用于极端事件估计的分布鲁棒优化框架

Yuanlu Bai、Henry Lam、Xinyu Zhang*哥伦比亚大学
工业工程与运筹学系，美国纽约州纽约市 10025

arXiv:2301.01360v1 [stat.ME] 3 Jan 2023

摘要

传统的极端事件估计方法依赖于精心选择的参数模型，这些模型可由极值理论（EVT）提供渐近理论依据。尽管这些方法功能强大且有坚实的理论基础，但可能面临棘手的偏差-方差权衡；当数据量过小时，这一问题尤其严重，进而降低尾部估计的可靠性。本文研究一种基于近期兴起的分布鲁棒优化文献的框架。该方法可视为传统极值理论的非参数替代方法：它对尾部施加一般的形状信念，而非采用参数假设，并利用最坏情形优化来处理非参数不确定性。我们阐释该方法如何规避极值理论中的偏差-方差权衡。另一方面，我们也面临保守性-方差权衡，并说明如何应对这一问题。我们还展示了针对相关优化问题的计算工具，并通过一系列数值示例，将我们的方法与传统极值理论进行比较。

关键词：极端事件估计、分布鲁棒优化、乔凯理论、形状约束、半定规划、保守性

1 引言

极端事件估计广泛应用于各种问题的风险评估和规划中，涵盖百年来的沿海洪灾、地震、金融危机，以及最近的疫情暴发。例如，在环境科学中，河流流量和波浪

*作者谨此感谢美国国家科学基金会通过 CA- REERCMMI-1834710 和 IIS-1849280 两项资助提供的支持。

高度数据用于分析和预测洪水（Davison 和 Smith，1990年）。在工程领域，通过估计故障或崩溃概率来评估系统可靠性（Hei- delberger，1995；Nicola 等人，1993；Jonasson 和 Rootz´en，2014；Zhao 等人，2016年）。在保险和金融领域，重大损失预测为定价和风险管理提供依据（Beirlant 和 Teugels，1992；Embrechts 等人，1997 年；Glasserman 等人，2007、2008；McNeil 等人，2015年）。

极端事件估计面临的一个初始挑战在于，顾名思义，数据中通常只有少量观测值能够直接提供有关分布尾部的信息。为此，主流方法是采用极值理论（EVT）。该方法基于渐近理论提出用于尾部外推的参数模型，主要包含两类方法。第一类是分块最大值法（Gumbel，1958），其依据是著名的 Fisher-Tippet-Gnedenko 定理（Fisher 和 Tippett，1928；Gnedenko，1943）。该定理指出，在最大吸引域假设和适当归一化的条件下，随着样本量增大，独立同分布随机变量 X_i 的最大值依分布收敛于广义极值分布（GEV）。广义极值分布仅包含三种更具体的分布：Gumbel、Fréchet 和 Weibull。拟合 n 个观测值的最大值（例如 M_n）后，我们便可根据关系 P($M_n \leq x$) $= \prod_{i=1}^{n}$P($X_i \leq x$) 推断数据分布的尾部。采用这种方法有其合理性，因为在某些自然情形下，数据是按特定时间段收集的最大值（Davison 和 Smith，1990）；否则，也可以将数据分成若干数据块，并求出每个数据块内的最大值（例如，Embrechts 等人（1997）第 8.1.2 节；McNeil 等人（2015）第 7.1.4 节）。第二类方法是超阈值峰值法（POT）（Smith，1984）。该方法依据 Pickands-Balkema-de Haan 定理（Balkema 和 de Haan，1974；Pickands III，1975），该定理指出，在与 Fisher-Tippet-Gnedenko 定理相同的吸引域假设下，超过某一阈值的所谓超阈值损失的分布函数会随着该阈值增大而收敛于广义帕累托分布（GPD）。

因此，广义帕累托分布适合作为拟合数据尾部的依据。对于这两类方法，无论是广义极值分布还是广义帕累托分布中的参数估计，以及其他实现细节，都已得到深入研究，例如最大似然法（Smith，1985）、概率加权矩法（Hosking 等人，1987）和 Hill 估计量（Davis 和 Resnick，1984；Hill，1975）。最后，除极值理论外，尾部建模还可采用多种探索性“拟合优度”工具，例如分位数图、平均超额图和最大值-和比率图（Embrechts 等人，1997）。

尽管极值理论（EVT）具有理论依据且在实践中颇具用处，但在平衡内在的偏差-方差权衡方面仍可能面临困难。具体而言，分块最大值法通常需要选择数据的区块大小，以获取各区块的最大值。区块数量越多，可用于估计广义极值分布（GEV）参数的最大值就越多，因此估计方差越小；但这不可避免地会减少每个区块中的样本量，从而导致偏差增大。同样，超阈值法（POT）需要选择用于定义超阈值损失的阈值。如果阈值选得较低，超过阈值的数据就更多，可用于估计广义帕累托分布（GPD）的参数，从而降低估计方差；但渐近近似不够准确会导致偏差增大。区块大小或阈值的选择可能与渐近理论的高阶行为密切相关（Smith, 1987年；Bladt 等人，2020年）。此外，当数据量较小时，可能无论如何选择区块大小或阈值，都无法同时使偏差和方差足够小，从而导致总体估计误差显著增大。

受上述传统极值理论（EVT）面临的挑战启发，本文提出一种利用近期兴起的分布鲁棒优化（*DRO*）工具进行极端事件估计的替代方法（Delage 和 Ye，2010年；Goh 和 Sim，2010 年；Kuhn 等人，2019 年）。分布鲁棒优化最初是一种用于不确定性下优化的方法，可视为经典鲁棒优化（RO）的一种推广（Ben-Tal 和 Nemirovski，1998年；Bertsimas 和 Sim，2004年；Ben-Tal 等人，2009年）。在包含不确定或模糊参数的问题中进行决策时，鲁棒优化主张针对最坏情形进行决策优化；这里的最坏情形是在一个称为不确定集或模糊集的可行域内确定的，该集合描述了参数可能取值的范围。因此，这类方法通常涉及一个极小极大问题，其中内层最大化用于计算最坏情形下的参数值。当随机问题中的参数是其底层概率分布时，便需要计算最坏情形分布，这种情形即为分布鲁棒优化。在本文中，我们对分布鲁棒优化采取更宽泛的理解，将其指代最坏情形分布或由此得到的最坏情况值的计算，而不一定涉及极小极大问题。

我们的分布鲁棒优化作为极值理论的一种非参数替代方法，运作方式如下：我们不再基于广义极值分布或广义帕累托分布等渐近合理的参数化假设对尾部进行外推，而是对尾部作出风格化的几何形状假设。这些几何假设包括单调性和凸性，它们有意设定得较为宽松，不仅涵盖广义极值分布和广义帕累托分布的尾部，实际上也涵盖了所有常见的参数化尾部分布。然而，挑战在于，这些宽松的几何假设并不能确定具体的尾部模型，换句话说，在这些前提下，尾部外推的方式可能有很多种。这正是分布鲁棒优化中最坏情况概念发挥作用之处——假设我们想要估计某个目标极值量，例如尾部概率，那么我们可以在所有满足几何假设的外推尾部中，找出使目标极值量达到最坏情况值的那个，

这就归结为在满足形状约束及其他辅助条件的尾部分布空间上求解优化问题，以确保外推一致性；这些约束恰好构成分布鲁棒优化框架中的不确定集。我们将证明，当这些约束经过正确校准时，所得的最坏情况值将为极值量给出统计上正确的界。

上述分布鲁棒优化框架在某种意义上避开了极值理论中的偏差-方差权衡：我们不再需要渐近分布近似，因此也不会因使用参数模型而产生模型设定错误。然而，分布鲁棒优化方法并未给出相合估计量，而是给出了目标极值量的界，这带来了保守性方面的挑战。也就是说，所得的界虽然正确，却可能较为宽松。具体而言，与超阈值法一样，我们的分布鲁棒优化方法需要选择一个用于界定尾部区域的阈值。在超阈值法中，阈值的选择面临偏差-方差权衡：偏差源于使用广义帕累托分布进行近似时产生的误差，而方差则源于其中的参数估计。在分布鲁棒优化中，阈值的选择面临保守性-方差权衡：如果阈值选得过低，保守性就会增大，因为此时有更多满足几何形状条件的尾部分布，从而使最坏情况值的界变得宽松；如果阈值选得过高，方差则会增大，因为可用于校准辅助约束的观测值会减少。为此，我们将通过量化所得界相对于底层尾部分布最大吸引域的宽松程度，研究分布鲁棒优化的保守性。这也为分布鲁棒优化与极值理论之间建立了数学联系。此外，我们还将研究缓解分布鲁棒优化保守性的方法。

这些方法包括增加辅助矩约束，以注入有关尾部的最大信息、缩小不确定集的规模；还包括用于选择界定尾部区域之阈值的程序。

在计算方面，我们还将介绍求解所提出分布鲁棒优化问题的重构方法和步骤。需要指出的是，分布鲁棒优化本质上是无限维优化问题，因为其决策变量是概率分布。为此，我们借鉴优化文献中的结果，将涉及几何形状约束的分布鲁棒优化问题重构为矩问题，并进一步将其对偶化为半定规划。最后，借助可求解且可进行统计校准的分布鲁棒优化表述，我们通过一系列数值示例将该方法与传统极值理论进行比较。具体而言，我们展示了分布鲁棒优化如何比极值理论更可靠地估计目标极端量：相较于极值理论，该方法生成的置信界更常满足正确覆盖。另一方面，分布鲁棒优化也要付出保守性方面的代价，这与我们的理论认识相符。

本文其余部分安排如下。第2节首先介绍用于极端事件分析的分布鲁棒优化框架，包括所涉优化问题中约束和阈值的选择，并讨论相关文献。第3节通过联系极值理论，量化该方法的保守性。第4节讨论所提出分布鲁棒优化问题的求解方法。第 5 节展示该方法的数值表现，并与极值理论进行比较。

2 框架与基本统计保证

我们感兴趣的是估计目标量 $\psi(P)$，它取决于分布 P；该分布未知，但可通过数据观测到。假设量 $\psi(P)$ 是一个极值

量，即 $\psi(P)$ 取决于 P 的尾部，例如某个较大 L 对应的尾部概率 $P(X \geq L)$，或某个区间 $[L, R]$ 对应的 $P(L \leq X \leq R)$，其中 X 服从 P。

我们通过建立一个优化问题来估计 $\psi(P)$。从高层次上看，该问题可写为：

$$\begin{aligned} &\max_P \quad \psi(P) \\ &\text{对 } P \text{ 施加的几何形状条件适用于 } x \geq a \\ &\text{对 } P \text{ 的辅助约束} \end{aligned} \tag{1}$$

其中，决策变量是未知的真实分布 P。在深入细节之前，我们先直观地解释公式 (1) 及其如何用于估计 $\psi(P)$。首先，我们不使用极值理论将 P 拟合为广义极值分布或广义帕累托分布等参数分布，而是对 P 的尾部施加几何形状条件，其中尾部区域由条件 $x \geq a$ 界定， a 是某个较大的阈值。施加形状条件时，我们还必须确保从非尾部区域进行一致的外推，并纳入所有有价值的额外信息，这些信息构成辅助约束。综合这些因素，公式 (1) 旨在为 $\psi(P)$ 提供置信上界。以下是一个直接的统计保证：

命题 2.1。 假设(1) 中的约束以统计置信度$(1-\alpha)$成立，即$\mathbb{P}\left(\begin{array}{l}\text{几何形状条件在 } P_{true} \text{ 上成立，对于 } x \geq a \\ \text{辅助约束在 } P_{true} \text{ 上成立，对于其中 } x \geq a\end{array}\right) \geq 1-\alpha$ 其中 P_{true} 表示真实的 P 分布。那么，优化问题(1)的最优值，记为 Z^*，是真实值 $\psi(P_{true})$的置信上界，记为Z_{true}，且置信度至少相同，即

$$\mathbb{P}(Z^* \geq Z_{true}) \geq 1-\alpha.$$

此外，对于渐近对应项，同样的结论也成立。也就是说，如果极限下确界 $\mathbb{P}\left(\text{几何形状条件在 } P_{true} \text{ 上成立，对于 } \begin{matrix} x \geq a \\ x \geq a \end{matrix} \text{ 辅助约束在 } P_{true} \text{ 上成立，那么} \right) \geq 1-\alpha$

$$\liminf \mathbb{P}(Z^* \geq Z_{true}) \geq 1-\alpha.$$

在上述内容中，$\mathbb{P}$ 表示关于数据的概率，而 liminf 是在样本量趋于无穷时取的。

证明。 假设 P_{true} 满足 (1) 中的所有约束。那么显然，它是优化问题 (1) 的一个可行解，因此 $Z^* \geq \psi(P_{true})$。所以 $\mathrm{P}(Z^* \geq \psi(P_{true})) \geq \mathrm{P}\left(\text{几何形状条件在 } P_{true} \text{ 上成立，而对于 } \begin{matrix} x \geq a \\ x \geq a \end{matrix} \text{ 辅助约束在 } P_{true} \text{ 上成立。渐近对应项也可用同样的方式得到。} \right) \geq 1-\alpha.$ □

问题 (1) 也可以表述为一个最小化问题，其相应的置信下界保证与命题 2.1 类似；但为避免重复，我们省略了这部分。根据命题 2.1，问题转变为：在 (1) 中加入哪些约束，才能为 $\psi(P)$ 获得良好的置信界？我们将在下一小节讨论这一问题。

2.1 优化目标与约束

为便于讨论，我们主要关注 $\psi(P)$，它是 X 的期望或分位数。前一种情况下，$\psi(P)$ 是 $\mathrm{E}[h(X)]$，其中 h 是某个函数，并且我们假设 $h(x)$ 仅在区域 $x \geq a$ 上非零（即尾部相关量；否则，我们总可以将 $\mathrm{E}[h(X)]$ 拆分为两部分：$\mathrm{E}[h(X); X < a]$ 和 $\mathrm{E}[h(X); X \geq a]$，前者是一个可用其他标准统计工具处理的非尾部估计问题）。后一种情况下，$\psi(P) = \min\{q : \mathrm{P}(X \leq q) \geq p\}$，其中 p 是目标概率水平。

约束 在（1）中施加约束是为了提供关于 P_{true}的信息；可行域越小，

所得的界就越紧。然而，在尾部区域，关于 P通常知之甚少。为此，我们施加了几何条件，这些条件为几乎所有常见参数分布所共有，且通常被认为成立。这两项条件是单调性和凸性，我们将其归入（1）中的“几何形状条件”。

更准确地说，令 $f(x)$ 和 $F(x)$ 分别表示概率分布 P 的密度和分布函数。单调性是指将 $f(x)$ 限制为在 $x \geq a$ 上右连续且非递增；凸性是指将 $f(x)$ 限制为在 $x \geq a$ 上凸（注意，此时可保证单侧导数存在；例如，Rockafellar（1970）定理 24.1）。为保持一致，我们可以借助 D 阶单调性的概念（Pestana 和 Mendonca（2001）；Van Parys 等（2019））来表示这两种情形。令 $\mathscr{P}^D(a)$ 表示所有满足以下条件的分布函数的集合：它们 $D-1$ 次可微，且 D 阶右导数存在、有限并在 $[a, \infty$ 上单调。那么，对于 $D=0$， $\mathscr{P}^D(a)$ 退化为 R 上所有不带任何单调性假设的概率分布； $\mathscr{P}^1(a)$ 对应单调尾部分布；标准凸分析表明， $\mathscr{P}^2(a)$ 恰好是凸尾部分布类。我们采用类 $\mathscr{P}^D(a)$ $D=1, 2$。

上述几何形状需要在阈值 a处配以特定的分布信息。在 a 处的这些信息有两个作用，相当于“边界”条件。其一，确保从非尾部区域向 a处尾部区域的外推自然合理，即在 x 达到 a之前的一个小邻域内， a 处的密度仍遵循所施加的形状条件。其二，避免优化及由此得到的置信界变得毫无意义，因为若没有这些边界条件，最坏情形密度可能会过度悲观。更确切地说，在单调情形下，这些信息

指的是 a处的密度值；而在凸情形下，则指 a处的密度值及其右导数。这些量可以利用标准统计工具以置信区间的形式估计（本节稍后将对此进行说明）。有了这些信息，单调情形和凸情形下的式（1）中的“几何形状条件”可写为

$$\mathscr{P}^1_{a,\eta} := \{P \in \mathscr{P}^1(a) | F'_+(a) \leq \eta\},\ \mathscr{P}^2_{a,\underline{\eta},\bar{\eta},\nu} := \{P \in \mathscr{P}^2(a) | \underline{\eta} \leq f(a) \leq \bar{\eta}, f'_+(a) \geq -\nu\}. \tag{2}$$

分别地，其中 $F'_+(\cdot)$ 和 $f'_+(\cdot)$ 表示 $F(\cdot)$ 和 $f(\cdot)$ 的右导数，而 $\eta, \underline{\eta}, \bar{\eta}, \nu$ 是可根据数据校准的参数。注意，在（2）中，单调情形下我们仅使用了 $F'_+(a)$ 的上界，而凸情形下仅使用了 $f'_+(a)$ 的下界。这是因为，当与集合 $\mathscr{P}^1(a)$ 或 $\mathscr{P}^2(a)$ 结合时， $F'_+(a)$ 的下界或 $f'_+(a)$ 的上界在优化问题中会变得冗余。

（1）中的辅助约束采用矩条件的形式，可利用 a以上的数据进行校准。这些约束旨在减少仅利用形状信念外推尾部所带来的保守性。一般而言，我们考虑以下形式的条件

$$\mathbb{E}[\boldsymbol{g}(X)] \in \mathbb{S} \tag{3}$$

其中 $\boldsymbol{g}(\cdot) = (g_j(\cdot))_{j=1,\ldots,d}$ 且 $g_j(\cdot)$ 可以采用示性函数的形式，例如 $g_j(x) = \mathbb{I}(a \leq x \leq b)$，也可以采用幂矩函数的形式，例如 $g_j(x) = x\mathbb{I}(x \geq a)$（后一种形式隐含地要求幂矩有限，但这并非总是合适）。此外，我们总是将其中一个 $g_j(\cdot)$ 设为 $\mathbb{I}(\cdot \geq a)$，以此编码 $x \geq a$ 上的概率质量，并可将其视为前文所述的边界条件。

对于 $\mathbb{S}$，我们考虑两种类型，即椭球 $\mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)$ 和矩形 $\mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}})$，它们分别由参数 $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r) \in \mathbb{R}^d \times \mathbb{R}^{d\times d} \times \mathbb{R}_+$ 和 $(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}) \in \mathbb{R}^d \times \mathbb{R}^d$ 参数化，如下所示

$$\mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r) := \left\{\boldsymbol{y} : [\boldsymbol{y} - \boldsymbol{\mu}]^\top \boldsymbol{\Sigma}^{-1} [\boldsymbol{y} - \boldsymbol{\mu}] \leq r\right\}, \tag{4}$$

$$\mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}) := \left\{\boldsymbol{y} : \underline{\boldsymbol{\mu}} \leq \boldsymbol{y} \leq \bar{\boldsymbol{\mu}}\right\}. \tag{5}$$

在我们.....之前 进一步讨论这些辅助约束的作用，

我们来更深入地讨论这些辅助约束。有人可能会质疑，是否可以将它们作为唯一的约束，而舍弃形状信息。采用这种方法的一个主要问题是，它会导致结果过于保守。这是因为，在无限维规划中存在这样一个事实：仅含矩约束的分布优化问题通常会得到一个有限支撑的最优解，其中支撑点的数量等于独立约束的数量（Winkler，1988）。这种离散分布并不现实，因此会给出一个非常悲观的界。

结合 (2) 中的几何形状条件和 (3) 中的辅助矩约束，(1) 现在可写为

$\mathfrak{P}(\mathscr{P}, \boldsymbol{g}, \mathrm{S})$：在满足 $P \in \mathscr{P}$、$\mathrm{E}_P[\boldsymbol{g}(X)] \in \mathrm{S}$ 的条件下，使 $\sup_P \psi(P)$ 最大化，其中 $\mathscr{P}$ 可以是 $\mathscr{P}_a^1$、η 或 $\mathscr{P}_a^2$、η、$\bar{\eta}$、ν；S 可以是上文所述的 $\mathrm{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)$ 或 $\mathrm{S}_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})$，而 E_P 表示关于 P 的期望。为方便起见，我们还将 $\mathrm{F}(\mathfrak{P}(\mathscr{P}, \boldsymbol{g}, \mathrm{S})) = \{P : \neg P \in \mathscr{P},\ \mathrm{E}_P[\boldsymbol{g}(X)] \in \mathrm{S}\}$ 记为问题 (6) 中的可行域。由命题 2.1 可直接得到以下推论：

定理 2.1。给定阈值 a，有如下结论。

(i) 假设真实分布 P_{true} 属于 $\mathscr{P}^1(a)$。假设所取的值 η 和集合 $\mathbb{S}$ 使真实分布函数满足 $F'_+(a) \le \eta$，且 $\mathbb{E}_{P_{true}}[\boldsymbol{g}(X)] \in \mathbb{S}$，二者同时具有置信度 $1-\alpha$。那么，在 $\mathscr{P} = \mathscr{P}^1_{a,\eta}$ 的条件下，(6) 的最优值也是 $\psi(P_{true})$ 的一个上界，且置信度为 $1-\alpha$。当置信度渐近满足时，也有类似结论。

(ii) 假设真实分布 P_{true} 属于 $\mathscr{P}^2(a)$。假设所取的值 η、η、ν 及集合 $\mathbb{S}$ 满足条件，使得真实密度函数满足 $\eta \le f(a) \le \eta$ 和 $f'_+(a) \ge -\nu$，且 $\mathbb{E}_{P_{true}}[\boldsymbol{g}(X)] \in \mathbb{S}$ 同时以置信度 $1-\alpha$ 成立。那么最优

(6)的值（其中 $\mathscr{P} = \mathscr{P}^2_{a,\eta,\bar{\eta},\nu}$）也是 $\psi(P_{true})$ 的上界，置信度为$1-\alpha$。当置信度渐近满足时，也有类似结论。

根据定理 2.1，只要我们对尾部的几何信念正确，且约束条件中的参数通过正确的置信界进行校准，优化问题 (6) 就能给出统计上有效的置信上界。在单调情形下，η 可作为密度在 a 处的置信上界。在凸情形下，η，η 可作为密度在 a 处置信区间的上下界，而 ν 可作为密度导数在 a 处的置信下界。

对于矩函数 $\boldsymbol{g}$ 和集合 S，选择并校准它们主要有两种方法，分别对应椭球和矩形。第一种方法中，我们可以考虑若干函数，例如区间示性函数或幂函数，并基于基本的多元中心极限定理，将 $\boldsymbol{\mu}$、$\boldsymbol{\Sigma}$ 和 r 分别校准为这些（广义）矩的点估计、其协方差矩阵估计（按样本量缩放）以及一个 χ^2 分位数。第二种方法中，我们可以使用柯尔莫哥洛夫-斯米尔诺夫统计量来校准 $\boldsymbol{\mu}$，$\bar{\boldsymbol{\mu}}$；此时，对于数据点 x_j，每个 $g_j(\cdot) = \mathbf{I}(\cdot \le x_j)$。

请注意，η或 η，η，ν，以及 $\mathrm{E}_{P_{true}}[\boldsymbol{g}(X)] \in \mathrm{S}$ 的条件需要同时成立。因此，需要使用邦费罗尼校正来校准所有这些参数值，以确保族置信水平 $1-\alpha$。这些步骤都是常规操作；为完整起见，我们在附录 A 中展示了详细内容。

最后，我们讨论如何选择阈值 a。一个简单的准则是舍弃 a 以上的少量数据，确保足以应用基本的中心极限定理，从而能够校准矩约束，并以足够的精度估计 a 处的分布信息。或者，我们可以考虑在多个 a 值下求解优化问题 (6)，

例如令 a_i 为 $i = 1, \ldots\ldots, m$。接着，假设每个问题在 a_i 下的可行域 $\mathsf{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))$，由 $\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i$ 配置，并经过校准，使得以置信度 $1-\alpha$ 有 $P_{true} \in \bigcap_{i=1,\ldots,m} \mathsf{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))$（例如，一种做法是确保对每个 i，以置信度 $1-\alpha/m$ 有 $P_{true} \in \mathsf{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))$，但这并非唯一做法）。这样，我们可以将 $\min_{i=1,\ldots,m} Z_i^*$ 设为置信上界，其中 Z_i^* 表示 $\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i)$ 的最优值。这个最终界是所有单独界中的最小值，因此可能更紧，但同时需要权衡额外的邦费罗尼校正对保证 $\mathsf{P}(P_{true} \in \bigcap_{i=1,\ldots,m} \mathsf{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))) \geq 1-\alpha$ 所产生的影响。我们将上述过程的保证总结如下。

定理 2.2。 给定一系列阈值 $a_i, i = 1, \ldots, m$，假设 $\psi(P)$ 仅在区域 $x \geq \max_i a_i$ 上依赖于 P。假设以下条件之一成立：

(i) 真实分布 $P_{true} \in \mathscr{P}^1(\min_i a_i)$。我们对每个满足 $P_{true} \in \bigcap_{i=1,\ldots,m} \mathbb{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))$ 且置信度为 $1-\alpha$ 的情形，设定 $\eta = \eta_i$ 和 $\mathbb{S} = \mathbb{S}_i$，并将其代入式 (6)。

(ii) 真实分布 $P_{true} \in \mathscr{P}^2(\min_i a_i)$。我们对每个满足 $P_{true} \in \bigcap_{i=1,\ldots,m} \mathbb{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))$ 且置信度为 $1-\alpha$ 的情形，设定 $\eta = \eta_i, \eta = \eta_i, \nu = \nu_i$ 和 $\mathbb{S} = \mathbb{S}_i$，并将其代入式 (6)。

那么，$\min_i Z_i^*$ 是 $\psi(P_{true})$ 的上界，置信度为 $1-\alpha$，其中 $Z_i^*, i = 1, \ldots, m$ 是 $\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i)$ 的最优值。当置信度渐近满足时，也有类似的结论。

2.2 相关文献

问题（1）可以视为分布鲁棒优化框架下的最坏情况优化问题。正如引言中所述，分布鲁棒优化倡导在存在模糊随机不确定性的情况下，以最坏情况视角进行决策。后者意味着决策者面对的是一个

随机优化问题，其中控制随机性的底层概率分布未知或存在模糊性。在这种情况下，分布鲁棒优化求解一个极小极大问题，其内层最大化是在被认为包含真实分布的一组分布中，寻找最坏情形分布；这组分布通常称为不确定集或模糊集。这一思想可追溯至Scarf（1957），并已在多个学科中得到日益广泛的应用，包括经济学（Hansen 和Sargent，2008）、随机控制（Petersen 等人，2000；Xu 和 Mannor，2012；Iyengar，2005）、金融（Glasserman 和 Xu，2014）、收益管理（Lim 等人，2006），以及近来兴起的机器学习（Rahimian 和 Mehrotra，2019；Kuhn 等人，2019；Blanchet 等，2021）。问题（1）对分布鲁棒优化采取了更一般的视角，指的是对分布进行最坏情况优化，但不一定涉及决策。命题 2.1 中的论断是数据驱动的分布鲁棒优化所直接提供的保证，即利用数据校准不确定集的分布鲁棒优化问题。具体而言，当该集合通过置信区域构造时，关于不确定性的置信保证会转化为分布鲁棒优化所得最优值的置信保证。

分布鲁棒优化的关键在于构造不确定集。为此，主要有两种方法。第一种是围绕基准分布构造一个邻域球，其大小由统计距离衡量，包括 ϕ- 散度（Gupta，2019；Bayraksan 和 Love，2015；Hu 和 Hong，2013；Gotoh 等，2018；Ghosh 和 Lam，2019；Atar 等，2015；Dey 和 Juneja，2010；Jiang 和 Guan，2018）以及Wasserstein距离（Esfahani 和 Kuhn，2018；Blanchet 和 Kang，2021；Gao 和 Kleywegt，2022；Xie，2019；Shafieezadeh-Abadeh 等，2019；Chen 和 Paschalidis，2018）。这种方法已被用作敏感性分析的非参数方法（Lam，2016，2018）。它还具有统计一致性，即当球的大小根据数据进行适当校准时，所得

最优值会收敛到真实值（Jiang 和 Guan，2016；Bertsimas 等，2018），并且与经验似然及其推广形式密切相关（Lam 和 Zhou，2017；Duchi 等，2021；Lam，2019；Blanchet 等，2019，2022）。构造不确定集的第二种方法是利用汇总分布信息，包括矩和支撑集（Delage 和 Ye，2010；Bertsimas 和 Popescu，2005；Wiesemann 等，2014；Goh 和 Sim，2010；Ghaoui 等，2003）、边际信息（Doan 等，2015；Dhara 等，2021）以及几何形状（Van Parys 等，2016；Li 等，2019；Chen 等，2021）。虽然这种方法不保证统计一致性，但对于分布不确定性显著的问题，它有助于获得界。在这些选择中，形状条件所需的数据输入最少，因此适用于极端事件分析，因为按定义，尾部的数据十分稀少。

近来有多项研究在极端事件分析中考虑分布鲁棒优化。与本文最相关的是 Lam 和 Mottet（2017），该研究提出利用凸性信息进行尾部外推。Lam 和 Mottet（2017）关注外推尾部的轻尾与重尾行为，并提出非线性优化程序，以区分这两种情况并计算最坏情形分布。Blanchet 等（2020）通过将采用 Rényi 散度邻域球作为不确定集的分布鲁棒优化纳入考量，研究广义极值分布的鲁棒化。他们关注最坏情形分布下广义极值分布最大吸引域的保持，并提出了在实践中校准邻域球大小的方法。我们的工作与这些研究不同：我们考虑一种无需广义极值分布的替代方法来估计极值量，因此不同于 Blanchet 等（2020）；此外，我们的不确定集构造（包括约束条件的选择和阈值的选取）及相关的计算方法，比 Lam 和

