摘　要

在实际投资组合优化问题中，资产收益率受各种因素影响具有不确定性。为了更好地提供决策方案以及衡量投资风险，如何将不确定收益率考虑到投资组合优化问题中是我们需要研究的问题。鲁棒优化是处理模型中具有不确定参数的一种流行方法。当不确定参数看作是具有未知分布的随机变量时，处理参数分布信息模糊性的一个重要方法是分布鲁棒优化 (DRO)。然而，一个不可避免的问题是 DRO 的最优解是非常保守的，因为 DRO 在分布模糊集中采取了最坏的情况，其中包含了所有的可能分布。一个简单方法是直接降低分布模糊集的大小，但是这种方法并不能解决问题，因为真正的分布可能脱离所选择的分布模糊集。在 DRO 问题中，我们也可以引入全局化的思想，以降低 DRO 的保守性并且放松不确定集的可行性要求。

本文主要基于 Wasserstein 度量和线性锥强对偶定理，将具有不确定分布的全局分布鲁棒优化问题 (GDRO) 转化为确定分布下的对应问题，并将全局分布鲁棒优化模型应用到 CVaR 投资组合问题中，构建了全局分布鲁棒投资组合问题，并通过模型转化得到确定分布下的投资组合对应问题。最后，对全局分布鲁棒投资组合优化问题模型进行数值模拟，证明了模型的可行性，通过模型对比说明了模型在降低分布鲁棒投资组合优化问题模型的保守性上是有效的，且对投资决策有一定的指导性。

关键词：投资组合；全局分布鲁棒优化；Wasserstein 度量；CVaR；线性锥

Abstract

In practical problems of portfolio optimization, the rate of return on assets is uncertain under the influence of various factors. In order to provide a better decision plan and measure investment risk, we need to study how to take uncertain rate of return into portfolio optimization. Robust optimization is a popular method to deal with uncertain parameters in a model. When an uncertain parameter is viewed as a random variable with an unknown distribution, an important way to deal with ambiguity in parameter distribution information is distributionally robust optimization (DRO). One unavoidable problem, however, is that the optimal solution to DRO is very conservative, because DRO takes the worst case in an ambiguity that includes all possible distributions. A simple approach is to reduce the distribution ambiguity, but this approach does not solve the problem because the true distribution may be detached from the selected ambiguity. Then, in the DRO problem, we can also introduce a global method to reduce the conservativeness of DRO and relax the feasibility requirements of ambiguity.

Based on Wasserstein metric and the strong duality theorem of linear cone, this dissertation transforms the globalized distributionally robust optimization problem with uncertain distribution into the globalized distributionally robust optimization counterpart with deterministic distribution. We apply the globalized distributionally robust optimization model to the CVaR portfolio problem, and construct a globalized distributionally optimization portfolio problem. Through model transformation, we obtain the globalized distributionally robust portfolio counterpart under deterministic distribution. Finally, numerical simulation is carried out to prove the feasibility of the model, and the comparison of the models shows that the model is effective in reducing the conservativeness of the model, and has certain guidance for investment decision.

Key Words: Portfolio; Globalized Distributionally Robust Optimization; Wasserstein Metric; CVaR; Linear Cone

目 录

1. 绪论 1
1.1 研究背景及意义 1
1.2 研究方法与创新点 5
1.3 研究内容与论文框架 5
2. 预备知识 8
2.1 鲁棒优化模型 8
2.2 线性锥对偶 9
2.3 CVaR 理论 10
3. Wasserstein 度量下的全局分布鲁棒优化问题 13
3.1 全局分布鲁棒优化问题 (GDRC) 13
3.2 模糊集及 Wasserstein 度量 14
3.3 Wasserstein 度量下的分布鲁棒优化问题 15
3.4 Wasserstein 度量下的全局分布鲁棒优化问题 21
4. 全局分布鲁棒优化投资组合问题 31
4.1 分布鲁棒 CVaR 模型等价形式 31
4.2 全局分布鲁棒 CVaR 模型等价形式 32
4.3 随机数值模拟 34
4.4 真实数值模拟 38
5. 结论 42
附录 45
参考文献 47

1. 绪论

1.1 研究背景及意义

在社会的生产生活中，风险无处不在，尤其在金融投资领域，风险更加是大家无法避开的话题。对于投资者来说，除了要关注投资带来的收益，也要考虑到金融投资中存在的各种风险，这使得大家的投资选择受到了明显的限制。在投资组合的选择中，如何平衡收益最大化和风险最小化这个天平，是投资者需要考虑的问题。由此对投资组合的风险管理显得尤为重要，其中合理且有效的风险度量是风险管理是否行之有效的关键。

美国经济学家 Markowitz(1952) 发表了《Portfolio selection》，被大家公认奠定了现代投资组合理论的基础。在文章中，Markowitz 提出了风险是来自收益率的不确定性，投资收益率的不确定性越强，风险也就越大。由此，Markowitz 用数量化方法来刻画风险，即通过计算收益率的波动性大小来衡量风险的大小，并建立了最优投资组合问题的模型。该模型目标函数为方差的最小化，约束条件为不小于预期收益率。Wiliam (1963) 提出了“单一指数模型”。该模型假设股票收益率只受到单一因素影响，并用计算机程序实现了该模型。同时论证了低成本分析的可能性，再加上相对较少的信息需要牺牲的可能性，大大减少了 Markowitz 理论中所用到的复杂计算。 Sharpe (1964) 提出了资本资产定价模型 (CAPM)，主要研究了证券市场中资产的预期收益率与风险资产之间的关系，以及均衡价格是如何形成的，是现代金融市场价格理论的支柱，广泛应用于投资决策和公司理财领域。 Ross (1976) 提出了一种新的资本资产定价模型，即套利定价模型 (ATP)。除了考虑投资组合内部风险，模型还考虑了一些综合因素对风险的影响。

对于风险的刻画，随着学者们研究的深入，有了很多新的风险度量方法。Markowitz (1959) 又提出了用下半方差来作为风险度量，建立了新的均值-半方差模型，该模型对风险的反应更加灵敏，可以对投资者提供更好的投资决策。 Konno 等 (1991, 1995) 提出了均值-绝对偏差模型和收益均值-方差-偏度

组合模型，开始考虑到偏度的不同。虽然风险度量的方法很多，但在具体的投资组合问题中，对风险的刻画还是有不足的地方。而 VaR (Morgan, 1994) 从一个全新的角度作为风险度量方法进入到大众的视野中。关于 VaR 的定义有许多版本，按字面解释就是在险价值，更确切的是指在一定概率水平下，某一金融资产或证券组合价值在未来特定时期内的最大可能损失。 Jorion (1996) 等都在这方面进行了研究，使得 VaR 理论逐渐趋向成熟。但是 VaR 不满足次可加性条件，使得模型是非凸的，为了修正 VaR 的缺点， Pflug (2000)，Acerbi (2002) 在 VaR 的基础上提出了 CVaR 的概念， CVaR 是指超过 VaR 部分的损失的条件期望。 Anderson 等 (2001) 研究了一种新的信用风险优化方法，模型基于 CVaR 风险度量。Prekopa (2003)，Yamai (2005) 等人分别对 VaR 和 CVaR 进行了详细的分析和比较。关于国内的 CVaR 研究，陈金龙和张维 (2002) 研究了 VaR 的修正模型 CVaR，并将其与投资组合优化统一模型联系起来。之后，刘小茂 (2004) 对资产组合的 CVaR 风险进行敏感度分析，并且研究了其经济意义。高全胜 (2005) 将 CVaR 与 MV 模型进行比较，发现 CVaR 更利于风险的分散与监管。

前面所述都是从风险度量角度对投资组合模型的改进，但是传统的投资组合模型有一些局限性，与实际情况可能存在一些偏离，使得模型的结果并不好。就模型本身来说，由于市场情况复杂多变，如果仅以历史数据和经验分析来预测风险，企业或者投资者可能会造成巨大损失。比如在股票市场投资中，如果将股票的收益率看作一个随机变量，那么历史数据只能包含这个随机变量部分分布特征，现实情况仍然是无法确定的。当大家都开始思考这个问题时，也就进入了不确定优化问题的研究。

