第 7 章 鲁棒预测与优化（Joint Estimation and Robustness Optimization）

朱桃增

7.1 鲁棒性优化（Robustness optimization）

在建模求解实际问题过程中，我们通常需要预测一些参数或者未知变量和优化模型。由于联合预测和优化（joint prediction and optimization）问题的复杂性，先预测后优化（predict-then-optimize）成为解决这类问题常用的方法。因此，决策者往往需要先使用历史数据估计模型所需参数，然后求解以下优化模型：

$$
\begin{aligned}
\hat{Z} = \min \quad & a_0(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) \\
\text{s.t.} \quad & a_i(\boldsymbol{x}, \hat{\boldsymbol{\beta}}) \le \tau \quad \forall i \in [I], \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
$$

其中，$\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^N$ 为决策变量，$\hat{\boldsymbol{\beta}}$ 为模型的输入参数，$a_i : \bar{\mathcal{X}} \times \bar{\mathcal{W}} \to \mathbb{R}$ 为定义在 $\bar{\mathcal{X}} \times \bar{\mathcal{W}} \subseteq \mathbb{R}^N \times \mathbb{R}^M$ 上的函数。本章着重于讨论当 a_i 是在 $\bar{\mathcal{X}} \times \bar{\mathcal{W}}$ 上的鞍函数（saddle function）的情形。有关鞍函数的定义如下：

定义 7.1: 鞍函数 (Saddle function)

当一个函数 $a : \bar{\mathcal{X}} \times \bar{\mathcal{W}} \to \mathbb{R}$ 满足以下性质时：

- 给定 $\boldsymbol{x} \in \bar{\mathcal{X}}$, $a(\boldsymbol{x}, \boldsymbol{\beta})$ 是关于 $\boldsymbol{\beta}$ 的凹函数，
- 给定 $\boldsymbol{\beta} \in \bar{\mathcal{W}}$，$a(\boldsymbol{x}, \boldsymbol{\beta})$ 是关于 $\boldsymbol{x}$ 的凸函数，

称之为定义在 $\bar{\mathcal{X}} \times \bar{\mathcal{W}}$ 上的鞍函数。♣

通常，我们可以通过统计方法得到上述优化模型的参数 $\hat{\boldsymbol{\beta}}$。然而，参数的估计值与真实值之间常常存在差异，从而可能导致优化问题的目标函数值严重偏离决策者可承受的范围。更有甚者，这会导致某些硬性约束条件不再满足以至于所得解不再具有可行性。有鉴于此，传统鲁棒优化考虑了如下优化模型：

$$
\begin{aligned}
Z_R(r) = \min \quad & \tau \\
\text{s.t.} \quad & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau \quad \forall \boldsymbol{\beta} \in \mathcal{U}(r), \\
& a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{U}(r),\ i \in [I], \\
& \boldsymbol{x} \in \mathcal{X}.
\end{aligned}
\tag{7.1}
$$

此处，$\mathcal{U}(r)$ 是参数 $\boldsymbol{\beta}$ 的不确定集，通常可以定义为

$$\mathcal{U}(r) \triangleq \left\{ \boldsymbol{\beta} \in \bar{\mathcal{W}} \;\middle|\; \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| \le r \right\}.$$

如何确定此集合的大小，也即 r 的值，是鲁棒优化面临的一个难题。虽然参数不确定集的大小和概率约束存在一定的联系，但是概率边界极度依赖于概率分布的假设以及函数 $a_i(\cdot,\cdot)$ 的形式。Zhu et al. (2021) 提出了如下鲁棒性优化模型：

$$\begin{aligned} Z_S(\tau_0) = \max \quad & r \\ \text{s.t.} \quad & a_0(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_0 && \forall \boldsymbol{\beta} \in \mathcal{U}(r), \\ & a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i && \forall \boldsymbol{\beta} \in \mathcal{U}(r),\, i \in [I], \\ & \boldsymbol{x} \in \mathcal{X},\; r \ge 0. \end{aligned} \tag{7.2}$$

此模型旨在提供一个能够使得满足条件的参数不确定集尽可能的大的解，从而降低参数不在所考虑集合中的可能性。与此同时，此解需要让成本控制在决策者设定的目标成本 $\tau_0(\le \hat{Z})$ 之内。由于参数不确定集的大小关于 r 是单调递增的，我们可以采用二分法来求解模型(7.2)。