Mottet（2017）更为一般。

除 Lam 和 Mottet（2017）以及 Blanchet 等（2020）外，其他将分布鲁棒优化用于极值问题的研究还包括 Engelke 和 Ivanovs（2017），该研究考察了以 χ^2 和一阶矩度量的邻域球约束下极值量的鲁棒界；以及 Birghila 等（2021），该研究考察了针对重尾分布、基于 Wasserstein距离和 f-散度邻域的界。此外，在多元情形中，一系列研究探讨了边缘分布已知、但依赖结构未知时的最坏情况界。其中包括 Embrechts 和 Puccetti（2006a,b）关于与金融风险相关的尾概率的研究；Puccetti 和 R¨ uschendorf（2013）关于与金融风险相关的尾概率的研究；Wang 和 Wang（2011）关于和的凸函数期望的研究；以及 Dhara 等（2021）关于条件风险价值的研究。此外，Yuen 等（2020）研究了受极值系数约束的最坏情况风险价值。这些研究与我们的侧重点不同，因为它们主要关注依赖结构，而较少关注尾部外推中的统计问题。

3 保守性

本节将建立关于保守性的理论保证，并借助极值理论推导渐近结果。具体而言，为了量化我们的分布鲁棒优化框架的保守性，我们研究阈值和目标量变得更加极端时相对误差（$:=$ [估计值]/[真实值]-1）的极限行为。首先，我们引入一种抽象形式化，它可视为一般框架的一个特例，并聚焦于两类问题设定：估计尾概率和尾部分位数。接下来，对于这两类问题，我们分别给出重尾分布和轻尾分布下相对误差的渐近结果。

最后，我们提供一些常用分布的示例，以便于理解。

我们推导得出，重尾/轻尾分布会导致所得界的松弛程度不同，并在表3.1中总结了结论。直观而言，与轻尾分布相比，在重尾分布下，我们的分布鲁棒优化方法给出的尾部概率上界保守性较低，因为真实的重尾往往衰减缓慢。反之，在轻尾分布下估计尾部分位数，比在重尾分布下保守性更低。

表3.1：保守性表现

	更重的尾部	更轻的尾部
估计尾概率	保守性较低	更保守
估计尾部分位数	更保守	较不保守

3.1 抽象形式化

由于一般情况下很难推导理论结果，我们聚焦于一种特定的抽象表述，其中几何假设选为凸性，矩区域选为单元素集合，并且阈值以下的信息是明确已知的，而非通过数据校准。虽然我们只考虑了本文所讨论的整个框架中的一个特殊情形，但我们注意到，这些渐近结果建立了与极值理论的联系，也为我们深入理解保守性提供了有益见解。

给定一个连续随机变量 X，其分布函数为 F，密度函数为f，我们将右端点记为 $x_F := \sup\{x \in \mathrm{R} : F(x) < 1\}$。我们假设，直到一个较大的阈值 $a < x_F$， F和f 均已知。我们的目标是估计与 a 以上尾部区域相关的量。

我们首先对真实概率分布作出一些假设。在 $1+\xi x>0$ 条件下，

$$H_\xi(x) := \begin{cases} \exp\left\{-(1+\xi x)^{-1/\xi}\right\} & \text{if } \xi \neq 0 \\ \exp\{-\exp\{-x\}\} & \text{if } \xi = 0 \end{cases}$$

表示广义极值分布（Embrechts 等人，1997 年）。

假设 3.1。 存在 $z < x_F$，使得在区间 (z, x_F)上， F二阶可微，且 f 为正、递减且凸。

假设3.2。 F 属于 H_ξ的最大吸引域（*MDA*），即$F \in MDA(H_\xi)$。换言之，存在标准化常数 $c_n > 0, d_n \in \mathbb{R}$，使得当 $n \to \infty$ 时，对于某个 $\xi \in \mathbb{R}$，有 $c_n^{-1}(M_n - d_n) \Rightarrow H_\xi$，其中 $M_n := \max(X_1, \cdots, X_n)$ 为样本最大值， $X_1, \cdots, X_n$ 是 X的随机实现。

我们回顾一下，广义极值分布有三种不同类型：Fr´echet 分布 ϕ_α ，其中 $\xi = \alpha^{-1} > 0$；Gumbel 分布 Λ ，其中 $\xi = 0$ ；以及 Weibull 分布 ψ_α ，其中 $\xi = -\alpha^{-1} < 0$。 ξ 是控制分布尾部行为的形状参数。事实上，根据 Embrechts 等人（1997），在 Fr´echet 情形下， $x_F = \infty$且 $\lim_{x\to\infty} \frac{\bar{F}(tx)}{\bar{F}(x)} = t^{-\frac{1}{\xi}}$， $t > 0$。这种渐近行为也称为在 ∞ 处以 $-1/\xi$ 为指数的正则变化，简记为 $F \in RV_{-1/\xi}$。 ξ越大， F 的尾部衰减越慢，因此 F 的尾部越重。类似地，当 $\xi = 0$ 且 $x_F = \infty$ （Gumbel 分布）时，其尾部比 Fr´echet 情形下的尾部更轻。

在 $\xi = 0$ 的情况下，我们增加一个假设；正态分布、伽马分布和指数分布等教科书分布通常都满足该假设。请注意，假设 3.3 与上述满足 $\xi = 0$ 的假设共同意味着 F 是一个冯·米塞斯函数。

假设 3.3. 极限 $\lim_{x\uparrow x_F} \frac{F(x) f'(x)}{\bar{F}^2(x)} = -1$.

已知 f 根据 Embrechts 等人（1997年），如果 $\xi > 0$，那么 $x_F = \infty$；如果 $\xi < 0$，那么

$x_F < \infty$；如果 $\xi = 0$，那么 x_F 可以是有限的，也可以是无限的。事实上，在我们的假设下，我们可以聚焦于 $x_F = \infty$ 的情况，以下两个命题为此提供了依据:

命题 3.1。 假设 F 满足假设 *3.1* 和假设*3.2*，其中 $\xi < 0$。令 $Y = 1/(x_F - X)$。用 F_Y 和 f_Y 分别表示 Y 的分布函数和密度函数。则存在 $z_Y < \infty$，使得在 (z_Y, ∞) 上， F_Y 二阶可微，且 f_Y 为正、递减且凸。此外， $F_Y \in MDA(H_{-\xi})$。

命题 3.2。 假设 F 满足假设 *3.1*、假设*3.2*（其中 $\xi = 0$），并且满足假设 *3.3*。此外，我们假设 $x_F < \infty$。令 $Y = 1/(x_F - X)$。用 F_Y 和 f_Y 分别表示 Y 的分布函数和密度函数。则存在 $z_Y < \infty$，使得在 (z_Y, ∞)上， F_Y 二阶可微，且 f_Y 为正、递减且凸。此外， $F_Y \in MDA(\Lambda)$ 具有无穷右端点。此外，我们有

$$\lim_{x \to \infty} \frac{\bar{F}_Y(x) f_Y'(x)}{f_Y^2(x)} = -1. \tag{7}$$

根据上述两个命题，在满足我们的假设时，如果 $x_F < \infty$，那么我们可以定义 $Y = 1/(x_F - X)$，它也满足我们的假设。已知 $F(x)$ $x \le a$ 意味着我们知道 Y 的分布函数 F_Y，直至 $1/(x_F - a)$。因此，我们可以将问题转化为一个等价的、具有无穷右端点的问题。此后，不失一般性，我们假设 $x_F = \infty$，且对于某个 $\xi > 0$，要么 $F \in MDA(H_\xi)$，要么$F \in MDA(\Lambda)$。

首先，我们考虑估计尾概率 $\psi(P) = E[h(X)]$，其中 $h(x) = \mathrm{I}(x > b)$，即 $\psi(P) = \mathrm{P}(X > b) = \int_b^\infty f(x)\mathrm{d}x$，其中 $b = b(a)$ 满足 $a \le b < \infty$。我们从已知信息 $f(x)$ $x \le a$ 中提取三个常数 β， η， ν:

$$\beta = \beta(a) := 1 - F(a) = \bar{F}(a);\ \eta = \eta(a) := f(a);\ \nu = \nu(a) := -f_+'(a).$$

为便于后续使用，我们定义

$$\mu = \mu(a) := \frac{\eta}{\nu}, \sigma = \sigma(a) := \frac{2\beta}{\nu}.$$

对于 β， η， $\nu > 0$，我们考虑以下形式的 RO问题：

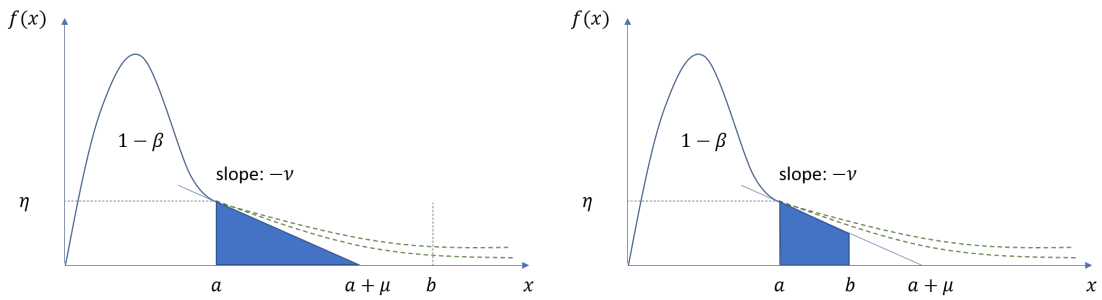
$$\begin{aligned} \max_f \ & \int_a^\infty \mathbb{I}(x > b) f(x) \mathrm{d}x \\ s.t. \ & \int_a^\infty f(x) \mathrm{d}x = \beta, \\ & f(a) = f(a+) = \eta, \\ & f'_+(a) \geq -\nu, \\ & f \text{ convex for } x \geq a, \\ & f(x) \geq 0 \text{ for } x \geq a. \end{aligned} \tag{8}$$

注意，(8) 等价于 $\mathfrak{P}(\mathscr{P}^2_{a,\ \eta,\ \eta,\ \nu}, \mathrm{I}(x > a)\ \{\beta\})$，因此我们实际上考虑的是此前定义框架的一个特例。从现在起，以上问题的最优值记为 $z^* = z^*(a,\ b)$。应用 Lam 和 Mottet（2017）的结果，可得以下定理：

定理 3.1。 若 $\eta^2 < 2\beta\nu$，则(8)的最优值记为 z^*，其为

$$z^* = \begin{cases} \frac{\nu}{2}(\sigma - \mu^2) & if\ \mu \leq b - a; \\ \frac{\nu}{2}[\sigma - 2(b-a)\mu + (b-a)^2] & if\ \mu > b - a. \end{cases} \tag{9}$$

为了直观理解定理 3.1，请注意，若从 (a, η) 处作一条斜率为$-\nu$的直线，则该直线会在 $(a + \mu, 0)$ 处与横轴相交。任何可行密度函数的尾部外推都必须位于这条直线之上。另一方面，它们可以尽可能接近这条直线。因此，直观而言， z^* 等于 β 减去图1中阴影区域的面积，这与定理中的结果完全一致。为了量化尾概率估计的保守性，我们需要指定一个关于 a 的适当函数 $b = b(a)$，并



(a) $b \geq a + \mu$ (b) $b < a + \mu$

图1：最优值的直观解释 z^*

考虑当 $a \to \infty$时相对误差的极限，其定义如下

$$\lim_{a\to\infty} \frac{z^*(a, b(a)) - \bar{F}(b(a))}{\bar{F}(b(a))}. \tag{10}$$

从直观上看，(10) 的值越大，鲁棒优化方法就越保守。我们将在后续小节中介绍 $b(a)$ 的选择。

现在我们考虑尾部分位数的估计。同样，我们可以应用 RO方法来获得最坏情况估计。具体而言，给定关于 $x \leq a$ 的 $F(x)$ 信息，假设我们的目标是估计 $q = F^{-1}(p)$，其中 $p \geq 1 - F(a) = 1 - \beta$。为了获得最坏情况估计，我们在所有可能的凸尾部外推中最大化 q。也就是说， q 的最坏情况估计通过求解以下问题得到

$$\begin{aligned}
\max_{f} \ & q \\
\text{s.t.} \ & \int_{-\infty}^{q} f(x)\mathrm{d}x = p, \\
& \int_{a}^{\infty} f(x)\mathrm{d}x = \beta, \\
& f(a) = f(a+) = \eta, \\
& f'_{+}(a) \geq -\nu, \\
& f \text{ convex for } x \geq a, \\
& f(x) \geq 0 \text{ for } x \geq a.
\end{aligned} \tag{11}$$

我们定义

$$q^* := \inf\{b \geq a : z^*(a,b) = 1-p\}. \tag{12}$$

图2中的曲线反映了 z^* 相对于 b的形状。对于满足 $\beta - \eta^2/(2\nu) \leq 1-p_1 \leq \beta$的 p_1，我们可以找到一个对应的点 q_1^*，使得 $z^*(a, q_1^*) = 1-p_1$。特别地，如果 $1-p_1 = \beta - \eta^2/(2\nu)$，那么根据式（12）的定义，我们有 $q_1^* = a + \mu$。然而，对于满足 $1-p_2 < \beta - \eta^2/(2\nu)$的 p_2，相应的 q_2^*定义为 ∞。下面的定理为 q^*这一直观定义提供了依据。

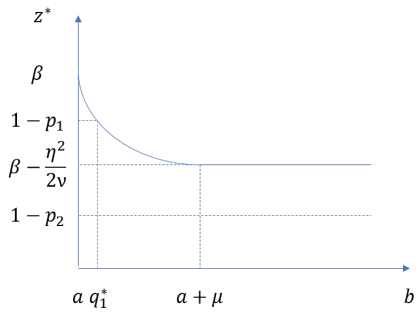


图2： z^* 相对于 b 的形状及 q^*的定义

定理 第3.2节 q^* 在（12）中定义，是（11）的最优值，可表示为

$$q^* = \begin{cases} a + \mu - \sqrt{\mu^2 - \sigma + \frac{2(1-p)}{\nu}} & if\ 1-\beta \leq p \leq 1-\beta+\frac{\eta^2}{2\nu}; \\ \infty & if\ p > 1-\beta+\frac{\eta^2}{2\nu}. \end{cases} \tag{13}$$

与尾概率估计类似，我们设法找到一个合适的函数 $p = p(a)$，它关于 a，并计算相对误差在 $a \to \infty$时的极限，其定义为

$$\lim_{a\to\infty} \frac{q^*(p(a)) - q(p(a))}{q(p(a))}. \tag{14}$$

回顾一下，$q = q(p(a)) := F^{-1}(p(a))$ 是真实的 p 分位数。同样，（14）的值越大，RO方法在估计尾部分位数时就越保守。

3.2 估计尾部概率时的保守性

我们考虑以下两种情况。

情形1： $\xi > 0$。假设 F（X 的真实分布函数和密度函数）满足假设 3.1 和 3.2，且 $\xi > 0$。由于 $F \in RV_{-1/\xi}$，根据 Karamata 表示定理（Embrechts 等人，1997 年），F 具有如下表示：

$$\bar{F}(x) = c(x) \exp\left\{ -\int_z^x \frac{1}{u(t)} \mathrm{d}t \right\}, z < x < \infty \tag{15}$$

其中 $c(x) \to c > 0$，$u(x)/x \to \xi$，当 $x \to \infty$ 时成立。此外，已知当阈值 a 充分大时，

$$P(X > x | X > a) \approx \bar{G}_{\xi;a,u(a)}(x) = \left(1 + \xi \frac{x-a}{u(a)}\right)^{-\frac{1}{\xi}}, x \geq a,$$

这正是广义帕累托分布方法的数学基础。

从直观上看，如果对于任意 $x \geq a$，$P(X > x | X > a)$ 恰好等于 $G_{\xi;a,\ u(a)}(x)$，那么我们可以得到

$$\bar{F}(x) = P(X > a)\bar{G}_{\xi;a,u(a)}(x) = \beta \left(1 + \xi \frac{x-a}{u(a)}\right)^{-\frac{1}{\xi}}.$$

因此，

$$f(x) = -\frac{\mathrm{d}}{\mathrm{d}x} P(X > x) = \frac{\beta}{u(a)} \left(1 + \xi \frac{x-a}{u(a)}\right)^{-\frac{1}{\xi}-1},$$

$$-f'(x) = -\frac{\mathrm{d}}{\mathrm{d}x} f(x) = \frac{(\xi+1)\beta}{u^2(a)} \left(1 + \xi \frac{x-a}{u(a)}\right)^{-\frac{1}{\xi}-2}.$$

如果我们用 a 替换 x，那么可以得到

$$\eta = \frac{\beta}{u(a)}, \nu = \frac{(\xi+1)\beta}{u^2(a)}.$$

因此，使用 β/u（a）和（$\xi+1$）β/u^2（a）分别近似 η 和 ν 似乎是合理的。事实上，$a \to \infty$ 时确实如此，以下命题可以证明这一点：

命题 3.3. 假设分布函数 F 及其对应的密度函数 f 满足假设 *3.1* 和 *3.2*，其中 $\xi > 0$，这意味着 F 具有如下表示：

式 (15)。以下陈述成立：

$$\lim_{x\to\infty}\frac{f(x)u(x)}{\bar{F}(x)}=1;\ \lim_{x\to\infty}-\frac{f'(x)u^2(x)}{(\xi+1)\bar{F}(x)}=1. \tag{16}$$

备注。 注意，该定理意味着，当 $a\to\infty$ 时，

$$\eta\sim\frac{\beta}{u(a)},\nu\sim\frac{(\xi+1)\beta}{u^2(a)},\mu\sim\frac{u(a)}{\xi+1},\sigma\sim\frac{2u^2(a)}{\xi+1}. \tag{17}$$

本节中，我们用 $g_1(a)\sim g_2(a)$ 表示 $\lim_{a\to\infty}g_1(a)/g_2(a)=1$。

为简便起见，我们首先考虑 $b=a+xu(a)$，其中 $x\geq 0$ 是一个固定数值。在这种情况下，已知当 $a\to\infty$ 时，$\bar{F}(b)/\bar{F}(a)\to(1+\xi x)^{-1/\xi}$（Embrechts 等人，1997 年）。利用上文命题 3.3 中的结论，我们可以得到以下定理：

定理 3.3。 假设分布函数 F 及其相应的密度函数 f 满足假设 *3.1* 和 *3.2*，且 $\xi>0$，这意味着 F 具有(15)所示的表示。 $b=b(a)$ 选为 $b=a+xu(a)$，其中 $x\geq 0$ 是一个固定数值。那么我们有

$$\lim_{a\to\infty}\frac{z^*(a,b)}{\bar{F}(b)}=\begin{cases}\left(1-\frac{1}{2(\xi+1)}\right)(1+\xi x)^{\frac{1}{\xi}} & \text{if } x\geq\frac{1}{\xi+1};\\ \left(1-x+\frac{\xi+1}{2}x^2\right)(1+\xi x)^{\frac{1}{\xi}} & \text{if } x<\frac{1}{\xi+1}.\end{cases} \tag{18}$$

事实上，我们可以将定理 3.3 中的结果推广到 $b=a+x(a)u(a)$ 的情形，其中 $\lim_{a\to\infty}x(a)=x_0\geq 0$。根据 Resnick (1987年) 中的命题0.5，可得极限 $\lim_{x\to\infty}\frac{\bar{F}(tx)}{\bar{F}(x)}=t^{-\frac{1}{\xi}}$ 在 $(0,\infty)$ 上局部一致成立。由于极限 $t^{-\frac{1}{\xi}}$ 在 t 中连续，因此连续收敛成立。我们有

$$\lim_{a\to\infty}\frac{a+x(a)u(a)}{a}=1+\xi x_0>0,$$

so

$$\lim_{a\to\infty}\frac{\bar{F}(a+x(a)u(a))}{\bar{F}(a)}=(1+\xi x_0)^{-\frac{1}{\xi}}.$$

此外，我们也可以按照上述证明中的讨论，求得 $z^*/F(a)$ 的极限，进而求得 $z^*/F(b)$ 的极限。例如，若选择 $b=2a$，则可以设定

$x(a) = a/u(a) \to 1/\xi$。由于 $1/\xi > 1/(\xi+1)$，我们得到

$$\lim_{a\to\infty} \frac{z^*}{\bar{F}(b)} = \left(1 - \frac{1}{2(\xi+1)}\right) 2^{\frac{1}{\xi}}. \tag{19}$$

我们注意到，对于 $\xi > 0$，(19) 关于 ξ 单调递减，并且当 $\xi \to \infty$ 时趋于 1。这表明，对于重尾分布，分布鲁棒优化方法在估计尾概率时较不保守。

情形 2： $\xi = 0$ 且 $x_F = \infty$。现在假设 F 和 f 满足假设 3.1、在 $\xi = 0$ 条件下的假设3.2，以及假设 3.3。将 (17) 中的 ξ 替换为 0，我们自然猜想，当 $a \to \infty$ 时，

$$\eta \sim \frac{\beta}{u(a)}, \nu \sim \frac{\beta}{u^2(a)}, \mu \sim u(a), \sigma \sim 2u^2(a). \tag{20}$$

下面的命题证实了这一猜想。

命题 3.4。假设分布函数 F 及其对应的密度函数 f 满足假设 *3.1*、在 $\xi = 0$ 条件下的假设 *3.2*，以及假设 *3.3*。我们有 ¯，其中 F 是冯·米塞斯函数，其表示如下：

$$\bar{F}(x) = c \exp\left\{-\int_z^x \frac{1}{u(t)} \mathrm{d}t\right\}, z < x < \infty \tag{21}$$

¯其中 c 是某个正数，且 $u(x) = F(x)/f(x)$ 为正且绝对连续，并且 极限$_{x\to\infty} u'(x) = 0$。此外，以下结论成立：

$$\lim_{x\to\infty} \frac{f(x)u(x)}{\bar{F}(x)} = 1; \tag{22}$$

$$\lim_{x\to\infty} -\frac{f'(x)u^2(x)}{\bar{F}(x)} = 1. \tag{23}$$

与情形1类似，我们可以选择 $b = a + xu(a)$，其中 $x \geq 0$ 是一个固定数值，由此可得以下结果：

定理 3.4. 假设分布函数 F 及其对应的密度函数f 满足假设 *3.1*、在 $\xi = 0$ 下的假设 *3.2*，并且满足假设 *3.3*。

$b = b(a)$ 选为 $b = a + xu(a)$，其中 $x \geq 0$ 为固定数值。于是我们有

$$\lim_{u\to\infty} \frac{z^*(a,b)}{\bar{F}(b)} = \begin{cases} \frac{1}{2}e^x & if\ x \geq 1; \\ \left(1 - x + \frac{1}{2}x^2\right)e^x & if\ x < 1. \end{cases} \tag{24}$$

事实上，$\lim_{a\to\infty} u(a)/a = 0$，因此无论 x多大，我们总有$(b-a)/a = xu(a)/a \to 0$，当 $a \to \infty$时成立。因此，如果改为选择 $b = 2a$，那么对于任意 $x \geq 0$，$\lim_{a\to\infty} F(b)/F(a)$ $\leq \lim_{a\to\infty} F(a + xu(a))/F(a) = e^{-x}$，从而 $F(b)/F(a) \to 0$，当$a \to \infty$时成立。我们还知道，在这种情况下，$z^*(a,b)/F(a) \to 1/2$。因此，在 Gumbel 情形下，$z^*(a,2a)/F(2a) \to \infty$，当 $a \to \infty$时成立。与式 (19) 相比，我们得出结论：在估计尾概率时，轻尾情形下的分布鲁棒优化方法比重尾情形下更保守。

3.3 估计尾部分位数的保守性

我们再次考虑以下两种情形。

情形1: $\xi > 0$. 假设 F和 f满足假设 3.1 和假设3.2，且$\xi > 0$。为简单起见，我们选择 $p = p(a)$，使得 $1 - p = x\beta$其中$1 - 2^{-1/(\xi+1)} < x \leq 1$是一个固定数。回顾一下，若 $p > 1 - \beta + \eta^2/(2\nu)$，则 $q^* = \infty$。利用 (17)，可得

$$\lim_{a\to\infty} \left(\beta - \frac{\eta^2}{2\nu}\right)/\beta = 1 - \frac{1}{2(\xi+1)}.$$

因此，要求$x > 1 - 2^{-1/(\xi+1)}$可确保当 a 足够大时，$1-p = x\beta > \beta - \eta^2/(2\nu)$，因而 RO 方法能够给出非平凡估计。

定理 3.5。假设分布函数 F 及其相应的密度函数f满足假设 *3.1* 和假设*3.2*，且 $\xi > 0$。p取为 $1 - p = x\beta$，其中$1 - 2^{-1/(\xi+1)} < x \leq 1$ 为固定数值。则有

$$\lim_{a\to\infty} \frac{q^*}{q} = x^{\xi}\left[\frac{\xi}{\xi+1}\left(1 - \sqrt{1 - 2(1-x)(\xi+1)}\right) + 1\right]. \tag{25}$$

我们注意到，随着 ξ增大， x的可行区间，即 $(1-2^{\frac{1}{(\xi+1)}}, 1]$，会变窄。我们还注意到，(25) 有一个只取决于 ξ的上界。粗略地说，对于任意 $\xi > 0$和 $1-2^{\frac{1}{(\xi+1)}} < x \leq 1$，(25) 的值始终有界于 $\frac{\xi}{\xi+1}+1$，且该值随 ξ增大而增大。因此，对于重尾分布，尾部分位数的估计更加保守。

情形 2： $\xi = 0$ **且** $x_F = \infty$。现在我们假设 F和 f满足假设 3.1、带有 $\xi = 0$的假设 3.2，以及假设 3.3。此外，我们选择 p使得 $1-p = x\beta$，其中 $1/2 < x \leq 1$是一个固定值，这保证了当 a足够大时， $q^* < \infty$。在这种情况下，我们有以下定理:

定理 3.6。假设分布函数 F及其相应的密度函数f满足假设 *3.1*、带有 $\xi = 0$ 的假设 *3.2*，以及假设 *3.3*。 p取为 $1-p = x\beta$，其中 $1/2 < x \leq 1$是一个固定值。那么我们有

$$\lim_{a\to\infty} \frac{q^*}{q} = 1. \tag{26}$$

与定理 3.5 相比，在轻尾情形下， x 的可行区间不仅更宽，而且相对误差的极限也更小，因此估计尾部分位数时保守性更低。

3.4 示例

对于案例 1 和 2，我们分别以帕累托分布和标准正态分布为例，验证了关于分布鲁棒优化方法的保守性与尾部厚重程度之间关系的主要结论。

示例3.1 (帕累托分布**)**。 假设 X 服从尺度参数为 $x_m > 0$ 、形状参数为 $\alpha > 0$的帕累托分布。那么尾分布函数为$\bar{F}(x) = (x_m/x)^{\alpha}$ *for* $x \in [x_m, \infty)$。已知 $F \in MDA(H_\xi)$，其中 $\xi = \alpha^{-1} > 0$。

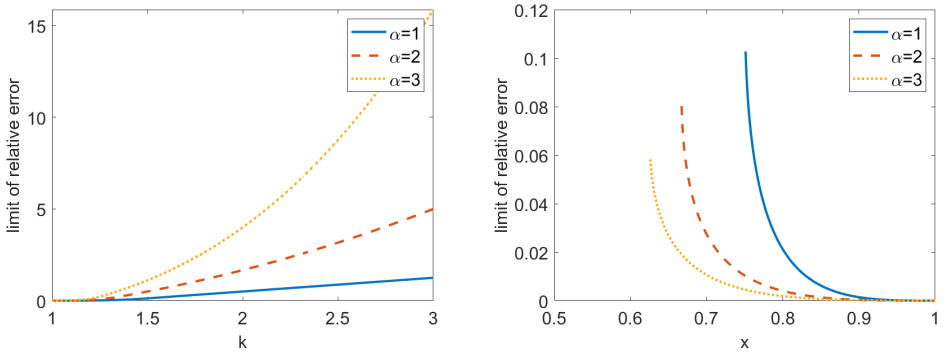
α 越大，ξ 就越小，尾部也越轻。在估计尾概率时，我们选择 $b = ka$，其中 $k \geq 1$。然后，根据定理 *3.3* 后的讨论可知，相对误差的极限为

$$\lim_{a\to\infty} \frac{z^*(a, b(a)) - \bar{F}(b(a))}{\bar{F}(b(a))} = \begin{cases} \left(1 - \frac{1}{2(\xi+1)}\right) k^{\frac{1}{\xi}} - 1 & if\ k \geq 1 + \frac{\xi}{\xi+1}; \\ \left(1 - \frac{k-1}{\xi} + \frac{(\xi+1)(k-1)^2}{2\xi^2}\right) k^{\frac{1}{\xi}} - 1 & if\ k < 1 + \frac{\xi}{\xi+1}. \end{cases}$$

在估计尾部分位数时，我们选择 $1 - p = x\beta$ 其中 $1 - \frac{1}{2(\xi+1)} < x < 1$。根据定理 *3.5*，可得相对误差的极限为

$$\lim_{a\to\infty} \frac{q^*(p(a)) - q(p(a))}{q(p(a))} = x^{\xi} \left[\frac{\xi}{\xi+1}\left(1 - \sqrt{1 - 2(1-x)(\xi+1)}\right) + 1\right] - 1.$$