随机优化 (SP) 是一种不确定优化问题。随机优化假定随机参数的概率分布已知，其目标函数通常是期望值形式，它假定随机参数的概率分布已知。但实际上即使有大量的历史数据作为随机变量真实分布的支撑，我们也难以得到足够精确的概率分布，这会导致模型结果与实际情况的偏离。针对这一问题，鲁棒优化 (RO) 提出了一种解决方法。鲁棒优化假定随机参数概率分布未知，但知道其波动范围，即参数的不确定集，目标函数是优化其不确定集中“最坏情况”。对于参数的不确定性，在不同问题中，都陆续有人提出了其对应的鲁棒方法，但鲁棒优化及分布鲁棒优化的保守性也一直是大家所探讨的问题。 Soyster(1974) 研究了线性规划问题中，约束中的系数矩阵具有列不确定性的问题的解，然而其方法所得到的解太过于保守。随后， Ben-Tal 和

Nemirovski (1998, 1999, 2000), Ghaoui 和 Lebret (1996, 2001) 等为降低保守性而引入了椭球型不确定集。但是椭球型不确定集一方面比较难跟现实数据结合，另外一方面，转化之后的模型大都是二阶锥规划或者半正定规划，求解都比较复杂。此外，这种方法依然比较保守。 Bertsimas 和 Sim (2005) 为了克服椭球型不确定集的缺点，引进了预算不确定集，将原问题转化成线性规划问题，并且得出了所得解可行概率的下限，也即所得解对所有约束不可行的概率的上限。

鲁棒优化中，不同的不确定集对结果的影响十分明显，当不确定集合越精细、模型复杂度越高，求解越困难。当不确定集合越宽泛时，所求出的最优解越保守，越不经济。为了权衡二者的关系，就有了上述对不确定集合选择的不同。分布鲁棒优化 (DRO) 正好改善了随机优化和鲁棒优化的缺点，综合了它们的优点。分布鲁棒优化在同样考虑“最坏情况”下，随机变量的分布是未知的，但是我们认为真实的分布与历史数据刻画的近似分布是非常接近的，那么利用近似分布的信息，来刻画一个分布模糊集，真实分布是包含其中的。分布鲁棒优化模型结合了统计学与优化理论，对模型结论也有了更好的解释性。

在全局鲁棒优化问题的研究上， Ben-Tal 等在 2006 年做出了开创性的工作。Ben-Tal 提出了一种处理不确定数据优化问题地新方法，在一般地鲁棒优化模型下, 在给定的有界不确定集中找到所有数据都能满足约束的决策。特别地, 当数据不在不确定集中时, 约束可以有控制的进行放松。2015 年，Ben-Tal 等人提出了全局鲁棒对应方法 (GRC)，允许在更大的不确定集中违反约束条件。约束条件的违反由参数到原始不确定性集的距离控制。同时扩展了初始 GRC 的可处理 GRC，使得全局鲁棒对应方法不仅适用于线性约束或锥约束，也适用于非线性约束，并且全局鲁棒优化对应方法对于用于控制约束违反的不确定集和距离度量函数都更灵活。 Zhao 等 (2019) 针对一个包含流量动态和需求不确定性的网络设计问题，提出了一种全局鲁棒优优化方法，并将这样一个全局鲁棒网络设计问题表述为一个可处理的线性规划模型，通过比较其求解性能证明该模型的鲁棒性和灵活性。

受全局鲁棒优化的启发，全局分布鲁棒优化问题的核心思想仍然是对约束条件进行放松，在原模糊集的基础上，设定一个被原模糊集包含且半径更小的模糊集。当未知分布处于内部核心模糊集中时，模型仍是对约束条件的严格满足，当未知分布处于核心模糊集外时，对约束条件进行放松，并由未

知分布到内部核心模糊集的最小距离控制。同样是矩信息刻画的模糊分布集，Ding 等 (2020) 与 Li 等 (2022) 从不同的侧重角度出发对全局分布鲁棒优化问题进行求解。前者侧重于矩信息下全局分布鲁棒解具有非保守性和灵活性，后者更关注建立合适的概率分布支撑集，根据数据选择一个核心集和包含该核心集的样本空间，是数据驱动下的结果。Liu 等 (2022) 强调了全局分布鲁棒对应方法丰富的建模表达性、吸引人的计算可处理性和潜在的应用前景。它可以根据数据的可用性来校准，甚至可以提供强有力的理论来选择其输入，例如全局鲁棒优化的影子价格。

在分布鲁棒优化中，同样的，对随机变量分布模糊集的选择进行了许多的探讨。早期研究往往假设随机参数的概率分布无法准确获取，但其部分广义矩信息可获取或估计，于是根据这些信息构建模糊集。例如有 Ghaoui 等 (2003), Popescu (2007), Chen 和 Sim (2010), Bertsimas 和 Popescu (2005), Natarajan 等 (2011), Wiesemann 等 (2014), Hanasusanto 等 (2015), Delage 和 Ye (2010) 等。在大数据时代背景下，问题环境中随机变量参数的历史数据越来越容易获取，为了获得随机参数概率分布，一种自然的想法是通过历史数据对应的经验分布来近似描述真实概率分布。但经验分布并不等同于真正概率分布，分布鲁棒优化的思想是假设真正概率分布与经验分布在概率空间中的统计距离不超过某一阈值，并以此构建模糊集。基于此思想，如何定义两个概率分布间的统计距离成为了关键，它不仅需要具有良好的统计学意义，而且需要保证相应的分布鲁棒优化模型可处理。有两种典型形式，其一是基于散度的模糊集，在 Ben-Tal 等 (2013)，Jiang 和 Guan (2016) 的文章中都有探讨不同的散度函数形式下问题的求解。虽然散度克服了未知矩信息所带来的困难，但是在描述模糊集的相关性上不充分，它只能描述同一水平下的分布之间的距离，对于不同水平下的分布距离往往不能表示。第二是基于 Wasserstein 度量的模糊集，此模糊集囊括了概率空间中 Wasserstein 距离为度量标准，以经验分布为球心，以某一可调参数为半径的球中所有概率分布。Esfahani 和 Kuhn(2018) 的研究表面从统计学上是可以置信的，对于更一般的 Wasserstein 度量定义以及更多的深入研究有 Zhao 和 Guan (2018)。

本文主要基于 Wasserstein 度量和线性锥强对偶定理，将具有不确定分布的全局分布鲁棒优化问题 (GDRO) 转化为确定分布下的对应问题，并将全局分布鲁棒优化模型应用到 CVaR 投资组合问题中，构建了全局分布鲁棒投资组合问题，并通过模型转化得到确定分布下的投资组合对应问题。最后，对全

局分布鲁棒投资组合优化问题模型进行数值模拟，证明了模型的可行性，通过模型对比说明了模型在降低分布鲁棒投资组合优化问题模型的保守性上是有效的，且对投资决策有一定的指导性。

1.2 研究方法与创新点

本文为了解决传统分布鲁棒优化问题中存在的问题，着重关注了在不确定因素下过度追求鲁棒性带来的保守性问题，因此在建模上选择了全局分布鲁棒优化模型。在模糊集的选择上选用了 Wasserstein 度量下的模糊集，并讨论了分布鲁棒优化问题的对应方法，在此基础上研究全局分布鲁棒优化问题。由于模型是基于 Wasserstein 模糊集下的“最坏情况”，不可直接求解，先将全局分布鲁棒优化问题进行分解成更小的优化问题，再分别讨论，利用线性锥对偶定理转化成可处理的确定分布下的全局分布鲁棒对应问题 (GDRC)。

在创新点方面，首先本文构建了一个具有全局概念的分布鲁棒优化模型，模型考虑到了传统的投资组合问题存在的保守性问题，能够通过引入放松核心模糊集外部分分布的约束来降低模型的保守性问题。其次，在模型的求解方面，基于 Wasserstein 度量的考虑，将模型转化成线性锥规划问题的形式，并利用其强对偶定理，进行模型转化，得到一个可处理的确定性问题。最后，本文在模型建立时是以金融证券中的投资组合问题作为背景的，在数值模拟中不仅进行了随机数值模拟，还用真实股票数据进行了模型对比。同时这样的全局分布鲁棒优化模型也适用于其他领域相关应用，如企业项目投资等。

1.3 研究内容与论文框架

本文将研究基于 Wasserstein 度量的全局分布鲁棒优化问题，找到其可处理的全局分布鲁棒对应问题。讨论在不同距离度量及目标函数的设定下，对应确定的全局分布鲁棒对应问题，把问题从一个不可解的问题，转化成一个可处理的问题。然后，我们构建全局分布鲁棒优化投资组合模型，进行模型转化，变成投资组合背景下的可处理的全局分布鲁棒优化对应问题。最后对进行全局分布鲁棒优化投资组合问题进行数值模拟。

