma_i) = \sum_{j=1}^{n_i} \underline{f}_i |\mathscr{I}_{ij}| p(\boldsymbol{z} - \boldsymbol{y}_{ij}|\mathbf{0}, \Sigma_i),$$

其中 $\boldsymbol{y}_{ij} \in \mathscr{I}_{ij}$，使得

$$\int_{\mathscr{B}} |\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) - \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z})| \, d\boldsymbol{z} < \frac{\delta}{16N}.$$

此外，上述不等式还意味着

$$\begin{aligned}
\int_{\mathbb{R}^m \setminus \mathscr{B}} \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z}) d\boldsymbol{z} &= 1 - \int_{\mathscr{B}} \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z}) d\boldsymbol{z} \\
&< 1 - \left(\int_{\mathscr{B}} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} - \frac{\delta}{16N} \right) \\
&= \int_{\mathbb{R}^m \setminus \mathscr{B}} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} + \frac{\delta}{16N} \\
&< \frac{\delta}{8N}.
\end{aligned}$$

最后，我们得到

$$\begin{aligned}
&\int_{\mathbb{R}^m} |\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) - \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z})| \, d\boldsymbol{z} \\
&< \int_{\mathscr{B}} |\chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) - \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z})| \, d\boldsymbol{z} + \int_{\mathbb{R}^m \setminus \mathscr{B}} \chi_{\mathscr{I}_i} * p(\boldsymbol{z}|\mathbf{0}, \Sigma_i) d\boldsymbol{z} + \int_{\mathbb{R}^m \setminus \mathscr{B}} \chi_{\mathscr{I}_i} * p_{n_i}(\boldsymbol{z}) d\boldsymbol{z} \\
&< \frac{\delta}{4N}.
\end{aligned} \tag{26}$$

我们已经证明，该卷积可以用高斯混合密度函数近似。下面我们将上述所有内容结合起来，完成证明的最后一步。

结合式（26）和式（25），我们推导得出

$$\int_{\mathbb{R}^m} \left| \chi_{\mathscr{I}_i}(\boldsymbol{z}) - \sum_{j=1}^{n_i} \underline{f}_i |\mathscr{I}_{ij}| p(\boldsymbol{z} - \boldsymbol{y}_{ij}|\mathbf{0}, \Sigma_i) \right| d\boldsymbol{z} < \frac{\delta}{2N}. \tag{27}$$

由于 $\boldsymbol{y}_{ij}$ 是确定的，因此 $p(\boldsymbol{z} - \boldsymbol{y}_{ij}|\mathbf{0}, \Sigma_i)$ 可视为一个高斯密度函数，其

均值为 $\boldsymbol{y}_{ij}$，协方差矩阵为 Σ_i。我们使用记号 $\boldsymbol{\mu}_{ij}$ 替换 $\boldsymbol{y}_{ij}$，于是 $p(\boldsymbol{z}-\boldsymbol{y}_{ij}|\boldsymbol{0},\Sigma_i)$ 可改写为 $p(\boldsymbol{z}|\boldsymbol{\mu}_{ij},\Sigma_i)$。由于不等式（27）对每个 $\mathscr{I}_i$、$i=1$、$\cdots$、N 都成立，因此有

$$\int_{\mathbb{R}^m}\left|\underline{f}_i\mathbf{1}_{\mathscr{I}_i}(\boldsymbol{z})-\sum_{j=1}^{n_i}\underline{f}_i|\mathscr{I}_{ij}|p(\boldsymbol{z}|\boldsymbol{\mu}_{ij},\Sigma_i)\right|d\boldsymbol{z}<\frac{\delta}{2N},\quad i=1,\cdots,N. \tag{28}$$

最后，记 $\pi_{ij}=\underline{f}_i|\mathscr{I}_{ij}|$和 $p^*(\boldsymbol{z})=\sum_{i=1}^N\sum_{j=1}^{n_i}\pi_{ij}p(\boldsymbol{z}|\boldsymbol{\mu}_{ij},\Sigma_i)$。则

$$\sum_{i=1}^N\sum_{j=1}^{n_i}\pi_{ij}=\sum_{i=1}^N\underline{f}_i|\mathscr{I}_i|=\int_{\mathbb{R}^m}f_N(\boldsymbol{z})d\boldsymbol{z}\to 1$$

如 $N\to+\infty$。此外，结合式 (20) 和 (28) 可得

$$\begin{aligned}
&\int_{\mathbb{R}^m}|f(\boldsymbol{z})-p^*(\boldsymbol{z})|d\boldsymbol{z}\\
&<\int_{\mathbb{R}^m}|f(\boldsymbol{z})-f_N(\boldsymbol{z})|d\boldsymbol{z}+\int_{\mathbb{R}^m}|f_N(\boldsymbol{z})-p^*(\boldsymbol{z})|d\boldsymbol{z}\\
&<\int_{\mathbb{R}^m}|f(\boldsymbol{z})-f_N(\boldsymbol{z})|d\boldsymbol{z}+\sum_{i=1}^N\int_{\mathbb{R}^m}\left|\underline{f}_i\mathbf{1}_{\mathscr{I}_i}(\boldsymbol{z})-\sum_{j=1}^{n_i}\underline{f}_i|\mathscr{I}_{ij}|p(\boldsymbol{z}|\boldsymbol{\mu}_{ij},\Sigma_i)\right|d\boldsymbol{z}\\
&<\frac{\delta}{2}+\sum_{i=1}^N\frac{\delta}{2N}=\delta
\end{aligned}$$

证明完毕。 □

定理 3 表明，条件数约束GMD仍具有拟合能力。值得一提的是，尽管可以降低基础适应度误差和渐近近似误差，但代价可能是分量数量增加。

3 GMD下QCCP的重构与算法

本节将在 $c(\boldsymbol{\xi},\boldsymbol{x})$ 的分布可以由相应的渐近GMD良好近似这一条件下，通过将原问题重构为一个易于求解的问题，提出求解方法。具体而言，根据定理1，$c(\boldsymbol{\xi},\boldsymbol{x})$ 的密度函数近似为

$$p(z)=\sum_{i=1}^K\pi_i p\left(z|\mu_i(\boldsymbol{x}),\sigma_i^2(\boldsymbol{x})\right),$$

其中 $\mu_i(\boldsymbol{x})=E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))$，且 $\sigma_i^2(\boldsymbol{x})=V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))$。

根据 Hu 等人（2022），问题 QCCP 可等价地表述为

$$\begin{aligned} \text{QCCP}_0 \quad \min_{\boldsymbol{x}\in\mathscr{X}} \quad & \rho(\boldsymbol{x}) \\ \text{s.t.} \quad & \sum_{i=1}^{K}\pi_i\Phi\left(-\frac{\mu_i(\boldsymbol{x})}{\sigma_i(\boldsymbol{x})}\right)\geq 1-\alpha, \end{aligned}$$

其中，$\Phi(\cdot)$ 是标准正态分布的累积分布函数。通过引入辅助向量 $\boldsymbol{y}$，上述问题可进一步等价地重构为

$$\begin{aligned} \text{QCCP}_1 \quad \min_{\boldsymbol{x}\in\mathscr{X},\boldsymbol{y}\in\mathscr{Y}} \quad & \rho(\boldsymbol{x}) \\ \text{s.t.} \quad & \Phi^{-1}(y_i)\sigma_i(\boldsymbol{x})+\mu_i(\boldsymbol{x})\leq 0,\ i=1,\cdots,K, \\ & \sum_{i=1}^{K}\pi_i y_i\geq 1-\alpha, \end{aligned}$$