实际上决策者可以根据自身偏好来改变参数不确定集的大小从而选择问题的解。然而，对于目标导向（target-oriented）类型的决策者，鲁棒性优化的方法则更具有吸引力。虽然在规范性分析（prescriptive analytics）中这种决策准则并不多见，但已有实证研究表明目标（target）在决策过程中起着至关重要的作用 (Payne et al., 1980, 1981; Merchant and Manzoni, 1989; Kőszegi and Rabin, 2006)。有关鲁棒性优化的起源以及相关介绍，读者可参考 Jaillet et al. (2022) 及其中的相关文献。

7.2 基于参数估计的鲁棒性优化

在给定数据集 $\mathcal{D}$ 情况下，参数 $\hat{\boldsymbol{\beta}}$ 可以通过求解如下优化问题得到：

$$\begin{aligned} \hat{\boldsymbol{\beta}} = \arg\min \quad & \rho(\boldsymbol{\beta}; \mathcal{D}) \\ \text{s.t.} \quad & \boldsymbol{\beta} \in \mathcal{W}. \end{aligned}$$

其中，$\rho(\boldsymbol{\beta}; \mathcal{D})$ 被称作为估计方法（estimation metric）。例如，在最小二乘估计中，$\rho(\boldsymbol{\beta}; \mathcal{D})$ 为残差（即观测值与预测值之间的差距）平方总和。此外，假定 $\mathcal{W}$ 是一个凸集（convex set）且其相对内点集（relative interior）非空，即 $\mathrm{relint}(\mathcal{W}) \neq \emptyset$。

7.2.1 基于参数估计的参数不确定集（Estimate uncertainty set）及其统计意义

定义 7.2: (Estimate Uncertainty Set)

基于参数估计方法，我们定义基于参数估计的参数不确定集：

$$\mathcal{E}(r;\mathcal{D}) \triangleq \left\{\boldsymbol{\beta} \in \mathcal{W} \,\middle|\, \rho(\boldsymbol{\beta};\mathcal{D}) \le \hat{\rho} + r\right\}. \tag{7.3}$$

其中，$\hat{\rho} = \rho(\hat{\boldsymbol{\beta}};\mathcal{D})$ 是估计方法的最优值，$r \ge 0$ 为 $\rho(\boldsymbol{\beta};\mathcal{D})$ 与最优值 $\hat{\rho}$ 之间的差距。♣

对于给定的显著性水平 α，上述定义的基于参数估计的参数不确定集可以近似地看作不能够拒绝原假设的参数的所有可能取值的集合。

命题 7.1

考虑一个线性模型，假设其扰动项是独立同分布的，并且其均值为 0，而方差 σ^2 未知。此外，存在由 P 个独立相互独立的观测组成的数据集 $\mathcal{D} = \{\boldsymbol{z}, \boldsymbol{Y}\}$，其中 $\boldsymbol{z} \in \mathbb{R}^P$、$\boldsymbol{Y} = \{\boldsymbol{1}, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{M-1}\} \in \mathbb{R}^{P\times M}$ 为列满秩矩阵。选取 $\rho(\boldsymbol{\beta};\mathcal{D}) = \|\boldsymbol{z} - \boldsymbol{Y}\boldsymbol{\beta}\|_2^2$：

- 如果
$$r = \hat{\rho}\frac{\chi^2_{M,1-\alpha}}{P-M}, \tag{7.4}$$
对于任意给定 $\boldsymbol{\theta} \in \mathcal{E}(r;\mathcal{D})$ 及原假设 $H_0 : \boldsymbol{\beta} = \boldsymbol{\theta}$，则在显著性水平 α 下通过卡方检验 (近似) 不能拒绝原假设；此处，$\chi^2_{M,1-\alpha}$ 是自由度为 M 的卡方分布的 $1-\alpha$ 分位点，M 为 $\boldsymbol{\beta}$ 的维度。
- 如果考虑的是一个正态线性模型且
$$r = \hat{\rho}F^{-1}_{M,P-M}(1-\alpha)\frac{M}{P-M}, \tag{7.5}$$
对于任意给定 $\boldsymbol{\theta} \in \mathcal{E}(r;\mathcal{D})$ 及原假设 $H_0 : \boldsymbol{\beta} = \boldsymbol{\theta}$，则在显著性水平 α 下通过 F 检验不能拒绝原假设；此处，$F^{-1}_{M,P-M}(1-\alpha)$ 是第一自由度为 M、第二自由度为 $P-M$ 的 F 分布的 $1-\alpha$ 分位点，M 为 $\boldsymbol{\beta}$ 的维度。♠