第一章为绪论部分，阐述了本文的研究背景与研究意义，以及论文研究思路、研究方法等，最后简述了本文研究内容和论文框架。

第二章为预备知识部分，主要按照建模思想，模型转化和优化问题归属三个方面来进行展开叙述。建模方面主要是鲁棒优化思想，本文着重用到了

分布鲁棒优化，以及其改进全局分布鲁棒优化模型。在模型转化方面主要简单介绍了线性锥对偶理论的相关概念。优化问题方面介绍了投资组合问题中的 CVaR 模型。

第三章为本文的核心部分，对基于 Wasserstein 度量的全局鲁棒优化问题模型进行了构建、等价转化。先是介绍了全局分布鲁棒优化问题以及 Wasserstein 模糊集的刻画，然后讨论了基于 Wasserstein 度量的分布鲁棒优化对应问题，在此基础上继续深入讨论了 Wasserstein 度量下的全局分布鲁棒对应问题，推导在计算上可处理的 GDRC，最后讨论了在特殊距离度量下的 GDRCs。

第四章为全局分布鲁棒优化模型在投资组合背景下的应用部分，将全局化的思想引入不确定投资组合问题中，构建全局分布鲁棒优化投资组合问题模型并进行模型转化变成可处理的对应问题。最后，进行了数值模拟并进行了模型对比和结果分析。

第五章为结论部分，对本文得到的结论进行一一叙述，并且分析了本文的优缺点，最后针对文中存在的不足提出了后续进一步优化的方向。

本文研究结构框架见图 1-1。

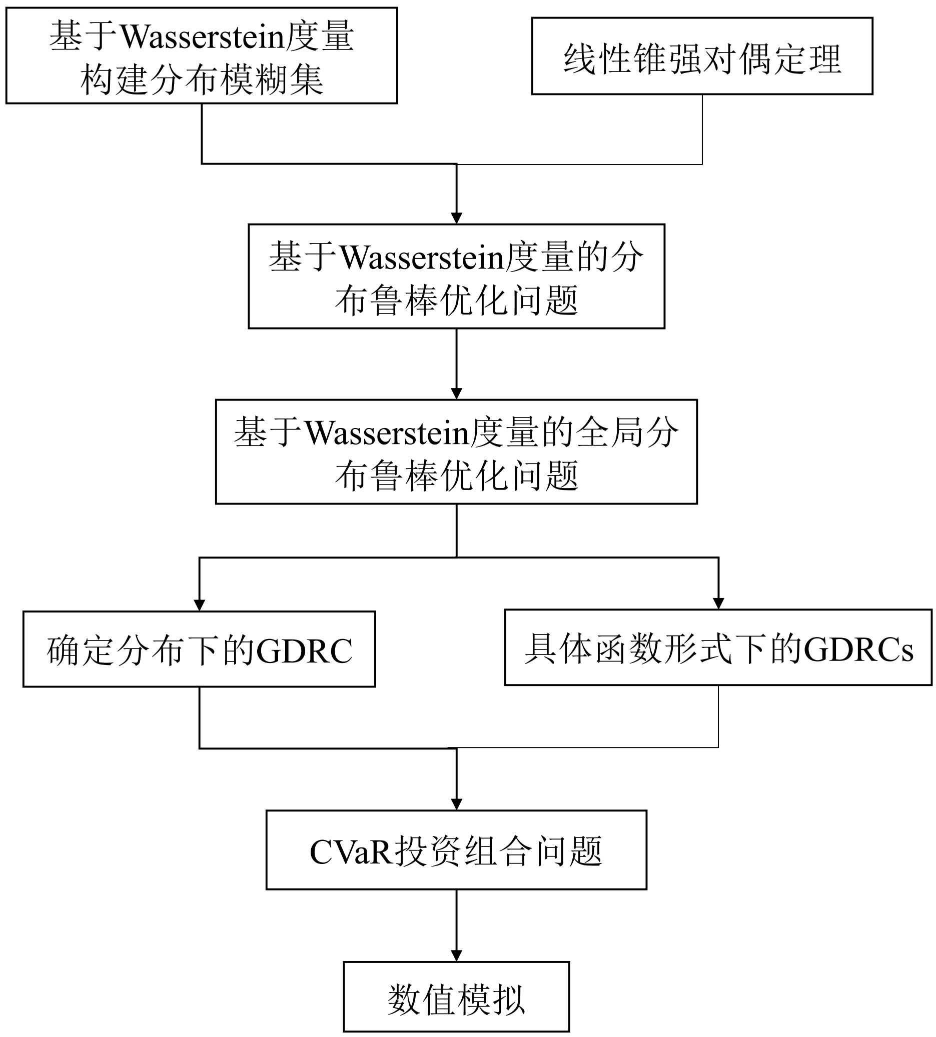
基于Wasserstein度量构建分布模糊集
线性锥强对偶定理
基于Wasserstein度量的分布鲁棒优化问题
基于Wasserstein度量的全局分布鲁棒优化问题
确定分布下的GDRC
具体函数形式下的GDRCs
CVaR投资组合问题
数值模拟

图 1-1: 本文结构关系图

2. 预备知识

2.1 鲁棒优化模型

设 $x \in \mathcal{X} \subseteq \mathbb{R}^n$ 为决策变量。目标函数定义为 $f(x): \mathcal{X} \to \mathbb{R}$；$g(x): \mathcal{X} \to \mathbb{R}^k$ 定义为一个函数向量，其中 $g(x) = (g_1(x), \ldots, g_k(x))^T, g_i(x): \mathcal{X} \to \mathbb{R}, i = 1, \ldots, k$. 那么一般优化问题的形式为

$$\min_{x \in \mathcal{X}} \{f(x) | g(x) \leq 0\}. \tag{2-1}$$

当问题中存在不确定参数 $\xi \in \mathbb{R}^m$ 时，问题 (2-1) 就变成了不确定优化问题。下面是解决不确定优化问题的几种模型。

随机优化假设 ξ 为随机变量，其服从的分布 $\mathbb{P}$ 已知。那么一个经典的随机优化模型为

$$\min_{x \in \mathcal{X}} \{\mathbb{E}_{\mathbb{P}}[f(x, \xi)]\}. \tag{2-2}$$

或

$$\min_{x \in \mathcal{X}} \{f(x) | \mathbb{E}_{\mathbb{P}}[g(x, \xi)] \leq 0\}. \tag{2-3}$$

随机优化是求解随机参数不确定优化问题的有力工具，但仍有两个不足之处。首先，涉及到多维积分的计算期望通常是难以处理的。其次，真实的分布总是很难估计，当使用带有偏差估计的分布时，SP 可能会产生过拟合决策。

鲁棒优化考虑了确定集合区域上的不确定性模型，而不是遵循随机分布。假设不确定参数 ξ 取不确定集 $\mathcal{U} \subseteq \mathbb{R}^m$ 中的任意值，也称不确定集。鲁棒优化模型为

$$\min_{x \in \mathcal{X}} \{f(x) | g(x, \xi) \leq 0, \forall \xi \in \mathcal{U}\}. \tag{2-4}$$

鲁棒优化模型的最大优点之一是其可处理性，这取决于优化问题和不确定集。鲁棒优化问题可以重新表述为线性规划 (LP)、二阶锥规划 (SOCP) 或半定规划 (SDP) 问题。然而，RO 也有缺点。在某些情况下，不确定集可能会给出非

常少见的“最坏情况”，以至于结果过于保守，无法在实践中使用。在不失可处理性优点的情况下，DRO 为有效解决 RO 的保守性限制提供了一种可行的方法。

分布鲁棒优化假设随机变量 ξ 的所有可能分布组成模糊集 $\mathcal{P}$，那么即使在“最坏情况”下，模型也在模糊集上表现良好。一般的 DRO 模型可表示为

$$\min_{x\in\mathcal{X}} \max_{\mathbb{P}\in\mathcal{P}} \{\mathbb{E}_{\mathbb{P}}[f(x,\xi)] | \max_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}[g(x,\xi)] \leq 0\}. \tag{2-5}$$

如果真实分布已知，那么模糊集 $\mathcal{P}$ 为单元素集合，则 DRO 可以变为 SP 问题。模糊集 $\mathcal{P}$ 在本质上决定了 DRO 模型的可处理性和最终结果的质量。

2.2 线性锥对偶

线性锥规划 (LCP) 是定义在锥集合上的线性规划问题，Boyd(2008)，Ben-Tal(2001)，方述诚 (2013) 等都在该问题上做了相应的研究。下面将介绍线性锥规划的概念及其对偶理论。