其中，$\mathscr{Y}=\{\boldsymbol{y}\in\mathbb{R}^K\colon\ 0\leq y_i\leq 1, i=1,\cdots,K\}$。

为得到局部最优解，可采用一些一阶算法求解 QCCP_1。下文考虑 $A(\boldsymbol{x})$、$\boldsymbol{a}(\boldsymbol{x})$ 和 $a(\boldsymbol{x})$ 的线性形式，并研究其全局最优解，即

$$A(\boldsymbol{x})=A_0+\sum_{i=1}^{n}x_iA_i,\ \boldsymbol{a}(\boldsymbol{x})=\boldsymbol{a}_0+\sum_{i=1}^{n}x_i\boldsymbol{a}_i \text{ and } a(\boldsymbol{x})=a_0+\sum_{i=1}^{n}x_ia_i. \tag{29}$$

随后，引入 Zhu 等人（2020）提出的一个引理来重构 QCCP_1。

引理 1. 若 $A(\boldsymbol{x})$、$\boldsymbol{a}(\boldsymbol{x})$ 和 $a(\boldsymbol{x})$ 满足式（29）的形式，则

$$\begin{aligned} E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \boldsymbol{\nu}_i^\top\widehat{\boldsymbol{x}}, \\ V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \widehat{\boldsymbol{x}}^\top\left(\Psi_i+\frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}, \end{aligned}$$

其中

$$\begin{aligned} \widehat{\boldsymbol{x}} &= (1;\boldsymbol{x}), \\ \boldsymbol{\nu}_i &= \left(\frac{1}{2}\text{tr}(A_j\Sigma_i)+\frac{1}{2}\boldsymbol{\mu}_i^\top A_j\boldsymbol{\mu}_i+\boldsymbol{a}_j^\top\boldsymbol{\mu}_i+a_j\right)_{j=0,\cdots,n}, \\ \Psi_i &= (A_0\boldsymbol{\mu}_i+\boldsymbol{a}_0,\cdots,A_n\boldsymbol{\mu}_i+\boldsymbol{a}_n)^\top\Sigma_i(A_0\boldsymbol{\mu}_i+\boldsymbol{a}_0,\cdots,A_n\boldsymbol{\mu}_i+\boldsymbol{a}_n), \\ \Phi_i &= (\text{tr}(A_k\Sigma_iA_l\Sigma_i))_{k,l=0,\cdots,n}. \end{aligned}$$

根据引理 1，问题 QCCP_1 等价于

$$\begin{aligned}\mathrm{QCCP}_{\mathrm{lin}} \quad \min_{\boldsymbol{x}\in\mathscr{X},\boldsymbol{y}\in\mathscr{Y}} \quad & \rho(\boldsymbol{x}) \\ \text{s.t.} \quad & \Phi^{-1}(y_i)\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i+\frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}}+\boldsymbol{\nu}_i^\top\widehat{\boldsymbol{x}}\le 0,\ i=1,\cdots,K, \\ & \sum_{i=1}^{K}\pi_i y_i\ge 1-\alpha.\end{aligned}$$

一般而言，问题 $\mathrm{QCCP}_{\mathrm{lin}}$ 是一个非凸规划。对于具有这种结构的问题，Hu 等人（2022）提出了一种空间分支定界（BB）算法，用以求得全局最优解，相关内容见附录 D。简而言之，BB算法在立方体 $\mathscr{Y}$ 划分出的子域上对约束进行松弛，以求解更简单的问题，并在每次迭代中为该子域生成一个下界。一旦松弛问题的最优值小于当前全局最优值，且松弛问题的最优解也满足原问题的可行性，就会更新原问题的全局解，并剪除局部下界大于全局下界的分支。通过在细分后的立方体上求解一系列松弛问题，这一迭代过程会生成一个非递增序列，并收敛至全局最优值；后文将对此进行详细讨论。

假设 $\widetilde{\mathscr{Y}}=\{\boldsymbol{y}\in\mathbb{R}^K:\ l_i\le y_i\le u_i, i=1,\cdots,K\}$ 是一个细分立方体。通过求解下述松弛问题，得到主问题$\widetilde{\ }$$\mathrm{QCCP}_{\mathrm{lin}}$ 在 $\mathscr{Y}$ 上的下界。

$$\begin{aligned}\mathrm{QCCP}_{\mathrm{relax}} \quad \min_{\boldsymbol{x}\in\mathscr{X},\boldsymbol{y}\in\widetilde{\mathscr{Y}}} \quad & \rho(\boldsymbol{x}) \\ s.t. \quad & \Phi^{-1}(l_i)\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i+\frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}}+\boldsymbol{\nu}_i^\top\widehat{\boldsymbol{x}}\le 0,\ i=1,\cdots,K, \\ & \sum_{i=1}^{K}\pi_i y_i\ge 1-\alpha.\end{aligned}$$

Hu 等人（2022）指出，如果 $\pi_i\ge 2\alpha$ 对 $i=1,\cdots,K$成立，那么 $\mathrm{QCCP}_{\mathrm{relax}}$ 是一个凸规划。

在实际实现中，由于代码中的终止准则，我们只能通过算法获得一个 ϵ-最优解。假设 $(\boldsymbol{x},\boldsymbol{y})$ 是问题 $\mathrm{QCCP}_{\mathrm{relax}}$ 的一个可行解。如果 $h_i(\boldsymbol{x},\boldsymbol{y})-h_i^r(\boldsymbol{x},\boldsymbol{y})\le\epsilon, i=1,\cdots,K$，其中

$$\begin{aligned}h_i(\boldsymbol{x},\boldsymbol{y})&=\Phi^{-1}(y_i)\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i+\frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}}+\boldsymbol{\nu}_i^\top\widehat{\boldsymbol{x}},\\ h_i^r(\boldsymbol{x},\boldsymbol{y})&=\Phi^{-1}(l_i)\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i+\frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}}+\boldsymbol{\nu}_i^\top\widehat{\boldsymbol{x}},\end{aligned}$$

那么我们称 $(\boldsymbol{x},\boldsymbol{y})$ 为 QCCP_{lin} 的一个 ϵ-可行解。如果 $(\boldsymbol{x},\boldsymbol{y})$ 是最优解且满足$h_i(\boldsymbol{x},\boldsymbol{y}) - h_i^r(\boldsymbol{x},\boldsymbol{y}) \leq \epsilon, i=1,\cdots,K$，则称其为 ϵ-最优解。值得一提的是，如果解 $(\boldsymbol{x},\boldsymbol{y})$ 是 ϵ-可行的，则表明 $0<y_i<1, i=1,\cdots,K$。否则，如果$y_i=1$, $h_i(\boldsymbol{x},\boldsymbol{y}) - h_i^r(\boldsymbol{x},\boldsymbol{y}) = +\infty$，这与 ϵ-可行的条件相矛盾。如果 $y_i=0$，则$h_i(\boldsymbol{x},\boldsymbol{y})$ 和 $h_i^r(\boldsymbol{x},\boldsymbol{y})$ 均等于 $-\infty$。在实际实现中，处理两个无穷大之差仍不方便。因此，在实际实现中，我们用近似集 $\mathscr{Y}^0 = \{\boldsymbol{y}\in\mathbb{R}^K$ 替换可行集 $\mathscr{Y}$: $\underline{y}_i \leq y_i \leq \overline{y}_i, i=1,\cdots,K\}$ 作为算法的输入，其中 $0<\underline{y}_i \leq \overline{y}_i<1$，且 $|\underline{y}_i|$ 和 $|1-\overline{y}_i|$ 是足够小的数。