图 *3* (a) 展示了在估计概率时，相对误差的极限如何随 k 和 α 变化；而(b) 展示了在估计分位数时，该极限如何随 x 和 α 变化。可以看出，在估计概率时，尾部越重，保守性越低；然而，在估计分位数时，尾部越重，相对误差的极限越大，同时能给出非平凡估计 q^* 的 x 范围越小，因此分布鲁棒优化方法更加保守。



(a) 估计尾概率　　(b) 估计尾部分位数

图 3：帕累托分布的相对误差极限

例 3.2（标准正态分布）。假设 X 服从标准正态分布。已知分布函数 $F \in MDA(H_\xi)$ 其中 $\xi = 0$。在

¯特别地， F是一个冯·米塞斯函数，其辅助函数为 $u(x) = F(x)/f(x) \sim 1/x, x \to \infty$。在估计尾概率时，若 $b = a + xu(a)$，则当 $a \to \infty$ 时，我们总有 $b - a \to 0$。然而，式 (24) 中的值仍随 x呈指数增长。因此，分布鲁棒优化方法在估计概率时非常保守。在估计尾部分位数时，我们可以对 $1/2 < x \leq 1$ 选择$1 - p = x\beta$。这个范围比任何 $\xi > 0$ 的范围都大，但相对误差的极限始终等于 *0*，这意味着与 *Fr´echet* 情形相比，分布鲁棒优化方法的保守性较低。

4 优化重构与求解可处理性

本节主要关注作为期望的 $\psi(P)$，即对于某个函数 h，$\mathrm{E}_P[h(X)]$。当目标 $\psi(P)$ 是分位数时，由于 $\psi(P) = \min\{q \colon \mathrm{P}(X \leq q) \geq p\}$ 可写为 $\psi(P) = \min\{q \colon \mathrm{E}[h(X)] \geq p - \mathrm{P}(X \leq a)\}$，其中 $h(X) = \mathrm{I}(a \leq X \leq q)$，而非尾部部分 $\mathrm{P}(X \leq a)$ 理应可以通过标准统计工具（例如经验分布）处理，因此分析可归结为期望情形。由此，在求分位数时，可以考虑对 q 采用二分法等线搜索方法，以获得满足以下条件的最小 q：

$$\max_{P \in \mathbb{F}(\mathfrak{P}(\mathscr{P}, \boldsymbol{g}, \mathbb{S}))} \mathbb{E}[\mathbb{I}(a \leq X \leq q)] \geq p - \mathbb{P}(X \leq a).$$

求解（6）存在一些挑战，因为它是一个无限维优化问题，而且几何形状约束带来了额外的复杂性。在接下来的小节中，我们将展示如何借助优化文献中的技术，将（6）化简为一个矩问题，再将其转化为对偶问题，从而归入可用标准方法求解的程序类别。

4.1 转化为矩问题

给定由 $\boldsymbol{g}$ 和 S 配置的 $\psi(P) = \mathrm{E}_P[h(X)]$, 矩约束，以及作为 $\mathscr{P}^1_{a,\ \eta}$ 或 $\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu}$ 的形状信息 $\mathscr{P}$，（6）可以写为

$\mathfrak{P}(h,\ \mathscr{P},\ \boldsymbol{g},\ \mathrm{S})$： $\max_P \mathrm{E}_P[h(X)]$ 满足 $P \in \mathscr{P},\ \mathrm{E}_P[\boldsymbol{g}(X)] \in \mathrm{S}$，其中我们现在(暂)重强调 h 在记号 $\mathfrak{P}(h,\ \mathscr{P},\ \boldsymbol{g},\ \mathrm{S})$ 中的作用。处理 (27) 的第一步是将其转换为等价的矩问题：

$\mathfrak{M}_a(H,\ \boldsymbol{G},\ \mathrm{S}')$： $\max_{X\sim Q} \mathrm{E}_Q[H(X)]$ 满足 $\mathrm{E}_Q[\boldsymbol{G}(X)] \in \mathrm{S}',\ \ Q \in \mathscr{P}[a,\infty)$，其中集合 S' 由 S 导出，$\mathscr{P}[a,\infty)$ 表示支撑集为 $[a,\infty)$ 的概率分布类，函数 H 和向量函数 $\boldsymbol{G}$ 分别由 h 和 $\boldsymbol{g}$导出。更确切地说，我们有以下结果。

定理 4.1. 假设 h 有界，且每个 g_i 都有下界，其支撑集为 $x \geq a$。

1. $\mathfrak{P}(h,\mathscr{P}^1_{a,\eta},\boldsymbol{g},\mathrm{S})$ 等价于 $\mathfrak{M}_a\left(\eta\tilde{H},\eta\tilde{\boldsymbol{G}},\mathrm{S}\right)$，其中 $\tilde{H}(x) = \int_a^x h(u)du$，且对于向量函数 $\tilde{\boldsymbol{G}}$ 中的每个分量 i，$\tilde{G}_i(x) = \int_a^x g_i(u)du$。可行解 P（属于 $\mathfrak{P}(h,\mathscr{P}^1_{a,\eta},\boldsymbol{g},\mathrm{S})$）与 Q（属于 $\mathfrak{M}_a(\eta\tilde{H},\eta\tilde{\boldsymbol{G}},\mathrm{S})$）之间的双射变换由 $P'_+(x) = \eta(1-Q(x))$ 给出（将 P 和 Q 视为分布函数）。

2. $\mathfrak{P}(h,\mathscr{P}^2_{a,\eta,\bar{\eta},\nu},\boldsymbol{g},\mathrm{S})$ 等价于 $\mathfrak{M}_a\left(\nu H,\left(\nu(x-a)\ \nu\boldsymbol{G}^\top\right)^\top,\mathrm{S}_R(\eta,\bar{\eta})\times\mathrm{S}\right)$，其中 $H(x) = \int_a^x\int_a^u h(v)dvdu$，且对于向量函数 $\boldsymbol{G}$ 中的每个分量 i，$G_i(x) = \int_a^x\int_a^u g_i(v)dvdu$。可行解 P（属于 $\mathfrak{P}(h,\mathscr{P}^2_{a,\eta,\bar{\eta},\nu},\boldsymbol{g},\mathrm{S})$）与 Q（属于 $\mathfrak{M}_a\left(\nu H,\left(\nu(x-a)\ \nu\boldsymbol{G}^\top\right)^\top,\mathrm{S}_R(\eta,\bar{\eta})\times\mathrm{S}\right)$）之间的双射变换由 $-P^{(2)}_+(x) = \nu(1-Q(x))$ 给出，其中 $P^{(2)}_+$ 表示 P 的二阶右导数（将 P 和 Q 视为分布函数）。

为了说明 定理 4.1，例如，考虑将尾区间概率作为目标，

其中，对于某个给定数值 $L, R,\ \ h(x) = \mathbf{I}(L \le x \le R)$。我们有

$$\tilde{H}(x) = (x - L)\mathbb{I}(L \le x \le R) + (R - L)\mathbb{I}(x \ge R);$$

$$H(x) = \frac{1}{2}(x - L)^2\mathbb{I}(L \le x \le R) + (R - L)(x - \frac{R + L}{2})\mathbb{I}(x \ge R).$$

如果 $g_i(x) = x^{i-1}\mathbf{I}(x \ge a)\ i \ge 1$，则有

$$\tilde{G}_i(x) = (x^i - a^i)/i,\ G_i(x) = x^{i+1}/(i^2 + i) - a^i x/i + a^{i+1}/(i + 1).$$

定理 4.1 表明，通过将决策变量（即概率分布）与支撑集为 $[a, \infty)$ 的概率分布函数建立一一映射，(6) 等价于一个矩约束规划。我们给出两种不同的方法来证明定理 4.1。第一种方法是分部积分法，即用决策变量中的分布函数的导数替代该分布函数。这种方法基于 Lam 和 Mottet（2017）的工作，该工作考虑了限制更严格的表述。第二种方法是 Choquet 理论；根据凸分析，该理论意味着紧凸集中的任意一点都可以表示为其极点的混合。在我们的语境中，规定的概率分布类可以写成较简单分布的混合表示。由此，可以将优化问题改写为以混合分布为决策变量的形式，随后消除形状约束。尽管这一方法的主要思想源自一些已有的分布鲁棒优化研究（Popescu, 2005; Van Parys 等, 2016; Li 等, 2019），但我们的定理适用于一般的矩集合 S，也允许形状约束 $\mathscr{P}^1_{a,\ \eta}$ 或 $\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu}$ 中相关参数满足不等式约束，而这些研究中使用的是单点集。

–

4.2 对偶化为半正定程序

矩问题 (28) 的约束条件数量有限，但决策变量是无限维的。接下来，我们将其转化为一个具有有限维

决策变量但约束条件无限多的对偶规划，并可进一步将其简化为更易处理的表述。由于矩问题 (28) 中的集合 S' 通常同时包含椭球集和矩形集，我们也在这一一般性下陈述定理。首先，我们引入以下假设以保证强对偶性：

假设 4.1 (斯莱特条件)。 在 $\mathfrak{M}_a(H, \mathbf{G}, \mathbb{S})$ 中，其中 $(G_j)_{j=1}^d$ 是可测函数，存在 $P \in \mathscr{P}[a, \infty)$ 使得 $\mathbb{E}_P[\boldsymbol{G}(X)]$ 位于$\mathbb{S}$的内部。

我们有以下对偶性结果：

定理 4.2。 给定常数值 $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \bar{\boldsymbol{\mu}}, \boldsymbol{\mu}, a$、函数 H 以及向量函数 $\boldsymbol{G_1}, \boldsymbol{G_2}$时，$\mathfrak{M}_a\big(H, (\boldsymbol{G_1}^\top, \boldsymbol{G_2}^\top)^\top, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r) \times \mathbb{S}_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})\big)$的对偶问题为

$$\min_{\substack{\kappa, \boldsymbol{\lambda_1} \geq 0, \boldsymbol{\lambda_2} \geq 0, \\ \|\boldsymbol{u}\|_2 \leq \lambda}} \quad \kappa + \lambda + r^{-1/2} \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\mu} + \boldsymbol{\lambda_1}^\top \bar{\boldsymbol{\mu}} - \boldsymbol{\lambda_2}^\top \boldsymbol{\mu} \tag{29a}$$

满足条件：$-H(x) + r^{-1/2} \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2} \boldsymbol{G_1}(x) + (\boldsymbol{\lambda_1} - \boldsymbol{\lambda_2})^\top \boldsymbol{G_2}(x) + \kappa \geq 0, \ \forall x \geq a.$ (29b) $\kappa, \boldsymbol{u}, \lambda, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}$ 为决策变量。*(29)* 的最优值至少与 $\mathfrak{M}_a\big(H, (\boldsymbol{G_1}^\top, \boldsymbol{G_2}^\top)^\top, S_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r) \times S_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})\big)$ 的最优值一样，并且在假设 *4.1* 下，二者取等。

定理 4.2 给出了当 S 为椭球与矩形的组合时，(28) 的对偶规划。假设 4.1 中确保强对偶性的斯莱特条件，可以通过逐个案例进行常规检验。此外，即使该条件不成立，定理 2.1 和 2.2 所提供的最终统计保证仍然有效，因为弱对偶性使我们能够得到一个更保守的界。定理 4.2 可由锥规划的对偶理论直接推出（例如，Shapiro（2001年））。

结合定理 4.1 和定理 4.2，我们可以将 (27) 转化为以下对偶规划，其中 (S $\mathscr{P}$) 为 (S_E, $\mathscr{P}^1_{a,\ \eta}$)、(S_R, $\mathscr{P}^1_{a,\ \eta}$)、(S_E, $\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu}$) 或 ($\mathrm{S}_R$, $\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu}$)。

推论 *4.1.* 给定满足定理 *4.1* 中假设的任意函数 $h, \boldsymbol{g}$，参数 $\eta, \nu, a, \boldsymbol{\mu}, \boldsymbol{\mu}, \bar{\boldsymbol{\mu}}, \boldsymbol{\Sigma} \succ 0$ 和 $\tilde{H}(x) = \int_a^x h(u)du$, $H(x) = \int_a^x \int_a^u h(v)dvdu$, $\tilde{G}_i(x) = \int_a^x g_i(u)du$, $G_i(x) = \int_a^x \int_a^u g_i(v)dvdu$，对于 $\boldsymbol{g}$ 的每个分量 i，我们有

1. 对于问题 $\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)\, \mathscr{P}^1_{a,\eta})$，转换后的矩问题的对偶问题为

$$\min_{\kappa, \|\boldsymbol{u}\|_2 \le \lambda} \kappa + \lambda + r^{-1/2}\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\mu} \text{ s.t. } -\eta\tilde{H}(x) + \eta\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-\frac{1}{2}}\tilde{\boldsymbol{G}}(x) + \kappa \ge 0,\ \forall x \ge a.$$

2. 对于问 题 $\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})\, \mathscr{P}^1_{a,\eta})$，转换后的矩问题的对偶问题为

$$\min_{\kappa, \boldsymbol{\lambda_1} \ge 0, \boldsymbol{\lambda_2} \ge 0} \kappa + \boldsymbol{\lambda_1}^\top \bar{\boldsymbol{\mu}} - \boldsymbol{\lambda_2}^\top \underline{\boldsymbol{\mu}},\ \text{s.t. } -\eta\tilde{H}(x) + \eta(\boldsymbol{\lambda_1} - \boldsymbol{\lambda_2})^\top \tilde{\boldsymbol{G}}(x) + \kappa \ge 0,\ \forall x \ge a.$$

3. 对于问题 $\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)\, \mathscr{P}^2_{a,\eta,\bar{\eta},\nu})$，转换后的矩问题的对偶问题为

$$\min_{\substack{\kappa, \|\boldsymbol{u}\|_2 \le \lambda \\ \delta_1 \ge 0, \delta_2 \ge 0}} \kappa + \lambda + r^{-1/2}\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} + \delta_1\bar{\eta} - \delta_2\underline{\eta},\ \text{s.t. } -\nu H(x) + \nu\boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{G}(x) + \nu(\delta_1 - \delta_2)(x-a) + \kappa \ge 0, x \ge a.$$

4. 对于问题 $\mathfrak{P}(h, \boldsymbol{g}, S_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})\, \mathscr{P}^2_{a,\eta,\bar{\eta},\nu})$，转换后的矩问题的对偶问题为 $\min_{\kappa, \delta_1 \ge 0, \delta_2 \ge 0}$，$\kappa + \boldsymbol{\lambda_1}^\top \bar{\boldsymbol{\mu}} - \boldsymbol{\lambda_2}^\top \boldsymbol{\mu} + \delta_1\bar{\eta} - \delta_2\eta$，满足条件：$-\nu H(x) + \nu(\boldsymbol{\lambda_1} - \boldsymbol{\lambda_2})^\top \boldsymbol{G}(x) + \nu(\delta_1 - \delta_2)(x-a) + \kappa \ge 0$，$\forall x \ge a \boldsymbol{\lambda_1} \ge 0, \boldsymbol{\lambda_2} \ge 0$在每种情况下，对偶问题的最优值至少与相应的$\mathfrak{P}(h, \boldsymbol{g}, S\ \mathscr{P})$的最优值一样大；并且在对相应矩问题应用假设 *4.1* 后，两者取等。

程序 (29) 及推论 4.1 中的特化版本包含无限多个约束条件，其中 (29b) 是对任意 $x \ge a$ 都成立的条件。在适当的假设下，我们可以将 (29) 转化为半正定程序。

定理 4.3。 如果约束 (29b) 可以写成一系列多项式不等式，即，

$$\sum_{i=0}^{k} y_{1i}x^i \ge 0,\ \forall x \ge a \quad \text{or} \quad \sum_{i=0}^{k} y_{2i}x^i \ge 0,\ \forall b \le x \le c, \tag{30}$$

那么，程序 (29) 等价于一个混合半正定与二阶锥规划

(半定规划-二阶锥规划)

$$
\begin{aligned}
\min_{\substack{\boldsymbol{u},\lambda,\kappa,\boldsymbol{\lambda_1}\geq 0,\boldsymbol{\lambda_2}\geq 0,\\ \boldsymbol{V}\succeq 0,\boldsymbol{W}\succeq 0,\|\boldsymbol{\Sigma}^{1/2}\boldsymbol{u}\|_2\leq\lambda}} \quad & \kappa+\lambda+\boldsymbol{u}^\top\boldsymbol{\mu}+\boldsymbol{\lambda_1}^\top\bar{\boldsymbol{\mu}}-\boldsymbol{\lambda_2}^\top\underline{\boldsymbol{\mu}} \\
\text{s.t.} \quad & \sum_{i,j:i+j=2l-1} v_{ij}=\sum_{i,j:i+j=2l-1} w_{ij}=0, \quad l=1,\cdots,k, \\
& \sum_{i,j:i+j=2l} v_{ij}=\sum_{r=l}^{k}\binom{r}{l}y_{1r}a^{r-l}, \quad l=0,\cdots,k, \\
& \sum_{i,j:i+j=2l} w_{ij}=\sum_{m=0}^{l}\sum_{r=m}^{k+m-l}\binom{r}{m}\binom{k-r}{l-m}y_{2r}b^{r-m}c^m, \quad l=0,\cdots,k
\end{aligned}
\tag{31}
$$

where $\boldsymbol{V}=[v_{ij}]_{i,j=0,\cdots,k}$ *and* $\boldsymbol{W}=[w_{ij}]_{i,j=0,\cdots,k}$.

定理 4.3 表明，当 $H(x)$ 和 $\boldsymbol{G}(x)$ 属于多项式函数类时，问题 (28) 可以表述为一个可处理的半定规划-二阶锥规划程序。如果 $h(x)$ 和 $\boldsymbol{g}(x)$ 是示性函数或分段多项式函数，则满足多项式函数类条件，从而保证问题具有可处理性。定理 4.3 可以利用矩的半定表示来证明（例如，Lasserre (2009)；Bertsimas and Popescu (2005)）。

在第 5 节中，我们利用推论 4.1 和定理 4.3 估计极值量并得出数值结果。在附录 B 中，我们还介绍了一些用于分布鲁棒优化问题的敏感性分析工具。

5 数值结果

我们展示了分布鲁棒优化框架的数值性能，并将其与传统的极值理论工具进行比较。在实验中，我们考虑样本量为 500 的合成数据。我们关注的量是尾部区间概率和分位数，稍后将对其进行具体说明。在实验中，我们从涵盖轻尾到重尾的“真实”分布中生成样本，包括：1）形状参数为 0.5、尺度参数为

1 的伽马分布；2）均值参数为 0、标准差参数为 1 的对数正态分布；3）形状参数为 1.5、尺度参数为 1 的帕累托分布。我们的目标是使用程序（6）获得单侧 95% 置信上界。为近似覆盖概率，我们将每项实验重复 200 次，并据此输出估计置信界的样本均值。参数校准通过自助法进行，重采样量为 500；对于密度及其导数，我们使用 R 软件包 ks 中的标准核估计量。

我们针对不同的 1）形状约束、2）矩约束、3）截断阈值 a 和 4）目标函数开展实验，并与超阈值法进行比较。这些实验旨在：1）通过展示我们的方法如何在广泛的设置下生成有效的置信界，验证我们的方法，这一点也得到定理 2.1 和 2.2 中统计保证的支持；2）为用户实施我们的方法提供指导；3）将我们的方法与超阈值法等现有方法进行比较。第 5.1–5.3 节将讨论与前两个目标相关的结果，而第 5.4 节将聚焦于上述第三个目标。

在下表中，我们用 (*D*， χ^2) 表示第 2.1 节所示的 D 阶单调性和椭球矩约束设置，其中 $D \in \{0, 1, 2\}$。类似地，(*D*，柯尔莫戈洛夫–斯米尔诺夫) 表示 D 阶单调性和矩形矩约束设置。

5.1 形状约束和矩约束的选择

在表 5.1 中，我们利用来自三个分布的合成数据集，估计尾部区间概率 P($q_{0.99} \leq X \leq q_{0.995}$)，其中 $q_{0.99}$ 和 $q_{0.995}$ 分别为理论上的 99^{th} 和 99.5^{th}百分位数。阈值 a 取为该合成数据集的 70^{th} 样本百分位数。在本表及后续各表中，“ 上界 ” 和 “ 覆盖概率”两列分别表示 200 次重复实验中的样本均值和置信上界的覆盖率比率，而“相对比率”列定义为置信上界的均值除以真实值。置信区间由 $q_{0.99} \leq X \leq q_{0.995}$ 给出。我们注意到，在某些情况下，平均覆盖概率为 1，即置信界在全部 200 次重复实验中都覆盖了真实值，此时样本标准差和置信区间宽度均为 0。

200 次重复实验中置信上界的覆盖率，“相对比率”列定义为置信上界的均值除以真实值。置信区间由 $sample\ mean \pm 1.96 * sample\ standard\ deviation/\sqrt{200}$ 给出。我们注意到，在某些情况下，平均覆盖概率为 1，即置信界在全部 200 次重复实验中都覆盖了真实值，此时样本标准差和置信区间宽度均为 0。

表 5.1：不同约束设置下的尾部概率估计。真实值为 0.005。

数据来源	约束设置	相对比率	上界	覆盖概率
伽马分布	$(0, \chi^2)$	14。46 ($\pm$0.57)	7.23×10^{-2} ($\pm 2.85 \times 10^{-3}$)	1。000 ($\pm$0.000)
	$(1, \chi^2)$	5.36 ($\pm$0.17)	2.68×10^{-2} ($\pm 8.26 \times 10^{-4}$)	1.000 ($\pm$0.000)
	$(2, \chi^2)$	3。05 ($\pm$0.08)	1.53×10^{-2} ($\pm 4.12 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(0, 柯尔莫戈洛夫–斯米尔诺夫)	14.17 ($\pm$0.12)	7.08×10^{-2} ($\pm 5.88 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(1, 柯尔莫戈洛夫–斯米尔诺夫)	6。81 ($\pm$0.07)	3.40×10^{-2} ($\pm 3.39 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(2, 柯尔莫戈洛夫–斯米尔诺夫)	4.12 ($\pm$0.04)	2.06×10^{-2} ($\pm 2.04 \times 10^{-4}$)	1。000 ($\pm$0.000)
对数正态	$(0, \chi^2)$	16。62 ($\pm$0.80)	8.31×10^{-2} ($\pm 3.99 \times 10^{-3}$)	1。000 ($\pm$0.000)
	$(1, \chi^2)$	6。52 ($\pm$0.26)	3.26×10^{-2} ($\pm 1.32 \times 10^{-3}$)	1.000 ($\pm$0.000)
	$(2, \chi^2)$	3.98 ($\pm$0.14)	1.99×10^{-2} ($\pm 7.11 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(0, 柯尔莫戈洛夫–斯米尔诺夫)	14.08 ($\pm$0.13)	7.04×10^{-2} ($\pm 6.49 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(1, 柯尔莫戈洛夫–斯米尔诺夫)	7。98 ($\pm$0.09)	3.99×10^{-2} ($\pm 4.33 \times 10^{-4}$)	1。000 ($\pm$0.000)
	(2, 柯尔莫戈洛夫–斯米尔诺夫)	5.09 ($\pm$0.05)	2.55×10^{-2} ($\pm 2.72 \times 10^{-4}$)	1.000 ($\pm$0.000)
帕累托	$(0, \chi^2)$	20.90 ($\pm$1.68)	1.05×10^{-1} ($\pm 8.39 \times 10^{-3}$)	1.000 ($\pm$0.000)
	$(1, \chi^2)$	9.85 ($\pm$0.76)	4.92×10^{-2} ($\pm 3.78 \times 10^{-3}$)	0.995 ($\pm$0.010)
	$(2, \chi^2)$	6.83 ($\pm$0.44)	3.41×10^{-2} ($\pm 2.21 \times 10^{-3}$)	0.995 ($\pm$0.010)
	(0, 柯尔莫戈洛夫–斯米尔诺夫)	14。13 ($\pm$0.13)	7.07×10^{-2} ($\pm 6.30 \times 10^{-4}$)	1.000 ($\pm$0.000)
	(1, 柯尔莫戈洛夫–斯米尔诺夫)	9.76 ($\pm$0.10)	4.88×10^{-2} ($\pm 5.15 \times 10^{-4}$)	1.000 ($\pm$0.000)
	(2, 柯尔莫戈洛夫–斯米尔诺夫)	6。81 ($\pm$0.07)	3.40×10^{-2} ($\pm 3.65 \times 10^{-4}$)	1。000 ($\pm$0.000)

在表 5.2 中，我们考虑使用来自三种分布的合成数据集估计分位数 $q_{0.99}$ 。阈值仍选为 70^{th} 样本百分位数。表中给出了每种分布的 99^{th}分位数的真实值。我们注意到，由于可能会在 D，柯尔莫戈洛夫–斯米尔诺夫处分配

概率质量，因此柯尔莫戈洛夫–斯米尔诺夫并非总能得到有效的分位数估计，所以我们没有将这些设置纳入此表。

表 5.2：不同约束设置下的分位数估计。

数据来源	约束设置	相对比率	上界	覆盖概率
伽马分布，真实分位点为 3.32。	$(0,\ \chi^2)$	2。13 ($\pm$0.04)	7.07×10^{0} ($\pm 1.34 \times 10^{-1}$)	1。000 ($\pm$0.000)
	$(1,\ \chi^2)$	1.50 ($\pm$0.03)	4.96×10^{0} ($\pm 9.66 \times 10^{-2}$)	1。000 ($\pm$0.000)
	$(2, \chi^2)$	1.39 ($\pm$0.03)	4.62×10^{0} ($\pm 9.20 \times 10^{-2}$)	0.995 ($\pm$0.010)
对数正态，真实分位点为 10.24。	$(0, \chi^2)$	2。54 ($\pm$0.10)	2.60×10^{1} ($\pm 1.05 \times 10^{0}$)	1。000 ($\pm$0.000)
	$(1, \chi^2)$	1.80 ($\pm$0.07)	1.84×10^{1} ($\pm 7.40 \times 10^{-1}$)	1.000 ($\pm$0.000)
	$(2,\ \chi^2)$	1。69 ($\pm$0.07)	1.73×10^{1} ($\pm 6.97 \times 10^{-1}$)	0。990 ($\pm$0.014)
帕累托，真实分位点为 21.54。	$(0,\ \chi^2)$	6.05 ($\pm$1.33)	1.30×10^{2} ($\pm 2.87 \times 10^{1}$)	1。000 ($\pm$0.000)
	$(1,\ \chi^2)$	3.92 ($\pm$0.71)	8.45×10^{1} ($\pm 1.54 \times 10^{1}$)	0.990 ($\pm$0.014)
	$(2, \chi^2)$	3。59 ($\pm$0.61)	7.73×10^{1} ($\pm 1.31 \times 10^{1}$)	0。985 ($\pm$0.017)

关于形状约束的选择，对于表 5.1 中的尾部概率估计问题，随着 D 增大，我们观察到上界和置信区间宽度均有所下降。例如，对于表 5.1 中的伽马数据，当约束设置从 $(0,\ \chi^2)$ 变为 $(2,\ \chi^2)$ 时，上界从 7.23×10^{-2} 降至 1.53×10^{-2}（真实值为 5.00×10^{-3}），置信区间宽度则从 2.85×10^{-3} 降至 2.04×10^{-4}，从而得到更紧的结果。类似地，对于表 5.2 中的分位数估计问题，形状约束越强，上界和置信区间宽度越小。同样，对于伽马数据，上界从 7.07 降至 4.62（真实值为 3.32），置信区间宽度则从 0.134 降至 0.092。

事实上，通过假设形状性质，与不采用形状假设的情况相比，我们将可行分布限制在一个更小的集合内（$D = 0$）。此外，凸性（$D = 2$）意味着单调性（$D = 1$），因此前者是后者的子集。我们的结果还表明，在更强的形状条件下校准参数所需的额外估计任务所带来的额外误差，似乎并未抵消

施加更强约束所带来的益处。在实践中，可以将阈值 a 附近的分布可视化，以评估形状假设的合理性。

现在我们比较用于尾部概率估计的矩约束。从表 5.1 可以看出，两种矩约束之间的差异并不像三种形状约束之间的差异那么显著。在没有任何形状约束的情况下（即 $D = 0$），对于所有三种分布，矩形约束的相对比率都小于椭球约束。在存在形状约束的情况下（即 $D = 1$，2），对于伽马分布和对数正态分布，椭球约束比矩形约束保守程度更低；对于帕累托分布，则仅略微更保守。事实上，对于帕累托数据和 $D = 1$，2，采用显著性水平为 0.05 的 Welch t 检验，我们无法拒绝 (D，χ^2) 和 (D，KS) 保守程度相同这一零假设。总的来说，如果在观察数据后决定选择 $D = 0$，那么矩形约束似乎是更好的矩约束选择。否则，椭球约束应当表现更好或与之相当。

5.2 阈值选择

本节比较不同的阈值选择。在表 5.3 中，我们考察约束设置 (2，χ^2) 下尾部区间概率 P($q_{0.99} \leq X \leq q_{0.995}$) 的估计。对于每种数据分布，我们测试四个单一截断阈值：60^{th}、70^{th}、80^{th}、90^{th} 样本百分位数。我们还测试将这四个阈值全部用作多个阈值的情况（有关多个阈值的使用详情，请参见补充材料 A）。在表 5.4 中，我们考察不同截断阈值选择下、同样在约束设置 (2，χ^2) 中的分位数 $q_{0.99}$ 估计。