接下来，我们可以通过下述命题得到基于极大似然估计（maximum likelihood estimation）构造的参数不确定集。这一集合可以近似地看作对应显著性水平下似然比检验（likelihood ratio test）不能够拒绝原假设的参数的所有可能取值的集合。

命题 7.2

考虑一个由 P 个独立同分布的观测组成的数据集 $\mathcal{D}$。假设其对应的对数似然函

数可以表示为

$$l(\boldsymbol{\beta};\mathcal{D}) \triangleq \ln\left(\prod_{i=1}^{P} f(z_i \mid \boldsymbol{\beta})\right).$$

此处，选取 $\rho(\boldsymbol{\beta};\mathcal{D}) = -l(\boldsymbol{\beta};\mathcal{D})/P$。在一些技术性假设条件下，如果

$$r = \frac{\chi^2_{M,1-\alpha}}{2P}, \tag{7.6}$$

对于任意给定 $\boldsymbol{\theta} \in \mathcal{E}(r;\mathcal{D})$ 及原假设 $H_0 : \boldsymbol{\beta} = \boldsymbol{\theta}$，则在显著性水平 α 下通过似然比检验 (近似) 不能拒绝原假设；此处，$\chi^2_{M,1-\alpha}$ 是自由度为 M 的卡方分布的 $1-\alpha$ 分位点，M 为 $\boldsymbol{\beta}$ 的维度。♠

由此可见，最大化基于参数估计的参数不确定集的大小可以近似等价于最大化其对应的置信水平。

7.2.2 基于参数估计的鲁棒性优化

根据上述定义可知，$\mathcal{E}(r;\mathcal{D})$ 表示在集合 $\mathcal{W}$ 中使得 $\rho(\boldsymbol{\beta};\mathcal{D})$ 与最优值 $\hat{\rho}$ 之间的差值小于 r 的所有参数的集合，从而有 $\hat{\boldsymbol{\beta}} \in \mathcal{E}(0;\mathcal{D})$。相应地，Zhu et al. (2021) 提出了如下基于参数估计的鲁棒性优化模型：

$$\begin{aligned} Z_E(\tau_0) = \max \quad & r \\ \text{s.t.} \quad & a_0(\boldsymbol{x},\boldsymbol{\beta}) \le \tau_0 \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}), \\ & a_i(\boldsymbol{x},\boldsymbol{\beta}) \le \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}),\ i \in [I], \\ & \boldsymbol{x} \in \mathcal{X},\ r \ge 0, \end{aligned} \tag{7.7}$$

称之为 JERO(joint estimation and robustness optimization) 模型。此模型旨在提供一个结合参数估计和后续优化问题的框架，来降低参数估计的误差以及不确定性对后续优化问题的影响。

虽然模型(7.7)的约束未必是关于 $\boldsymbol{x}$ 和 r 的联合凸函数，但其是关于 $\boldsymbol{x}$ 的凸函数。因此，可以通过求解以下一系列子问题得到模型的最优解：

$$\begin{aligned} Z_E^r(\tau_0) = \min \quad & t - \tau_0 \\ \text{s.t.} \quad & a_0(\boldsymbol{x},\boldsymbol{\beta}) \le t \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}), \\ & a_i(\boldsymbol{x},\boldsymbol{\beta}) \le \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D}),\ i \in [I], \\ & \boldsymbol{x} \in \mathcal{X}. \end{aligned} \tag{7.8}$$

假设存在一个方法可以得到模型(7.8)的最优解，则可以通过二分法求得模型(7.7)的最优解。有关二分法的具体步骤可参考 Zhu et al. (2021)，故在此不再赘述。

命题 7.3

假设存在算法可以得到模型(7.8)的最优解且存在 $\bar{r} > 0$ 使得 $Z_E^{\bar{r}}(\tau_0) > 0$。此外，如果模型(7.7)存在可行解，则对任意 $\Delta > 0$, 至多需要求解 $\lceil \log_2(\bar{r}/\Delta) \rceil$ 次模型(7.8)，从而得到一个解 $\boldsymbol{x}$ 及对应的 $r^\dagger$，使得 $r^\dagger$ 与模型(7.7)的最优值 r^* 之间的差值小于等于 Δ, 也即 $|r^\dagger - r^*| \le \Delta$。♠