命题 3。假设 $\mathscr{X}$ 是紧的，且算法1的初始立方体 $\mathscr{Y}^0$ 满足$0<\underline{y}_i \leq \overline{y}_i<1$, $i=1,\cdots,K$。为了生成 QCCP_{lin} 的 ϵ-最优解或验证其不可行性，所需求解的松弛子问题 $\text{QCCP}_{\text{relax}}$ 数量至多为

$$\left\lfloor \frac{x_{\max} d^*}{\epsilon}[y_{k^*}]\sqrt{\lambda^*} \right\rfloor^K,$$

其中，$\lfloor\cdot\rfloor$ 表示小于或等于给定数值的最大整数，$x_{\max} = \max_{\boldsymbol{x}\in\mathscr{X}} \{||\widehat{\boldsymbol{x}}||\}$,$[y_{k^*}] = \max \{\overline{y}_i - \underline{y}_i, i=1,\cdots,K\}$, $d^* = \max\left\{ \frac{1}{\phi\ \Phi^{-1}(\underline{y}^*)}, \frac{1}{\phi\ \Phi^{-1}(\overline{y}^*)} \right\}$, $\phi(\cdot)$ 是标准正态密度函数，$\underline{y}^* = \min\left\{\underline{y}_i, i=1,\cdots,K\right\}$, $\overline{y}^* = \max\ \{\overline{y}_i, i=1,\cdots,K\}$，而 λ^* 是 $\Psi_i + \frac{1}{2}\Phi_i$, $i=1,\cdots,K$中的最大特征值。

证明。对于任意 $\boldsymbol{x}\in\mathscr{X}$ 和 $\boldsymbol{y}\in\mathscr{Y}^0$,

$$\begin{aligned} h_i(\boldsymbol{x},\boldsymbol{y}) - h_i^r(\boldsymbol{x},\boldsymbol{y}) &= \left(\Phi^{-1}(y_i) - \Phi^{-1}(\underline{y}_i)\right)\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i + \frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}} \\ &= (y_i - \underline{y}_i)\left.\frac{d}{dy}\Phi^{-1}(y)\right|_{y=y_i^0}\sqrt{\widehat{\boldsymbol{x}}^\top\left(\Psi_i + \frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}} \\ &\leq (\overline{y}_i - \underline{y}_i) d^* x_{\max}\sqrt{\lambda^*}, \end{aligned}$$

其中 $y_i^0 \in [\underline{y}_i, \overline{y_i}]$。最后一个不等式成立，因为 $\Phi^{-1}(y)$ 的导数为 $\frac{1}{\phi\ \Phi^{-1}((y))}$，且 $\lambda^* I - \left(\Psi_i + \frac{1}{2}\Phi_i\right)$ 是半正定的，因此 $\lambda^* x_{\max}^2 \geq \lambda^* ||\widehat{\boldsymbol{x}}||^2 \geq \widehat{\boldsymbol{x}}^\top\left(\Psi_i + \frac{1}{2}\Phi_i\right)\widehat{\boldsymbol{x}}$。

假设立方体 $\widetilde{\mathscr{Y}} = \prod_{i=1}^K [l_i, u_i]$ 满足

$$u_i - l_i \leq \frac{\epsilon}{d^* x_{\max}\sqrt{\lambda^*}},\ i=1,\cdots,K.$$

如果问题 $\text{QCCP}_{\text{relax}}$ 在 $\mathscr{Y}$ 上可行，记其最优解为 $(\widetilde{\boldsymbol{x}}^*, \widetilde{\boldsymbol{y}}^*)$。由于$h_i(\widetilde{\boldsymbol{x}}^*, \widetilde{\boldsymbol{y}}^*) - h_i^r(\widetilde{\boldsymbol{x}}^*, \widetilde{\boldsymbol{y}}^*) \leq \epsilon$，该解是在 $\mathscr{Y}$ 上的一个 ϵ-最优解。如果问题 $\text{QCCP}_{\text{relax}}$ 在 $\mathscr{Y}$ 上不可行，则问题 QCCP_{lin} 在 $\mathscr{Y}$ 上必定不可行。在这两种情况下，$\mathscr{Y}$ 都不会被划分为两个更小的立方体。

根据算法，初始立方体 $\mathscr{Y}^0$ 将从

最长边的中点处划分为两个较小的立方体，前提是最长边大于 $\frac{\epsilon}{d^* x_{\max}\sqrt{\lambda^*}}$。现在考虑长度为 $[y_i] = \overline{y}_i - \underline{y}_i \gg \frac{\epsilon}{d^* x_{\max}\sqrt{\lambda^*}}$的立方体的第 i 条边。根据中点切分规则，要生成这条边长度等于 $\left(\frac{1}{2}\right)^1 [y_i]$的所有实例，只需划分 $2^0 = 1$ 次。如果这条边上的所有实例都等于 $\left(\frac{1}{2}\right)^2 [y_i]$，则需要划分 $2^0 + 2^1 = 3$ 次。不难得出，划分 $2^n - 1$ 次即可获得这条边长度等于 $\left(\frac{1}{2}\right)^n [y_i]$ 的所有实例。

令 $n^* = \log_2 \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon} + 1 \right\rfloor$。那么，边长为 ith 的实例数等于

$$[y_i] \left(\frac{1}{2}\right)^{\log_2 \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon}+1 \right\rfloor} = [y_i] \frac{1}{\left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon} + 1 \right\rfloor} < [y_i] \frac{1}{\frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon}} = \frac{\epsilon}{d^* x_{\max}\sqrt{\lambda^*}}.$$

因此，不再划分这条边，对应的划分数为

$$2^{n^*} - 1 = \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon} + 1 \right\rfloor - 1 = \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_i]}{\epsilon} \right\rfloor.$$

注意，每次划分对应于一次 $\text{QCCP}_{\text{relax}}$问题的计算。这意味着划分次数等于 $\text{QCCP}_{\text{relax}}$的计算次数。因此，求解的 $\text{QCCP}_{\text{relax}}$ 问题数至多为

$$\left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_1]}{\epsilon} \right\rfloor \times \cdots \times \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_K]}{\epsilon} \right\rfloor \leq \left\lfloor \frac{d^* x_{\max}\sqrt{\lambda^*}[y_{k^*}]}{\epsilon} \right\rfloor^K.$$

证明完毕。 □

命题 3 给出了 BB算法中所求解的 QCCP_{lin}子问题数量的最坏情况估计。算法的复杂度关于分量数 K呈指数级。幸运的是， K 通常是一个较小的数。此外，由于算法只考虑与超平面 $\sum_{i=1}^K \pi_i y_i = 1 - \alpha$ 相交的子立方体，因此需要求解的子问题数远少于命题 3 中的最坏情况估计。