我们观察到，随着截断阈值增大，结果对于

表 5.3：不同截断阈值下的尾部概率估计。真实值为 0.005。

数据来源	约束设置	相对比率	上界	覆盖概率
伽马分布	60^{th}	3.130 ($\pm$0.088)	1.57×10^{-2} ($\pm 4.39 \times 10^{-4}$)	1.000 ($\pm$0.000)
	70^{th}	3.053 ($\pm$0.082)	1.53×10^{-2} ($\pm 4.12 \times 10^{-4}$)	1.000 ($\pm$0.000)
	80^{th}	2.948 ($\pm$0.077)	1.47×10^{-2} ($\pm 3.85 \times 10^{-4}$)	0.995 ($\pm$0.010)
	90^{th}	2.823 ($\pm$0.077)	1.41×10^{-2} ($\pm 3.85 \times 10^{-4}$)	0。995 ($\pm$0.010)
	(60^{th}, 70^{th}, 80^{th}, 90^{th})	2.946 ($\pm$0.079)	1.47×10^{-2} ($\pm 3.93 \times 10^{-4}$)	0。995 ($\pm$0.010)
对数正态	60^{th}	4。113 ($\pm$0.147)	2.06×10^{-2} ($\pm 7.37 \times 10^{-4}$)	1。000 ($\pm$0.000)
	70^{th}	3.984 ($\pm$0.142)	1.99×10^{-2} ($\pm 7.11 \times 10^{-4}$)	1。000 ($\pm$0.000)
	80^{th}	3.831 ($\pm$0.140)	1.92×10^{-2} ($\pm 6.99 \times 10^{-4}$)	1.000 ($\pm$0.000)
	90^{th}	3。629 ($\pm$0.129)	1.81×10^{-2} ($\pm 6.45 \times 10^{-4}$)	0。995 ($\pm$0.010)
	(60^{th}, 70^{th}, 80^{th}, 90^{th})	3.792 ($\pm$0.132)	1.90×10^{-2} ($\pm 6.61 \times 10^{-4}$)	1.000 ($\pm$0.000)
帕累托	60^{th}	7.093 ($\pm$0.500)	3.55×10^{-2} ($\pm 2.50 \times 10^{-3}$)	0.995 ($\pm$0.010)
	70^{th}	6.825 ($\pm$0.443)	3.41×10^{-2} ($\pm 2.21 \times 10^{-3}$)	0.995 ($\pm$0.010)
	80^{th}	6.428 ($\pm$0.367)	3.21×10^{-2} ($\pm 1.84 \times 10^{-3}$)	0.995 ($\pm$0.010)
	90^{th}	5.559 ($\pm$0.233)	2.78×10^{-2} ($\pm 1.17 \times 10^{-3}$)	0.995 ($\pm$0.010)
	(60^{th}, 70^{th}, 80^{th}, 90^{th})	5.821 ($\pm$0.240)	2.91×10^{-2} ($\pm 1.20 \times 10^{-3}$)	0。995 ($\pm$0.010)

尾概率估计和分位数估计而言都趋于不那么保守。例如，对于伽马分布，当阈值从 60^{th} 样本百分位数增至 90^{th} 样本百分位数时，尾概率估计的相对比率从3.130降至2.823，分位数估计的相对比率则从1.426降至1.331。随着阈值增大，能够利用更多信息，因此这一现象是合理的。多个阈值的表现介于单个阈值的表现之间。理想情况下，只要参数能够得到良好校准，我们就应选择较大的阈值。然而，在实践中，通常很难判断哪个阈值满足这一条件。因此，使用多个阈值也是合理的选择，因为这种方法对阈值选择不那么敏感。

表 5.4：不同截断阈值下的分位数估计。

数据来源	约束设置	相对比率	上界	覆盖概率
真实分位点为 3.32 的伽马分布	60^{th}	1.426 ($\pm$0.027)	4.73×10^{0} ($\pm 8.95 \times 10^{-2}$)	1.000 ($\pm$0.000)
	70^{th}	1。392 ($\pm$0.028)	4.62×10^{0} ($\pm 9.20 \times 10^{-2}$)	0。995 ($\pm$0.010)
	80^{th}	1.359 ($\pm$0.029)	4.51×10^{0} ($\pm 9.47 \times 10^{-2}$)	0。990 ($\pm$0.014)
	90^{th}	1.331 ($\pm$0.029)	4.41×10^{0} ($\pm 9.76 \times 10^{-2}$)	0.980 ($\pm$0.019)
	$(60^{th}, 70^{th}, 80^{th}, 90^{th})$	1。354 ($\pm$0.031)	4.49×10^{0} ($\pm 1.02 \times 10^{-1}$)	0.980 ($\pm$0.019)
对数正态 w. 真实分位点 10.24	60^{th}	1。702 ($\pm$0.067)	1.74×10^{1} ($\pm 6.89 \times 10^{-1}$)	0。995 ($\pm$0.010)
	70^{th}	1.686 ($\pm$0.068)	1.73×10^{1} ($\pm 6.97 \times 10^{-1}$)	0.990 ($\pm$0.014)
	80^{th}	1.670 ($\pm$0.069)	1.71×10^{1} ($\pm 7.04 \times 10^{-1}$)	0.985 ($\pm$0.017)
	90^{th}	1。650 ($\pm$0.068)	1.69×10^{1} ($\pm 6.97 \times 10^{-1}$)	0.970 ($\pm$0.024)
	$(60^{th}, 70^{th}, 80^{th}, 90^{th})$	1.690 ($\pm$0.071)	1.73×10^{1} ($\pm 7.30 \times 10^{-1}$)	0。980 （$\pm$0.019）
帕累托与真实分位点 21.54	60^{th}	3。637 ($\pm$0.648)	7.83×10^{1} ($\pm 1.40 \times 10^{1}$)	0。990 ($\pm$0.014)
	70^{th}	3。587 （$\pm$0.607）	7.73×10^{1} ($\pm 1.31 \times 10^{1}$)	0.985 ($\pm$0.017)
	80^{th}	3。542 ($\pm$0.588)	7.63×10^{1} ($\pm 1.27 \times 10^{1}$)	0.985 ($\pm$0.017)
	90^{th}	3。347 ($\pm$0.450)	7.21×10^{1} ($\pm 9.69 \times 10^{0}$)	0.985 ($\pm$0.017)
	$(60^{th}, 70^{th}, 80^{th}, 90^{th})$	3。584 ($\pm$0.552)	7.72×10^{1} ($\pm 1.19 \times 10^{1}$)	0。985 ($\pm$0.017)

5.3 不同目标函数下的性能

本节旨在了解我们的方法在不同目标函数下的性能变化。在表5.5中，我们展示了约束设置 (2, χ^2) 和 (2,柯尔莫戈洛夫–斯米尔诺夫) 下不同区间的尾概率估计结果。具体而言，目标概率选为 P($q_{\text{LHS}} \leq X \leq q_{\text{LHS}+0.005}$)，其中拉丁超立方抽样取值各不相同，范围为 0.90 至 0.99。也就是说，我们将真实概率值保持为 0.005，而随着拉丁超立方抽样值增大，区间 $[q_{\text{LHS}}, q_{\text{LHS}+0.005}]$ 向尾部更远处移动。截断阈值选为 70^{th} 样本百分位数。

从表中可以看出，区间越远离阈值、越接近更远尾部，相对比率往往越大，也就是说，拉丁超立方抽样越大，相对比率也越大。例如，对于伽马数据，在 (2， χ^2) 约束条件下，随着左端点的拉丁超立方抽样从 0.90 增至 0.99，相对比率从 1.714 增至 3.053。所有数据分布和

表5.5：不同目标函数下的尾部概率估计。目标概率为 P($q_{\text{LHS}} \leq X \leq q_{\text{LHS}+0.005}$)。

		(2, χ^2)			(2.柯尔莫戈洛夫–斯米尔诺夫)		
数据来源	LHS 分位数	相对比率	上界	覆盖概率	相对比率	上界	覆盖概率
伽马分布	0.900	1。714 (±0.010)	8.57×10^{-3} ($\pm 4.94 \times 10^{-5}$)	1.000 (±0.000)	1。766 (±0.014)	8.83×10^{-3} ($\pm 7.14 \times 10^{-5}$)	1。000 (±0.000)
	0.910	1.782 (±0.010)	8.91×10^{-3} ($\pm 4.87 \times 10^{-5}$)	1.000 (±0.000)	1.780 (±0.014)	8.90×10^{-3} ($\pm 7.12 \times 10^{-5}$)	1。000 (±0.000)
	0.920	1。860 (±0.010)	9.30×10^{-3} ($\pm 4.92 \times 10^{-5}$)	1。000 (±0.000)	1。806 (±0.015)	9.03×10^{-3} ($\pm 7.40 \times 10^{-5}$)	1。000 (±0.000)
	0.930	1.950 (±0.010)	9.75×10^{-3} ($\pm 5.24 \times 10^{-5}$)	1.000 (±0.000)	1.844 (±0.016)	9.22×10^{-3} ($\pm 7.88 \times 10^{-5}$)	1。000 (±0.000)
	0.940	2.051 (±0.012)	1.03×10^{-2} ($\pm 6.14 \times 10^{-5}$)	1。000 (±0.000)	1.899 (±0.017)	9.50×10^{-3} ($\pm 8.55 \times 10^{-5}$)	1。000 (±0.000)
	0.950	2.163 (±0.016)	1.08×10^{-2} ($\pm 8.09 \times 10^{-5}$)	1。000 (±0.000)	1.982 (±0.019)	9.91×10^{-3} ($\pm 9.35 \times 10^{-5}$)	1。000 (±0.000)
	0.960	2.278 (±0.023)	1.14×10^{-2} ($\pm 1.16 \times 10^{-4}$)	1。000 (±0.000)	2.111 (±0.021)	1.06×10^{-2} ($\pm 1.03 \times 10^{-4}$)	1。000 (±0.000)
	0.970	2.390 (±0.033)	1.20×10^{-2} ($\pm 1.63 \times 10^{-4}$)	1。000 (±0.000)	2.326 (±0.023)	1.16×10^{-2} ($\pm 1.15 \times 10^{-4}$)	1。000 (±0.000)
	0.980	2.565 (±0.041)	1.28×10^{-2} ($\pm 2.04 \times 10^{-4}$)	1.000 (±0.000)	2.755 (±0.028)	1.38×10^{-2} ($\pm 1.39 \times 10^{-4}$)	1.000 (±0.000)
	0.990	3.053 (±0.082)	1.53×10^{-2} ($\pm 4.12 \times 10^{-4}$)	1。000 (±0.000)	4。115 (±0.041)	2.06×10^{-2} ($\pm 2.04 \times 10^{-4}$)	1。000 (±0.000)
对数正态	0.900	1。860 (±0.011)	9.30×10^{-3} ($\pm 5.52 \times 10^{-5}$)	1.000 (±0.000)	1.951 (±0.016)	9.75×10^{-3} ($\pm 7.75 \times 10^{-5}$)	1。000 (±0.000)
	0.910	1。959 (±0.011)	9.80×10^{-3} ($\pm 5.45 \times 10^{-5}$)	1.000 (±0.000)	1。987 (±0.016)	9.93×10^{-3} ($\pm 7.91 \times 10^{-5}$)	1.000 (±0.000)
	0.920	2。080 (±0.011)	1.04×10^{-2} ($\pm 5.49 \times 10^{-5}$)	1.000 (±0.000)	2.038 (±0.017)	1.02×10^{-2} ($\pm 8.39 \times 10^{-5}$)	1.000 (±0.000)
	0.930	2.225 (±0.011)	1.11×10^{-2} ($\pm 5.71 \times 10^{-5}$)	1。000 (±0.000)	2.102 (±0.018)	1.05×10^{-2} ($\pm 9.04 \times 10^{-5}$)	1。000 (±0.000)
	0.940	2.402 (±0.013)	1.20×10^{-2} ($\pm 6.50 \times 10^{-5}$)	1。000 (±0.000)	2。185 (±0.020)	1.09×10^{-2} ($\pm 9.93 \times 10^{-5}$)	1。000 (±0.000)
	0.950	2.613 (±0.018)	1.31×10^{-2} ($\pm 8.86 \times 10^{-5}$)	1。000 (±0.000)	2.298 (±0.022)	1.15×10^{-2} ($\pm 1.10 \times 10^{-4}$)	1。000 (±0.000)
	0.960	2.854 (±0.030)	1.43×10^{-2} ($\pm 1.48 \times 10^{-4}$)	1。000 (±0.000)	2。471 (±0.025)	1.24×10^{-2} ($\pm 1.23 \times 10^{-4}$)	1.000 (±0.000)
	0.970	3.109 (±0.051)	1.55×10^{-2} ($\pm 2.57 \times 10^{-4}$)	1.000 (±0.000)	2。748 (±0.029)	1.37×10^{-2} ($\pm 1.43 \times 10^{-4}$)	1.000 (±0.000)
	0.980	3.399 (±0.078)	1.70×10^{-2} ($\pm 3.89 \times 10^{-4}$)	1.000 (±0.000)	3.305 (±0.036)	1.65×10^{-2} ($\pm 1.80 \times 10^{-4}$)	1。000 (±0.000)
	0.990	3.984 (±0.142)	1.99×10^{-2} ($\pm 7.11 \times 10^{-4}$)	1。000 (±0.000)	5。091 (±0.054)	2.55×10^{-2} ($\pm 2.72 \times 10^{-4}$)	1.000 (±0.000)
帕累托	0.900	2.204 (±0.014)	1.10×10^{-2} ($\pm 6.80 \times 10^{-5}$)	1.000 (±0.000)	2.252 (±0.018)	1.13×10^{-2} ($\pm 9.23 \times 10^{-5}$)	1。000 (±0.000)
	0.910	2.338 (±0.013)	1.17×10^{-2} ($\pm 6.42 \times 10^{-5}$)	1.000 (±0.000)	2。306 (±0.019)	1.15×10^{-2} ($\pm 9.70 \times 10^{-5}$)	1.000 (±0.000)
	0.920	2.511 (±0.013)	1.26×10^{-2} ($\pm 6.28 \times 10^{-5}$)	1.000 (±0.000)	2.370 (±0.020)	1.19×10^{-2} ($\pm 1.02 \times 10^{-4}$)	1.000 (±0.000)
	0.930	2.739 (±0.013)	1.37×10^{-2} ($\pm 6.39 \times 10^{-5}$)	1.000 (±0.000)	2.456 (±0.022)	1.23×10^{-2} ($\pm 1.10 \times 10^{-4}$)	1.000 (±0.000)
	0.940	3.042 (±0.014)	1.52×10^{-2} ($\pm 6.83 \times 10^{-5}$)	1.000 (±0.000)	2.578 (±0.024)	1.29×10^{-2} ($\pm 1.19 \times 10^{-4}$)	1.000 (±0.000)
	0.950	3.453 (±0.017)	1.73×10^{-2} ($\pm 8.53 \times 10^{-5}$)	1.000 (±0.000)	2.739 (±0.026)	1.37×10^{-2} ($\pm 1.32 \times 10^{-4}$)	1.000 (±0.000)
	0.960	4。013 (±0.032)	2.01×10^{-2} ($\pm 1.58 \times 10^{-4}$)	1.000 (±0.000)	2.975 (±0.030)	1.49×10^{-2} ($\pm 1.50 \times 10^{-4}$)	1.000 (±0.000)
	0.970	4.759 (±0.077)	2.38×10^{-2} ($\pm 3.85 \times 10^{-4}$)	1.000 (±0.000)	3。375 (±0.034)	1.69×10^{-2} ($\pm 1.71 \times 10^{-4}$)	1.000 (±0.000)
	0.980	5.691 (±0.186)	2.85×10^{-2} ($\pm 9.28 \times 10^{-4}$)	1.000 (±0.000)	4.178 (±0.045)	2.09×10^{-2} ($\pm 2.23 \times 10^{-4}$)	1.000 (±0.000)
	0.990	6.825 (±0.443)	3.41×10^{-2} ($\pm 2.21 \times 10^{-3}$)	0.995 (±0.010)	6。808 (±0.073)	3.40×10^{-2} ($\pm 3.65 \times 10^{-4}$)	1.000 (±0.000)

约束设置均呈现相同趋势。因此，当我们根据非尾部数据推断尾部区域时，如果目标量与更远尾部相关，估计会更加保守。在这种情况下，如第5.2节所述，如果条件允许，可以尝试提高阈值，从而在一定程度上降低保守性。否则，至少可以通过这种方法得到保守且安全的估计。

5.4 与超阈值法的比较

最后，我们将本文的方法与传统的超阈值法进行比较。传统方法通过对超额损失数据进行最大似然拟合得到广义帕累托分布（例如 Smith（1987年）），然后利用德尔塔法获得尾区间概率的 95% 置信上界。表5.6展示了超阈值法的结果；我们根据均值超额函数的线性特征，采用图形法选择拟合广义帕累托分布的阈值（参见 Embrechts 等人（1997年））。

比较表5.5和表5.6可知，虽然我们的分布鲁棒优化方法得到的界比超阈值法宽松，但其覆盖率是正确的。相比之下，超阈值法在许多情况下覆盖不足（表5.6中以粗体标示）。例如，对于目标函数 P($q_{0.90} \leq X \leq q_{0.905}$) 和 Gamma数据集，超阈值法给出的置信上界 4.64×10^{-3} 甚至小于真实值，而分布鲁棒优化给出的结果为 8.57×10^{-2}，其配置为 (2，χ^2)。另一方面，超阈值法的覆盖率仅为 65%，而分布鲁棒优化的覆盖率为 100%。超阈值法覆盖率欠佳，表明数据量太小，无法进行恰当的估计。

总体而言，超阈值法给出的估计更接近真实目标量，但其置信界可能达不到规定的覆盖率。我们的建议是：如果建模者优先关注数量级，那么选择广义帕累托分布会更合适；而如果风险厌恶的建模者希望获得具有正确置信保证的界，那么选择我们的分布鲁棒优化方法会更合适。

参考文献

Atar, R., Chowdhary, K., andDupuis, P.(2015年). 基于 Rényi 散度的风险敏感函数的鲁棒界。*SIAM/ASA不确定性量化期刊*, 3(1):18–33.

表5.6：使用 POT 方法并采用不同目标函数进行尾部概率估计。目标概率为 P($q_{\text{LHS}} \leq X \leq q_{\text{LHS+0.005}}$)。表中以粗体标示名义置信水平下的覆盖概率。

数据来源	LHS 分位数	相对比率	上界	覆盖概率
伽马分布	0.900	0.928 (±0.060)	4.64×10^{-3} ($\pm 3.01 \times 10^{-4}$)	**0.650 (±0.066)**
	0.910	1.096 (±0.046)	5.48×10^{-3} ($\pm 2.32 \times 10^{-4}$)	**0.780 (±0.057)**
	0.920	1.159 (±0.040)	5.79×10^{-3} ($\pm 2.01 \times 10^{-4}$)	**0.850 (±0.049)**
	0.930	1.177 (±0.035)	5.88×10^{-3} ($\pm 1.75 \times 10^{-4}$)	**0.880 (±0.045)**
	0.940	1.170 (±0.032)	5.85×10^{-3} ($\pm 1.59 \times 10^{-4}$)	**0.905 (±0.041)**
	0.950	1。164 (±0.030)	5.82×10^{-3} ($\pm 1.49 \times 10^{-4}$)	**0.895 (±0.042)**
	0.960	1。169 (±0.030)	5.85×10^{-3} ($\pm 1.50 \times 10^{-4}$)	**0.865 (±0.047)**
	0.970	1.210 (±0.032)	6.05×10^{-3} ($\pm 1.58 \times 10^{-4}$)	**0.900 (±0.042)**
	0.980	1.346 (±0.037)	6.73×10^{-3} ($\pm 1.86 \times 10^{-4}$)	0。950 (±0.030)
	0.990	1.705 (±0.056)	8.52×10^{-3} ($\pm 2.81 \times 10^{-4}$)	0.970 (±0.024)
对数正态	0.900	1.002 (±0.051)	5.01×10^{-3} ($\pm 2.55 \times 10^{-4}$)	**0.730 (±0.062)**
	0.910	1。110 (±0.037)	5.55×10^{-3} ($\pm 1.87 \times 10^{-4}$)	**0.830 (±0.052)**
	0.920	1。185 (±0.023)	5.93×10^{-3} ($\pm 1.16 \times 10^{-4}$)	0.910 (±0.040)
	0.930	1.200 (±0.020)	6.00×10^{-3} ($\pm 1.01 \times 10^{-4}$)	0.935 (±0.034)
	0.940	1。219 (±0.019)	6.09×10^{-3} ($\pm 9.45 \times 10^{-5}$)	0.965 (±0.025)
	0.950	1.241 (±0.020)	6.20×10^{-3} ($\pm 9.86 \times 10^{-5}$)	0.955 (±0.029)
	0.960	1.272 (±0.022)	6.36×10^{-3} ($\pm 1.10 \times 10^{-4}$)	0。950 (±0.030)
	0.970	1。331 (±0.024)	6.65×10^{-3} ($\pm 1.19 \times 10^{-4}$)	0.975 (±0.022)
	0.980	1.461 (±0.029)	7.31×10^{-3} ($\pm 1.43 \times 10^{-4}$)	0.990 (±0.014)
	0.990	1.731 (±0.045)	8.65×10^{-3} ($\pm 2.23 \times 10^{-4}$)	0.985 (±0.017)
帕累托	0.900	1。061 (±0.029)	5.30×10^{-3} ($\pm 1.44 \times 10^{-4}$)	**0.830 (±0.052)**
	0.910	1.100 (±0.020)	5.50×10^{-3} ($\pm 1.00 \times 10^{-4}$)	**0。875 (±0.046)**
	0.920	1.131 (±0.015)	5.66×10^{-3} ($\pm 7.45 \times 10^{-5}$)	0.910 (±0.040)
	0.930	1.156 (±0.015)	5.78×10^{-3} ($\pm 7.37 \times 10^{-5}$)	0.940 (±0.033)
	0.940	1。183 (±0.016)	5.92×10^{-3} ($\pm 7.81 \times 10^{-5}$)	0。950 (±0.030)
	0.950	1。216 (±0.017)	6.08×10^{-3} ($\pm 8.45 \times 10^{-5}$)	0.980 (±0.019)
	0.960	1.260 (±0.018)	6.30×10^{-3} ($\pm 9.19 \times 10^{-5}$)	0.985 (±0.017)
	0.970	1.323 (±0.021)	6.62×10^{-3} ($\pm 1.04 \times 10^{-4}$)	0.990 (±0.014)
	0.980	1.428 (±0.026)	7.14×10^{-3} ($\pm 1.30 \times 10^{-4}$)	0.990 (±0.014)
	0.990	1.627 (±0.046)	8.13×10^{-3} ($\pm 2.28 \times 10^{-4}$)	0.980 (±0.019)

Balkema, A. A. 和 de Haan, L.（1974）。高龄时的剩余寿命。《概率年刊》，2(5):792–804。

Bayraksan, G. 和 Love, D. K.（2015）。使用 phi-散度的数据驱动随机规划。载于《运筹学革命》，第 1–19 页。INFORMS。

Beirlant, J. 和 Teugels, J.L.（1992）。非寿险大额索赔建模。《保险数学与经济学》，11(1):17–29。

Ben-Tal, A.、El Ghaoui, L. 和 Nemirovski, A.（2009）。鲁棒优化。普林斯顿应用数学丛书。普林斯顿大学出版社。

Ben-Tal, A. 和 Nemirovski, A.（1998）。鲁棒凸优化。《运筹学数学》，23(4):769–805。

Bertsimas, D.、Gupta, V. 和 Kallus, N.（2018）。鲁棒样本均值近似。《数学规划》，171(1):217–282。

Bertsimas, D. 和 Popescu, I.（2005）。概率论中的最优不等式：一种凸优化方法。《*SIAM*优化杂志》，15(3):780–804。

Bertsimas, D. 和 Sim, M.（2004）。鲁棒性的代价。《运筹学》，52(1):35– 53。

Birghila, C.、Aigner, M. 和 Engelke, S.（2021年）。基于 Wasserstein距离和 f-散度的分布鲁棒尾部界。*arXiv*预印本 *arXiv:2106.06266*。

Bladt, M.、Albrecher, H. 和 Beirlant, J.（2020）。极值中的阈值选择与修剪。《极值》，23(4):629–665。

Blanchet, J., He, F. 和 Murthy, K.（2020）。关于分布鲁棒极值分析。极值，第 1–31 页。

Blanchet, J. 和 Kang, Y.（2021）。基于 Wasserstein距离的样本外推断。《style id='1'>《运筹学》，69(3):985–1013。

Blanchet, J., Kang, Y., and Murthy, K. (2019年). 稳健Wasserstein剖面推断及其在机器学习中的应用。《应用概率学杂志》, 56(3):830–857.

Blanchet, J., Murthy, K., and Nguyen, V. A. (2021年). Wasserstein分布鲁棒估计量的统计分析。载于《运筹学教程：新兴优化方法与建模技术及其应用》，第227–254页。INFORMS.

Blanchet, J., Murthy, K., and Si, N. (2022). Wasserstein分布鲁棒估计中的置信区域。《*Biometrika*》, 109(2):295–315.

Chen, R. and Paschalidis, I. C. (2018). 一种基于分布鲁棒优化的回归模型稳健学习方法。《机器学习研究杂志》, 19(13):1–48.

Chen, X., He, S., Jiang, B., Ryan, C. T., and Zhang, T. (2021年). 带有非凸形状约束的离散矩问题。《运筹学》, 69(1):279–296.