线性锥规划问题 LCP 的一般形式为

$$\begin{aligned} V_{LCP} = \min \quad & a \cdot x \\ s.t. \quad & a^i \cdot x = t_i \quad i = 1, \dots, m \\ & x \in \mathcal{K}, \end{aligned} \tag{2-6}$$

其中 $\mathcal{K}$ 为闭凸锥，“$\cdot$” 为定义在 $\mathcal{K}$ 所在的欧氏空间中的内积运算。Boyd(2008)，Gao(2004) 等通过 Lagrange 对偶、Fenchel 对偶以及共轭对偶等对偶理论分别建立了线性锥规划对偶模型（LCD）：

$$\begin{aligned} V_{LCD} = \max \quad & t^T y \\ s.t. \quad & \sum_{i=1}^{m} y_i a^i + s = a \\ & s \in \mathcal{K}^*, y \in \mathbb{R}^m, \end{aligned} \tag{2-7}$$

其中 $\mathcal{K}^*$ 为 $\mathcal{K}$ 的对偶锥当 $\mathcal{K} = \mathbb{R}^n_+$ 时，LCP 为标准线性规划问题，此时 $\mathcal{K}^* = \mathbb{R}^n_+$；当 $\mathcal{K}$ 为一系列二阶锥的笛卡尔乘积时，LCP 为二阶锥规划问题；当 $\mathcal{K} = \mathcal{S}^n_+$ 时，LCP 为标准的半定规划问题，此时 $\mathcal{K}^* = \mathcal{S}^n_+$。由此，线性锥规划的标

准型涵盖了经典的线性规划、二阶锥规划和半定规划。

类似于其他规划问题的对偶理论，可以建立线性锥规划问题的弱对偶定理和强对偶定理。

定理 2.1. (方述诚，2013)(弱对偶定理) 若线性锥规划标准模型 (2-6) 及其对偶问题 (2-7) 都是可行的，则对于 (2-6) 的任何可行解 x 和 (2-7) 的任何可行解 (y, s)，都有 $a \cdot x \geq t^T y$。

定理 2.2. (方述诚，2013)(弱对偶定理) 对于线性锥规划标准模型 (2-6) 及其对偶问题 (2-7)，当原问题 (2-6) 严格可行且目标函数值有下界时，存在对偶问题 (2-7) 的可行解 (y^*, s^*) 使得 $t^T y^* = V_{LCP}$，即此时对偶问题的最优解是可达的；同样的，当对偶问题 (2-7) 严格可行且目标值有上界时，存在原问题 (2-6) 的可行解 x^* 使得 $a \cdot x^* = V_{LCD}$，即此时原问题的最优解是可达的。

2.3 CVaR 理论

CVaR 理论的研究和发展是建立在 VaR 理论之上的，了解 CVaR 理论首先需要了解 VaR 理论。在上世纪九十年代 VaR 方法被提出并用来进行风险度量，此方法很快成为了金融市场度量市场风险的主流方法，其相关应用也迅速发展。VaR 即风险价值模型，意味着风险中的价值，其确切含义是指，在一定的置信水平和一定的持有期内，某一个金融资产或证券组合在未来资产价格波动的情况下所面临的最大损失额 (Jorion, 2007)。公式表示为:

$$\mathbb{P}_r(\Delta P \Delta t \leq VaR) = \alpha,$$

其中 $\mathbb{P}$ 为资产价值损失小于可能损失上限的概率，ΔP 为在一定持有期 Δt 内的价值损失额，α 为给定的置信水平。另可表示为:

$$VaR_\alpha(x) = \min\{x \mid F_x(x) \geq \alpha\}, \tag{2-8}$$

其中 X 是一个损失的随机变量，$VaR_\alpha(x)$ 是离散分布的置信水平 α 的非凸和不连续函数，$F_X(x)$ 是 x 的分布函数。

由于 VaR 模型不具有次可加性，因此在 VaR 的基础上发展出 CVaR，并且由于其满足一致性公理的风险度量方法，被学术界认为是一种比 VaR 方法更为准确合理的风险度量方法。

CVaR 即条件风险价值，是由 Rockafellar 和 Uryasev(1999) 提出，其含义为在投资组合的损失超过某个给定 VaR 值的条件下，该投资组合的平均损失值，数学定义如下：

$$CVaR_\alpha = \mathbb{E}\left[X \mid X \geq VaR_\alpha\right]. \tag{2-9}$$

下面简单介绍一下 CVaR 的转化过程 (Rockafellar 和 Uryasev, 1999)：

设决策向量 $x \in \mathbb{R}^n$ 的损失函数为 $f(x,y)$, x 代表一个投资组合，向量 $y \in \mathbb{R}^m$ 为不确定随机变量；$p(y)$ 为密度函数，$f(x,y)$ 在不超过阈值 α 的情况下可以表示为：

$$\psi(x,\alpha) = \int_{f(x,y)\leq\alpha} p(y)d(y). \tag{2-10}$$

一般来说，$\psi(x,\alpha)$ 相对于 α 是不递减的，并且是右连续的，但不一定左连续，因为有跳跃的可能性。对于决策变量 x 和指定的概率水平 β, β-VaR 和 β-CVaR 分别可以表示为：

$$\alpha_\beta(x) = \min\{\alpha \in R : \psi(x,\alpha) \geq \beta\}, \tag{2-11}$$

$$\phi_\beta(x) = \frac{1}{1-\beta}\int_{f(x,y)\geq\alpha_\beta(x)} f(x,y)p(y)d(y). \tag{2-12}$$

在上式中，$\alpha_\beta(x)$ 是由 α 值组成的非空区间的左端点，因此 $\psi(x,\alpha)=\beta$，则 $f(x,y) \geq \alpha_\beta(x)$ 的概率等于 $1-\beta$. 由于 $\phi_\beta(x)$ 中含有 $\alpha_\beta(x)$，并且 $\alpha_\beta(x)$ 难以根据定义直接求解，$\phi_\beta(x)$ 可以用 $\min_{\alpha\in\mathbb{R}} F_\beta(x,\alpha)$ 来替代

$$F_\beta(x,\alpha) = \alpha + \frac{1}{1-\beta}\int_{y\in\mathbb{R}^m} \left[f(x,y)-\alpha\right]^+ p(y)d(y), \tag{2-13}$$

即

$$F_\beta(x,\alpha) = \mathbb{E}\left[y \mid y \geq \alpha_\beta(y)\right] = \mathbb{E}\left[y \mid y \geq VaR_\beta\right].$$

并且有以下结论：

(1) $F_\beta(x,\alpha)$ 关于 α 是凸的和连续可微的；

(2) 最小化 CVaR 有

$$\min_{x\in\mathbb{R}^n} \phi_\beta(x) = \min_{(x,\alpha)\in\mathbb{R}^n\times\mathbb{R}} F_\beta(x,\alpha). \tag{2-14}$$

结论 (1) 是优化问题中一个非常好的特性，可以确保得到优化问题的全局最优解；结论 (2) 则表明求解 CVaR 不用像定义 (4-12) 一样先求 VaR，大大简化了求解过程，同时还可以得到相应的 VaR 值。若优化问题式 (4-13) 的最优解为 (x^*, α^*)，则 x^* 为最优投资组合，α^* 为 VaR，$F_\beta(x^*, \alpha^*)$ 即是 CVaR。

如果概率密度函数 $p(y)$ 的解析式不可得或 (4-13) 积分不易求解时，可用历史数据或蒙特卡洛方法模拟取样，由此得到 N 个数据 $y_1, y_2, \ldots, y_N$，则式 (4-13) 可由下式近似：

$$\tilde{F}_\beta(x, \alpha) = \alpha + \frac{1}{N(1-\beta)} \sum_{i=1}^{N} [f(x, y_i) - \alpha]^+ . \tag{2-15}$$

若令 $w_i = [f(x, y_i) - \alpha]^+$，则有 $w_i \geq 0$ 且 $w_i \geq f(x, y_i) - \alpha$。这种近似简化并不取决于 y 的特定分布（如正态分布等），对于非正态分布也同样有效。

3. Wasserstein 度量下的全局分布鲁棒优化问题

3.1 全局分布鲁棒优化问题 (GDRC)

在经典的分布鲁棒优化问题中，设决策变量 $x \in \mathbb{R}^n$，损失函数 $g(x, \xi)$，由于约束中存在未知参数 $\xi \in \mathbb{R}^n$ ，为了达到鲁棒性，我们取定一个未知参数 ξ 的分布 $\mathbb{P}$ 可能模糊分布集 $\mathcal{P}$ ，使得所有在该集合中的可能分布都满足约束条件，分布鲁棒优化问题可写作如下形式：

$$\mathbb{E}_{\mathbb{P}}[g(x, \xi)] \leq 0, \ \ \forall \mathbb{P} \in \mathcal{P}. \tag{3-1}$$