4 数值模拟

本节首先进行数值实验，以验证理论结果。随后测试BB算法的效率。

4.1 渐近分布条件检验

我们从方程（13）出发，通过直接比较 Z_h 的密度函数与相应的理论渐近密度函数，测试均值 α_i和权重 ω_i对渐近误差的影响。

我们从 $h=10$ 个相互独立的单位方差正态分布变量中模拟生成 20000 个样本，并设定不同的均值，然后计算这些样本的平方加权和，以获得 Z_h的样本。 Z_h 的样本直方图由一条平滑曲线近似表示，该曲线表示 Z_h的密度函数。通过设定 $K=1$，可直接根据定理1推导渐近分布。图1展示了三个子图，分别采用不同的均值和权重设置：左图中为 $\alpha_i=0$、 $i=1$、$\cdots$、h和$\omega_i=\frac{1}{2^i}$、 $i=1$、$\cdots$、h；中图中为 $\alpha_i=0$、 $i=1$、$\cdots$、h和 $\omega_i=1$、 $i=1$、$\cdots$、m；右图中α_i、 $i=1$、$\cdots$、h从(0,10)均匀抽样，且 $\omega_i=\frac{1}{2^i}$、 $i=1$、$\cdots$、h。可以看出，当均值非零时，即使所有随机变量的权重不相等，渐近近似误差仍然很小。与权重不相等的情况相比，权重相等时渐近近似误差会减小。这些发现与定理2的结果一致。

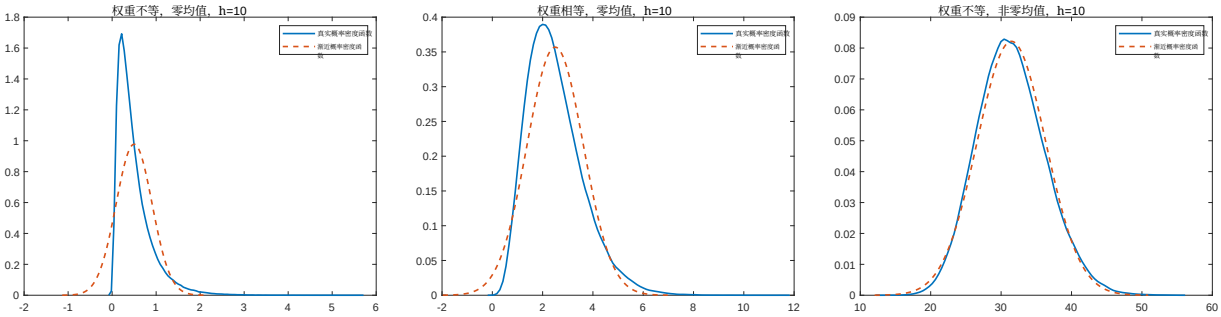


图1：不同因素对渐近近似误差的影响（含 $K=1$的 GMD）

随后，我们通过其他数值实验直接验证定理1的结果，再次采用比较密度函数的方法。我们设定 $K=3$、 $A=\mathrm{I}$、 $\boldsymbol{a}=\mathbf{0}$ 和 $a=0$。每个分量的协方差矩阵由 $\Sigma_i=D\Lambda D^{\top}$生成，其中 D 是随机生成的正交矩阵， Λ 是一个对角矩阵，其元素在(0,1)范围内均匀生成。混合权重 π_i、$i=1$、$\cdots$、K均匀生成后再进行归一化。

结果如图 2 所示。比较左图和中图可见，随着 m 增大， $c(\boldsymbol{\xi},\boldsymbol{x})$ 的真实分布逐渐收敛于其渐近分布。在右图中，我们将 GMD 分量的均值设为非零。具体而言，每个 $\boldsymbol{\mu}_i$ 均从 (0,10) 中均匀生成。结果表明，非零均值情形的渐近近似误差小于零均值情形。

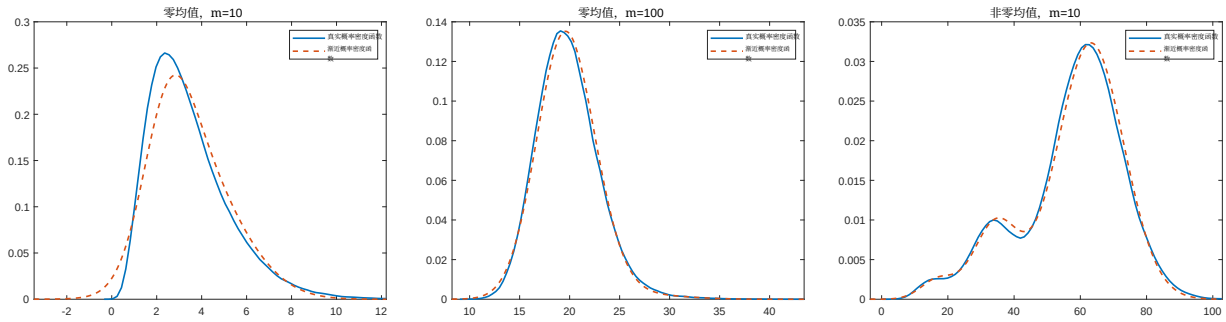


图 2：不同因素对渐近近似误差的影响（GMD， $K=3$）

最后，我们通过仿真验证定理 2、命题 2 和定理 3 的含义：限制 $\Sigma_i,\ i=1,\cdots,K$ 的条件数，可以同时降低基础适应度误差和渐近近似误差。仿真步骤如下。首先，从服从特定分布的 30 维随机向量 $\boldsymbol{\xi}$ 中随机生成 20000 个样本，作为比较基准；该分布在现实中通常是未知的。然后，利用这些样本估计一个 GMD，或一个带有条件数约束的 GMD。记服从估计所得 GMD 的随机向量为 $\boldsymbol{\xi}$。接下来，使用蒙特卡洛方法绘制 $c(\boldsymbol{\xi},\boldsymbol{x})\hat{}$ 的密度函数和 $c(\boldsymbol{\xi},\boldsymbol{x})$ 的密度函数。通过比较这两个密度函数，我们可以检验基础适应度误差。最后，根据定理1绘制 $c(\boldsymbol{\xi},\boldsymbol{x})$ 的渐近密度函数，并通过将其与 $c(\boldsymbol{\xi},\boldsymbol{x})\hat{}$ 的密度函数进行比较，评估渐近近似误差。

更具体地说，第一步，我们将 $\boldsymbol{\xi}$ 的分布指定为具有 $K=2$个分量的 GMD。对于第 i个分量，$\boldsymbol{\mu}_i$ 在(0,1)上均匀生成，$\Sigma_i=LL^{\top}$，其中 L 的元素在(0.5,1.5)上均匀生成，混合权重 π_i 则随机生成。在图3所示的模拟结果中，协方差矩阵的条件数分别为436510和1742000。第二步，使用可参考 McLachlan and Krishnan (1997) 的 EM 算法来估计 GMD 的参数。此外，$A=\mathrm{I}$，$\boldsymbol{a}$ 在(-100,100)上随机生成，并且$a=0$。

在图3中，“真实概率密度函数”、“估计概率密度函数”和“渐近概率密度函数”分别表示 $c(\boldsymbol{\xi},\boldsymbol{x})$、$c(\boldsymbol{\xi},\boldsymbol{x})$ 的密度函数以及 $c(\boldsymbol{\xi},\boldsymbol{x})$ 的理论渐近分布。在左侧子图中，我们在第二步使用无约束的 GMD。含有两个分量的估计密度能够很好地拟合真实密度，但渐近密度与估计密度相差甚远。这表明基础适应度误差较小，但渐近近似误差较大。如中间子图所示，当我们对条件数施加约束时，估计密度与真实密度相差较大，却接近其渐近密度，这表明基础适应度误差较大，但渐近近似误差较小。右侧子图显示，随着