事实上，JERO 模型是否可解依赖于能否将如下一般鲁棒约束：

$$a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i \quad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}),\ i \in [I] \cup \{0\}, \tag{7.9}$$

或者

$$\max_{\boldsymbol{\beta} \in \mathcal{E}(r;\mathcal{D})} a_i(\boldsymbol{x}, \boldsymbol{\beta}) \le \tau_i \qquad i \in [I] \cup \{0\},$$

转化为可求解的等价形式。

下述命题提供了一种求鲁棒约束等价形式的方法。

命题 7.4

假设 $\mathcal{E}(r; \mathcal{D})$ 是一个闭合凸集（closed convex set），$a_i(\boldsymbol{x}, \cdot)$ 是一个闭凹函数（closed concave function），并且 relint $(\mathcal{E}(r; \mathcal{D})) \cap \text{relint}(\bar{\mathcal{W}}) \neq \emptyset$。当且仅当存在 $\boldsymbol{x}$ 和 $\boldsymbol{\nu}$ $(\boldsymbol{x} \in \bar{X},\ \boldsymbol{\nu} \in \mathbb{R}^M)$ 使得

$$\delta_r^*(\boldsymbol{\nu}) - a_i^*(\boldsymbol{x}, \boldsymbol{\nu}) \le \tau_i$$

时，$\boldsymbol{x}$ 是鲁棒约束(7.9)的一个可行解。其中，

$$a_i^*(\boldsymbol{x}, \boldsymbol{\nu}) \triangleq \inf_{\boldsymbol{\beta} \in \bar{\mathcal{W}}} \left\{\boldsymbol{\beta}'\boldsymbol{\nu} - a_i(\boldsymbol{x}, \boldsymbol{\beta})\right\}$$

和

$$\delta_r^*(\boldsymbol{\nu}) \triangleq \sup_{\boldsymbol{\beta} \in \mathcal{W}} \left\{\boldsymbol{\beta}'\boldsymbol{\nu} \mid \rho(\boldsymbol{\beta}; \mathcal{D}) \le \hat{\rho} + r\right\}$$

分别表示函数 $a_i(\boldsymbol{x}, \boldsymbol{\nu})$ 的凹共轭函数（concave conjugate function）以及集合 $\mathcal{E}(r; \mathcal{D})$ 的支撑函数（support function）。♠

下述命题提供了一种计算集合 $\mathcal{E}(r; \mathcal{D})$ 的支撑函数 $\delta_r^*(\boldsymbol{\nu})$ 的方法。

命题 7.5

假设 $r > 0$，relint$(\mathcal{E}(r; \mathcal{D})) \neq \emptyset$，并且 $\rho(\boldsymbol{\beta}; \mathcal{D})$ 是关于 $\boldsymbol{\beta}$ 的凸函数（convex function）。则集合 $\mathcal{E}(r; \mathcal{D})$ 的支撑函数可以表示为：

$$\delta_r^*(\boldsymbol{\nu}) = \inf_{\mu > 0} \left\{(\hat{\rho} + r)\mu + \mu\rho^*(\boldsymbol{\nu}/\mu; \mathcal{D})\right\}, \tag{7.10}$$

其中，

$$\rho^*(\boldsymbol{\nu}; \mathcal{D}) \triangleq \sup_{\boldsymbol{\beta} \in \mathcal{W}} \left\{\boldsymbol{\beta}'\boldsymbol{\nu} - \rho(\boldsymbol{\beta}; \mathcal{D})\right\}$$

表示 $\rho(\boldsymbol{\beta}; \mathcal{D})$ 的凸共轭函数（convex conjugate function）。♠

有关共轭函数的计算，感兴趣的读者可参考 Boyd and Vandenberghe (2004)、Ben-Tal et al. (2015) 及其所涉及的相关文献，故在此不再展开。

7.2.3 基于参数估计的鲁棒性优化的应用

下面通过一个鲁棒投资组合优化（robust portfolio optimization）问题来展示 JERO 模型在实际问题中的应用。有关其他可能的应用场景可参考 Zhu et al. (2021)。

考虑一个有 N 种风险资产的投资组合优化问题。此外，存在由风险资产的历史收益组成的数据集 $\mathcal{D} = \{z_1, \ldots, z_P\}$，其中 $z_i \in \mathbb{R}^N$ 表示 N 种风险资产的一组历史收益的观测值。通过历史数据可计算如下统计量：

$$\hat{\boldsymbol{\beta}} = \frac{1}{P}\sum_{i=1}^{P} z_i, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{P}\sum_{i=1}^{P}(z_i - \hat{\boldsymbol{\beta}})(z_i - \hat{\boldsymbol{\beta}})^{\top}.$$