这里我们放松约束条件，在模型 (3-1) 中加入一个被原模糊分布集 $\mathcal{P}$ 包含且比原模糊分布集更小的分布集合 $\mathcal{P}'$ ，在不等式约束右边加入一个可控的距离函数加以放松条件，就是本文讨论的全局分布鲁棒优化问题，具体形式如下：

$$\mathbb{E}_{\mathbb{P}}[g(x, \xi)] \leq \min_{\mathbb{P}' \in \mathcal{P}'} H(\mathbb{P}, \mathbb{P}'), \qquad \forall \mathbb{P} \in \mathcal{P}, \tag{3-2}$$

其中 H 为距离函数，$H(\mathbb{P}, \mathbb{P}) = 0$，且该函数值非负。若 $\mathbb{P} \in \mathcal{P}'$, 则不等式约束变为 $\mathbb{E}_{\mathbb{P}}[g(x, \xi)] \leq 0$，即为原始问题 (3-1)；若 $\mathbb{P} \in \mathcal{P} \backslash \mathcal{P}'$, 则允许违反原始不等式 (3-1)，这由分布 $\mathbb{P}$ 到模糊分布集 $\mathcal{P}'$ 中分布的最小距离控制。从模型 (3-2) 中可以看出，当随机变量分布取到更核心的模糊集 $\mathcal{P}'$ 中的分布时，此时的距离函数 H 为 0，严格遵循不等式约束，因为真实分布更可能处于核心集合内，与 $\mathcal{P}'$ 集合中的分布信息有更多的重合。那么，对于随机变量分布取到核心集合之外的分布时，我们根据该分布距离核心集合 $\mathcal{P}'$ 的最短距离来对其约束进行放松，问题 (3-2) 的不等号右边为一个正数，距离越近，数值越小，放松的越少；距离越远，说明其分布信息越不重要，可以适当将约束放松的越多。这样模型可以使得分布鲁棒优化问题的保守性降低，故而全局分布鲁

棒优化模型既具有鲁棒性，同时又具有可控的保守性。

3.2 模糊集及 Wasserstein 度量

模糊集 $\mathcal{P}$ 是分布鲁棒优化模型的关键组成部分。一个好的模糊集应该足够大，由具有高置信度的真实数据组成。另一方面，模糊集应该足够小，以排除极端情况下的分布，但这将导致模型过度保守。此外，模糊集也应该很容易从数据中参数化，并且在理想情况下，有助于将分布鲁棒优化问题作为一个结构化的数学程序重新表述，以便于现成的优化软件解决。

本文中模糊集是以 Wasserstein 度量作为概率分布距离函数，定义为概率分布空间中的一个球。这种基于度量的模糊集包含了所有接近给定概率度量的经验分布或最可能的分布。通过调整模糊集半径，可以控制优化问题的保守性程度。如果半径减少到 0，则模糊集缩小为仅仅包含经验分布的单元素集合，在这种情况下，分布鲁棒问题就变为无模糊的随机优化。

下面首先介绍 Wasserstein 度量的定义，再介绍基于 Wasserstein 度量的模糊集，最后说明本文中关于 Wasserstein 度量及分布模糊集的设定。

定义 3.1. （*Kantorovich, 1957*）（*Wasserstein* 度量）设 ξ 和 ξ' 为概率空间 $\mathcal{M}(\Xi)$ 上的两个随机变量，Ξ 为随机变量 ξ 和 ξ' 的支撑集，$\mathcal{M}(\Xi)$ 表示在支撑集上所有可能分布的集合。$\mathbb{P}_1$ 是随机变量 ξ 的分布，$\mathbb{P}_2$ 是随机变量 ξ' 的分布。$d(\xi,\xi')$ 为概率空间上的一个度量，则这两个概率分布之间的 p-*Wasserstein* 度量为：

$$W_p(\mathbb{P}_1,\mathbb{P}_2)=\inf_{\Pi\in\mathcal{M}(\Xi\times\Xi)}\left(\int_{\Xi\times\Xi}[d(\xi,\xi')]^p\Pi(d\xi,d\xi')\right)^{1/p},$$

$$\Pi(\Xi,d\xi')=\mathbb{P}_1(d\xi'),\ \Pi(d\xi,\Xi)=\mathbb{P}_2(d\xi),\ \Pi(d\xi,\Xi)\triangleq\int_{\xi'\in\Xi}\Pi(d\xi,d\xi').$$

其中 $\Pi(d\xi,d\xi')$ 是随机变量 ξ 和 ξ' 的联合分布，$\mathcal{M}(\Xi\times\Xi)$ 是所有可能联合分布的集合。

根据上述定义 3.1，若以 $\mathbb{P}_2$ 作为模糊集分布球的球心，以 ε 为半径，与分布 $\mathbb{P}_2$ 的 Wasserstein 距离 $W_p(\mathbb{P}_1,\mathbb{P}_2)$ 小于 ε 的任意分布 $\mathbb{P}_1$ 组成的分布集合，即为 Wasserstein 模糊集，可以写作如下形式：

$$\mathbb{B}_\varepsilon(\mathbb{P}_2)=\{\mathbb{P}_1\in\mathcal{M}(\Xi):W_p(\mathbb{P}_1,\mathbb{P}_2)\le\varepsilon\}.$$

在实际模型中，考虑到模糊集是随机变量 ξ 真实分布 $\mathbb{P}$ 的一个可能集合，关于 ξ 可观测到的分布信息为 N 个独立同分布的样本数据，故以经验分布 $\mathbb{P}_N$ 为模糊集中心。同时，在本文中我们取 $p=1$, $\Xi=\mathbb{R}^n$, $d(\xi,\xi')=\|\xi-\xi'\|$, 则 1-Wasserstein 度量及模糊集 $\mathcal{P}$ 可以分别表示为:

$$W(\mathbb{P},\mathbb{P}_N)=\inf_{\Pi\in\mathcal{M}(\Xi\times\Xi)}\int_{\Xi\times\Xi}d(\xi,\xi')\Pi(d\xi,d\xi'),\tag{3-3}$$

$$\mathcal{P}=\mathbb{B}_\varepsilon(\mathbb{P}_N)=\left\{\mathbb{P}\in\mathcal{M}(\Xi):W(\mathbb{P},\mathbb{P}_N)\leq\varepsilon\right\}.\tag{3-4}$$

我们注意到两个分布 $\mathbb{P}$ 和 $\mathbb{P}_N$ 的 Wasserstein 距离可以被视为将分布 $\mathbb{P}$ 移动到分布 $\mathbb{P}_N$ 的最小运输成本，其中将单位质量 ξ 移动到 ξ' 的代价等于 $d(\xi,\xi')$。此外，Wasserstein 模糊集包含所有连续或离散分布，这些分布相对于 Wasserstein 度量足够接近离散经验分布 $\mathbb{P}_N$。Bolley 等 (2007)，Nicolas 等 (2015) 从统计上证明实际分布 $\mathbb{P}$ 属于 $\mathbb{P}_N$ 附近的 Wasserstein 模糊集，如果其半径是 $log(1/\beta)/N$ 的次线性增长函数，则其可信度为 $1-\beta$，因此保证了可达到的样本外的置信区间上界。

3.3 Wasserstein 度量下的分布鲁棒优化问题

在讨论全局分布鲁棒优化问题之前，先考虑基于 Wasserstein 度量的分布鲁棒优化问题，根据式 (3-1) 以及式 (3-4)，可以将问题描述为如下不等式约束:

$$\mathbb{E}_{\mathbb{P}}[g(x,\xi)]\leq 0,\ \ \forall\mathbb{P}\in\mathbb{B}_\varepsilon(\mathbb{P}_N).\tag{3-5}$$

任意属于模糊集中的分布 $\mathbb{P}$ 都满足不等式，那么等价于模糊集中分布的期望最大值小于 0，即上式 (3-5) 等价于

$$\max_{\mathbb{P}\in\mathbb{B}_\varepsilon(\mathbb{P}_N)}\left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)]\right\}\leq 0.\tag{3-6}$$