分量数量增加，基于条件数受约束的 GMD 得到的渐近分布能够很好地逼近真实分布，这表明基础适应度误差和渐近近似误差同时减小。

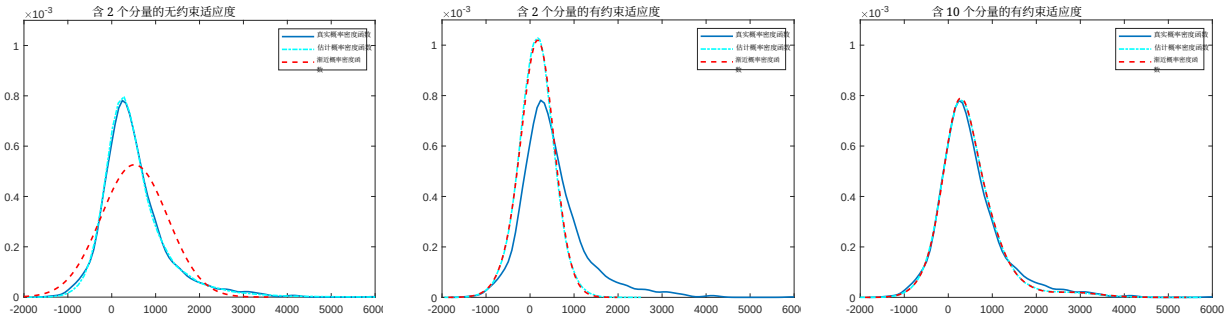


图3：条件数约束GMD的拟合能力

4.2 BBAlgorithm测试

在本小节中，我们通过一些数值示例来测试BB算法的效率。我们将目标函数设为线性函数 $\rho(\boldsymbol{x}) = b_0 + \sum_{i=1}^{n} b_i x_i$，其中 b_i、$i = 0$、$\cdots$、n均从(-5,5)中均匀生成。为了构造非凸QCCP$_{\text{lin}}$，对于第 i个分量，$\boldsymbol{\mu}_i$的各项从$(9i-8, 9i+1)$中随机生成，并通过 $LL^\top$计算协方差矩阵，其中 L 的各项从$(i, i+1)$中随机生成。此外，我们设定 $m = n$、$A_i = O$、$i = 1$、$\cdots$、n，$\boldsymbol{a}_i$的第 i 项为-1，其余项为0，$i = 1$、$\cdots$、n、$\boldsymbol{a}_0 = -\mathbf{1}$，以及 $a_i = 0$、$i = 0$、$\cdots$、n。对于决策向量 $\boldsymbol{x}$，令 $x_1 = 1$，其他项在 [-100,100]之间变化。置信水平设为 $\alpha = 0.05$。

我们比较 BB算法 与 MATLAB 中的优化函数“fmincon”的效率。在“fmincon”的选项中，将算法设置为“SQP”，并将函数最大求值次数和最大迭代次数均设为 30000。对于 BB算法，如果下界与当前最优值之间的相对误差小于 10^{-2}，则终止算法。如果某个分支的下界 $\boldsymbol{l}$ 和上界 $\boldsymbol{u}$ 对于 $\boldsymbol{y}$ satisfies$\|\boldsymbol{u} - \boldsymbol{l}\|/\|\boldsymbol{l}\| \leq 10^{-2}$，则不再进一步划分该分支。我们使用“fmincon”得到的最优解作为 BB算法的初始点。

表1和表2展示了不同问题规模设置下模拟问题的最优值和计算时间。在这项模拟中，混合权重是随机生成的，并满足 $\pi_i \geq \alpha$, $i = 1, \cdots, K$，因此子问题是凸的。对于每种问题规模，实验均执行5次，“Max”“Ave”和“Min”分别表示五次实验中的最大值、平均值和最小值。由于BB算法的子问题均为凸问题，BB算法可以在较短时间内获得全局最优解。此外，我们可以看到，与BB算法相比，“fmincon”耗时更少，但只能获得局部最优解。

表1：BB 和 Fmincon 得到的最优值：凸子问题

		BB			Fmincon		
K	n	Max	Ave	Min	Max	Ave	Min
2	50	−611.03	−3341.81	−6345.70	−407.19	−2962.81	−5694.38
	100	−4493.19	−8268.63	−12758.77	−3920.77	−7743.47	−12193.84
	200	−12522.61	−18967.46	−35916.11	−10690.08	−17122.09	−33868.77
3	50	−989.28	−2324.47	−4538.31	−72.49	−1618.60	−4152.89
	100	−7722.69	−12366.06	−17369.61	−6550.45	−10956.14	−16073.30
	200	−8932.67	−17104.38	−33098.43	−5949.81	−14680.56	−30000.16
4	50	−2980.62	−4949.57	−7741.34	−1927.50	−4065.83	−6743.16
	100	−6094.91	−10166.83	−18501.75	−4537.32	−8533.31	−16319.39
	200	−8143.13	−18400.38	−31427.90	−3281.08	−14422.20	−28218.68

表2：BB 和 Fmincon 的计算时间：凸子问题（秒）

		BB			Fmincon		
K	n	Max	Ave	Min	Max	Ave	Min
2	50	1.10	0.96	0.83	0.70	0.57	0.45
	100	3.05	2.77	2.53	2.02	1.67	1.44
	200	15.08	13.34	10.58	9.40	8.18	5.83
3	50	7.59	4.08	2.04	0.84	0.54	0.36
	100	17.80	9.88	5.35	6.24	2.73	1.54
	200	77.37	58.30	23.77	12.46	11.66	10.09
4	50	12.87	7.56	3.44	0.69	0.62	0.48
	100	86.15	40.69	9.58	2.82	2.01	1.65
	200	297.62	184.20	100.57	15.72	12.66	10.42

随后，我们在其他参数设置保持不变的情况下执行了其他实验，但 $\pi_i,\ i = 1, \cdots, K$ 除外；这些参数是随机生成的，且至少有一个 π_i 满足 $\pi_i < 2\alpha$。在这种情况下，BB算法的子问题是非凸的。表3和表4展示了最优值及相应的计算时间。可以看到，尽管问题规模更小，BB算法的计算时间仍远远长于子问题为凸问题时的计算时间。

表3：BB 和 Fmincon 求得的最优值：非凸子问题

		BB			Fmincon		
K	n	Max	Ave	Min	Max	Ave	Min
2	10	19.01	−797.56	−2441.26	20.66	−784.15	−2441.26
	20	−876.37	−3886.61	−5533.25	−845.59	−3871.81	−5533.25
	30	−42.20	−2987.99	−6297.43	−9.35	−2825.91	−6253.98
3	10	−923.75	−1560.27	−2265.11	−789.21	−1498.22	−2265.11
	20	−3602.49	−4117.10	−4488.56	−3231.04	−3979.91	−4488.56
	30	−1816.84	−3128.75	−5495.17	−1666.50	−2911.59	−5407.16
4	10	−1016.62	−1468.16	−1760.37	−961.03	−1428.68	−1739.06
	20	−47.44	−1981.41	−4281.85	−25.34	−1884.08	−4177.93
	30	−1691.45	−5069.70	−6711.06	−1356.68	−4767.04	−6610.53

表4：BB 和 Fmincon 的计算时间：非凸子问题（秒）

		BB			Fmincon		
K	n	Max	Ave	Min	Max	Ave	Min
2	10	250.94	156.82	8.88	0.33	0.17	0.06
	20	563.30	157.49	0.43	0.20	0.16	0.10
	30	838.02	556.79	363.75	3.28	0.92	0.32
3	10	75.74	33.84	0.29	0.10	0.08	0.06
	20	1411.34	542.89	19.10	0.14	0.14	0.13
	30	8656.94	3030.72	724.30	0.74	0.39	0.30
4	10	568.24	196.93	0.34	0.20	0.13	0.10
	20	1047.51	445.32	156.69	1.42	0.60	0.26
	30	27044.59	8830.70	748.77	0.48	0.43	0.38