Davis, R. 和 Resnick, S.（1984年）。受极值理论启发的尾部估计。《统计年刊》，12(4):1467–1487。

Davison, A. C. 和 Smith, R. L.（1990年）。高阈值以上超阈值的模型。《皇家统计学会杂志：*B*辑（方法论）》，52(3):393–425。

Delage, E. 和 Ye, Y.（2010年）。具有矩不确定性的分布鲁棒优化及其在数据驱动问题中的应用。《运筹学》，58(3):595–612。

Dey, S. 和 Juneja, S.（2010年）。将厚尾约束纳入金融模型的熵方法。可在 *SSRN 1647048* 获取。

Dhara, A.、Das, B. 和 Natarajan, K.（2021年）。具有单变量和双变量边缘分布的最坏情形预期尾部损失。*INFORMS*计算期刊，33(1)：370–389。

Dharmadhikari, S. 和 Joag-Dev, K.（1988）。单峰性、凸性及其应用。《概率与数理统计》。爱思唯尔科学。

Dik, J. 和 de Gunst, M.（1985）。正态分布下的一般二次型的分布。《荷兰统计学》，39(1)：14–26。

Doan, X. V.、Li, X. 和 Natarajan, K.（2015）。利用重叠边缘分布，在投资组合优化中实现对依赖关系的稳健性。《运筹学》，63(6)：1468–1488。

Duchi, J. C.、Glynn, P. W. 和 Namkoong, H.（2021）。稳健优化的统计学：一种广义经验似然方法。《运筹学数学》，46(3)：946–969。

Embrechts, P.、Kl¨uppelberg, C. 和 Mikosch, T.（1997）。极端事件建模：保险与金融。数学应用丛书第33卷。施普林格，柏林、海德堡，第1版。

Embrechts, P. 和 Puccetti, G.（2006年a）。相依风险函数的界。《金融与随机过程》，10(3)：341–352。

Embrechts, P. 和 Puccetti, G.（2006年b）。多变量风险函数的界。《多元分析杂志》，97(2)：526–547。

Engelke, S. 和 Ivanovs, J.（2017）。多变量极值中的鲁棒界。《应用概率年刊》，27(6):3706–3734。

Esfahani, P. M. 和 Kuhn, D.（2018）。使用 Wasserstein度量的数据驱动分布鲁棒优化：性能保证与可处理的重构。《数学规划》，171(1):115–166。

Fisher, R. A. 和 Tippett, L. H. (1928). 样本中最大或最小成员的频率分布的极限形式。《剑桥哲学学会数学学报》，24(2):180–190。

Gao, R. 和 Kleywegt, A. (2022). 基于Wasserstein距离的分布鲁棒随机优化。《运筹学数学》。

Ghaoui, L. E.、Oks, M. 和 Oustry, F. (2003). 最坏情况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。

Ghosh, S. 和 Lam, H. (2019年). 随机模拟中的稳健分析：计算与性能保证。《运筹学》，67(1):232–249。

Glasserman, P.、Kang, W. 和 Shahabuddin, P. (2007). 多因子投资组合信用风险中的大偏差。《数理金融》，17(3):345–379。

Glasserman, P., Kang, W. 和 Shahabuddin, P.（2008）。多因子投资组合信用风险的快速仿真。《运筹学》，56(5):1200–1217。

Glasserman, P. 和 Xu, X.（2014）。稳健风险度量与模型风险。《定量金融》，14(1):29–58。

Gnedenko, B.（1943）。关于随机序列最大项的极限分布。《数学年刊》，44(3):423。

Goberna, M.、G´omez, S.、Guerra, F. 和 Todorov, M.（2007）。线性

半无限规划中的敏感性分析：扰动目标函数和右端项系数。《欧洲运筹学杂志》，181(3)：1069 – 1085。

Goberna, M. A. 和 L´opez, M.A.（2000）。线性半无限优化。《实践中的数学方法》。John Wiley。

Goberna, M. A.、L´opez, M. A. 和 Pastor, J.（1981）。半无限规划中的 Farkas-Minkowski 系统。《应用数学与优化》，7(1)：295–308。

Goh, J. 和 Sim, M.（2010）。分布鲁棒优化及其可处理的近似。《运筹学》，58(4-part-1)：902–917。

Gotoh, J.、Kim, M. J. 和 Lim, A. E. B.（2018）。鲁棒经验优化与均值–方差优化几乎相同。《运筹学快报》，46(4)：448 – 452。

Gumbel, E. J.（1958）。《极值统计》。哥伦比亚大学出版社，纽约，NY。

Gupta, V.（2019年）。用于分布鲁棒优化的近最优贝叶斯歧义集。《管理科学》，65(9):4242–4260。

Hansen, L. P. 和 Sargent, T. J.（2008）。《稳健性》。普林斯顿大学出版社。

Heidelberger, P.（1995）。排队与可靠性模型中稀有事件的快速仿真。《*ACM*建模与计算机仿真汇刊》（*TOMACS*），5(1):43–85。

Hill, B. M.（1975）。关于分布尾部推断的一种简单通用方法。《统计年刊》，3(5):1163–1174。

Hosking, J. R.、Wallis, J. R. 和 Hosking, J. R.（1987）。广义帕累托分布的参数和分位数估计。《技术计量学》，29(3)：339–349。

Hu, Z. 和 Hong, L. J.（2013）。受 Kullback-Leibler散度约束的分布鲁棒优化。载于 *Optimization Online*。

Iyengar, G. N.（2005）。鲁棒动态规划。《运筹学数学》，30(2):257–280。

Jiang, R. 和 Guan, Y.（2016）。数据驱动的机会约束随机规划。数学规划，158(1):291–327。

Jiang, R. 和 Guan, Y.（2018）。具有分布模糊性的风险规避型两阶段随机规划。《运筹学》，66(5):1390–1405。

Jonasson, J. K. 和 Rootz´en, H.（2014）。自然驾驶研究中近碰撞事件的内部验证：连续多变量方法。事故分析与预防，62:102–109。

Kuhn, D., Esfahani, P. M., Nguyen, V. A. 和 Shafieezadeh-Abadeh, S.（2019）。Wasserstein 分布鲁棒优化：机器学习中的理论与应用。载于《分析时代的运筹学与管理科学》，第 130–166 页。INFORMS。

Lam, H.（2016年）。随机系统的鲁棒敏感性分析。《运筹学数学》，41(4)：1248–1275。

Lam, H.（2018）。输入过程对序列依赖性的敏感性：一种鲁棒方法。《管理科学》，64(3)：1311–1327。

Lam, H.（2019年）。通过基于经验散度的分布鲁棒优化恢复最佳统计保证。《运筹学》，67(4):1090–1105。

Lam, H. 和 Mottet, C.（2017）。无需参数模型的尾部分析：最坏情形视角。《运筹学》，65(6):1696–1711。

Lam, H. 和 Zhou, E.（2017）。用于量化样本均值近似中不确定性的经验似然方法。《运筹学快报》，45(4):301 – 307。

Lasserre, J. B.（2009）。矩、正多项式及其应用，第 1 卷。世界科学出版社。

Li, B.、Jiang, R. 和 Mathieu, J. L.（2019）。具有矩和单峰性信息的模糊风险约束。《数学规划》，173(1):151–192。

Lim, A. E.、Shanthikumar, J. G. 和 Shen, Z. M.（2006）。模型不确定性、鲁棒优化与学习。载于《创新决策制定的模型、方法与应用》，第 66–94 页。INFORMS。

McNeil, A. J.、Frey, R. 和 Embrechts, P.（2015）。《定量风险管理：概念、技术与工具》。普林斯顿金融丛书。普林斯顿大学出版社，修订版。

Nicola, V. F., Nakavama, M. K., Heidelbereer, P., 和Goyal, A.（1993年）。一般故障和修复过程的高可靠性系统快速仿真。《*IEEE*计算机汇刊》，42(12):1440–1452。

Noether, G. E.（1963年）。离散情形下柯尔莫哥洛夫统计量的注记。《计量学》，7(1):115–116。

Pestana, D. D. 和 Mendonca, S.（2001）。高阶单调函数与概率论。收录于《广义凸性与广义单调性》，第317–331页。施普林格。

Petersen, I., James, M. 和 Dupuis, P.（2000）。具有相对熵约束的随机不确定系统的极小极大最优控制。*IEEE* 自动控制汇刊，45(3):398–412。

Pickands III, J.（1975）。使用极值次序统计量进行统计推断。《统计年刊》，3(1):119–131。

Popescu, I.（2005）。一种针对凸分布类最优矩界的半定规划方法。《运筹学数学》，30(3):632–657。

Puccetti, G. 和R¨uschendorf, L.（2013）。相依风险之和的精确界。《应用概率学杂志》，50(1):42–53。

Rahimian, H. 和 Mehrotra, S.（2019）。分布鲁棒优化：综述。*arXiv* 预印本 *arXiv:1908.05659*。

Resnick, S. I.（1987年）。《极值、正则变差与点过程》。施普林格运筹学与金融工程丛书。施普林格，纽约，纽约州，第1版。

Rockafellar, R. T.（1970）。凸分析。普林斯顿大学出版社。

Scarf, H. E.（1957年）。《库存问题的极小极大解》。兰德公司，圣莫尼卡。

Shafieezadeh-Abadeh, S., Kuhn, D., and Esfahani, P. M.（2019年）。基于质量传输的正则化。《机器学习研究杂志》，20(103):1–68。

Shapiro, A.（2001年）。关于锥线性问题对偶理论。载于半无限规划，第135–165页。施普林格。

Smith, R. L.（1984年）。样本极值的阈值方法。载于统计极值与应用，第621–638页。施普林格。

Smith, R. L. (1985年). 非正则情形中的极大似然估计。《*Biometrika*》, 72(1):67.

Smith, R. L. (1987年). 概率分布尾部的估计。《统计年刊》, 15(3):1174–1207.

Van Parys, B. P., Goulart, P. J., and Kuhn, D. (2016年). 通过半定规划得到的广义高斯不等式。《数学规划》, 156(1-2):271–302.

Van Parys, B. P. G., Goulart, P. J., and Morari, M. (2019年). 结构化分布的分布鲁棒期望不等式。《数学规划》, 173(1):251– 280.

Wang, B. andWang, R. (2011年). 单调边缘密度下的完全可混合性与凸最小化问题。《多元分析杂志》, 102(10):1344–1360.

Wiesemann, W., Kuhn, D., and Sim, M. (2014). 分布鲁棒凸优化。《运筹学》, 62(6):1358–1376.

Winkler, G. (1988). 矩集合的极点。《运筹学数学》, 13(4).

Xie, W. (2019年). 分布鲁棒两阶段随机规划的可处理重构，使用 ∞–wasserstein距离。*arXiv*预印本 *arXiv: 1908.08454*。

Xu, H. 和 Mannor, S.（2012年）。分布鲁棒马尔可夫决策过程。《运筹学》数学, 37(2):288–300。

Yuen, R.、Stoev, S. 和 Cooley, D.（2020）。极值风险价值的分布鲁棒推断。保险：数学与经济学，92:70–89。

Zhao, D., Lam, H., Peng, H., Bao, S., LeBlanc, D. J., Nobukawa, K., andPan, C.S.（2016年）。基于重要性抽样技术加速评估自动驾驶车辆在换道场景中的安全性。*IEEE Transactions on Intelligent Transportation Sys-tems*，18(3):595–607。

补充材料

补充材料 A 校准程序。

补充材料B 敏感性分析。

补充材料C 结果证明。

A 补充材料：校准程序

式（4）和（5）的构造受到以下两个弱收敛结果的启发。

1. χ^2分布。给定来自具有正定方差-协方差矩阵 $\mathbf{\Sigma}$ 的 d 维随机向量 n 个样本点 $\mathbf{X}$，根据弱收敛结果（例如 Dik 和 de Gunst（1985）中的推论 2.1），我们有

$$n(\mathbb{E}[\boldsymbol{X}]-\hat{\boldsymbol{\mu}})^{\top}\hat{\mathbf{\Sigma}}_n^{-1}(\mathbb{E}[\boldsymbol{X}]-\hat{\boldsymbol{\mu}})\Rightarrow\chi_d^2 \tag{A.1}$$

其中，$\hat{\boldsymbol{\mu}}$ 和 $\hat{\mathbf{\Sigma}}_n$ 分别为样本均值和样本协方差矩阵，χ_d^2 为具有 d 个自由度的卡方分布。

2. 柯尔莫哥洛夫分布。给定随机变量 X 的 n 个独立同分布实现的有序点 $\{x_1,\cdots,x_n\}$，其连续累积分布函数为 F，对应的经验分布为 F_n，柯尔莫哥洛夫-斯米尔诺夫统计量 $D_n\big(:=\sup_x|F_n(x)-F(x)|\big)$ 收敛于柯尔莫哥洛夫分布，例如 Noether（1963年），即

$$\sqrt{n}\max_{i=1,\cdots,n}\Big(\max\big(\frac{i-n}{n}+\mathbb{E}[\mathbb{I}(X\geq x_i)],\frac{n+1-i}{n}-\mathbb{E}[\mathbb{I}(X\geq x_i)]\big)\Big)\Rightarrow\sup_{t\in[0,1]}|B(t)-tB(1)| \tag{A.2}$$

其中 $B(t)$ 是标准维纳过程。

我们讨论基于经验观测的校准方法，以实现定理 2.1 和 2.2 中的统计保证结果。为了构造以高概率满足真实分布 P_{true} 覆盖率的可行区域，可以从统计学角度校准参数 $\mathbf{\Gamma}$，η，ν。总体而言，可以基于上述弱收敛结果构造 S。η 和 ν 的估计可以通过核密度估计和自助法实现。应用邦费罗尼校正，以保证这些参数估计的同时置信水平。我们分别在椭球情形和矩形情形下说明该流程。考虑一个样本实现

$\{x_1, x_2, \cdots, x_n\}$ 来自具有累积分布函数 F的随机变量 X。

椭球约束 记$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(x_i)$、$\hat{\boldsymbol{\Sigma}} = \frac{1}{n-1}\sum_{i=1}^{n} \left(\boldsymbol{g}(x_i) - \hat{\boldsymbol{\mu}}\right)\left(\boldsymbol{g}(x_i) - \hat{\boldsymbol{\mu}}\right)^{\top}$，并假设 $\boldsymbol{g}$ 的维度为 d。为校准椭球区域 S_E，我们利用 (A.1) 中所示的 χ^2 弱收敛结果。具体而言，我们需要确定 z，使得

$$n(\mathbb{E}[\boldsymbol{g}(X)] - \hat{\boldsymbol{\mu}})^{\top} \hat{\boldsymbol{\Sigma}}_n^{-1} (\mathbb{E}[\boldsymbol{g}(X)] - \hat{\boldsymbol{\mu}}) \leq z,$$

从而构造所需的置信区域。

为此，在$\mathscr{P}^1_{a,\ \eta}$ 或$\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu}$ 的条件下，分别将值 z 选为自由度为 d 的卡方分布的 $\left(1 - \frac{\alpha}{2}\right)^{\text{th}}$ 或 $\left(1 - \frac{\alpha}{3}\right)^{\text{th}}$ 分位数。对于前一种情况，将 η 估计为 a处自助法密度的 $1 - \frac{\alpha}{2}$ 百分位数。对于后一种情况，分别将 η 和 $\bar{\eta}$ 选为 a 处自助法密度的 $\frac{\alpha}{6}$ 百分位数和 $1 - \frac{\alpha}{6}$ 百分位数，并将 $-\hat{}\nu$ 选为自助系数 $F_+^{(2)}(a)$的 $\frac{\alpha}{3}$ 百分位数。

对于多个阈值，我们希望获得一个在不同 a_i、$i = 1$、...、m 下所有目标值中最优的结果，并以概率 $1 - \alpha$覆盖真实值。在这种情况下，在 $\mathscr{P}^1_{a_i}$、η 下，值 z 分别选为具有 d 个自由度的卡方分布的 $\left(1 - \frac{\alpha}{m+1}\right)^{\text{th}}$ 或 $\left(1 - \frac{\alpha}{2m+1}\right)^{\text{th}}$ 分位数；在 $\mathscr{P}^2_{a_i}$、η、$\bar{\eta}$、ν 下亦然。或者，我们也可以通过自助法选择 z，如算法 $\hat{}\,1$ 所示。对于每个 a_i、$i = 1$、...、m，$\hat{\boldsymbol{\mu}}$、$\hat{\boldsymbol{\Sigma}}$ 分别为经验均值和协方差。η选为 a_i 处自助法密度的第 $1 - \frac{\alpha}{m+1}$ 百分位数，其中 $\mathscr{P}^1_{a_i}$、η。对于 $\mathscr{P}^2_{a_i}$、η、$\bar{\eta}$、ν、η 和 $\bar{\eta}$，分别选为 a_i 处自助法密度的第 $\frac{\alpha}{4m+2}$ 百分位数和第 $1 - \frac{\alpha}{4m+2}$ 百分位数；而 $-\hat{}\nu$ 为自助法系数 $F_+^{(2)}$ (a_i) 的第 $\frac{\alpha}{2m+1}$ 百分位数。

算法1 用于计算 z 的自助法程序。

1: **对于** 500 **次**

2: 从 $\{x_1, \cdots, x_n\}$ 中有放回地随机抽取 n 个数据点。

3: 每次抽样后计算 $\hat{z}$：

$$\hat{z} = \max_{a_i, i=1,\ldots,m} \left\{ n\left(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}\right)^T \hat{\boldsymbol{\Sigma}}^{-1} \left(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\mu}}\right) \right\}$$

其中，$\hat{\boldsymbol{\mu}}$ 是 $\{\boldsymbol{g}(x_i^b)\}_{i=1}^n$ 的经验均值，而 $\{x_1^b, \cdots, x_n^b\}$ 是重新抽样后得到的数据集。

4: **循环结束**

5: 分别选择上述在 $\mathscr{P}^1_{a_i,\eta}$ 或 $\mathscr{P}^2_{a_i,\eta,\bar{\eta},\nu}$ 下获得的 $\hat{z}$ 的 $(1-\frac{\alpha}{m+1})^{\text{th}}$ 或 $(1-\frac{\alpha}{2m+1})^{\text{th}}$ 分位数。

矩形约束 为校准矩形区域 S_R，我们使用（A.2）中所示的柯尔莫哥洛夫-斯米尔诺夫弱收敛结果。具体而言，我们需要确定 z 在

$$\sqrt{n} \max_{i=1,\cdots,n} \|F(x_i) - \widehat{F}(x_i)\|_\infty \leq z, \tag{A.3}$$

其中，F 是经验分布，用于构造置信区域。为此，值 z 是柯尔莫哥洛夫分布在 $\mathscr{P}^1_{a,\eta}$ 或 $\mathscr{P}^2_{a,\eta,\hat{\eta},\nu}$ 下的 $\left(1-\frac{\alpha}{2}\right)^{\text{th}}$ 或 $\left(1-\frac{\alpha}{3}\right)^{\text{th}}$ 分位数。对于前一种情况，η 是自助法密度 $F^{(1)}(a)$ 的 $1-\frac{\alpha}{2}$ 百分位数。对于后一种情况，$\bar{\eta}$ 和 η 分别是具有联合概率 $1-\frac{\alpha}{3}$ 的 $F^{(1)}(a)$ 的上界和下界，而 $\hat{\nu}$ 是 $F_+^{(2)}(a|x \geq a)$ 在概率 $1-\frac{\alpha}{3}$ 下的下界。

对于在一系列阈值 $a_i, i = 1, ..., m$ 中进行选择的情形，在 $\mathscr{P}^1_{a_i,\eta}$ 或 $\mathscr{P}^2_{a_i,\eta,\hat{\eta},\nu}$ 条件下，取柯尔莫哥洛夫分布的 $\left(1-\frac{\alpha}{m+1}\right)^{\text{th}}$ 或 $\left(1-\frac{\alpha}{2m+1}\right)^{\text{th}}$ 分位数除以 $\sqrt{n}$，作为 z。或者，我们也可以通过自助法选择 z，如算法 2 所示，其中 $\Delta = 1-\frac{\alpha}{m+1}$ 对于 $\mathscr{P}^1_{a_i,\eta}$，而 $\Delta = 1-\frac{\alpha}{2m+1}$ 对于 $\mathscr{P}^2_{a_i,\eta,\bar{\eta},\nu}$。对于第一种情况，将 η 设为自助法密度 $F^{(1)}(a_i)$ 的第 $1-\frac{\alpha}{m+1}$ 百分位数。对于第二种情况，$\bar{\eta}$ 和 η 分别是 $F_+^{(1)}(a_i)$ 的上界和下界，其联合概率为 $1-\frac{\alpha}{2m+1}$；而 $\hat{\nu}$ 是 $F_+^{(2)}(a_i|x \geq a_i)$ 的下界，其概率为 $1-\frac{\alpha}{2m+1}$。

算法 2：用于计算 z 的自助法程序。1：对于 $b = 1, \ldots, 500$，执行以下操作。2：从 $\{x_1, \cdots, x_n\}$ 中有放回地随机抽取 n 个数据点。3：每次抽样后计算 $\hat{z}$：$\hat{z} = \max_{a_b, i=1,\ldots,m} \left\{ \sqrt{n} \max_{j=1,2,\cdots,n} \left\{ \max \left\{ \frac{j}{n} - \#\{k: \frac{a_i \leq x_k \leq x_j^b\}}{n}, \#\{k: a_i \leq x_k \leq x_j^b\} - \frac{j-1}{n} \right\} \right\} \right\}$，其中 $\{x_1^b, \cdots, x_n^b\}$ 是每次抽样后得到的非递减有序数据集。4：结束循环。5：选取上述得到的 $\hat{z}$ 中的第 $(\Delta)^{\text{th}}$ 分位数。6：将第 5 步选出的 $(\Delta)^{\text{th}}$ 分位数除以 $\sqrt{n}$，并将结果作为 z 输出。

B 敏感性分析

从较高层面来看，我们对带参数 $\boldsymbol{\mu}$ 的参数化问题 $\mathfrak{P}_{\boldsymbol{\mu}}$ （A.4）中的最优值进行了扰动分析。具体而言，对于给定的扰动方向$\boldsymbol{dr}$，我们给出了一个公式，将问题扰动后的最优值表示为关于 $\boldsymbol{dr}$ 的线性函数。这项分析展示了分布鲁棒优化估计如何随校准精度变化，从而帮助我们理解该估计对数据和校准程序中的随机性有多稳健。

在程序 $\mathfrak{P}_{\boldsymbol{\mu}}$ 中，给定函数 $H(x)$ 和 $G_j(x)$ $\forall j \in \{0\} \cup [d]$，对于某个正整数 d，其中 $[d] := \{1, \ldots, d\}$，我们作如下假设：(1). $G_0(x) = 1$，$\mu_0 = 1$，将非负有界测度限制为概率测度；(2). 参数 $\mu_j \in \mathrm{R}$ $\forall j \in [d]$。注意，$\mathfrak{P}_{\boldsymbol{\mu}}$的对偶问题，即 $\mathfrak{D}_{\boldsymbol{\mu}}$，是一个线性半无限规划，其中包含有限数量的决策变量和无限数量的线性约束。根据 Goberna 等人（1981），如果约束系统的每个可行点都满足关系 $\boldsymbol{a}^\top \boldsymbol{x} \geq \beta$ （其关联向量为（$\boldsymbol{a}^\top$，β）$^\top$），则称该关系为程序 $\mathfrak{P}$ 中的线性推论关系。若约束系统的每个线性推论关系都是 $\mathfrak{P}$ 某个有限子系统的线性推论关系，则称程序 $\mathfrak{P}$ 为 $Farkas\text{-}Minkowski$ （FM）程序。最后，对于任意程序 $\mathfrak{P}$，F($\mathfrak{P}$)、F*($\mathfrak{P}$) 和 $v(\mathfrak{P})$ 分别表示可行

域、最优解区域和 $\mathfrak{P}$ 的最优目标值。

$$
\begin{aligned}
(\mathfrak{P}_{\boldsymbol{\mu}}): \max_{X\sim P\in\mathscr{M}^{+}(\Omega)} \quad & \mathbb{E}_P[H(X)] \\
\text{s.t.} \quad & \mathbb{E}_P[\boldsymbol{G}(X)] = \boldsymbol{\mu}.
\end{aligned}
\qquad
\begin{aligned}
(\mathfrak{D}_{\boldsymbol{\mu}}): \min_{\boldsymbol{y}} \quad & \boldsymbol{y}^{\top}\boldsymbol{\mu} \\
\text{s.t.} \quad & \boldsymbol{y}^{\top}\boldsymbol{G}(x) \geq H(x), \forall x \in \Omega.
\end{aligned}
\tag{A.4}
$$

其中 $\mathscr{M}^{+}(\Omega)$ 表示 Ω 上非负有界测度的空间 .

定理 A.1. 假设 $\mathfrak{P}_{\boldsymbol{\mu}}$ 可行，且 $\mathfrak{D}_{\boldsymbol{\mu}}$ 是一个可行 *FM* 系统，满足 $\boldsymbol{\mu} \in$ 内$\big(\mathbb{M}(\mathfrak{D}_{\boldsymbol{\mu}})\big)$，那么对于任意满足以下条件的方向 $\boldsymbol{dr}$：约束系统 $\big\{\mathbb{E}_P[G_j(X)]+\mu_j\pi = dr_j, \forall j \in\{0\}\cup[d]\big\}$ 对于$(P,$ $\pi)$ 非空，存在 $\epsilon > 0$ 使得 $\forall\rho$： $0 \leq \rho < \epsilon$,

$$
v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) = v(\mathfrak{P}_{\boldsymbol{\mu}}) + \rho \min\{\boldsymbol{dr}^{\top}\boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^{*}(\mathfrak{D}_{\boldsymbol{\mu}})\}. \tag{A.5}
$$

该推导主要遵循 Goberna 和 L´opez（2000）关于线性半无限线性优化的标准对偶性与敏感性分析，见 Goberna 等人（2007）的定理 2。要使 $\mathfrak{D}_{\boldsymbol{\mu}}$ 成为 FM 系统，有若干充分条件。例如，若 Ω 是紧集，函数 H 和向量函数 $\boldsymbol{G}$ 连续，且 $\exists \boldsymbol{y}$， $\boldsymbol{y}^{\top}\boldsymbol{G}(x)$ $> H(x)$ $\forall x \in \Omega$，则根据引理 A.1， $\mathfrak{D}_{\boldsymbol{\mu}}$ 是 FM 系统。一般而言，这种一阶展开需要求解一个辅助优化问题，以获得差值 $v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) - v(\mathfrak{P}_{\boldsymbol{\mu}})$。然而，若 $\boldsymbol{dr}$ 是 $\boldsymbol{\mu}$的倍数，即对某个 $c \neq 0$有 $\rho\boldsymbol{dr} = c\boldsymbol{\mu}$，则约束系统 $\big\{\mathrm{E}_P[G_j(X)] + \mu_j\pi = dr_j, \ \forall j \in\{0\}\cup[d]\big\}$ 对于 $(\mathcal{P}, \ \pi)$ 非空，因此 (A.5) 可简化为

$$
v(\mathfrak{P}_{\boldsymbol{\mu}+\boldsymbol{dr}}) = v(\mathfrak{P}_{(1+c)\boldsymbol{\mu}}) = v(\mathfrak{P}_{\boldsymbol{\mu}}) + c\min\{\boldsymbol{\mu}^{\top}\boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^{*}(\mathfrak{D}_{\boldsymbol{\mu}})\} = (1+c)v(\mathfrak{P}_{\boldsymbol{\mu}}).
$$

C 补充材料：结果证明

C.1 定理 2.1 的证明

证明。对于程序 $\mathfrak{P}(\mathscr{P}^1_{a,\ \eta},\ \boldsymbol{g},\ \mathbb{S})$，若 $P_{true} \in \mathbf{F}(\mathfrak{P}(\mathscr{P}^1_{a,\ \eta},\ \boldsymbol{g},\ \mathbb{S}))$，则有 $\psi(P_{true}) \leq v(\mathfrak{P}(\mathscr{P}^1_{a,\ \eta},\ \boldsymbol{g},\ \mathbb{S}))$，其中 $v(\mathfrak{P})$ 是程序 $\mathfrak{P}$的最优值。因此，

$$\mathbb{P}(v(\mathfrak{P}(\mathscr{P}^1_{a,\eta}, \boldsymbol{g}, \mathbb{S})) \geq \psi(P_{true})) \geq \mathbb{P}(P_{true} \in \mathbb{F}(\mathfrak{P}(\mathscr{P}^1_{a,\eta}, \boldsymbol{g}, \mathbb{S}))) = 1 - \alpha.$$

对于 $\mathfrak{P}(\mathscr{P}^2_{a,\ \eta,\ \bar{\eta},\ \nu},\ \boldsymbol{g},\ \mathbb{S})$，类似的论证同样成立。 □

C.2 定理 2.2 的证明

证明。若 $P_{true} \in \bigcap_{i=1,\ ...,\ m} \mathbf{F}(\mathfrak{P}(\mathscr{P}_i,\ \boldsymbol{g}_i,\ \mathbb{S}_i))$，则有

$$\psi(P_{true}) \leq v\big(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i)\big), i = 1, ..., m.$$

因此

$$\begin{aligned}
&\mathbb{P}\Big(\min_{a_i, i=1,...,m} v\big(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i)\big) \geq \psi(P_{true})\Big) \\
\geq &\mathbb{P}\Big(P_{true} \in \bigcap_{i=1,...,m} \mathbb{F}(\mathfrak{P}(\mathscr{P}_i, \boldsymbol{g}_i, \mathbb{S}_i))\Big) \\
= &1 - \alpha.
\end{aligned}$$

对于 $P_{true} \in \mathscr{P}^2$（$\min_i a_i$），类似的论证同样成立。 □

C.3 命题 3.1 的证明

证明。根据定义，我们知道 $Y > 0$ 以概率 1 成立，并且

$$F_Y(x) = \mathbb{P}\left(Y \leq x\right) = P\left(X \leq x_F - x^{-1}\right) = F\left(x_F - x^{-1}\right). \tag{A.6}$$

$F_Y \in MDA(H_{-\xi})$ 这一事实已由 Embrechts 等人（1997）证明。现在我们证明其余结论。通过求导可得

$$f_Y(x) = f\left(x_F - x^{-1}\right) x^{-2}, \tag{A.7}$$

$$f'_Y(x) = f'\left(x_F - x^{-1}\right) x^{-4} - 2f\left(x_F - x^{-1}\right) x^{-3}. \tag{A.8}$$

根据假设 3.1，对于 $x > z_Y$， $f(x_F - x^{-1})$ 和 $f'(x_F - x^{-1})$ 均存在：$= 1/(x_F - z) > 0$。因此， F_Y 在 (z_Y, ∞) 上二阶可微。我们还知道，对于 $x > z_Y$， $f(x_F - x^{-1}) > 0$。于是由 (A.7) 可得， $f_Y > 0$ 在 (z_Y, ∞) 上。此外， f_Y 在 (z_Y, ∞) 上单调递减，因为对于 $x > z_Y$， $f(x_F - x^{-1})$ 和 x^{-2} 均为正且单调递减。现在我们只需证明 f_Y 在 (z_Y, ∞) 上也为凸函数。事实上，对于任意 $z_Y < x_1 < x_2$ 和 $0 < \lambda < 1$，有

$$
\begin{aligned}
& f_Y\left(\lambda x_1 + (1-\lambda)x_2\right) \\
=& f\left(x_F - \frac{1}{\lambda x_1 + (1-\lambda)x_2}\right) \frac{1}{\left(\lambda x_1 + (1-\lambda)x_2\right)^2} \\
\leq& f\left(x_F - \frac{\lambda}{x_1} - \frac{1-\lambda}{x_2}\right)\left[\frac{\lambda}{x_1^2} + \frac{1-\lambda}{x_2^2}\right] \\
\leq& \left[\lambda f\left(x_F - x_1^{-1}\right) + (1-\lambda) f\left(x_F - x_2^{-1}\right)\right]\left[\frac{\lambda}{x_1^2} + \frac{1-\lambda}{x_2^2}\right] \\
=& \lambda^2 f\left(x_F - x_1^{-1}\right) x_1^{-2} + (1-\lambda)^2 f\left(x_F - x_2^{-1}\right) x_2^{-2} + \lambda(1-\lambda)\left[f\left(x_F - x_1^{-1}\right) x_2^{-2} + f\left(x_F - x_2^{-1}\right) x_1^{-2}\right] \\
\leq& \lambda^2 f\left(x_F - x_1^{-1}\right) x_1^{-2} + (1-\lambda)^2 f\left(x_F - x_2^{-1}\right) x_2^{-2} + \lambda(1-\lambda)\left[f\left(x_F - x_1^{-1}\right) x_1^{-2} + f\left(x_F - x_2^{-1}\right) x_2^{-2}\right] \\
=& \lambda f_Y(x_1) + (1-\lambda) f_Y(x_2).
\end{aligned}
$$

因此， F_Y 也满足假设 3.1。 □

C.4 命题 3.2 的证明

证明。 显然，(A.6)、(A.7) 和 (A.8) 仍然成立。接下来，我们可以参照命题 3.1 的证明来定义 z_Y，并证明在 (z_Y, ∞) 上， F_Y 二阶可微，且 f_Y 为正、递减且凸。本证明的其余部分借鉴了 Embrechts 等人（1997）。由于 F 满足假设 3.1 和 3.3，众所周知， F 是一个冯·米塞斯函数，其辅助函数为 F/f。更具体地说， F 有如下表示:

$$
\bar{F}(x) = c \exp\left\{-\int_z^x \frac{f(t)}{\bar{F}(t)} \mathrm{d}t\right\}, z < x < x_F
$$