“最坏情况”期望问题 (3-6) 构成了概率分布上的无限维优化问题，因此难以处理。接下来我们将证明 (3-6) 可以利用鲁棒优化的工具重新表示为有限维凸优化问题。这里 $\mathbb{P}_N$ 是离散的经验分布。当取 N 个样本时，不妨设取到的样本数据是 $\hat{\xi_1},\hat{\xi_2},\ldots,\hat{\xi_N}$, 那么离散分布 $\mathbb{P}_N(\xi)=\frac{1}{N}\sum_{i=1}^N\delta_{\hat{\xi}_i}(\xi)$, 即有 $\mathbb{P}_N(\hat{\xi_i})=\frac{1}{N}$, $\forall i=1,\ldots,N$。同时对模型做出如下凸性假设。

假设 3.2. （*Esfahani, 2018*）（凸性）随机变量 ξ 的支撑集 Ξ 是闭凸集，且负的损失函数 $-g(x,\xi)$ 是下半连续的正常凸函数。此外，我们假设 $-g(x,\xi)$ 不等于 $-\infty$ 对于任意 ξ 在支撑集 Ξ 上。

定理 3.3. 如果假设 3.2 成立，那么对于任意 $\varepsilon \geq 0$，存在决策向量 $x \in \mathbb{R}^n$ 不等式约束 (3-6) 等价于如下不等式

$$\frac{1}{N}\sum_{i=1}^{N} g(x,\hat{\xi}_i) + \varepsilon\|x\|_* \leq 0, \tag{3-7}$$

其中 $\|\cdot\|_*$ 是 $\|\cdot\|$ 的共轭范数。

证明.

要证明上述结论，先将不等式 (3-6) 左边的问题展开为分布鲁棒优化问题一般形式，再对该问题进行分析。

$$\max_{\mathbb{P}\in\mathbb{B}_\varepsilon(\mathbb{P}_N)} \mathbb{E}_{\mathbb{P}}[g(x,\xi)] = \begin{cases} \max\limits_{\xi\in\mathcal{M}(\Xi)} & \int_{\Xi} g(x,\xi)\Pi(d\xi,\Xi) \\ s.t. & \int_{\Xi\times\Xi} d(\xi,\xi')\Pi(d\xi,d\xi') \leq \varepsilon, \\ & \Pi(\Xi,d\xi') = \mathbb{P}_N(d\xi'). \end{cases} \tag{3-8}$$

(i) 根据 $\mathbb{P}_N$ 是离散的经验分布化简模型

当取 N 个样本时，取到的样本是 $\hat{\xi_1},\hat{\xi_2},\ldots,\hat{\xi_N}$，那么离散分布 $\mathbb{P}_N(\xi) = \frac{1}{N}\sum_{i=1}^{N}\delta_{\hat{\xi}_i}(\xi)$，即有 $\mathbb{P}_N(\hat{\xi}_i) = \frac{1}{N}, \forall i = 1,\ldots,N$。另外由随机变量 $\xi \sim \mathbb{P}$，随机变量 $\xi' \sim \mathbb{P}_N$，则可将 $\mathbb{P}_N$ 概率密度有关的积分变成概率质量之和。下面对上述问题 (3-8) 等式右边部分中的联合分布及边际分布函数进行一个离散化的处理。

$$\begin{aligned}\Pi(d\xi,\Xi) &= \int_{\xi'\in\Xi}\Pi(d\xi,d\xi') \\ &= \sum_{i=1}^{N}\Pi(d\xi\,|\,\xi'=\hat{\xi}_i)\cdot\mathbb{P}_N(\hat{\xi}_i) \\ &= \sum_{i=1}^{N}\Pi(d\xi\,|\,\xi'=\hat{\xi}_i)\cdot\frac{1}{N} \\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{P}^i(d\xi),\end{aligned} \tag{3-9}$$

在上式 (3-9) 边际分布的转化结果中，令 $\mathbb{P}^i(d\xi)$ 为关于随机变量 ξ 的条件概率分布函数，其定义为 $\mathbb{P}^i(d\xi)=\Pi(d\xi\,|\,\xi'=\hat{\xi}_i)$，是指给定 $\xi'=\hat{\xi}_i$ 条件下，关于 ξ 的概率分布。同理，下面为联合分布转化过程：

$$\begin{aligned}\Pi(d\xi,d\xi') &= \Pi(d\xi\,|\,\xi')\cdot\mathbb{P}_N(\xi') \\ &= \frac{1}{N}\sum_{i=1}^{N}\delta_{\hat{\xi}_i}(\xi')\mathbb{P}^i(d\xi).\end{aligned} \tag{3-10}$$

根据 (3-9) 和 (3-10) 的转化, 所以最大化问题 (3-8) 可化为如下问题 (3-11):

$$\begin{aligned}\max_{\mathbb{P}^i\geq 0}\quad & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}g(x,\xi)\mathbb{P}^i(d\xi) \\ s.t.\quad & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}d(\xi,\hat{\xi}_i)\mathbb{P}^i(d\xi)\leq\varepsilon, \\ & \int_{\Xi}\mathbb{P}^i(d\xi)=1,\quad i=1,2,\ldots,N.\end{aligned} \tag{3-11}$$

在上述模型中，决策变量变为条件概率 $\mathbb{P}^i(i=1,2,\ldots,N)$，目标函数和第一个不等式约束相应的变为与条件概率 $\mathbb{P}^i$ 有关的积分，并根据概率的正则性增加了等式约束 $\int_{\Xi}\mathbb{P}^i(d\xi)=1(i=1,2,\ldots,N)$，同时根据概率的非负性对 $\mathbb{P}^i$ 增加非负不等式约束。

(ii) 问题 (3-11) 的对偶问题

为求得问题 (3-11) 的对偶问题，首先引入松弛变量 $\lambda, s_i(i=1,2,\ldots,N)$，

并将支撑集 $\Xi=\mathbb{R}^n$ 代入，得到原问题 (3-11) 的 Lagrange 函数为

$$\begin{aligned}L(\mathbb{P}^i,\lambda,s_i)=&\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}^n}g(x,\xi)\mathbb{P}^i(d\xi)\\&+\lambda\cdot\left\{\varepsilon-\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}^n}d(\xi,\hat{\xi}_i)\mathbb{P}^i(d\xi)\right\}\\&+\frac{1}{N}\sum_{i=1}^N s_i\cdot\Big[1-\int_{\mathbb{R}^n}\mathbb{P}^i(d\xi)\Big].\end{aligned}$$

那么问题 (3-11) 可用 Lagrange 函数表示为

$$\max_{\mathbb{P}^i\geq 0}\min_{\substack{\lambda\geq 0\\ s_i}}L(\mathbb{P}^i,\lambda,s_i). \tag{3-12}$$

其对偶问题可表示为

$$\min_{\substack{\lambda\geq 0\\ s_i}}\max_{\mathbb{P}^i\geq 0}L(\mathbb{P}^i,\lambda,s_i). \tag{3-13}$$

考虑问题 (3-13) 的内部最大化问题 $\max_{\mathbb{P}^i\geq 0}L(\mathbb{P}^i,\lambda,s_i)$, 即

$$\begin{aligned}&\max_{\mathbb{P}^i\geq 0}\left\{\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}^n}g(x,\xi)\mathbb{P}^i(d\xi)+\lambda\cdot\Big[\varepsilon-\frac{1}{N}\sum_{i=1}^N\int_{\mathbb{R}^n}d(\xi,\hat{\xi}_i)\mathbb{P}^i(d\xi)\Big]\right.\\&\qquad\qquad\left.+\frac{1}{N}\sum_{i=1}^N s_i\cdot\Big[1-\int_{\mathbb{R}^n}\mathbb{P}^i(d\xi)\Big]\right\}\\&=\max_{\mathbb{P}^i\geq 0}\left\{\frac{1}{N}\left[\int_{\mathbb{R}^n}\sum_{i=1}^N\Big(g(x,\xi)-\lambda d(\xi,\hat{\xi}_i)-s_i\Big)\mathbb{P}^i(d\xi)\right]+\lambda\varepsilon+\frac{1}{N}\sum_{i=1}^N s_i\right\}\\&=\max_{\mathbb{P}^i\geq 0}\left\{\frac{1}{N}\left[\int_{\mathbb{R}^n}\sum_{i=1}^N\Big(g(x,\xi)-\lambda d(\xi,\hat{\xi}_i)-s_i\Big)\mathbb{P}^i(d\xi)\right]\right\}+\lambda\varepsilon+\frac{1}{N}\sum_{i=1}^N s_i\\&=\begin{cases}\lambda\varepsilon+\frac{1}{N}\sum_{i=1}^N s_i, & \forall i: g(x,\xi)-\lambda d(\xi,\hat{\xi}_i)-s_i\leq 0,\ \forall\xi\in\mathbb{R}^n;\\ +\infty, & \text{其他}.\end{cases}\end{aligned}$$