5 结论

本文研究了具有二次随机性的机会约束规划，其中构成二次随机性的随机变量假设服从 GMD。求解该问题的关键发现是，在某些温和条件下，二次随机性的渐近分布是一元广义多元分布，可用于近似二次随机性的分布，从而能够采用有效的 BB 方法求得全局解。此外，得益于 GMD 的拟合能力，所提出的方法是一种统一方法，可应用于许多不同情形。

本文的一个有趣发现是，在实际应用中使用条件数约束GMD不仅可以降低渐近近似误差，还能很好地学习数据的特征。然而，这种约束GMD的代价是

需要更多分量，从而导致BB算法需要更多计算时间。尽管我们已经证明，条件数约束GMD的密度函数可以以任意精度逼近任意密度函数，但在给定逼近精度的情况下究竟需要多少个分量，目前仍不清楚。此外，还需要进一步研究如何高效估计约束GMD的参数。我们相信，这些问题值得在未来深入探索。

附录

附录A：命题1的证明

我们首先介绍 Zhu 等人（2020）中的一个相关引理。

引理2。给定决策向量 $\boldsymbol{x}$,

$$\begin{aligned} E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \frac{1}{2}\mathrm{tr}(A\Sigma_j)+\frac{1}{2}\boldsymbol{\mu}_j^\top A\boldsymbol{\mu}_j+\boldsymbol{a}^\top\boldsymbol{\mu}_j+a \\ V_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x})) &= \frac{1}{2}\mathrm{tr}((A\Sigma_j)^2)+(A\boldsymbol{\mu}_j+\boldsymbol{a})^\top\Sigma_j(A\boldsymbol{\mu}_j+\boldsymbol{a}) \end{aligned}$$

分别。这里，tr(·) 是一个矩阵的迹。

证明。对于均值，我们有

$$E_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x}))=\sum_{i=1}^{K}\pi_i E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})).$$

对于方差，我们有

$$\begin{aligned} V_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})) &= E_{p^*}(c^2(\boldsymbol{\xi},\boldsymbol{x}))-E^2_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})) \\ &= \sum_{i=1}^{K}\pi_i E_{p_i}(c^2(\boldsymbol{\xi},\boldsymbol{x}))-E^2_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})) \\ &= \sum_{i=1}^{K}\pi_i(E^2_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))+V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})))-E^2_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x})). \end{aligned}$$

注意到

$$
\begin{aligned}
&\sum_{i=1}^{K}\pi_i E_{p_i}^2(c(\boldsymbol{\xi},\boldsymbol{x}))-E_{p^*}^2(c(\boldsymbol{\xi},\boldsymbol{x}))\\
&=\sum_{i=1}^{K}\pi_i E_{p_i}^2(c(\boldsymbol{\xi},\boldsymbol{x}))-(\sum_{i=1}^{K}\pi_i E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x})))^2\\
&=\sum_{i=1}^{K}\pi_i(1-\pi_i)E_{p_i}^2(c(\boldsymbol{\xi},\boldsymbol{x}))-\sum_{i\neq j}\pi_i\pi_j E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x}))\\
&=\sum_{i=1}^{K}\pi_i(\sum_{j\neq i}\pi_j)E_{p_i}^2(c(\boldsymbol{\xi},\boldsymbol{x}))-\sum_{i\neq j}\pi_i\pi_j E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x}))\\
&=\sum_{1\leq i<j\leq K}\pi_i\pi_j(E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))-E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x})))^2.
\end{aligned}
$$

因此，我们有

$$
V_{p^*}(c(\boldsymbol{\xi},\boldsymbol{x}))\;=\;\sum_{i=1}^{K}V_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))+\sum_{1\leq i<j\leq K}\pi_i\pi_j(E_{p_i}(c(\boldsymbol{\xi},\boldsymbol{x}))-E_{p_j}(c(\boldsymbol{\xi},\boldsymbol{x})))^2.
$$

证明完毕。 □

附录 B：L´évy 连续性引理

引理 3. 对于一列随机变量 $\{X_n\}_{n=1}^{+\infty}$ 和 X，$X_n \xrightarrow{d} X$ 当且仅当 $\lim_{n\to+\infty} E\left(e^{\mathrm{i}tX_n}\right)=E\left(e^{\mathrm{i}tX}\right)$ 对每个 $t\in\mathbb{R}$ 均成立。这里，$\xrightarrow{d}$ 表示依分布收敛。

附录 C：定理 3 的一个引理介绍

在介绍该引理之前，我们先回顾一些记号。由于该引理的结论对任意 $\mathscr{I}_i$、$i\in\{1,\cdots,N\}$ 均成立，不失一般性，我们将 Σ_i、$\mathscr{I}_i$ 和 $\underline{f}_i$ 改写为 Σ、$\mathscr{I}$ 和 $\underline{f}$。相应的记号为

$$
\begin{aligned}
p(\boldsymbol{z}) &= p(\boldsymbol{z}|\boldsymbol{0},\Sigma)=\frac{1}{(2\pi)^{\frac{m}{2}}|\Sigma|^{-\frac{1}{2}}}e^{-\frac{1}{2}\boldsymbol{z}^\top\Sigma^{-1}\boldsymbol{z}},\\
\chi_{\mathscr{I}}(\boldsymbol{z}) &= \underline{f}\mathbf{1}_{\mathscr{I}}(\boldsymbol{z})=\begin{cases}\underline{f} & \boldsymbol{z}\in\mathscr{I}\\ 0 & \text{otherwise}\end{cases},\\
\chi_{\mathscr{I}}*p(\boldsymbol{z}) &= \int_{\mathbb{R}^m}\chi_{\mathscr{I}}(\boldsymbol{y})p(\boldsymbol{z}-\boldsymbol{y})d\boldsymbol{y}=\underline{f}\int_{\mathscr{I}}p(\boldsymbol{z}-\boldsymbol{y})d\boldsymbol{y}.
\end{aligned}
$$

令 $\mathscr{I}_1$、$\cdots$、$\mathscr{I}_n$ 为 $\mathscr{I}$ 的划分，即 $\mathscr{I}$ 的任意元素都属于 $\mathscr{I}_j$、$j=1$、$\cdots$、n 中且仅属于其中一个。记函数 $\chi_{\mathscr{I}}*p_n(\boldsymbol{z})=\sum_{j=1}^{n}\underline{f}|\mathscr{I}_j|p(\boldsymbol{z}-\boldsymbol{y}_j)$，其中 $\boldsymbol{y}_j\in\mathscr{I}_j$、$j=1$、$\cdots$、$n$。记 $|\mathscr{I}_{\max}|=\max_{j\in\{1,\cdots,n\}}\{|\mathscr{I}_j|\}$。根据黎曼积分的定义，

$\chi_{\mathscr{I}} * p(\boldsymbol{z})$ 与 $\chi_{\mathscr{I}} * p_n(\boldsymbol{z})$ 之间的关系是，对于任意给定的 $\boldsymbol{z}$,

$$\lim_{|\mathscr{I}_{\max}|\to 0} \chi_{\mathscr{I}} * p_n(\boldsymbol{z}) = \chi_{\mathscr{I}} * p(\boldsymbol{z}).$$

下面的引理表明，这种收敛是一致收敛 .