若 $\hat{\boldsymbol{\Sigma}} \succ \mathbf{0}$，则可选取

$$\rho(\boldsymbol{\beta}; \mathcal{D}) = \frac{1}{2}\Big(M + \ln\det\hat{\boldsymbol{\Sigma}} + \ln\big(1 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\big)\Big).$$

从而，决策者可求解如下 JERO 模型得到相应的投资方案：

$$\begin{aligned}
\max \quad & r \\
\text{s.t.} \quad & \boldsymbol{\beta}'\boldsymbol{x} \geq \tau \qquad\qquad \forall \boldsymbol{\beta} \in \mathcal{E}(r; \mathcal{D}), \\
& \boldsymbol{x}'\boldsymbol{e} = 1, \\
& \boldsymbol{x} \in \mathbb{R}^N_+,\ r \in \mathbb{R}_+.
\end{aligned} \tag{7.11}$$

其中，$\boldsymbol{x}$ 代表投资组合决策，$\boldsymbol{\beta}$ 代表风险资产的期望收益，τ 代表投资组合的最低期望收益。

命题 7.6

模型(7.11)可转化为

$$\begin{aligned}
\max \quad & r \\
\text{s.t.} \quad & -\hat{\boldsymbol{\beta}}'\boldsymbol{x} + \sqrt{e^{2r} - 1}\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{x}\right\|_2 + \tau \leq 0, \\
& \boldsymbol{x}'\boldsymbol{e} = 1, \\
& \boldsymbol{x} \in \mathbb{R}^N_+,\ r \in \mathbb{R}_+.
\end{aligned}$$

♠

可以看出，

$$-\hat{\boldsymbol{\beta}}'\boldsymbol{x} + \sqrt{e^{2r} - 1}\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{x}\|_2 + \tau \leq 0,$$

可表示为

$$S(\boldsymbol{x}) \triangleq \frac{\hat{\boldsymbol{\beta}}'\boldsymbol{x} - \tau}{\left\|\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{x}\right\|_2} \geq \sqrt{e^{2r} - 1}.$$

其中，$S(\boldsymbol{x})$ 为夏普指数（sharpe ratio），有关夏普指数的介绍可参考 Sharpe (1966, 1994)。此处，JERO 模型等价于最大化投资组合对应的夏普指数。本章节旨在通过一个简单的鲁棒投资组合优化问题来向读者展示 JERO 模型在实际问题中的应用；有关更一般化的鲁棒投资组合优化问题，感兴趣的读者可参考 Goldfarb and Iyengar (2003)。

本章参考文献

Ben-Tal, Aharon, Dick Den Hertog, and Jean-Philippe Vial, "Deriving robust counterparts of nonlinear uncertain inequalities," *Mathematical Programming*, 2015, *149* (1-2), 265–299.

Boyd, Stephen and Lieven Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

Goldfarb, Donald and Garud Iyengar, "Robust portfolio selection problems," *Mathematics of Operations Research*, 2003, *28* (1), 1–38.

Jaillet, Patrick, Sanjay Dominik Jena, Tsan Sheng Ng, and Sim Melvyn, "Satisficing Models under Uncertainty.," *Informs Journal on Optimization*, 2022.

Kőszegi, Botond and Matthew Rabin, "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

Merchant, Kenneth A and Jean-François Manzoni, "The Achievability of Budget Targets in Profit Centers: A Field Study," *Accounting Review*, 1989, pp. 539–558.

Payne, John W, Dan J Laughhunn, and Roy Crum, "Translation of gambles and aspiration level effects in risky choice behavior," *Management Science*, 1980, *26* (10), 1039–1060.

__ , __ , and __ , "Further Tests of Aspiration Level Effects in Risky Choice Behavior," *Management Science*, 1981, *27* (8), 953–958.

Sharpe, William F, "Mutual fund performance," *The Journal of Business*, 1966, *39* (1), 119–138.

__ , "The Sharpe ratio," *Journal of Portfolio Management*, 1994, *21* (1), 49–58.

Zhu, Taozeng, Jingui Xie, and Melvyn Sim, "Joint estimation and robustness optimization," *Management Science*, 2021.