那么对偶问题 (3-13) 可化为

$$\begin{aligned}\min_{\lambda,s_i}\quad & \lambda\varepsilon+\frac{1}{N}\sum_{i=1}^{N}s_i\\ s.t.\quad & \max_{\xi\in\mathbb{R}^n}\left\{g(x,\xi)-\lambda d(\xi,\hat{\xi}_i)\right\}\le s_i,\quad \forall i=1,2,\ldots,N,\\ & \lambda\ge 0.\end{aligned}\tag{3-14}$$

(iii)$\forall\varepsilon>0$, 证明原问题 (3-11) 和对偶问题 (3-14) 满足强对偶定理

首先回顾问题 (3-11):

$$\begin{aligned}\max_{\mathbb{P}^i\ge 0}\quad & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}g(x,\xi)\mathbb{P}^i(d\xi)\\ s.t.\quad & \frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}d(\xi,\hat{\xi}_i)\mathbb{P}^i(d\xi)\le\varepsilon,\\ & \int_{\Xi}\mathbb{P}^i(d\xi)=1,\quad i=1,2,\ldots,N.\end{aligned}$$

将问题 (3-11) 写成如下线性锥优化问题形式：

$$\max_{\mathbb{P}^i\ge 0}\quad \frac{1}{N}\sum_{i=1}^{N}\left\langle g(x,\cdot),\mathbb{P}^i\right\rangle \tag{3-15}$$

$$s.t.\quad \mathcal{A}(\mathbb{P}^1,\mathbb{P}^2,\ldots,\mathbb{P}^N)-\begin{bmatrix}1\\ \vdots\\ 1\\ -\varepsilon\end{bmatrix}\in\{0\}\times\cdots\times\{0\}\times\mathcal{S}_+^1, \tag{3-16}$$

$$\mathbb{P}^i\in\mathcal{C}. \tag{3-17}$$

其中 $\mathcal{A}$ 是线性映射

$$\mathcal{A}:(\mathbb{P}^1,\mathbb{P}^2,\ldots,\mathbb{P}^N)\longmapsto\left\{\int_{\Xi}\mathbb{P}^1(d\xi),\ldots,\int_{\Xi}\mathbb{P}^N(d\xi),-\frac{1}{N}\sum_{i=1}^{N}\int_{\Xi}d(\xi,\hat{\xi}_i)\mathbb{P}^i(d\xi)\right\}^T,$$

且 $\mathcal{C}=cone(\mathcal{M}),\mathcal{M}=\{\mathbb{P}:\mathbb{P}\ge 0,\mathbb{P}(\mathbb{R}^n)=1\}$.

为证明线性锥强对偶成立，找到上述原问题的严格可行解，不妨取决策

变量条件概率 $\mathbb{P}^i$ 的特解为

$$\mathbb{P}^i = \Pi(d\xi \,|\, \xi' = \hat{\xi}_i) = \delta_{\hat{\xi}_i}(\xi') = \begin{cases} 1, & \xi = \hat{\xi}_i; \\ 0, & \xi \neq \hat{\xi}_i. \end{cases}$$

由于 $\mathbb{P}^i \in \mathcal{M}$, 所以有 $\lambda \mathbb{P}^i \in cone(\mathcal{M}) = \mathcal{C}$, 故 $\mathbb{P}^i$ 在可行域内部而不是边界，且满足原问题的第二个约束条件 (3-17)。

根据 $d(\cdot,\cdot)$ 以及 $\mathbb{P}^i$ 的定义，有

$$\begin{cases} \xi = \hat{\xi}_i, & d(\xi, \hat{\xi}_i) = 0; \\ \xi \neq \hat{\xi}_i, & \mathbb{P}^i(d\xi) = 0. \end{cases}$$

由此易知

$$\frac{1}{N}\sum_{i=1}^{N}\int_{\Xi} d(\xi, \hat{\xi}_i)\mathbb{P}^i(d\xi) = 0.$$

将 $\mathbb{P}^i$ 代入原问题的第一个约束 (3-16)

$$\begin{aligned} &\mathcal{A}(\mathbb{P}^1, \mathbb{P}^2, \ldots, \mathbb{P}^N) - [1, \ldots, 1, -\varepsilon]^T \\ &= [1, \ldots, 1, 0]^T - [1, \ldots, 1, -\varepsilon]^T \\ &= [0, \ldots, 0, \varepsilon]^T \in \{0\} \times \cdots \times \{0\} \times \mathcal{S}_+^1, \end{aligned}$$

则 $\mathbb{P}^i$ 满足原问题的约束，是原问题的严格可行解。由 Shapiro(2005) 中的命题 3.4，强对偶定理成立。那么对偶问题 (3-14) 的最优解即为原问题 (3-11) 的最优解，下面继续分析对偶问题 (3-14) 并求解。

(iv) 对偶问题 (3-14) 的求解

先考虑对偶问题 (3-14) 第一个约束, 对于每一个 $\lambda > 0$, 根据 Abadeh(2015) 中的引理 1 有如下结论

$$\max_{\xi \in \mathbb{R}^n} \left\{ g(x, \xi) - \lambda d(\xi, \hat{\xi}_i) \right\} = \begin{cases} g(x, \hat{\xi}_i), & \text{若} \|x\|_* \leq \lambda; \\ -\infty, & \text{其他}. \end{cases}$$

所以 (3-14) 可化为

$$\begin{aligned} \min_{\lambda, s_i} \quad & \lambda\varepsilon + \frac{1}{N}\sum_{i=1}^{N} s_i \\ s.t. \quad & g(x, \hat{\xi}_i) \leq s_i, \quad \forall i = 1, 2, \ldots, N, \\ & \|x\|_* \leq \lambda. \end{aligned} \tag{3-18}$$

在问题 (3-18) 中，最小化目标函数，选取决策变量 $\lambda, s_i(i = 1, 2, \ldots, N)$ 时，考虑到两个不等式约束可以取到 $s_i = g(x, \hat{\xi}_i)(i = 1, 2, \ldots, N), \lambda = \|x\|_*$，代入到目标函数中，得到最小化的结果为 $\frac{1}{N}\sum_{i=1}^{N} g(x, \hat{\xi}_i) + \varepsilon\|x\|_*$,同时对偶问题的解也是原问题的解，即为问题 (3-8) 的解。那么对于不等式约束 (3-6)，我们有

$$\frac{1}{N}\sum_{i=1}^{N} g(x, \hat{\xi}_i) + \varepsilon\|x\|_* \leq 0. \quad \square$$

注：从定理 3.3 中，我们注意到不确定性问题通过对偶理论求解变为经验分布下的目标函数期望再加上一个正则项，正则项的系数由分布模糊集的大小决定。这是从模型形式上的解释，当我们将逻辑回归模型中的损失函数 $l_\beta(x, y) = log(1 + exp(-y\beta^T x))$ 代入定理 3.3 的函数 $g(.,.)$ 中，其中 β 为决策变量即逻辑回归模型中的回归参数，(x, y) 为随机变量，向量 x 代表了数据特征，y 为此数据特征下的分类标签仅可取到 +1 和 -1 代表了两种类型。代入以上具体损失函数时，模型退化为 Abadeh(2015) 中的分布鲁棒逻辑回归模型。相比较于一般的逻辑回归模型所存在的过度拟合，分布鲁棒模型下给了优化问题一个正则项，使得模型具有鲁棒性，同时也能从分布鲁棒优化问题视角下解释了机器学习中进行逻辑回归时加入正则项的合理性。

3.4 Wasserstein 度量下的全局分布鲁棒优化问题

全局分布鲁棒优化问题 (3-2) 中 $\mathcal{P}, \mathcal{P}'$ 为概率分布集，这里用 Wasserstein 度量作为分布距离的测度，那么有模糊集假设如下:

$$\mathcal{P} = \mathbb{B}_{\varepsilon_1}(\mathbb{P}_N) = \left\{\mathbb{P} \in \mathcal{M}(\Xi) : W(\mathbb{P}, \mathbb{P}_N) \leq \varepsilon_1\right\},$$

$$\mathcal{P}' = \mathbb{B}_{\varepsilon_2}(\mathbb{P}_N) = \left\{\mathbb{P}' \in \mathcal{M}(\Xi) : W(\mathbb{P}', \mathbb{P}_N) \leq \varepsilon_2\right\}.$$