引理 4. 如果 $\mathscr{B}$ 是一个紧集，那么对于任意给定的 $\epsilon > 0$，存在一个划分 $\mathscr{I}_1$、$\cdots$、$\mathscr{I}_n$ of $\mathscr{I}$，使得

$$\int_{\mathscr{B}} |\chi_{\mathscr{I}} * p(\boldsymbol{z}) - \chi_{\mathscr{I}} * p_n(\boldsymbol{z})| d\boldsymbol{z} < \epsilon.$$

证明。首先我们证明，对于任意给定的 $\epsilon > 0$，存在一个仅与 ϵ 有关的 δ，使得如果 $\boldsymbol{y}_1$、$\boldsymbol{y}_2 \in \mathbb{R}^m$、$|\boldsymbol{y}_1 - \boldsymbol{y}_2|_\infty < \delta$，其中 $|\cdot|_\infty$ 表示向量各个分量中的最大绝对值，则 $|p(\boldsymbol{z} - \boldsymbol{y}_1) - p(\boldsymbol{z} - \boldsymbol{y}_2)| < \frac{\epsilon}{\underline{f}|\mathscr{I}||\mathscr{B}|}$。

给定 $\boldsymbol{z}$ 时，$p(\boldsymbol{z} - \boldsymbol{y}) = p(\boldsymbol{y} - \boldsymbol{z})$ 是一个均值向量为 $\boldsymbol{z}$ 的高斯密度函数，它是一致连续的，因此根据定义，存在一个 $\delta(\boldsymbol{z})$，使得如果 $\boldsymbol{y}_1$、$\boldsymbol{y}_2 \in \mathbb{R}^m$、$|\boldsymbol{y}_1 - \boldsymbol{y}_2|_\infty < \delta(\boldsymbol{z})$，则

$$|p(\boldsymbol{z} - \boldsymbol{y}_1) - p(\boldsymbol{z} - \boldsymbol{y}_2)| < \frac{\epsilon}{\underline{f}|\mathscr{I}||\mathscr{B}|}.$$

考虑另一个点 $\boldsymbol{z}'$ 并令 $\Delta\boldsymbol{z} = \boldsymbol{z}' - \boldsymbol{z}$。若 $|\boldsymbol{y}_1 - \boldsymbol{y}_2|_\infty < \delta(\boldsymbol{z})$，则

$$|(\boldsymbol{y}_1 - \Delta\boldsymbol{z}) - (\boldsymbol{y}_2 - \Delta\boldsymbol{z})|_\infty < \delta(\boldsymbol{z})$$

进而

$$|p(\boldsymbol{z}' - \boldsymbol{y}_1) - p(\boldsymbol{z}' - \boldsymbol{y}_2)| = |p(\boldsymbol{z} - (\boldsymbol{y}_1 - \Delta\boldsymbol{z})) - p(\boldsymbol{z} - (\boldsymbol{y}_2 - \Delta\boldsymbol{z}))| < \frac{\epsilon}{\underline{f}|\mathscr{I}||\mathscr{B}|},$$

这意味着，给定 ϵ，存在一个关于 $\boldsymbol{z}$ 和 $\boldsymbol{y}$ 不变的 δ，使得若 $\boldsymbol{y}_1$，则有 $\boldsymbol{y}_2 \in \mathbb{R}^m$、$|\boldsymbol{y}_1 - \boldsymbol{y}_2|_\infty < \delta$,

$$|p(\boldsymbol{z} - \boldsymbol{y}_1) - p(\boldsymbol{z} - \boldsymbol{y}_2)| < \frac{\epsilon}{\underline{f}|\mathscr{I}||\mathscr{B}|}.$$

然后令 t 划分 $\mathscr{I}_1, \cdots, \mathscr{I}_n$ 满足：对于任意 $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathscr{I}_j, j = 1, \cdots, n, |\boldsymbol{y}_1 - \boldsymbol{y}_2|_\infty$ <

δ。于是我们有

$$\begin{aligned}
|\chi_{\mathscr{I}} * p(\boldsymbol{z}) - \chi_{\mathscr{I}} * p_n(\boldsymbol{z})| &= \left| \underline{f} \sum_{j=1}^{n} \int_{\mathscr{I}_j} (p(\boldsymbol{z}-\boldsymbol{y}) - p(\boldsymbol{z}-\boldsymbol{y}_j)) d\boldsymbol{y} \right| \\
&\leq \underline{f} \sum_{j=1}^{n} \int_{\mathscr{I}_j} |p(\boldsymbol{z}-\boldsymbol{y}) - p(\boldsymbol{z}-\boldsymbol{y}_j)| \, d\boldsymbol{y} \\
&\leq \underline{f} \sum_{j=1}^{n} \frac{|\mathscr{I}_j|\epsilon}{\underline{f}|\mathscr{I}||\mathscr{B}|} = \frac{\epsilon}{|\mathscr{B}|}.
\end{aligned}$$

最后，我们得到

$$\int_{\mathscr{B}} |\chi_{\mathscr{I}} * p(\boldsymbol{z}) - \chi_{\mathscr{I}} * p_n(\boldsymbol{z})| d\boldsymbol{z} < \epsilon.$$

证明完成。 □

附录 D：Hu 等人（2022）提出的分支定界算法

算法1：分支定界（BB）

输入：$i=0, v^*=+\infty, \mathscr{Y}^0=\{\boldsymbol{y}\in\mathbb{R}^K:\ \underline{y}_{\ i}\le y_i\le\overline{y}_i, i=1,\cdots,K\}$。

输出：全局最优解 $\boldsymbol{x}^*$ 和全局最优值 v^*。

1: 求解问题 $\text{QCCP}_{\text{relax}}$。

如果 $\text{QCCP}_{\text{relax}}$ 不可行，则问题 QCCP_{lin} 不可行，停止。

否则，$v_0:=$ 最优值，$(\boldsymbol{x}^c,\boldsymbol{y}^c):=$ 最优解。

如果 $\boldsymbol{x}^c$ 对 QCCP_{lin}可行，$v^*:=v_0$，$\boldsymbol{x}^*:=\boldsymbol{x}^c$，停止。

否则，$\Omega=\{(\mathscr{Y}^i,v_i)\}$。设置 $i:=i+1$。转到步骤 2。

结束如果

结束如果

2: 若 $\Omega=\emptyset$, $\boldsymbol{x}^*$ 是最优解，则停止。

否则，转到步骤 3。

结束 if

3: 从 Ω 中选取并移除一个使 v_i 最小的分支 $(\mathscr{Y}^i,v_i)$。将分支 $\mathscr{Y}^i$ 细分为 $\mathscr{Y}_1^i$ 和 $\mathscr{Y}_2^i$，方法是将 $\mathscr{Y}^i$ 的最长边在其中点处分开。记 $\mathscr{Y}_1^i$为上方矩形，它包含上角点 $(\overline{y}_1,\cdots,\overline{y}_K)$；记 $\mathscr{Y}_2^i$为下方矩形，它包含上角点 $(\underline{y}_1,\cdots,\underline{y}_K)$.