其中 c 是一个正数常量。因此，我们有

$$
\bar{F}_Y(x) = \bar{F}\left(x_F - x^{-1}\right) = c \exp\left\{-\int_z^{x_F - x^{-1}} \frac{f(t)}{\bar{F}(t)} \mathrm{d}t\right\}
$$

$$= c\exp\left\{-\int_{1/(x_F-z)}^{x}\frac{f\left(x_F-s^{-1}\right)s^{-2}}{\bar{F}\left(x_F-s^{-1}\right)}\mathrm{d}s\right\} = c\exp\left\{-\int_{z_Y}^{x}\frac{f_Y(s)}{\bar{F}_Y(s)}\mathrm{d}s\right\}.$$

根据定义， F_Y 也是一个冯·米塞斯函数，其辅助函数为 F_Y/f_Y，这意味着 $F_Y \in MDA(\Lambda)$，且式 (7) 成立。

□

C.5 定理 3.1 的证明

证明。 定义 $h(x) = \mathbb{I}(x > b)$。相应地,

$$H(x) = \int_0^x \int_0^t h(s+a)\mathrm{d}s\mathrm{d}t = \frac{(x-b+a)^2}{2}\mathbb{I}(x > b-a).$$

于是，Lam 和 Mottet（2017）中的假设 1 和假设 2 成立。事实上， $h : \mathbb{R} \to \mathbb{R}^+$ 有界，且关于 $[a, b)$ 单调不减，关于 (b, ∞) 单调不增。此外， $\lambda = \lim_{x\to\infty} H(x)/x^2 = \frac{1}{2}$。因此，我们可以应用 Lam 和 Mottet（2017）中的定理 4。特别地， $z^* = \max_{x\in[0,\mu]} W(x)$, 其中

$$W(x) = \begin{cases} \nu\left(\frac{\sigma-\mu^2}{\sigma-2\mu x+x^2}H(x) + \frac{(\mu-x)^2}{\sigma-2\mu x+x^2}H(\frac{\sigma-\mu x}{\mu-x})\right) & \text{if } x \in [0,\mu); \\ \nu(H(\mu) + \lambda(\sigma-\mu^2)) & \text{if } x = \mu. \end{cases}$$

如果 $\mu \le b-a$，则可以证明 $\arg\max\ W(x) = \mu$，从而 $z^* = W(\mu) = \frac{\nu}{2}(\sigma-\mu^2)$。如果 $\mu > b-a$，则可以证明 $\arg\max\ W(x) = [b-a, \mu]$，从而$z^* = \frac{\nu}{2}[\sigma - 2(b-a)\mu + (b-a)^2]$。

□

C.6 定理3.2的证明

证明。 我们将（11）的最优值记为 q_{opt} ，接下来要证明$q^* = q_{opt}$。首先，我们证明 $q_{opt} \le q^*$。事实上，对于任意可行函数 f， p-分位数q 是满足 $\int_q^\infty \tilde{f}(x)\mathrm{d}x = 1-p$的值。我们看到， $\tilde{f}$对于（8）也是可行的，因此 $z^*(a, q) \ge 1-p$。根据 q^*的定义，我们知道 $q \le q^*$。因此， $q_{opt} \le q^*$。现在我们证明 $q^{opt} = q^*$。

如果 $\beta - \eta^2/(2\nu) < 1-p \le \beta$，那么 $a \le q^* < a+\mu$。考虑一个可行函数 $\tilde{f}$，它

˜在 $[a, q^*]$ 上递减，导数为 $-\nu$。那么， f 的 p 分位数恰好是 q^*。

如果 $1-p=\beta-\eta^2/(2\nu)$，那么 $q^*=a+\mu$。考虑一个可行函数 $\tilde{f}$，它在 $[a, a+\mu-\varepsilon]$ 上以导数 $-\mu$ 递减，随后变得极其平坦。这里， ε 是一个很小的正数。因此， p-分位数介于 $a+\mu-\varepsilon$ 和 $a+\mu$之间。当 $\varepsilon \to 0$时，我们得到 $q_{opt}=q^*$。

如果 $1-p<\beta-\eta^2/(2\nu)$，那么 $q^*=\infty$。不过，我们仍考虑与上述情况相同的可行函数。现在，分位数 q 可以任意大，因此 $q_{opt}=\infty=q^*$。

总之，(12) 中定义的 q^* 正是 (11) 的最优值。

利用 (9)，我们可以很容易地得出 q^* 的显式表达式，因为 z^* 的非常数部分实际上是一个二次函数。

□

C.7 命题 3.3 的证明

¯证明。 已知 $F \in RV_{-1/\xi}$。我们还有 F 绝对连续，且 $(F)'=-f$ 单调。因此，我们可以应用 Resnick（1987）中的命题 0.7，得到 $f \in RV_{-1/\xi-1}$，并且 $\lim_{x\to\infty} -\frac{xf(x)}{F(x)} = -\frac{1}{\xi}$。此外，由于根据我们的假设， f是凸的， f 绝对连续，且 f' 单调。于是我们可以再次应用命题 0.7，得到 $\lim_{x\to\infty} \frac{xf'(x)}{f(x)} = -\frac{1}{\xi} - 1$.

——　　-

在表示（15）中，我们要求 $\lim_{x\to\infty} u(x)/x=\xi$，从而得到

$$\lim_{x\to\infty} \frac{f(x)u(x)}{\bar{F}(x)} = \lim_{x\to\infty} \frac{xf(x)}{\bar{F}(x)}\frac{u(x)}{x} = 1,$$

$$\lim_{x\to\infty} -\frac{f'(x)u^2(x)}{(\xi+1)\bar{F}(x)} = \lim_{x\to\infty} -\frac{1}{\xi+1}\frac{xf'(x)}{f(x)}\frac{xf(x)}{\bar{F}(x)}\frac{u^2(x)}{x^2} = 1.$$

□

C.8 定理 3.3 的证明

证明。 如果 $x > 1/(\xi+1)$，则 $\lim_{a\to\infty}(b-a)/\mu = (\xi+1)x > 1$。因此，对于充分大的$a$，必有 $b - a > \mu$。由定理 3.1，我们得到 $z^* = \frac{\nu}{2}(\sigma - \mu^2)$。因此，

$$\lim_{a\to\infty}\frac{z^*(a,b)}{\bar{F}(a)} = \lim_{a\to\infty}\frac{z^*(a,b)}{\beta} = \lim_{a\to\infty}\left(1 - \frac{\nu\mu^2}{2\beta}\right) = 1 - \frac{1}{2(\xi+1)}.$$

若 $x < 1/(\xi+1)$，则类似地，对于充分大的 a，有 $b - a < \mu$，因此 $z^* = \frac{\nu}{2}[\sigma - 2(b-a)\mu + (b-a)^2]$。于是，

$$\lim_{a\to\infty}\frac{z^*(a,b)}{\bar{F}(a)} = \lim_{a\to\infty}\frac{z^*(a,b)}{\beta} = \lim_{a\to\infty}\left(1 - xu(a)\frac{\nu\mu}{\beta} + x^2u^2(a)\frac{\nu}{2\beta}\right) = 1 - x + \frac{\xi+1}{2}x^2.$$

显然，若 $x = 1/(\xi+1)$，则其极限与 $x > 1/(\xi+1)$ 的情形相同。

因此，结合 $\lim_{a\to\infty}\bar{F}(b)/\bar{F}(a) = (1+\xi x)^{-1/\xi}$，可得 (18) . □

C.9 命题 3.4 的证明

证明。 表示式 (21) 的证明见 Embrechts 等人（1997）中的例 3.3.23。

（22）直接由 $u(x)$ 的表达式得出。（23）由假设 3.3 得出。 □

C.10 定理 3.4 的证明

证明。 与定理 3.3 的证明类似，利用命题 3.4 的结果，可得

$$\lim_{a\to\infty}\frac{z^*(a,b)}{\bar{F}(a)} = \lim_{a\to\infty}\frac{z^*(a,b)}{\beta} = \begin{cases} \frac{1}{2} & \text{if } x \geq 1; \\ 1 - x + \frac{1}{2}x^2 & \text{if } x < 1. \end{cases}$$

此外，已知 $\lim_{a\to\infty}F(b)/F(a) = e^{-x}$(Embrechts 等人，1997 年)，因此我们得到式 (24)。

□

C.11 定理 3.5 的证明

证明。 我们有

$$\lim_{a\to\infty}\frac{q^*}{q}=\lim_{a\to\infty}\frac{(q^*-a)/a+1}{(q-a)/a+1}.$$

首先，我们 处理 $\lim_{a\to\infty}(q^*-a)/a$。我们已证明，对于充分大的 a,

$$q^*=a+\mu-\sqrt{\mu^2-\sigma+\frac{2(1-p)}{\nu}}.$$

Then

$$\lim_{a\to\infty}\frac{q^*-a}{a}=\lim_{a\to\infty}\frac{\mu-\sqrt{\mu^2-\sigma+\frac{2x\beta}{\nu}}}{\mu}\frac{\mu}{u(a)/(\xi+1)}\frac{u(a)}{(\xi+1)a}=\frac{\xi}{\xi+1}\left(1-\sqrt{1-2(1-x)(\xi+1)}\right).$$

接下来，我们处理 $\lim_{a\to\infty}(q-a)/a$。定义 $U(t)=F^{-1}(1-t^{-1})$ $t>0$。于是可以将 q 和 a 写为

$$q=F^{-1}(p)=U\left(\frac{1}{1-p}\right),a=F^{-1}(1-\beta)=U\left(\frac{1}{\beta}\right).$$

已知（Embrechts 等人，1997 年）

$$\lim_{s\to\infty}\frac{U(st)-U(s)}{u(U(s))}=\frac{t^\xi-1}{\xi}.$$

注意到 $\frac{1}{1-p}/\frac{1}{\beta}=1/x$，因此令 $s=1/\beta$, $t=1/x$，得到

$$\lim_{a\to\infty}\frac{q-a}{u(a)}=\frac{t^\xi-1}{\xi}=\frac{x^{-\xi}-1}{\xi}.$$

Hence $\lim_{a\to\infty}(q-a)/a=x^{-\xi}-1$。结合这两部分，我们得到 (25)。 □

C.12 定理 3.6 的证明

证明。 当 a 足够大时，我们知道 $q^*\le a+\mu$。另一方面，由于 $p\ge 1-\beta$,

我们知道真实分位数 $q\ge a$。因此

$$1\le\lim_{a\to\infty}\frac{q^*}{q}\le\lim_{a\to\infty}\frac{a+\mu}{a}=1+\lim_{a\to\infty}\frac{\mu}{u(a)}\frac{u(a)}{a}.$$

根据命题 3.4，我们知道 $\mu/u(a)\to 1$。此外，我们还知道 $u(a)\to a=0$。因此，我们得到（26）。

□

C.13 定理 4.1 的证明：分部积分法

证明。对于定理的第一项，不失一般性，我们假设 $\boldsymbol{g} = (\mathbf{I}(x \geq a)\ g(x))^\top$ 以及 $\mathbf{S} = \{\boldsymbol{\Gamma}\} = \{(\beta,\ \Gamma)^\top\}$，其中 $\beta,\ \Gamma$ 为标量，且 $g(x)$：$[a, \infty) \to \mathbf{R}$ 是定义在 $[a, \infty)$ 上的可积函数；稍后我们将说明如何将该结果推广到任意区域 S。我们将程序 $\mathfrak{P}(h,\ \boldsymbol{g},\ \{\boldsymbol{\Gamma}\},\ \mathscr{P}^1_{a,\ \eta})$ 重写为

$$\max_f \int_a^\infty h(x)f(x)dx \tag{A.9a}$$

$$\text{s.t.} \int_a^\infty f(x)dx = \beta, \tag{A.9b}$$

$$\int_a^\infty g(x)f(x)dx = \Gamma, \tag{A.9c}$$

$$f(a) \leq \eta, \tag{A.9d}$$

$$f(x) \text{ exists, non-increasing and right-continuous for } x \geq a, \tag{A.9e}$$

$$f(x) \geq 0 \text{ for } x \geq a. \tag{A.9f}$$

根据假设，$h(x)$ 和 $g(x)$ 仅在 $[a, \infty)$ 上取非零值，因此我们可以只考虑从 a 开始的积分。我们将 $f(x)$ 视为累积分布函数 F 的右导数。由于单调函数的不连续点至多可数，且这些点不会影响在 $[a, \infty)$ 上的积分值，因此我们假设对于 $x \geq a$，$f(x)$ 在所有这些不连续点处右连续。

记

$$\tilde{H}(x) = \int_a^x h(u)du, \qquad \tilde{G}(x) = \int_a^x g(u)du.$$

~ ~由于 H 连续，按定义有 $H(a) = 0$，且由 (A.9d)、(A.9e) 可知 f 有界变差。

以及（A.9f），利用分部积分可得，（A.9a）等于 t o

$$\begin{aligned}\int_a^\infty f(x)h(x)dx &= \int_a^\infty f(x)d\tilde{H}(x) \\ &= f(x)\tilde{H}(x)|_a^\infty - \int_a^\infty \tilde{H}(x)df(x) \\ &= -\int_a^\infty \tilde{H}(x)df(x)\end{aligned}$$

其中，第三个等式由稍后给出的引理 $A.2$ 得出，结合 $\alpha = 0$ 和 $H(x) = O(x)$ 作为 $x \to \infty$，因为 h 有界。

（A.9b）可改写为

$$\begin{aligned}\int_a^\infty f(x)dx &= \int_a^\infty f(x)d(x-a) \\ &= f(x)(x-a)|_a^\infty - \int_a^\infty (x-a)df(x) \\ &= -\int_a^\infty (x-a)df(x)\end{aligned}$$

其中，第三个等式再次由引理 $A.2$ 结合 $\alpha = 0$ 得出。

对于（A.9c），由于 G 连续，且根据定义，$G(a) = 0$，我们可以写为

$$\begin{aligned}\int_a^\infty f(x)g(x)dx &= \int_a^\infty f(x)d\tilde{G}(x) \\ &= f(x)\tilde{G}(x)|_a^\infty - \int_a^\infty \tilde{G}(x)df(x) \\ &= -\int_a^\infty \tilde{G}(x)df(x)\end{aligned}$$

其中，第三个等式由引理 $A.3$ 得出。

最后，由于 根据引理 A.2，结合 $\alpha = 0$，$f(x) \to 0$ 可得 $x \to \infty$，因此我们可以将（A.9d）写为

$$f(a) = -\int_a^\infty df(x).$$

因此，(A.9) 等价于

$$\max_f \; -\int_a^\infty \tilde{H}(x)df(x), \tag{A.10a}$$

$$\text{s.t.} \quad -\int_a^\infty (x-a)df(x) = \beta, \tag{A.10b}$$

$$-\int_a^\infty \tilde{G}(x)df(x) = \Gamma, \tag{A.10c}$$

$$-\int_a^\infty df(x) \le \eta, \tag{A.10d}$$

$$f(x) \text{ exists, non-increasing and right-continuous for } x \ge a, \tag{A.10e}$$

$$f(x) \ge 0 \text{ for } x \ge a, \tag{A.10f}$$

$$xf(x) \to 0 \text{ as } x \to \infty, \tag{A.10g}$$

$$\tilde{G}(x)f(x) \to 0 \text{ as } x \to \infty. \tag{A.10h}$$

可按如下方式验证 (A.9) 与 (A.10) 的等价性：将 (A.9) 的可行域记为 F_1 ，将 (A.10) 的可行域记为 F_2。根据上述讨论，很容易得出 $\mathrm{F}_1 \subseteq \mathrm{F}_2$。对于另一个方向，我们对 (A.10a)、(A.10b)、(A.10c)、(A.10d) 分别进行分部积分，得到 (A.9a)、(A.9b)、(A.9c)、(A.9d)。因此，$\mathrm{F}_2 \subseteq \mathrm{F}_1$。这表明 (A.9) 与 (A.10) 等价。

最后，令 $p(x) = -\frac{f(x)}{\eta} + 1$。于是，(A.10) 可改写为

$$\max_f \eta \int_a^\infty \tilde{H}(x)dp(x) \tag{A.11a}$$

$$\text{s.t.} \quad \eta \int_a^\infty (x-a)dp(x) = \beta, \tag{A.11b}$$

$$\eta \int_a^\infty \tilde{G}(x)dp(x) = \Gamma, \tag{A.11c}$$

$$\eta \int_a^\infty dp(x) \leq \eta, \tag{A.11d}$$

$$p(x) \text{ exists, non-decreasing and right-continuous for } x \geq a, \tag{A.11e}$$

$$0 \leq p(x) \leq 1 \text{ for } x \geq a, \tag{A.11f}$$

$$p(x) \to 1 \text{ as } x \to \infty, \tag{A.11g}$$

$$p(x) = 0 \text{ for } x < a, \tag{A.11h}$$

$$(1-p(x))x \to 0 \text{ for } x \to \infty, \tag{A.11i}$$

$$\tilde{G}(x)(1-p(x)) \to 0 \text{ as } x \to \infty. \tag{A.11j}$$

由于 $\tilde{H}(x) = (x-a) = \tilde{G}(x) = 0$ 在 $x = a$处，因此可以唯一确定（除去零测集），一个非递减、右连续的 p ，使得 $\lim_{x\to\infty} p(x) = 1$ ，且 $p(x) = 0$对 $x < a$ 成立，并且其概率测度的支撑集位于 $[a, \infty)$ 。最后，根据引理 A.4，约束条件 (A.11i) 和 (A.11j) 可以由该优化问题中的其他约束条件推导得出。约束条件 (A.11d) 包含在 (A.11f) 中。我们得到如下等价问题

如下

$$
\begin{aligned}
\max_{f}\ & \eta\int_{-\infty}^{\infty}\tilde{H}(x)dp(x)\\
\text{s.t.}\quad & \eta\int_{-\infty}^{\infty}(x-a)dp(x)=\beta,\\
& \eta\int_{-\infty}^{\infty}\tilde{G}(x)dp(x)=\Gamma,\\
& p(x)\text{ exists, non-decreasing and right-continuous for } x\in\mathbb{R},\\
& 0\le p(x)\le 1 \text{ for } x\in\mathbb{R},\\
& p(x)\to 1 \text{ as } x\to\infty,\\
& p(x)=0 \text{ for } x<a.
\end{aligned}
$$

至此，定理前半部分的证明结束。

对于定理的后半部分，我们首先考虑 $\eta=\bar{\eta}=\eta$，并将程序$\mathfrak{P}\big(h,\ \boldsymbol{g},\ \{\boldsymbol{\Gamma}\},\ \mathscr{P}^{2}_{a,\ \eta,\ \eta,\ \nu}\big)$改写为

$$
\begin{aligned}
\max_{f}\ & \int_{a}^{\infty}h(x)f(x)dx\\
\text{s.t.}\quad & \int_{a}^{\infty}f(x)dx=\beta,\\
& \int_{a}^{\infty}g(x)f(x)dx=\Gamma,\\
& f(a)=f(a+)=\eta,\\
& f'_{+}(a)\ge-\nu,\\
& f \text{ convex for } x\ge a,\\
& f(x)\ge 0 \text{ for } x\ge a.
\end{aligned}
\tag{A.12}
$$

根据 Lam 和 Mottet（2017），表述 (A.12) 等价于

$$\max_{f}\int_{a}^{\infty}h(x)f(x)dx \tag{A.13a}$$

$$\text{s.t.}\quad \int_{a}^{\infty}f(x)dx=\beta, \tag{A.13b}$$

$$\int_{a}^{\infty}g(x)f(x)dx=\Gamma, \tag{A.13c}$$

$$f(a) = \eta, \tag{A.13d}$$

$$f'_+(x) \text{ exists and is non-decreasing and right-continuous for } x \geq a, \tag{A.13e}$$

$$-\nu \leq f'_+(x) \leq 0 \text{ for } x \geq a, \tag{A.13f}$$

$$f'_+(x) \to 0 \text{ a } x \to \infty, \tag{A.13g}$$

$$f(x) = \int_a^x f'_+(t)dt + \eta \text{ for } x \geq a. \tag{A.13h}$$

这里，$f(a+)$ 表示在 a 处的右极限，而 $f(a) = f(a+)$ 表示 f 在 a 处右连续，这意味着在 a 处进行连续外推。

记

$$\tilde{H}(x) = \int_a^x h(u)du, \qquad H(x) = \int_a^x \tilde{H}(u)du,$$
$$\tilde{G}(x) = \int_a^x g(u)du, \qquad G(x) = \int_a^x \tilde{G}(u)du.$$

˜考虑目标函数 (A.13a)。由于 H 和 H 是连续的，根据 (A.13h)，f 是绝对连续的；又由于 (A.13f) 和 (A.13g)，f'_+ 具有有界变差。因此，利用分部积分，得

$$\begin{aligned}
\int_a^\infty f(x)h(x)dx &= \int_a^\infty f(x)d\tilde{H}(x) \\
&= f(x)\tilde{H}(x)|_a^\infty - \int_a^\infty \tilde{H}(x)f'_+(x)dx \\
&= -\int_a^\infty \tilde{H}(x)f'_+(x)dx \\
&= -H(x)f'_+(x)|_a^\infty + \int_a^\infty H(x)df'_+(x) \\
&= \int_a^\infty H(x)df'_+(x)
\end{aligned} \tag{A.14}$$

˜其中，第三个等式由引理 A.2 结合 $\alpha = 0$ 得出，并且由于 h 有界，当$x \to \infty$ 时，$H(x) = O(x)$。第五个等式再次由引理 A.2 结合 $\alpha = 0$ 得出，并且当 $x \to \infty$时，$H(x) = O(x^2)$。

对于 (A.13b)，我们可以写成

$$\int_a^\infty f(x)dx = \int_a^\infty f(x)d(x-a)$$

$$
\begin{aligned}
&= f(x)(x-a)|_a^\infty - \int_a^\infty (x-a) f'_+(x)dx \\
&= -\int_a^\infty (x-a) f'_+(x)dx \\
&= -\frac{(x-a)^2}{2} f'_+(x)|_a^\infty + \int_a^\infty \frac{(x-a)^2}{2} df'_+(x) \\
&= \int_a^\infty \frac{(x-a)^2}{2} df'_+(x)
\end{aligned}
$$

只需将 (A.14) 中的 $h \equiv 1$ 视为.....即可。

对于 (A.13d)，注意到由引理 A.2 可知，当 $x \to \infty$ 时，$f(x) \to 0$，且 $\alpha = 0$，我们可以再次使用分部积分，得到

$$
\begin{aligned}
f(a) &= -\int_a^\infty f'_+(x)dx = -\int_a^\infty f'_+(x)d(x-a) \\
&= -f'_+(x)(x-a)|_a^\infty + \int_a^\infty (x-a)df'_+(x) \\
&= \int_a^\infty (x-a)df'_+(x).
\end{aligned}
$$

˜对于 (A.13c)，由于 G 和 $G(x)$ 连续，使用分部积分可得，

$$
\begin{aligned}
\int_a^\infty f(x)g(x)dx &= \int_a^\infty f(x)d\tilde{G}(x) \\
&= f(x)\tilde{G}(x)|_a^\infty - \int_a^\infty \tilde{G}(x) f'_+(x)dx \\
&= -\int_a^\infty \tilde{G}(x) f'_+(x)dx \\
&= -G(x) f'_+(x)|_a^\infty + \int_a^\infty G(x)df'_+(x) \\
&= \int_a^\infty G(x)df'_+(x)
\end{aligned}
$$

其中第三个等式和第五个等式由引理 A.3 得出。因此，(A.13) 为

等价于

$$
\max_f \int_a^\infty H(x)df'_+(x) \tag{A.15a}
$$

$$
\text{s.t.} \quad \int_a^\infty \frac{(x-a)^2}{2} df'_+(x) = \beta, \tag{A.15b}
$$

$$
\int_a^\infty (x-a)df'_+(x) = \eta, \tag{A.15c}
$$

$$\int_a^\infty G(x)df'_+(x) = \Gamma, \tag{A.15d}$$

$$f'_+(x) \text{ exists and is non-decreasing and right-continuous for } x \geq a, \tag{A.15e}$$

$$-\nu \leq f'_+(x) \leq 0 \text{ for } x \geq a, \tag{A.15f}$$

$$f'_+(x) \to 0 \text{ as } x \to \infty, \tag{A.15g}$$

$$f(x) = \int_a^x f'_+(t)dt + \eta \text{ for } x \geq a, \tag{A.15h}$$

$$xf(x), x^2 f'_+(x) \to 0 \text{ as } x \to \infty, \tag{A.15i}$$

$$\tilde{G}(x)f(x), G(x)f'_+(x) \to 0 \text{ as } x \to \infty \tag{A.15j}$$

且约束 (A.15h) 表明，可从 $f(x) = \int_a^x f'_+(t)dt + \eta$ 中恢复 f。请注意，f 的定义具有与所得 $f'_+(x)$ 相一致的右导数。

可以如下验证 (A.13) 与 (A.15) 的等价性：记 (A.13) 的可行域为 F_1，(A.15) 的可行域为 F_2。由上述讨论可直接得出 $\mathrm{F}_1 \subseteq \mathrm{F}_2$。反过来，对 (A.15a)、(A.15b)、(A.15c)、(A.15d) 分别进行分部积分，可得到 (A.13a)、(A.13b)、(A.13d)、(A.13c)。因此 $\mathrm{F}_2 \subseteq \mathrm{F}_1$。这说明 (A.13) 与 (A.15) 等价。

现在证明 (A.15i) 和 (A.15j) 是冗余的。对于满足 (A.15e)、(A.15f) 和 (A.15g) 的 $f'_+(x)$，由 (A.15b) 可知

$$\int_x^\infty \frac{(t-a)^2}{2} df'_+(t) \to 0 \text{ as } x \to \infty.$$

注意，根据 (A.15e)，利用 $f'_+(x)$ 的非递减性质，我们有以下不等式

$$\int_x^\infty \frac{(t-a)^2}{2} df'_+(t) \geq \frac{(x-a)^2}{2} \int_x^\infty df'_+(t) \geq 0.$$

因此，借助 (A.15g)，令 $f'_+(x) \to 0$ 为 $x \to \infty$，可得

$$\frac{(x-a)^2}{2} f'_+(x) \to 0 \text{ as } x \to \infty. \tag{A.16}$$

对于 (A.15c)，我们可以写成

$$\eta = \int_a^\infty (x-a)df'_+(x) = (x-a)f'_+(x)|_a^\infty - \int_a^\infty f'_+(x)dx = \int_a^\infty -f'_+(x)dx$$

w这里的第三个等式由 (A.16) 得出。容易得出

$$\int_x^\infty -f'_+(t)dt \to 0 \text{ as } x \to \infty. \tag{A.17}$$

由 (A.15h)，结合 (A.15f) 和 (A.17)，我们可以轻易推导出

$$f(x) = \int_a^x f'_+(t)dt + \eta = \int_a^x f'_+(t)dt + \int_a^\infty -f'_+(x)dx = \int_x^\infty -f'_+(t)dt \geq 0 \tag{A.18}$$

$\implies$ $f(x) \to 0$ 当 $x \to \infty$.