其中 $\varepsilon_2 \le \varepsilon_1$ 即有 $\mathcal{P}' \subset \mathcal{P}$。同样地，模糊集定义式中 $\mathbb{P}_N$ 为经验分布，Ξ 为随机变量 ξ 的支撑集，$\mathcal{M}(\Xi)$ 表示在支撑集上所有可能分布的集合。那么问题 (3-2) 变为如下的全局分布鲁棒优化问题 (GDRC)：

$$\mathbb{E}_{\mathbb{P}}[g(x,\xi)] \le \min_{\mathbb{P}' \in \mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} H(\mathbb{P},\mathbb{P}'), \quad \forall \mathbb{P} \in \mathbb{B}_{\varepsilon_1}(\mathbb{P}_N). \tag{3-19}$$

将不确定性问题 (GDRC) 通过对偶转化成确定分布下的问题，然后考虑函数 g 具体形式以及距离函数 H 具体形式下的 GDRCs，并讨论转化后的具体确定性优化问题的类型。

首先考虑约束 (3-19) 是一个半无限约束，因为它必须适用于所有 $\mathbb{P} \in \mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)$，这使得它很难实际应用。在下面的定理中，我们推导了 GDRC(3-19) 的等价形式。在此之前，先对函数 g 以及 H 的函数形式进行一些说明。关于函数 g，前文设定中 $-g$ 是下半连续的正常凸函数，为了计算上的简便，一般讨论线性函数或分片线性函数形式。关于函数 H 的函数形式，$H(\cdot,\cdot)$ 本质为距离函数，这里用 Wasserstein 度量中的联合分布求积分形式来作为距离函数 $H(\cdot,\cdot)$ 的表示

$$H(\mathbb{P},\mathbb{P}') = \int_{\Xi\times\Xi} \phi(\xi,\xi')\Pi(d\xi,d\xi'), \tag{3-20}$$

其中 Π 是概率分布 $\mathbb{P},\mathbb{P}'$ 的联合分布。对于确定的变量 v,z，距离函数可退化为 $H(v,z)=\phi(v,z)$。对于 $H(\cdot,\cdot)$ 函数形式的讨论，即为对 $\phi(\cdot,\cdot)$ 函数形式的讨论，将在定理 3.4 之后展开。

定理 3.4. 设 $g(x,\cdot)$ 是 $\mathbb{R}^n$ 上的凹函数，且 H 是非负函数。$\mathcal{P}'$ 是非空紧凸集，且 $\varepsilon_2 \le \varepsilon_1$，即有 $\mathcal{P}' \subset \mathcal{P}$. 那么 $x \in \mathbb{R}^n$ 满足不等式 (3-19) 当且仅当存在 $z \in \mathbb{R}^n$ 满足如下不等式

$$\frac{1}{N}\sum_{i=1}^{N} g(x,\hat{\xi}_i) + \varepsilon_1\|x-z\|_* + \varepsilon_2\|z\|_* + H^*(z;-z) \le 0. \tag{3-21}$$

证明.

首先考虑上述问题 (3-19)，可转化成如下不等式

$$\max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \left\{ \mathbb{E}_{\mathbb{P}}[g(x,\xi)] - \min_{\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} H(\mathbb{P},\mathbb{P}') \right\} \le 0.$$

不妨假设

$$
\begin{aligned}
F(x) &\equiv \max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - \min_{\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} H(\mathbb{P},\mathbb{P}')\right\} \\
&= \max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N),\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - H(\mathbb{P},\mathbb{P}')\right\},
\end{aligned} \tag{3-22}
$$

那么求解问题 (3-19) 等价于求解 $F(x)\le 0$.

(i) 先对 $F(x)$ 进行初步处理

为了将概率分布 $\mathbb{P},\mathbb{P}'$ 分离，在 (3-22) 中引入两个概率分布 $\mathbb{Q},\mathbb{Q}'$，令其分别相等，加入约束条件中。再引入松弛向量 v,z，把约束条件取最小值加入到目标最大值函数中。

$$
\begin{aligned}
F(x) &= \max_{\substack{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)\\ \mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)}} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - H(\mathbb{P},\mathbb{P}')\right\} \\
&= \max_{\substack{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)\\ \mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)\\ \mathbb{Q},\mathbb{Q}'}} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - H(\mathbb{Q},\mathbb{Q}') \mid \mathbb{Q}=\mathbb{P},\mathbb{Q}'=\mathbb{P}'\right\} \\
&= \max_{\substack{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)\\ \mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)\\ \mathbb{Q},\mathbb{Q}'}} \min_{v,z} \Bigg\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - H(\mathbb{Q},\mathbb{Q}') - \Big(\mathbb{E}_{\mathbb{Q}}(v^T\xi) - \mathbb{E}_{\mathbb{P}}(v^T\xi)\Big) \\
&\qquad - \Big(\mathbb{E}_{\mathbb{Q}'}(z^T\xi) - \mathbb{E}_{\mathbb{P}'}(z^T\xi)\Big)\Bigg\}.
\end{aligned}
$$

我们将上式中的最大值、最小值对调，得到如下对偶问题:

$$
\min_{v,z} \max_{\substack{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)\\ \mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)\\ \mathbb{Q},\mathbb{Q}'}} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] - H(\mathbb{Q},\mathbb{Q}') - \mathbb{E}_{\mathbb{Q}}(v^T\xi) + \mathbb{E}_{\mathbb{P}}(v^T\xi) - \mathbb{E}_{\mathbb{Q}'}(z^T\xi) + \mathbb{E}_{\mathbb{P}'}(z^T\xi)\right\}.
$$

此时，问题变为内部最大化问题，再将内部最大化问题分割成三个小的最大化问题并分别进行分析

$$
\begin{aligned}
\min_{v,z} \Bigg\{ &\max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \left\{\mathbb{E}_{\mathbb{P}}[g(x,\xi)] + \mathbb{E}_{\mathbb{P}}(v^T\xi)\right\} + \max_{\mathbb{Q},\mathbb{Q}'} \left\{-H(\mathbb{Q},\mathbb{Q}') - \mathbb{E}_{\mathbb{Q}}(v^T\xi) - \mathbb{E}_{\mathbb{Q}'}(z^T\xi)\right\} \\
&+ \max_{\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} \left\{\mathbb{E}_{\mathbb{P}'}(z^T\xi)\right\}\Bigg\}.
\end{aligned} \tag{3-23}
$$

三个内部最大化问题假设如下

$$h_1(v,x) = \max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \Big\{ \mathbb{E}_{\mathbb{P}}[g(x,\xi)] + \mathbb{E}_{\mathbb{P}}(v^T\xi) \Big\}, \tag{3-24}$$

$$h_2(v,z) = \max_{\mathbb{Q},\mathbb{Q}'} \Big\{ -H(\mathbb{Q},\mathbb{Q}') - \mathbb{E}_{\mathbb{Q}}(v^T\xi) - \mathbb{E}_{\mathbb{Q}'}(z^T\xi) \Big\}, \tag{3-25}$$

$$h_3(z) = \max_{\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} \Big\{ \mathbb{E}_{\mathbb{P}'}(z^T\xi) \Big\}. \tag{3-26}$$

那么 $F(x)$ 有如下等价形式：

$$\min_{v,z} \big\{ h_1(v,x) + h_2(v,z) + h_3(z) \big\}.$$

(ii) 求解 h_1, h_2, h_3

观察最大化问题 (3-24) 和 (3-26) 不难看出，通过我们的转化这已经是两个单纯的基于 Wasserstein 度量的分布鲁棒优化问题，由定理 3.3，我们得到

$$\begin{aligned} h_1(v,x) &= \max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \Big\{ \mathbb{E}_{\mathbb{P}}[g(x,\xi)] + \mathbb{E}_{\mathbb{P}}(v^T\xi) \Big\} \\ &= \max_{\mathbb{P}\in\mathbb{B}_{\varepsilon_1}(\mathbb{P}_N)} \Big\{ \mathbb{E}_{\mathbb{P}}[g(x,\xi) + v^T\xi] \Big\} \\ &= \frac{1}{N}\sum_{i=1}^{N} \big(g(x,\hat{\xi}_i) + v^T\hat{\xi}_i \big) + \varepsilon_1 \|x+v\|_*. \end{aligned}$$

同理可得

$$\begin{aligned} h_3(z) &= \max_{\mathbb{P}'\in\mathbb{B}_{\varepsilon_2}(\mathbb{P}_N)} \Big\{ \mathbb{E}_{\mathbb{P}'}(z^T\xi) \Big\} \\ &= \frac{1}{N}\sum_{i=1}^{N} z^T\hat{\xi}_i + \varepsilon_2 \|z\|_*. \end{aligned}$$