如果 $\sum_{j=1}^K\pi_j\underline{y}_j^1>1-\alpha$，其中（$\underline{y}_1^1$，$\cdots$，$\underline{y}_K^1$）是 $\mathscr{Y}_1^i$ 的下角点，则设置 $\underline{v}_2^i=v_i$ 并 Ω: $=\Omega\bigcup(\mathscr{Y}_2^i,\ \underline{v}_2^i)$ 。设置 $i:=i+1$。转到步骤 2。

否则，如果 $\sum_{j=1}^K\pi_j\overline{y}_j^2<1-\alpha$，其中（$\overline{y}_1^2$，$\cdots$，$\overline{y}_K^2$）是 $\mathscr{Y}_2^i$ 的上角点，则求解机会约束规划$_{\text{relax}}$ 其中 $\mathscr{Y}_1$。

若其不可行，Ω: $=\Omega$，$i:=i+1$，转到步骤 2。

否则，记最优值 $\underline{v}_1^i$ 并记最优解（$\boldsymbol{x}_1^i$，$\boldsymbol{y}_1^i$）

若 $\underline{v}_1^i>v^*$, $\Omega:=\Omega$，$i:=i+1$，转到步骤 2。

否则若 $\boldsymbol{x}_1^i$ 对 QCCP 可行$_{\text{lin}}$, $v^*:=\underline{v}_1^i$, $\boldsymbol{x}^*:=\boldsymbol{x}_1^i$，并从 Ω 中删除所有分支 $(\mathscr{Y}_i,v_i)$ 其中 $v_i\ge v^*$。设置 $i:=i+1$，转到步骤 2。

否则 $\Omega:=\Omega\bigcup(\mathscr{Y}_1^i,\underline{v}_1^i)$，$i:=i+1$，转到步骤 2。

结束 if

结束 if

否则令 $\underline{v}_2^i=v_i$。设 $\Omega=\Omega\bigcup(\mathscr{Y}_2,\underline{v}_2^i)$。求解问题 $\text{QCCP}_{\text{relax}}$，其中 $\mathscr{Y}_1$，并求得最优值 $\underline{v}_1^i$。

如果其不可行，$i:=i+1$，转到步骤 2。

否则，记最优值为 $\underline{v}_1^i$，最优解为 $(\boldsymbol{x}_1^i,\boldsymbol{y}_1^i)$。

如果 $\underline{v}_1^i>v^*$, $i:=i+1$，转到步骤 2。

否则，如果 $\boldsymbol{x}_1^i$ 对 QCCP_{lin} 可行，$v^*:=\underline{v}_1^i$, $\boldsymbol{x}^*:=\boldsymbol{x}_1^i$，并从 Ω 中删除所有实例 $(\mathscr{Y}_i,v_i)$ 满足 $v_i\ge v^*$。令 $i:=i+1$，转到步骤 2。

否则，$\Omega:=\Omega\bigcup(\mathscr{Y}_1^i,\underline{v}_1^i)$，$i:=i+1$，转到步骤 2。

结束 if

结束 if

结束 if

4:返回 $\boldsymbol{x}^*$, v^*.

参考文献

Calafiore, G., M. C. Campi. 2005. 不确定凸规划：随机化解与置信水平。《数学规划》102(1) 25–46. Calafiore, G. C., L. El Ghaoui. 2006. 关于分布鲁棒机会约束线性规划。《优化理论与应用杂志》130(1) 1–22. Campi, M. C., S. Garatti. 2008. 不确定凸规划中随机化解的精确可行性。SIAM 优化期刊 19 1211–1230. Campi, M. C., S. Garatti. 2011. 机会约束优化的抽样与舍弃方法：可行性与最优性。《优化理论与应用杂志》148 257–280. Charnes, A., W. W. Cooper, G. H. Symonds. 1958. 成本期限与确定性等价物：取暖油随机规划的一种方法。《管理科学》4 235–263. Chen, G., H. C. Zhang, Y. H. Song. 2022. 非高斯分布不确定性下的机会约束直流最优潮流。2022 IEEE 电力与能源学会年会。1–5. doi:10.1109/PESGM48719.2022.9916658. Chen, Z. P., S. Peng, J Liu. 2018. 数据驱动鲁棒机会约束问题：一种混合模型方法。《优化理论与应用杂志》179 1065–1085. Cui, X. T., S. S. Zhu, X. L. Sun, D. Li. 2013. 使用近似参数化风险价值的非线性投资组合选择。《银行与金融期刊》37(6) 2124–2139. El Ghaoui, L., M. Oks, F. Oustry. 2003. 最坏情形风险价值与鲁棒投资组合优化：一种锥规划方法。《运筹学》51(4) 543–556. Golub, G. H., C. F. Van Loan. 2013.《矩阵计算》。第4版。约翰斯·霍普金斯大学，巴尔的摩。Hanasusanto, G. A., V. Roitch, D. Kuhn, W. Wiesemann. 2017. 均值与离差信息下的模糊联合机会约束。《运筹学》65(3) 751–767. Henrion, R., A. Moller. 2012. 高斯分布下线性机会约束的梯度公式。《运筹数学》37(3) 475–488. Hu, Z. L., W. J. Sun, S. S. Zhu. 2022. 高斯混合模型下的机会约束规划。《IISE 汇刊》54(12) 1117–1130. Hull, J. 2009.《期权、期货及其他衍生品》。培生普伦蒂斯·霍尔，新泽西州。Jorion, P. 2007.《风险价值：金融风险管理的新基准》。McGraw-Hill Companies，纽约。

Kishida, M., M. Nagahara. 2023. 使用最坏情形条件风险价值的风险感知最大无操作控制。IEEE 自动控制汇刊 doi:10.1109/TAC.2023. 3235246. Luedtke, J., S. Ahmed. 2008. 带概率约束优化的样本近似方法。SIAM 优化期刊 19 674–699. Marron, J. S., M. P. Wand. 1992. 精确均值积分平方误差。统计年刊 20 712–736. McLachlan, G., D Peel. 2000. 《有限混合模型》。威利出版社，纽约。McLachlan, G. J., T. Krishnan. 1997. 《EM 算法及其扩展》。约翰·威利父子出版公司，纽约。Miller, L. B., H. Wagner. 1965. 带联合约束的机会约束规划。《运筹学》13 930–945. Nemirovski, A., A. Shapiro. 2006. 机会约束规划的凸近似。SIAM 优化期刊 17(4) 969–996. Pearson, K. 1894. 对进化数学理论的贡献。伦敦皇家学会哲学汇刊 54 326–330. Pr´ekopa, A. 1970. 关于概率规划。载于《普林斯顿数学规划研讨会论文集》，普林斯顿大学出版社，普林斯顿，新泽西州，113–138. Ren, K., H. Ahn, M. Kamgarpour. 2022. 多模态环境不确定性下的机会约束轨迹规划。IEEE 控制系统快报 7 13–18. Shapiro, A., D. Dentcheva, A. Ruszczynski. 2009. 《随机规划讲义：建模与理论》。工业与应用数学学会，费城，宾夕法尼亚州。Van der Vaart, A. W. 1998. 《渐近统计学》。剑桥大学出版社。Wilson, R. 2000. MGMM：用于计算机视觉的多分辨率高斯混合模型。第15届国际模式识别会议论文集。ICPR-2000，第1卷。212–215. doi:10.1109/ICPR.2000.905305. Zhu, S. S., W. Zhu, X. Pei, X. T. Cui. 2020. 对最优投资组合选择中的崩盘风险进行对冲。《银行与金融期刊》119 105905. Zymler, S., D. Kuhn, B. Rustem. 2013. 基于二阶矩信息的分布鲁棒联合机会约束。《数学规划》137 167–198。