接着，由 (A.15b)，我们可以写成

$$\begin{aligned}\beta &= \int_a^\infty \frac{(x-a)^2}{2} df'_+(x) \\ &= f'_+(x)\frac{(x-a)^2}{2}|_a^\infty - \int_a^\infty f'_+(x)(x-a)dx \\ &= \int_a^\infty -f'_+(x)(x-a)dx\end{aligned}$$

其中第三个等式由 (A.16) 得出。接着，容易得出

$$\int_x^\infty -f'_+(t)(t-a)dt \to 0 \text{ as } x \to \infty.$$

注意，利用 (A.15f) 中 $f'_+(x)$ 的非正性，我们有以下不等式

$$\int_x^\infty -f'_+(t)(t-a)dt \geq (x-a)\int_x^\infty -f'_+(t)dt \geq 0.$$

因此，结合式 (A.18) 中的 $f(x)$，我们有

$$(x-a)f(x) \to 0 \text{ as } x \to \infty$$

由此可知，约束 (A.15i) 存在冗余。

现在我们证明约束 (A.15j) 存在冗余。首先考虑 $g(x)$ 是一个非-˜非负函数的情况。此时， $G(x)$ 和 $G(x)$ 是非递减的非负连续函数。

由（A.15d），我们有

$$\int_x^\infty G(t)df'_+(t) \to 0 \text{ as } x \to \infty.$$

注意，根据（A.15e）和 $G(x)$， $f'_+(x)$ 具有非递减性质，因此我们有以下不等式

$$\int_x^\infty G(t)df'_+(t) \geq G(x)\int_x^\infty df'_+(t) \geq 0.$$

因此，根据（A.15g），当 $f'_+(x) \to 0$ 趋于 $x \to \infty$ 时，我们有

$$G(x)f'_+(x) \to 0 \text{ as } x \to \infty. \tag{A.19}$$

对于（A.15d），我们可以写成

$$\begin{aligned}\Gamma &= \int_a^\infty G(x)df'_+(x) \\ &= f'_+(x)G(x)|_a^\infty - \int_a^\infty f'_+(x)\tilde{G}(x)dx \\ &= \int_a^\infty -f'_+(x)\tilde{G}(x)dx\end{aligned}$$

其中第三个等式由（A.19）得出。因此可得

$$\int_x^\infty -f'_+(t)\tilde{G}(t)dt \to 0 \text{ as } x \to \infty.$$

注意，根据（A.15f）， $f'_+(x)$具有非正性，且 $G(x)$具有非递减˜性质，因此我们有以下不等式

$$\int_x^\infty -f'_+(t)\tilde{G}(t)dt \geq \tilde{G}(x)\int_x^\infty -f'_+(t)dt \geq 0.$$

因此，结合（A.18）中 $f(x)$ 的等式，我们有

$$\tilde{G}(x)f(x) \to 0 \text{ as } x \to \infty.$$

现在考虑这样的情形： $g(x)$ 是有下界函数，且 $g(x)$ 的值可能为负。我们考虑 $\tilde{g}(x) = g(x) + |\min_{x\geq a} g(x)|$。显然， $\tilde{g}(x)$ 和$|\min_{x\geq a} g(x)|$ 是非负函数，这意味着我们可以使用上述结果。因此，关于 $g(x)$的结果可由积分的线性性和极限的和法则得出。详细推导如下:

给定

$$\begin{aligned}\tilde{G}(x) &= \int_a^x \tilde{g}(t)dt + \int_a^x -|\min_{x\geq a} g(x)|dt \\ &= \int_a^x \tilde{g}(t)dt - (x-a)|\min_{x\geq a} g(x)|,\end{aligned}$$

$$G(x) = \int_a^x \int_a^v \tilde{g}(u)dudv - \frac{(x-a)^2}{2}|\min_{x\geq a} g(x)|,$$

我们有

$$\begin{aligned}
&\lim_{x\to\infty} \int_a^x \tilde{g}(t)dt f(x) \to 0,\ \lim_{x\to\infty} -(x-a)|\min_{x\geq a} g(x)|f(x) \to 0 \\
\Longrightarrow\ &\lim_{x\to\infty} \tilde{G}(x)f(x) \to 0, \\
&\lim_{x\to\infty} \int_a^x \int_a^v \tilde{g}(u)dudv f'_+(x) \to 0,\ \lim_{x\to\infty} -\frac{(x-a)^2}{2}|\min_{x\geq a} g(x)|f'_+(x) \to 0 \\
\Longrightarrow\ &\lim_{x\to\infty} G(x)f(x) \to 0,
\end{aligned}$$

从而得出约束（A.15j）存在冗余。

最后，令 $p(x) = \frac{f'_+(x)}{\nu} + 1$。那么 (A.15) 等价于

$$\begin{aligned}
&\max_p \int_a^\infty \nu H(x)dp(x) \\
\text{s.t.}\quad &\int_a^\infty \nu\frac{(x-a)^2}{2}dp(x) = \beta, \\
&\int_a^\infty \nu(x-a)dp(x) = \eta, \\
&p(x) \text{ exists and is non-decreasing and right-continuous for } x \geq a, \\
&0 \leq p(x) \leq 1 \text{ for } x \geq a, \\
&p(x) \to 1 \text{ for } x \to \infty, \\
&\int_a^\infty \nu G(x)dp(x) = \Gamma
\end{aligned} \tag{A.20}$$

或等价地 y

$$
\begin{aligned}
&\max_{p} \int_{-\infty}^{\infty} \nu H(x) dp(x) \\
\text{s.t.} \quad &\int_{-\infty}^{\infty} \nu \frac{(x-a)^2}{2} dp(x) = \beta, \\
&\int_{-\infty}^{\infty} \nu (x-a) dp(x) = \eta, \\
&p(x) \text{ exists and is non-decreasing and right-continuous for } x \in \mathbb{R}, \\
&0 \le p(x) \le 1 \text{ for } x \in \mathbb{R}, \\
&p(x) \to 1 \text{ for } x \to \infty, \\
&p(x) = 0 \text{ for } x < a, \\
&\int_{-\infty}^{\infty} \nu G(x) dp(x) = \Gamma.
\end{aligned}
\tag{A.21}
$$

由于 $H(x) = (x-a)^2 = (x-a)a = G(x) = 0$ 在 $x = a$处，因此可以唯一确定一个非递减、右连续的 p，除测度为零的集合外，满足 $\lim_{x\to\infty} p(x) = 1$ 且 $p(x) = 0$，对于 $x < a$，其支撑集上有一个概率测度 $[a, \infty)$。因此，(A.21) 等价于 (A.20)。

当 $\eta \neq \bar{\eta}$时，可以将 (A.21) 中的等式约束 $\int_{-\infty}^{\infty} \nu(x-a)dp(x) = \eta$ 替换为 $\eta \le \int_{-\infty}^{\infty} \nu(x-a)dp(x) \le \bar{\eta}$，从而形成一个矩形约束。注意，上述推导对 β 和 Γ 的任意选择均成立，因此将 $\{\boldsymbol{\Gamma}\} = \{(\beta, \Gamma)^\top\}$ 替换为一般的 S 时，结论仍然成立。由此得证。

□

C.14 定理 4.1 的证明：Choquet 方法

证明。 不失一般性，假设 $\boldsymbol{g} = \big(\mathbf{I}(x \ge a)\ g(x)\big)^\top$ 且 $\mathbf{S} = \{\boldsymbol{\Gamma}\} = \{(\beta, \Gamma)^\top\}$ 对于某些标量 β, Γ 其中 $g(x)\colon [a, \infty) \to \mathbf{R}$ 是$[a, \infty$上的可积函数，然后我们说明如何将该结果推广到任意区域 S。注意，

对于程序 (A.9) 中的任何可行解 f，我们可以利用定理 A.1 得到

$$\begin{aligned}
\int_a^\infty h(x)f(x)dx &= \beta\int_a^\infty\int_{0+}^\infty h(x)\frac{\mathbb{I}(a\le x<a+z)}{z}dQ(z)dx\\
&= \beta\int_{0+}^\infty\int_a^\infty h(x)\frac{\mathbb{I}(a\le x<a+z)}{z}dxdQ(z)\\
&= \beta\int_{0+}^\infty\frac{\int_a^{a+z}h(x)dx}{z}dQ(z) = \beta\int_{0+}^\infty\frac{\tilde{H}(z+a)}{z}dQ(z),\\
\int_a^\infty g(x)f(x)dx &= \beta\int_{0+}^\infty\frac{\tilde{G}(z+a)}{z}dQ(z),\\
\int_a^\infty f(x)dx &= \beta\int_a^\infty\int_{0+}^\infty\frac{\mathbb{I}(a\le x<a+z)}{z}dQ(z)dx\\
&= \beta\int_{0+}^\infty\int_a^\infty\frac{\mathbb{I}(a\le x<a+z)}{z}dxdQ(z)\\
&= \beta\int_{0+}^\infty dQ(z) = \beta,\\
f(a) &= \int_{0+}^\infty\frac{\mathbb{I}(a\le x<a+z)}{z}dQ(z) = \int_{0+}^\infty\frac{1}{z}dQ(z).
\end{aligned}$$

因此，程序 (A.9) 等价于以下程序，

$$\begin{aligned}
&\max_{Z\sim Q}\ \beta\mathbb{E}_Q\Big[\frac{\tilde{H}(Z+a)}{Z}\Big]\\
&\text{s.t.}\quad \beta\mathbb{E}_Q[1]=\beta,\\
&\qquad \mathbb{E}_Q\Big[\frac{\tilde{G}(Z+a)}{Z}\Big]=\frac{\Gamma}{\beta},\\
&\qquad \mathbb{E}_Q\big[\frac{1}{Z}\big]\le\frac{\eta}{\beta},\\
&\qquad Q\in\mathscr{M}^+(0,\infty)
\end{aligned}$$

w其中，$\mathscr{M}^+(0,\infty)$ 是 $(0,\infty)$ 上非负有界测度的空间。

由于 $\mathrm{E}_Q[\frac{1}{Z}]<\infty$，我们可以通过 $\frac{dQ}{dQ}(z)=\frac{\beta}{\eta}\frac{1}{z}$ 定义一个关于 Q 绝对连续的分布函数 $\tilde{Q}\in\mathscr{M}^+(0,\infty)$。我们将决策变量从 Q 转换为 Q。此外，可行集还可以进一步限制为 $\mathscr{P}[0,\infty)$。这是因为函数 $H(z+a)$ $G(z+a)$ z 根据构造在 $z=0$ 处等于 0。因此，我们总可以在 0 处添加任意质量，以达到 $\mathrm{E}_{\tilde{Q}}[1]$ 的上界。由此可推得，在对测度进行适当归一化后，我们可以施加如下约束：

$\mathbb{E}_{\tilde{Q}}[1] = 1$。最后，我们令 $x = z + a$，并得到以下程序与程序 (A.9) 等价：

$$\begin{aligned} \max_{X \sim \tilde{Q}} \quad & \eta \mathbb{E}_{\tilde{Q}}[\tilde{H}(X)] \\ \text{s.t.} \quad & \eta \mathbb{E}_{\tilde{Q}}[X - a] = \beta, \\ & \eta \mathbb{E}_{\tilde{Q}}[\tilde{G}(X)] = \Gamma, \\ & \tilde{Q} \in \mathscr{P}[a, \infty). \end{aligned}$$

对于定理的第二部分，我们首先考虑 $\eta = \bar{\eta} = \eta_{\circ}$ 对于程序（A.12）中的任意可行解f，Lam 和 Mottet（2017）的引理 1 表明，f 是非递增的。接着，根据 Popescu（2005）的引理 5.1，$f(x)\ \forall x \geq a$ 可以写成右 a-三角密度的广义混合，即

$$f(x) = \beta \int_0^{\infty} \frac{(a + z - x)^+}{z^2/2} dQ(z)$$

其中 Q 是定义在 $[0, \infty)$ 上的概率测度。由于对于 $x \geq a$，f 处处存在，因此我们有

$$Q(z = 0) = 0.$$

˜ ˜然后，由于 $H(x)$ $G(x)$，$(a + z - x)$ 都是绝对连续的，利用分部积分可得，

$$\begin{aligned} \int_a^{\infty} h(x) f(x) dx &= \beta \int_a^{\infty} h(x) \int_0^{\infty} \frac{(a + z - x)^+}{z^2/2} dQ(z) dx \\ &= \beta \int_0^{\infty} \frac{\int_a^{\infty} h(x)(a + z - x)^+ dx}{z^2/2} dQ(z) \\ &= \beta \int_0^{\infty} \frac{\int_a^{a+z} h(x)(a + z - x) dx}{z^2/2} dQ(z) \\ &= \beta \int_0^{\infty} \frac{\tilde{H}(x)(a + z - x)|_{x=a}^{x=a+z} + \int_a^{a+z} \tilde{H}(x) dx}{z^2/2} dQ(z) \\ &= \beta \int_0^{\infty} \frac{\int_a^{a+z} \tilde{H}(x) dx}{z^2/2} dQ(z) = \beta \int_{0+}^{\infty} \frac{H(a + z)}{z^2/2} dQ(z), \\ \int_a^{\infty} g(x) f(x) dx &= \beta \int_{0+}^{\infty} \frac{G(a + z)}{z^2/2} dQ(z), \\ \int_a^{\infty} f(x) dx &= \beta \int_a^{\infty} \int_0^{\infty} \frac{(a + z - x)^+}{z^2/2} dQ(z) dx = \beta \int_0^{\infty} \frac{\int_a^{\infty} (a + z - x)^+ dx}{z^2/2} dQ(z) \end{aligned}$$

$$
\begin{aligned}
&= \beta \int_0^\infty \frac{\int_a^{a+z} x(a+z-x)dx}{z^2/2} dQ(z) \\
&= \beta \int_0^\infty \frac{(x-a)(a+z-x)|_{x=a}^{x=a+z} + \int_a^{a+z}(x-a)dx}{z^2/2} dQ(z) \\
&= \beta \int_0^\infty \frac{\int_a^{a+z}(x-a)dx}{z^2/2} dQ(z) = \beta \int_{0+}^\infty \frac{z^2/2}{z^2/2} dQ(z) = \beta, \\
f(a) &= \beta \int_{0+}^\infty \frac{z}{z^2/2} dQ(z), \\
-f'_+(a) &= \lim_{\delta_n \downarrow 0} \frac{-f(a+\delta_n)+f(a)}{\delta_n} = \beta \lim_{\delta_n \downarrow 0} \int_0^\infty \frac{z-(z-\delta_n)^+}{\delta_n z^2/2} dQ(z) \\
&= \beta \lim_{\delta_n \downarrow 0} \int_{0+}^\infty \frac{z-(z-\delta_n)^+}{\delta_n z^2/2} dQ(z) = \beta \int_{0+}^\infty \frac{1}{z^2/2} dQ(z).
\end{aligned}
$$

由于 $0 \le \frac{z-(z-\delta_n)^+}{\delta_n z^2/2} \le \frac{z-(z-\delta_{n+1})^+}{\delta_{n+1} z^2/2}$，根据单调收敛定理，可以交换极限与积分。因此，程序（A.12）等价于以下程序，

$$
\begin{aligned}
\max_{Z\sim Q}\ & \beta \mathbb{E}_Q\Big[\frac{H(Z+a)}{Z^2/2}\Big] \\
\text{s.t.}\quad & \beta \mathbb{E}_Q[1] = \beta, \\
& \beta \mathbb{E}_Q\Big[\frac{G(Z+a)}{Z^2/2}\Big] = \Gamma, \\
& \beta \mathbb{E}_Q[\frac{Z}{Z^2/2}] = \eta, \\
& \beta \mathbb{E}_Q[\frac{1}{Z^2/2}] \le \nu, \\
& Q \in \mathscr{M}^+(0,\infty).
\end{aligned}
$$

由于 $\mathbb{E}_Q[\frac{1}{Z^2/2}] < \infty$，我们可以通过 $\frac{dQ}{d\tilde{Q}}(z) = \frac{\beta}{\nu} \frac{1}{z^2/2}$ 定义一个分布函数 $\tilde{Q} \in \mathscr{M}^+(0,\infty)$，它关于 Q 绝对连续。我们将决策变量从 $\tilde{Q}$ 转换为 Q。此外，可将可行集进一步限制为 $\mathscr{P}[0,\infty)$。这是因为函数 $H(z+a)\,G(z+a)\,z, z^2/2$ 按构造在 $z=0$ 处等于 0。因此，我们总可以在 0 处添加任意质量，以达到 $\mathbb{E}_{\tilde{Q}}[1]$ 的上界。进而，通过对测度进行适当归一化，我们可以施加约束 $\mathbb{E}_{\tilde{Q}}[1] = 1$。最后，令 $x = z + a$，即可得到以下程序与之等价：

程序（A.12）：

$$\begin{aligned}\max_{X\sim\tilde{Q}}\ & \nu\mathbb{E}_{\tilde{Q}}[H(X)]\\ \text{s.t.}\ & \nu\mathbb{E}_{\tilde{Q}}\Big[\frac{(X-a)^2}{2}\Big]=\beta,\\ & \nu\mathbb{E}_{\tilde{Q}}[G(X)]=\Gamma,\\ & \nu\mathbb{E}_{\tilde{Q}}[X-a]=\eta,\\ & \tilde{Q}\in\mathscr{P}[a,\infty).\end{aligned}$$

当 $\eta\neq\bar{\eta}$ 时，可以将等式约束 $\nu\mathbf{E}_{\tilde{Q}}[X-a]=\eta$ 替换为 $\eta\leq\nu\mathbf{E}_{\tilde{Q}}[X-a]\leq\bar{\eta}$，从而得到矩形约束。注意，上述推导对于 β 和 Γ 的任意取值均成立，因此将 $\{\mathbf{\Gamma}\}=\{(\beta,\ \Gamma)^\top\}$ 替换为一般的 S 时，结论仍然成立。证明完毕。

□

C.15 定理 4.2 的证明

证明。不失一般性，我们考虑 $r=1$。对于除 1 以外的正 r，可以将 $\mathbf{\Sigma}$ 替换为 $r\mathbf{\Sigma}$，其余推导保持不变。我们可以分别考虑 S 为纯超椭球或纯超矩形的情形，最终结果是两者的组合。对于超椭球情形，我们将 $\mathfrak{M}_a\big(H,\ \boldsymbol{G},\ \mathsf{S}_E(\boldsymbol{\mu},\ \mathbf{\Sigma},\ 1)\big)$重写，其中给定常数值 $\boldsymbol{\mu},\ \mathbf{\Sigma},\ \bar{\boldsymbol{\mu}},\ \boldsymbol{\mu},\ a$ 为

$$\begin{aligned}-\\ \max_{P}\ & \mathbb{E}_P[H(X)]\\ \text{s.t.}\ & \begin{bmatrix}\mathbf{\Sigma}^{-1/2}\big(\mathbb{E}_P[\boldsymbol{G}(X)]-\boldsymbol{\mu}\big)\\ 1\end{bmatrix}\in\mathbb{K},\\ & P\in\mathscr{P}[a,\infty)\end{aligned}\tag{A.22}$$

其中，$\mathbf{\Sigma}^{-1/2}$ 是通过 Cholesky 分解得到的 $\mathbf{\Sigma}^{-1}$ 的下三角平方根矩阵，K 表示二阶锥 $\{(x_1,\ \ldots\ldots,\ x_d,\ x_{d+1})^\top\in\mathrm{R}^{d+1}:\ \ x_{d+1}\geq$

$\sqrt{x_1^2+\cdots+x_d^2}\}$.

对偶问题推导如下：第一步是为该程序 (A.22) 构造拉格朗日函数。注意，对于变量 $\lambda \geq 0,\ \kappa,\ P \in \mathscr{M}^+[a,\infty)$，有如下关系

$$\begin{aligned}
&\mathbb{E}_P[H(X)] + \lambda\big(1 - \|\boldsymbol{\Sigma}^{-1/2}(\mathbb{E}_P[\boldsymbol{G}(X)] - \boldsymbol{\mu})\|_2\big) + \kappa(1 - \mathbb{E}_P[1]) \\
=&\mathbb{E}_P[H(X)] + \lambda\big(1 - \max_{\|\boldsymbol{u}\|_2 \leq 1} \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbb{E}_P[\boldsymbol{G}(X)] - \boldsymbol{\mu})\big) + \kappa(1 - \mathbb{E}_P[1]) \\
=&\mathbb{E}_P[H(X)] + \lambda - \max_{\|\boldsymbol{u}\|_2 \leq \lambda} \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbb{E}_P[\boldsymbol{G}(X)] - \boldsymbol{\mu}) + \kappa(1 - \mathbb{E}_P[1]) \\
=&\min_{\|\boldsymbol{u}\|_2 \leq \lambda} \lambda + \kappa + \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} + \mathbb{E}_P[H(X) - \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{G}(X) - \kappa]
\end{aligned}$$

其中 $\|\cdot\|_2$ 表示 $\mathbb{R}^p$ 中的欧几里得范数。

我们可以写出如下拉格朗日函数：

$$L(P, \lambda, \boldsymbol{u}, \kappa) = \lambda + \kappa + \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} + \mathbb{E}_P[H(X) - \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{G}(X) - \kappa].$$

可以验证，$\mathfrak{P}_a\big(H,\ \boldsymbol{G},\ \mathsf{S}_E(\boldsymbol{\mu},\ \boldsymbol{\Sigma})\big)$ 可由以下内容推导得出

$$\max_{P \in \mathscr{M}^+[a,\infty)} \min_{\|\boldsymbol{u}\|_2 \leq \lambda, \kappa} L(P, \lambda, \boldsymbol{u}, \kappa)$$

相应的对偶问题可由以下内容推导得出

$$\min_{\kappa, \|\boldsymbol{u}\|_2 \leq \lambda} \max_{P \in \mathscr{M}^+[a,\infty)} L(P, \lambda, \boldsymbol{u}, \kappa).$$

接下来，我们将对偶问题写成如下形式：

$$\begin{aligned}
\min_{\kappa, \boldsymbol{u}, \lambda} \quad & \lambda + \kappa + \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} \\
\text{s.t.} \quad & -H(x) + \boldsymbol{u}^\top \boldsymbol{\Sigma}^{-1/2}\boldsymbol{G}(x) + \kappa \geq 0, x \in [a,\infty), \\
& \begin{bmatrix} \boldsymbol{u} \\ \lambda \end{bmatrix} \in \mathbb{K}.
\end{aligned}$$

接着，对于超矩形情形，我们将 $\mathfrak{M}_a\ H, \boldsymbol{G}, \mathsf{S}_R(\boldsymbol{\mu}, \bar{\boldsymbol{\mu}})\big)$ 改写为

$$\begin{aligned}
\max_{X \sim P} \quad & \mathbb{E}_P[H(X)] \\
\text{s.t.} \quad & \mathbb{E}_P[\boldsymbol{G}(X)] \leq \bar{\boldsymbol{\mu}}, \\
& -\mathbb{E}_P[\boldsymbol{G}(X)] \leq -\underline{\boldsymbol{\mu}}, \\
& P \in \mathscr{P}[a,\infty).
\end{aligned}$$

考虑以下拉格朗日函数，其变量为 $P, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \kappa$

$$L(P, \boldsymbol{\lambda_1}, \boldsymbol{\lambda_2}, \kappa) = \mathbb{E}_P[H(X)] + \boldsymbol{\lambda_1}^\top(\bar{\boldsymbol{\mu}} - \mathbb{E}_P[\boldsymbol{G}(X)]) + \boldsymbol{\lambda_2}^\top(-\underline{\boldsymbol{\mu}} + \mathbb{E}_P[\boldsymbol{G}(X)]) + \kappa(1 - \mathbb{E}_P[1])$$

$$= \boldsymbol{\lambda_1}^\top\bar{\boldsymbol{\mu}} - \boldsymbol{\lambda_2}^\top\underline{\boldsymbol{\mu}} + \kappa + \mathbb{E}_P[-\boldsymbol{\lambda_1}^\top\boldsymbol{G}(X) + \boldsymbol{\lambda_2}^\top\boldsymbol{G}(X) - \kappa + H(X)].$$

因此，我们得到如下对偶问题：

$$\begin{aligned} \min_{\boldsymbol{\lambda_1}\geq 0, \boldsymbol{\lambda_2}\geq 0, \kappa} \quad & \boldsymbol{\lambda_1}^\top\bar{\boldsymbol{\mu}} - \boldsymbol{\lambda_2}^\top\underline{\boldsymbol{\mu}} + \kappa \\ \text{s.t.} \quad & -H(x) + (\boldsymbol{\lambda_1} - \boldsymbol{\lambda_2})^\top\boldsymbol{G}(x) + \kappa \geq 0, x \in [a, \infty). \end{aligned}$$

由此得证。 □

C.16 推论 4.1 的证明

证明。 利用定理 4.1，我们可以轻松得到如下矩约束变换：

$$\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r), \mathscr{P}^1_{a,\eta}) \to \mathfrak{M}_a(\eta\tilde{H}, \eta\tilde{\boldsymbol{G}}, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)),$$

$$\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}), \mathscr{P}^1_{a,\eta}) \to \mathfrak{M}_a(\eta\tilde{H}, \eta\tilde{\boldsymbol{G}}, \mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}})),$$

$$\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r), \mathscr{P}^2_{a,\underline{\eta},\bar{\eta},\nu}) \to \mathfrak{M}_a(\nu H, \nu(x-a, \boldsymbol{G}^\top)^\top, \mathbb{S}_R(\underline{\eta}, \bar{\eta}) \times \mathbb{S}_E(\boldsymbol{\mu}, \boldsymbol{\Sigma}, r)),$$

$$\mathfrak{P}(h, \boldsymbol{g}, \mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}}), \mathscr{P}^2_{a,\underline{\eta},\bar{\eta},\nu}) \to \mathfrak{M}_a(\nu H, \nu(x-a, \boldsymbol{G}^\top)^\top, \mathbb{S}_R(\underline{\eta}, \bar{\eta}) \times \mathbb{S}_R(\underline{\boldsymbol{\mu}}, \bar{\boldsymbol{\mu}})).$$

Then using Theorem 4.2, we obtain the results in the corollaries. □

C.17 定理 4.3 的证明

证明。 应用 Bertsimas and Popescu (2005) 中的命题 3.1，约束条件 (30) 等价于存在两个半正定矩阵 $\boldsymbol{V} = [v_{ij}]_{i,j=0,\cdots,k}$ 和

$\boldsymbol{W} = [w_{ij}]_{i,j=0,\cdots,k}$ 使得

$$
\begin{aligned}
&\boldsymbol{V}, \boldsymbol{W} \succeq 0, \\
&\sum_{i,j:i+j=2l-1} v_{ij} = 0, \quad l = 1, \cdots, k, \\
&\sum_{i,j:i+j=2l} v_{ij} = \sum_{r=l}^{k} \binom{r}{l} y_{1r} a^{r-l}, \quad l = 0, \cdots, k, \\
&\sum_{i,j:i+j=2l-1} w_{ij} = 0, \quad l = 1, \cdots, k, \\
&\sum_{i,j:i+j=2l} w_{ij} = \sum_{m=0}^{l} \sum_{r=m}^{k+m-l} \binom{r}{m} \binom{k-r}{l-m} y_{2r} b^{r-m} c^{m}, \quad l = 0, \cdots, k.
\end{aligned}
\tag{A.23}
$$

将 29 替换为 A.23，即可得到结果。 □

C.18 定理 A.1 的证明

证明。 该证明的思路源自 Goberna 等人（2007）的定理 2。在后续证明中，对于任意集合 S，我们用 int(S) 表示 S 的内部，$|S|$ 表示其基数，cone(S) $:=_m\{\sum \lambda_i \boldsymbol{x}_i :\ \boldsymbol{x}_i \in S$ $\lambda_i \in R^+,\ m \in N\}$ 表示其锥包。所有自然数组成的集合为 N。$_{i=1}F^*(\mathfrak{P})$ 表示程序 $\mathfrak{P}$ 的最优区域。我们首先回顾半无限线性规划的一些定义以及稍后将使用的相关结果。更多信息见 Goberna 等人（1981）；Goberna 和 L´opez（2000）。根据 Goberna 等人（1981）的定理 3.1，$\mathfrak{D}_\mu$ 是 FM，当且仅当 特征锥 $K_c(\mathfrak{D}_\mu)$ 是闭集，

$$
K_c(\mathfrak{D}_\mu) := \operatorname{cone}\Big(\big(G_0(x), G_1(x), \ldots, G_n(x), H(x)\big)^\top, x \in \Omega; (\mathbf{0}_{n+1}, -1) \Big).
$$

线性约束系统 $\{\boldsymbol{a}(x)^\top \boldsymbol{y} \geq b(x)\ x \in S\}$ 若满足规范闭，则当 $(1) \exists \boldsymbol{y}^0, \boldsymbol{a}(x)^\top \boldsymbol{y}^0 > b(x)\ \forall x \in S$; $(2)\ \exists \alpha(x) : \Omega \to R^{++}$, 集合 $\{\alpha(x) \cdot (\boldsymbol{a}(x)^\top, b(x))^\top, x \in \Omega\}$ 紧致时成立。 $\mathfrak{D}_\mu$ 是 FM，当约束系统规范闭时（Goberna 等人（1981）的推论 3.1.1）。这是因为规范闭意味着

$$
\operatorname{cone}\Big(\big(G_0(x), G_1(x), \ldots, G_n(x), H(x)\big)^\top, x \in \Omega \Big)
$$

是闭的，这是 $K_c(\mathfrak{D}_\mu)$ 闭性的充分条件。因此我们有

引理 A.1. 程序 $\mathfrak{P}$ 是 *FM*，当且仅当约束系统是规范闭的。

这里讨论 $\mathfrak{P}_{\boldsymbol{\mu}}$ 与 $\mathfrak{D}_{\boldsymbol{\mu}}$之间的最优性和对偶理论。将 $\mathfrak{D}_{\boldsymbol{\mu}}$的一阶矩锥记为 $\mathrm{M}(\mathfrak{D}_{\boldsymbol{\mu}}) := \mathrm{cone}\Big(\big(G_0(x)\ G_1(x)\ldots, G_n(x)\big)^\top, x \in \Omega\Big)$。那么，对于 $\mathrm{F}(\mathfrak{D}_{\boldsymbol{\mu}}) \neq \emptyset$，$\mathrm{F}(\mathfrak{D}_{\boldsymbol{\mu}})$ 有界，当且仅当 $\mathrm{M}(\mathfrak{D}_{\boldsymbol{\mu}}) = \mathrm{R}^{n+1}$ （Goberna 和 L´opez（2000）的定理 9.1）；$\mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})$ 是非空有界集，当且仅当 $\boldsymbol{\mu} \in \mathrm{int}(\mathrm{M}(\mathfrak{D}_{\boldsymbol{\mu}}))$（Goberna 和 L´opez（2000）的推论 9.3.1）。根据 Goberna 等人（1981）的推论 4.1.1，给定任意问题 $\mathfrak{P}$ 及其相应的对偶问题 $\mathfrak{D}$，如果问题$\mathfrak{D}$ 是一个可行 FM，则以下结论成立：(1) $v(\mathfrak{D}) = -\infty$ 当且仅当 $\mathrm{F}(\mathfrak{P}) = \emptyset$；(2) $v(\mathfrak{D}) > -\infty$ 当且仅当 $v(\mathfrak{D}) = v(\mathfrak{P})$。

记作 $Z^* = v(\mathfrak{P}_{\boldsymbol{\mu}})$。假设 $\boldsymbol{dr} \neq \mathbf{0}$ 和 $\rho > 0$以及 $\boldsymbol{dr} = \mathbf{0}$ 或 $\rho = 0$ 均为平凡情形。由于 $\boldsymbol{\mu} \in \mathrm{int}\big(\mathrm{M}(\mathfrak{D}_{\boldsymbol{\mu}})\big)$，$\mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})$ 有界。结合 $\mathfrak{D}_{\boldsymbol{\mu}}$ 是一个 FM 系统这一事实，我们得到 $\mathfrak{P}_{\boldsymbol{\mu}}$ 和 $\mathfrak{D}_{\boldsymbol{\mu}}$ 均有解，且强对偶性成立（参见 Goberna 等人（2007）中定理 2 后的第二条注释）。

此外，

$$v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) \leq \min\{\sum_{j\in\{0\}\cup[d]} y_j(\mu_j + \rho\boldsymbol{dr}_j) | \boldsymbol{y} \in \mathbb{F}(\mathfrak{D}_{\boldsymbol{\mu}})\} \leq \min\{\sum_{j\in\{0\}\cup[d]} y_j(\mu_j + \rho\boldsymbol{dr}_j) | \boldsymbol{y} \in \mathbb{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})\}$$

$$= v(\mathfrak{D}_{\boldsymbol{\mu}}) + \rho\min\{\boldsymbol{dr}^\top\boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})\} = v(\mathfrak{P}_{\mu}) + \rho\min\{\boldsymbol{dr}^\top\boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})\} \qquad (A.24)$$

其中，第一个不等式基于弱对偶，最后一个等式则由 $\mathfrak{P}_{\mu}$ 与 $\mathfrak{D}_{\mu}$之间的强对偶性得出。考虑辅助问题 $\mathfrak{P}'$ 及其对应的对偶问题 $\mathfrak{D}'$ 如下。

$$
\begin{aligned}
(\mathfrak{P}'): \max_{P,\pi}\ & \mathbb{E}_P[H(X)] + Z^*\pi && (\mathfrak{D}'): \min_{\boldsymbol{y}}\ && \boldsymbol{dr}^\top \boldsymbol{y} \\
\text{s.t.}\ & \mathbb{E}_P[G_j(X)] + \mu_j \pi = \boldsymbol{dr}_j, \forall j \in \{0\}\cup[d], && \text{s.t.}\ && \sum_{\{0\}\cup[d]} y_j G_j(x) \geq H(x),\ \ \forall x \in \mathbb{\Omega}, \\
& P \in \mathscr{M}^+(\mathbb{\Omega}). && && \sum_{\{0\}\cup[d]} y_j \mu_j = Z^*.
\end{aligned}
$$

由于 $\mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})$ 非空且有界，并且是 $\mathfrak{D}'$的可行域，因此 $\mathrm{M}(\mathfrak{D}') = \mathrm{R}^{n+1}$，这意味着每个 $\boldsymbol{dr} \in \mathrm{int}\big(\mathrm{M}(\mathfrak{D}')\big)$，从而强对偶性成立。此外，$\mathfrak{P}'$ 可解，因为作为决策变量的概率测度具有紧支撑。上述两个论证表明，$\boldsymbol{dr}^\top \boldsymbol{y}$ 可在可行域 $\mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})$ 上取得其最小值。设 $(P, \bar{\pi})$ 为 $\mathfrak{P}'$的一个最优解，P^*为 $\mathfrak{P}_{\boldsymbol{\mu}}$的一个最优解。我们证明，对于任意充分小的 $\rho > 0$，$P^* + \rho(P + \bar{\pi}P^*)$ 是 $\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}$的可行解。若 $\bar{\pi} \geq 0$，ρ可以是任意正数。若 $\bar{\pi} < 0$，$\rho \leq -\frac{1}{\bar{\pi}}$。为¯说明这一点，(1) $P^* + \rho(P + \bar{\pi}P^*) \in \mathcal{M}^+(\Omega)$ 且 (2). $\mathrm{E}_{P^*+\rho(\bar{P}+\bar{\pi}P^*)}[G_j(X)] = \mathrm{E}_{P^*}[G_j(X)] + \mathrm{E}_{\rho(\bar{P}+\bar{\pi}P^*)}[G_j(X)] = \mathrm{E}_{P^*}[G_j(X)] + \rho\mathrm{E}_{\bar{P}}[G_j(X)] + \rho\bar{\pi}\mathrm{E}_{P^*}[G_j(X)] = \mu_j + \rho(\boldsymbol{dr}_j - \mu_j\bar{\pi} + \bar{\pi}\mu_j) = \mu_j + \rho\boldsymbol{dr}_j, j \in \{0\}\cup[d]$。

–

利用 $\mathfrak{P}'$ 和 $\mathfrak{D}'$的强对偶性，我们有

$$
\begin{aligned}
v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) &\geq \mathbb{E}_{P^*+\rho(\bar{P}+\bar{\pi}P^*)}[H(X)] \\
&= \mathbb{E}_{P^*}[H(X)] + \rho\mathbb{E}_{\bar{P}}[H(X)] + \rho\bar{\pi}\mathbb{E}_{P^*}[H(X)] \\
&= v(\mathfrak{P}_{\boldsymbol{\mu}}) + \rho\mathbb{E}_{\bar{P}}[H(X)] + \rho\bar{\pi}Z^* = v(\mathfrak{P}_{\boldsymbol{\mu}}) + v(\mathfrak{P}') \\
&= v(\mathfrak{P}_{\boldsymbol{\mu}}) + \rho\min\{\boldsymbol{dr}^\top \boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})\}. \qquad \text{(A.25)}
\end{aligned}
$$

结合 (A.24) 和 (A.25)，我们有

$v(\mathfrak{P}_{\mu}) + \rho\mathrm{min}\{\boldsymbol{dr}^\top \boldsymbol{y} | \boldsymbol{y} \in \mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})\} \geq v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) \geq v(\mathfrak{P}_{\boldsymbol{\mu}}) + \rho\mathrm{min}\{\boldsymbol{dr}^\top \boldsymbol{y} | \boldsymbol{y} \in \mathrm{F}^*(\mathfrak{D}_{\boldsymbol{\mu}})$.}于是有

$$
v(\mathfrak{P}_{\boldsymbol{\mu}+\rho\boldsymbol{dr}}) = v(\mathfrak{P}_{\mu}) + \rho\min\{\boldsymbol{dr}^\top \boldsymbol{y} | \boldsymbol{y} \in \mathbb{F}^*(\mathfrak{D}_{\boldsymbol{\mu}}), \}
$$

至此，证明结束。 □

C.19 一些有用的定理和引理

引理 A.2。 给定 $\alpha \in \mathbb{R}$，若 (i) $\int_a^\infty x^\alpha f(x)dx = \Gamma$，(ii) $f(x)$ 关于 $x \geq a$ 非递增，(iii) $f(x) \geq 0, \forall x \geq a$，则有 $x^{\alpha+1}f(x) \to 0$，当 $x \to \infty$。

除上述条件外，若 $f'_+(x)$ 是非递减的，且对 $x \geq a$为非正，并且 $f(x) = \int_a^x f'_+(t)dt + \eta$，则当 $x \to \infty$时，我们有 $x^{\alpha+2}f'_+(x) \to 0$。

引理 *A.2* 的证明。 对于 $\alpha \leq 0$， x^α 成为有界函数，因此很容易得到这些结果。

现在我们假设 $\alpha > 0$。考虑函数

$$T(x) = x^{\alpha+1}f(x) - (\alpha+1)\int_a^x t^\alpha f(t)dt.$$

对于任意 $(a \vee 0) \leq x_1 \leq x_2$,

$$\begin{aligned} T(x_2) - T(x_1) &= x_2^{\alpha+1}f(x_2) - x_1^{\alpha+1}f(x_1) - \int_{x_1}^{x_2} f(t)dt^{\alpha+1} \\ &\leq x_2^{\alpha+1}f(x_2) - x_1^{\alpha+1}f(x_1) - f(x_2)(x_2^{\alpha+1} - x_1^{\alpha+1}) \\ &= -x_1^{\alpha+1}f(x_1) + x_1^{\alpha+1}f(x_2) \\ &= x_1^{\alpha+1}(f(x_2) - f(x_1)) \leq 0. \end{aligned}$$

由上可知， $T(x)$ 关于 $x \geq a \vee 0$ 非递增。由于对于 $x \geq a \vee 0$， $x^{\alpha+1}f(x) \geq 0$和 $0 \leq \int_a^x t^\alpha f(t)dt \leq \Gamma$，因此 $T(x)$ 有下界，进而收敛于某个极限。鉴于当 $x \to \infty$ 时，$\int_a^x t^\alpha f(t)dt \to \Gamma$，我们必有 $x^{\alpha+1}f(x)$ 趋于某个有限的非负值。若对所有足够大的 x，$x^{\alpha+1}f(x) \geq \epsilon > 0$，则对所有足够大的 x，$x^\alpha f(x) \geq \frac{\epsilon}{x}$，因此 $\int_a^\infty x^\alpha f(x)dx = \infty$。这与上述条件 (ii) 矛盾。因此， $x^{\alpha+1}f(x)$ 必须收敛于 0。

–

为证明第二部分，考虑一个函数

$$P(x) = -x^{\alpha+2}f'_+(x) + (\alpha+2)[-\int_x^\infty t^{\alpha+1}f'_+(t)dt].$$

注意，由于 f 绝对连续，且正如我们刚刚证明的，$\lim_{x\to\infty} x^{\alpha+1}f(x) \to 0$，分部积分可得

$$-\int_x^\infty t^{\alpha+1}f'_+(t)dt = x^{\alpha+1}f(x) + (\alpha+1)\int_x^\infty t^\alpha f(t)dt < \infty, \forall x \geq (a\vee 0).$$

对于任意 $(a\vee 0) \leq x_1 \leq x_2$,

$$\begin{aligned} P(x_2) - P(x_1) &= x_1^{\alpha+2}f'_+(x_1) - x_2^{\alpha+2}f'_+(x_2) + \int_{x_1}^{x_2} f'_+(t)dt^{\alpha+2} \\ &\leq x_1^{\alpha+2}f'_+(x_1) - x_2^{\alpha+2}f'_+(x_2) + f'_+(x_2)(x_2^{\alpha+2} - x_1^{\alpha+2}) \\ &= x_1^{\alpha+2}(f'_+(x_1) - f'_+(x_2)) \leq 0 \end{aligned}$$

由上可知，对于 $x \geq (a\vee 0)$， $P(x)$ 是非递增的。由于 $-x^{\alpha+2}f'_+(x) \geq 0$ 对于 $x \geq (a\vee 0)$，且 $-\int_x^\infty t^\alpha f'_+(t)dt$ 是非负的，因此 $P(x)$ 有下界，并收敛到一个极限。因为 $\lim_{x\to\infty} x^{\alpha+1}f(x) \to 0$ 且 $\lim_{x\to\infty}\int_x^\infty f(t)t^\alpha dt = 0$，所以 $\lim_{x\to\infty}(-\int_x^\infty x^{\alpha+1}f'_+(t)dt) = 0$。这意味着，当 $x\to\infty$ 时， $-x^{\alpha+2}f'_+(x)$ 必须收敛到某个有限的非负值。

若对于所有充分大的 x， $-x^{\alpha+2}f'_+(x) \geq \epsilon > 0$。这意味着对于所有充分大的 x, $-x^{\alpha+1}f'_+(x) \geq \frac{\epsilon}{x}$，因此当 $x \geq a$ 时， $-\int_x^\infty t^{\alpha+1}f'_+(t)dt = \infty$。这与上文所述 $-t^{\alpha+1}f'_+(t)$ 的积分有限相矛盾。因此， $-x^{\alpha+1}f'_+(x)$ 必须收敛到 0。

□

引理 *A.3.* 给定函数 $g(x)$，其在 $x \geq a$ 上有下界。若 *(i)* $\int_a^\infty g(x)f(x)dx = \Gamma$，*(ii)* $f(x)$ 在 $x \geq a$ 上非递增，*(iii)* $f(x) \geq 0, \forall x \geq a$ 是密度函数，则当 $x\to\infty$ 时，我们有 $\tilde{G}(x)f(x) \to 0$，其中 $\tilde{G}(x) = \int_a^x g(u)du$。

除了上述条件外，若 $f'_+(x)$ 是非递减的，且对 $x \geq a$ 非正，并且 $f(x) = \int_a^x f'_+(t)dt + \eta$，则有 $G(x)f'_+(x) \to 0$，其中 $x\to\infty$，且 $G(x) = \int_a^x \tilde{G}(u)du$。引理 *A.3* 的证明。我们首先假设函数 $g(x)$ 对 $x \geq a$ 是非负函数。

考虑函数

$$T(x) = \tilde{G}(x)f(x) - \int_a^x g(t)f(t)dt.$$

对于任意 $a \le x_1 \le x_2$,

$$\begin{aligned} T(x_2) - T(x_1) &= \tilde{G}(x_2)f(x_2) - \tilde{G}(x_1)f(x_1) - \int_{x_1}^{x_2} g(t)f(t)dt \\ &= \tilde{G}(x_2)f(x_2) - \tilde{G}(x_1)f(x_1) - \int_{x_1}^{x_2} f(t)d\tilde{G}(t) \\ &\le \tilde{G}(x_2)f(x_2) - \tilde{G}(x_1)f(x_1) - f(x_2)(\tilde{G}(x_2) - \tilde{G}(x_1)) \\ &= -\tilde{G}(x_1)f(x_1) + f(x_2)\tilde{G}(x_1) \\ &= \tilde{G}(x_1)(f(x_2) - f(x_1)) \le 0. \end{aligned}$$

˜由上可知，对于 $x \ge a$， $T(x)$ 是非递增的。由于对于 $x \ge a$， $G(x)f(x) \ge 0$ 且$0 \le \int_a^x g(t)f(t)dt \le \Gamma$，因此 $T(x)$ 有下界并收敛到一个极限。由于当 $x \to \infty$ 时，$\int_a^x g(t)f(t)dt \to \Gamma$，所以当 $x \to \infty$ 时， $\tilde{G}(x)f(x)$ 必须趋于某个有限的非负˜值。现在我们证明，当 $x \to \infty$ 时， $G(x)f(x) \to 0$ 。

˜首先，由于 $G(x)$ 是非负且非递增函数，因此我们必有˜$G(x)$ 要么发散，要么收敛到某个正的有限值。我们先考虑以下情形：当 $x \to \infty$ 时， $G(x)$ 收敛到某个正的有限值。由于$x \to \infty$时 $f(x)$ 趋于零（事实上， $f(x) \ge 0$， $f(x)$ 对于 $x \ge a$ 是非递增的，且 $f(x)$ 是一个密度函数），因此˜当 $x \to \infty$ 时，我们有 $G(x)f(x) \to 0$ 。

˜现在考虑第二种情形，即当 $x \to \infty$ 时， $G(x)$ 发散。我们用反证法证明。由于对于 $x \ge a$， $G(x)f(x)$ 是非负的，我们假设当 $x \to \infty$ 时， $G(x)f(x)$ 将收敛到一个正值。这意味着存在 $\epsilon > 0$ 和 $x_\epsilon > a$ 使得˜$G(x)f(x) \ge \epsilon$， $\forall x \ge x_\epsilon$

因此，我们有

$$\begin{aligned} \int_{x_\epsilon}^{\infty} f(x)g(x)dx &\ge \int_{x_\epsilon}^{\infty} \frac{\epsilon}{\tilde{G}(x)} g(x)dx \\ &= \int_{x_\epsilon}^{\infty} \frac{\epsilon}{\tilde{G}(x)} d\tilde{G}(x) \\ &= \int_{x_\epsilon}^{\infty} \epsilon \, d\ln \tilde{G}(x) \end{aligned}$$

$$= \epsilon \ln \tilde{G}(x)|_{\infty} - \epsilon \ln \tilde{G}(x_\epsilon) = \infty$$

然而，我们有

$$\int_{x_\epsilon}^{\infty} f(x)g(x)dx \leq \int_{a}^{\infty} f(x)g(x)dx = \Gamma < \infty.$$

˜我们得出矛盾。因此，当 $x \to \infty$ 时，必有 $G(x)f(x) \to 0$ 。

为证明第二部分，我们考虑函数

$$P(x) = -G(x)f'_+(x) + [-\int_{x}^{\infty} \tilde{G}(t)f'_+(t)dt].$$

˜注意，由于 f 绝对连续，且我们刚刚证明了 $\lim_{x\to\infty} G(x)f(x) \to 0$，通过分部积分可得

$$-\int_{x}^{\infty} \tilde{G}(t)f'_+(t)dt = \tilde{G}(x)f(x) + \int_{x}^{\infty} f(t)g(t)dt < \infty.$$

对于任意 $a \leq x_1 \leq x_2$，我们有

$$\begin{aligned} P(x_2) - P(x_1) &= -G(x_2)f'_+(x_2) + G(x_1)f'_+(x_1) + [\int_{x_1}^{x_2} \tilde{G}(t)f'_+(t)dt] \\ &= -G(x_2)f'_+(x_2) + G(x_1)f'_+(x_1) + f'_+(x_2)(G(x_2) - G(x_1)) \\ &= G(x_1)(f'_+(x_1) - f'_+(x_2)) \leq 0. \end{aligned}$$

由上可知，对于 $x \geq a$， $P(x)$ 非递增。由于 $-G(x)f'_+(x) \geq 0$ 且 $-\int_x^\infty \tilde{G}(t)f'_+(t)dt \geq 0$ 对于 $x \geq a$，我们有 $P(x)$ 有下界并收敛到一个极限。因为 $\lim_{x\to\infty} \tilde{G}(x)f(x)$ 和 $\lim_{x\to\infty} \int_x^\infty f(t)g(t)dt$ 均等于零，所以 $\lim_{x\to\infty}(-\int_x^\infty \tilde{G}(t)f'_+(t)dt) = 0$。这意味着，当 $x \to \infty$ 时， $-G(x)f'_+(x)$ 必须趋于某个有限的非负值。现在我们证明，当 $x \to \infty$ 时， $-G(x)f'_+(x) \to 0$ 。

首先，由于 $G(x)$是非负且非递减的函数，因此$G(x)$必定会发散至无穷大，或收敛至某个正的有限值。我们先考虑 $G(x)$随着 $x \to \infty$趋于某个正的有限值的情形。假设 $f'_+(x) \not\to 0$。那么，根据 $f'_+(x)$ 的非正性和非递减性，我们有 $f'_+(x) \to c < 0$ 随着 $x \to \infty$。然而， $f(x) = \int_a^x f'_+(t)dt + \eta \to -\infty$ 随着 $x \to \infty$，这违反了 $f(x)$的非负条件。因此， $f'_+(x)$随着 $x \to \infty$趋于零，从而意味着

$-G(x)f'_+(x) \to 0$ 作为 $x \to \infty$。

现在考虑第二种情形：当 $G(x)$ 随着 $x \to \infty$ 趋于无穷时。我们用反证法证明。由于 $-G(x)f'_+(x)$ 对 $x \geq 0$ 非负，我们假设 $-G(x)f'_+(x)$ 随着 $x \to \infty$ 趋于某个正值。这意味着存在 $\epsilon > 0$ 和 $x_\epsilon > a$，使得 $-G(x)f'_+(x) \geq \epsilon, \forall x \geq x_\epsilon$

因此，我们有

$$\begin{aligned}\int_{x_\epsilon}^{\infty} -f'_+(x)\tilde{G}(x)dx &\geq \int_{x_\epsilon}^{\infty} \frac{\epsilon}{G(x)}\tilde{G}(x)dx \\ &= \int_{x_\epsilon}^{\infty} \frac{\epsilon}{G(x)}dG(x) \\ &= \int_{x_\epsilon}^{\infty} \epsilon \; d\ln G(x) \\ &= \epsilon \ln G(x)|_\infty - \epsilon \ln G(x_\epsilon) = \infty.\end{aligned}$$

然而，我们有

$$\int_{x_\epsilon}^{\infty} -f'_+(x)\tilde{G}(x)dx = \tilde{G}(x_\epsilon)f(x_\epsilon) + \int_{x_\epsilon}^{\infty} f(t)g(t)dt < \infty.$$

W得到矛盾。因此，当 $x \to \infty$ 时，我们必有 $G(x)f'_+(x) \to 0$。

现在考虑这样的情形：对 $x \geq a$， $g(x)$ 是有下界函数，且 $g(x)$ 的值可以为负。考虑 $\tilde{g}(x) = g(x) + |\min_{x \geq a} g(x)|$。显然， $\tilde{g}(x)$ 和$|\min_{x \geq a} g(x)|$ 是非负函数，因此可以使用上文的结果。由积分的线性性和极限的和法则，可得关于 $g(x)$ 的结果。

□

引理 A.4. 如果$\mathbb{E}[G(X)] = \beta$，其中 β 是某个有限常数，且 $G(x) : \mathbb{R} \to \mathbb{R}$ 是非递减函数，并且当 $x \geq a$ 时有下界，界为某个常数 a。对于任意$P \in \mathscr{P}[a, \infty)$，我们有 $\mathbb{E}_P[\mathbb{I}(X \geq x)]G(x) \to 0$，当 $x \to \infty$时。

引理 *A.4* 的证明。 我们首先假设 $G(x)$ 是一个非负函数。根据 $\mathrm{E}[G(X)] = \beta$，不难得到 $\mathrm{E}[G(X)\mathrm{I}(X \geq x)] \to 0$ as $x \to \infty$。于是，$\mathrm{E}[\mathrm{I}(X \geq x)]G(x) \leq \mathrm{E}[G(X)\mathrm{I}(X \geq x)] \to 0$ as $x \to \infty$。第一个不等式由

$G(x)$ 的非递减性质得出。并且，根据 $G(x)$ 的非负性质，我们有 $\mathrm{E}[\mathrm{I}(X \geq x)]G(x) \geq 0$。这意味着 $\mathrm{E}[\mathrm{I}(X \geq x)]G(x) \to 0$ as $x \to \infty$。

现在考虑 $G(x)$ 是有下界函数且 $G(x)\tilde{}$ 的值可能为负的情形。我们考虑 $G(x) = G(x) + |\min_{x \geq a} G(x)|$。这里，最小值是$\tilde{}$相对于概率空间的支撑集取得的。显然，$G(x)$ 和 $|\min_{x \geq a} G(x)|$ 是非负函数，因此我们可以使用上述结果。由极限的和法则可得 $G(x)$ 的结果。已知 $\lim_{x \to \infty} \mathrm{E}[\mathrm{I}(X \geq x)][G(x) + |\min_{x \geq a} G(x)|] = 0$ 且 $\lim_{x \to \infty} \mathrm{E}[\mathrm{I}(X \geq x)][-|\min_{x \geq a} G(x)|] = 0$，因此有 $\lim_{x \to \infty} \mathrm{E}[\mathrm{I}(X \geq x)]G(x) = 0$。

□

定理 A.1。以下两个陈述等价:

(1). 函数 f 在 $x \geq a$ 上有限、非递增且右连续，并且 $\int_a^\infty f(x)dx = \beta$。

(2). $f(x)$ $x \geq a$ 是示性函数 $\mathbb{I}(a \leq x < a+z)\, z$, $z > \infty$的广义混合，即: ————

$$f(x) = \beta \int_{0+}^{\infty} \frac{\mathbb{I}(a \leq x < a+z)}{z} dQ(z)$$

其中 Q 是 $(0, \infty)$上的概率测度。

定理 *A.1* 的证明。$(1) \implies (2)$:

定义的概率分布函数 $F(x) := 1 - \beta + \int_a^x f(x)dx, \forall x \geq a\tilde{}$ 对于 $x \geq a$ 绝对连续。我们考虑另一个概率分布 F，其中 $\tilde{F}(x) = {}^{F}(x)_{\beta}^{+\beta-1}, x \geq a$ 且 $F(x) = 0, x < a$。那么 $\tilde{F}$ 在 $x \geq a$ 上仍然绝对连续，且密度 $\tilde{f}(x)$ 在 $x < a$ 时为零，而 ${}^{f}(x)\,{}_{\beta}$ 对于 $x \geq a$。因此，$\tilde{F}$ 在点 a 处为单峰。此外，$\tilde{F}(x) = \int_a^x \tilde{f}(s)ds$。对于 $x \geq a$，$\tilde{F}(x)$ 的右导数存在。我们还注意到，$\tilde{f}(x+\delta) \leq {}^{\int_x^{x+\delta} \tilde{f}}(\,{}_{s\ ds} \leq \tilde{f})\,(x)$。由于 f 对于 $x \geq a$ 右连续，因此有 $\tilde{f}(x) = \lim_{\delta \downarrow 0} \tilde{f}(x+\delta) \leq \lim_{\delta \downarrow 0} {}^{\int_x^{x+\delta} \tilde{f}}(\,{}_{\delta}^{s\ ds} \leq \tilde{f})\,(x)$。因此，

————

————

$$\tilde{F}'_+(x) = \lim_{\delta \downarrow 0} \frac{\tilde{F}(x+\delta) - \tilde{F}(x)}{\delta} = \lim_{\delta \downarrow 0} \frac{\int_x^{x+\delta} \tilde{f}(s)ds}{\delta} = \tilde{f}(x).$$

˜根据 Dharmadhikari 和 Joag-Dev (1988) 的定理 1.2， F 是分布函数 $W_{a,\ z}$ 的广义混合，其中 $W_{a,\ z}$ 表示区间 $(a,\ a+z)$ $z>0$ 上的均匀分布函数˜。我们不考虑 $z\leq 0$ 的原因是，对于 $x<a$˜， $F(x)=0$，且 F 在 $x\geq a$ 上绝对连续。特别地，我们有

$$\tilde{F}(x)=\int_{0+}^{\infty}W_{a,z}(x)dQ(z)$$

其中 Q 是定义在 $(0,\infty)$ 上的概率测度。

于是我们有

$$\tilde{f}(x)=\tilde{F}'_{+}(x)=\lim_{\delta_n\downarrow 0}\frac{\tilde{F}(x+\delta_n)-\tilde{F}(x)}{\delta_n}=\lim_{\delta_n\downarrow 0}\int_{0+}^{\infty}\frac{W_{a,z}(x+\delta_n)-W_{a,z}(x)}{\delta_n}dQ(z) \tag{A.26}$$

$$=\lim_{\delta_n\downarrow 0}\int_{0+}^{(x-a)-}\frac{W_{a,z}(x+\delta_n)-W_{a,z}(x)}{\delta_n}dQ(z)$$
$$+\lim_{\delta_n\downarrow 0}\frac{W_{a,z}(x+\delta_n)-W_{a,z}(x)}{\delta_n}Q(x-a)+\lim_{\delta_n\downarrow 0}\int_{(x-a)_+}^{\infty}\frac{W_{a,z}(x+\delta_n)-W_{a,z}(x)}{\delta_n}dQ(z) \tag{A.27}$$

$$=\lim_{\delta_n\downarrow 0}\frac{W_{a,x-a}(x+\delta_n)-W_{a,x-a}(x)}{\delta_n}Q(z=x-a)+\int_{(x-a)}^{\infty}\frac{1}{z}dQ(z), \tag{A.28}$$

并且 $\lim_{\delta_n\downarrow 0}\frac{W_{a,x-a}(x+\delta_n)-W_{a,\delta_n}}{\delta_n}=0_{(x)}$。因此我们有 $\tilde{f}(x)=\int_{(x-a)}^{\infty}\frac{1}{z}dQ(z)$。

交换极限与求和，可由以下观察得出：等式 (A.26) 中的极限以及 (A.27) 中的前两项均存在。由于 $0\leq\frac{W_{a,z}(x+\delta_n)-W_{a,\delta_n}}{\delta_n}\leq_{(x)}\frac{W_{a,z}(x+\delta_{n+1})-W_{a,z}(x)}{\delta_{n+1}}$，等式 (A.28) 中交换极限与积分可由单调收敛定理得出。

进而可得

$$\int_{0+}^{\infty}\frac{\mathbb{I}(a\leq x<a+z)}{z}dQ(z)=\int_{(x-a)_+}^{\infty}\frac{1}{z}dQ(z)=\tilde{f}(x)$$
$$\Longrightarrow\quad f(x)=\beta\int_{0+}^{\infty}\frac{\mathbb{I}(a\leq x<a+z)}{z}dQ(z)$$

(2) $\Longrightarrow$ (1)：

由于

$$\int_a^{\infty}f(x)dx=\int_a^{\infty}\beta\int_{0+}^{\infty}\frac{\mathbb{I}(a\leq x<a+z)}{z}dQ(z)dx=\beta\int_{0+}^{\infty}\int_a^{\infty}\frac{\mathbb{I}(a\leq x<a+z)}{z}dxdQ(z)$$

$$= \beta \int_{0+}^{\infty} 1 dx dQ(z) = \beta.$$

对于任意 $x_2 \geq x_1 \geq a$，我们有 $\forall z > 0$，$\mathbb{I}(a \leq x_1 < a + z) \geq \mathbb{I}(a \leq x_2 < a + z)$。容易得到

$$f(x_1) = \beta \int_{0+}^{\infty} \frac{\mathbb{I}(a \leq x_1 < a + z)}{z} dQ(z) \geq \beta \int_{0+}^{\infty} \frac{\mathbb{I}(a \leq x_2 < a + z)}{z} dQ(z) = f(x_2).$$

最后，$f(x)$ 关于 $x \geq a$ 右连续，因为 $\forall x \geq a$,

$$\begin{aligned}
\lim_{\delta \downarrow 0} f(x + \delta) &= \lim_{\delta \downarrow 0} \beta \int_{0+}^{\infty} \frac{\mathbb{I}(a \leq x + \delta < a + z)}{z} dQ(z) \\
&= \beta \int_{0+}^{\infty} \lim_{\delta \downarrow 0} \frac{\mathbb{I}(a \leq x + \delta < a + z)}{z} dQ(z) \\
&= \beta \int_{0+}^{\infty} \frac{\mathbb{I}(a \leq x < a + z)}{z} dQ(z) = f(x).
\end{aligned}$$

由于 $\mathbb{I}(a \leq x < a + z)$ 是右连续函数，且 $\left| \frac{\mathbb{I}(a \leq x < a + z)}{z} \right|$ 被 Q-可积函数 $\frac{1}{z}$ 所界定。$\frac{1}{z}$ 的可积性源于以下事实：$f(a) = \beta \int_{0+}^{\infty} \frac{1}{z} dQ(z) < \infty$。积分与极限的交换依据控制收敛定理。

□