# 第 4 章 分布鲁棒优化（Distributionally robust optimization）

章宇

现实世界的优化问题指在满足相关约束条件的前提下，确定一组决策变量的值，使预设的目标函数值最优。相关研究成果（理论、模型、算法、应用）在管理科学、金融工程、军事指挥等领域发挥着巨大指导作用，创造了巨大的社会经济价值。

由于本书定位为入门级、科普级，本章将以优化问题中一类简单但重要的线性优化（亦称线性规划）问题为例，阐述**分布鲁棒优化**的基本思想、基本模型、基本结论。感兴趣的读者需细读相关文献以获得更深入广泛的理解。

线性规划问题 (4.1) 中，决策变量为 $\boldsymbol{x} \in \mathbb{R}^I$，环境参数包括费用向量 $\boldsymbol{a}_0 \in \mathbb{R}^I$、约束条件左端项系数向量 $\boldsymbol{a}_m \in \mathbb{R}^I$、右端项系数 $b_m \in \mathbb{R}$，这些参数为确定值。线性规划可用于解决许多现实问题，例如投资组合优化、生产计划、最短路等问题。

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{a}_0^\top \boldsymbol{x}, \\ \text{s.t.} \quad & \boldsymbol{a}_m^\top \boldsymbol{x} \le b_m \quad m \in [M]. \end{aligned} \tag{4.1}$$

现实世界中描述未来发生事件的环境参数在优化/规划/计划阶段往往不确定，例如未来某商品需求量、两地间旅行时长、某股票回报率等。为了让优化结果对现实更具指导意义，在优化模型中考虑环境参数的不确定性至关重要。

在不确定环境下，(4.1) 中对于某一 $m \in [M]$ 的约束式变成了

$$\boldsymbol{a}(\tilde{\boldsymbol{\varepsilon}})^\top \boldsymbol{x} \le b(\tilde{\boldsymbol{\varepsilon}}). \tag{4.2}$$

其中，为了阐述方便，忽略下标 $m$；随机变量 $\tilde{\boldsymbol{\varepsilon}}$ 表示影响环境参数的随机因素（例如，旅行时长受天气、交通灯时长等随机因素影响），假设 $\boldsymbol{a}(\tilde{\boldsymbol{\varepsilon}})$ 和 $b(\tilde{\boldsymbol{\varepsilon}})$ 皆为 $\tilde{\boldsymbol{\varepsilon}}$ 的仿射函数，即 $\boldsymbol{a}(\tilde{\boldsymbol{\varepsilon}}) \triangleq \boldsymbol{a}^0 + \sum_{j\in[J]} \boldsymbol{a}^j \tilde{\varepsilon}_j, b(\tilde{\boldsymbol{\varepsilon}}) \triangleq b^0 + \sum_{j\in[J]} b^j \tilde{\varepsilon}_j$, 则有

$$\boldsymbol{a}(\tilde{\boldsymbol{\varepsilon}})^\top \boldsymbol{x} - b(\tilde{\boldsymbol{\varepsilon}}) = \underbrace{(\boldsymbol{a}^0)^\top \boldsymbol{x} - b^0}_{=y^0(\boldsymbol{x})} + \sum_{j\in[J]} \big(\underbrace{(\boldsymbol{a}^j)^\top \boldsymbol{x} - b^j}_{=y^j(\boldsymbol{x})}\big)\tilde{\varepsilon}_j = y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})'\tilde{\boldsymbol{\varepsilon}}.$$

因此，约束式 (4.2) 等价于

$$y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\varepsilon}} \le 0. \tag{4.3}$$

不确定环境下的线性规划从技术上主要关注如何处理约束式 (4.3)。因其左端项为随机变量而右端项为实数，故通常意义上无法直接比较大小，“≤”符号用在此处不够

严谨。对此，本章主要介绍两种典型处理方式，第4.2节基于**分布鲁棒机会约束规划**思想，介绍如何处理

$$\mathbb{P}[y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\varepsilon}} \le 0] \ge 1 - \epsilon \quad \forall \mathbb{P} \in \mathcal{F}, \tag{4.4}$$

也就是约束 (4.3) 成立的概率不小于 $1 - \epsilon$，其中阈值 $\epsilon \in [0, 1]$ 典型取值为 1% 或 5%。第4.3节基于**分布鲁棒线性优化**范式，介绍如何处理

$$\mathbb{E}_{\mathbb{P}}[y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^\top \tilde{\boldsymbol{\varepsilon}}] \le 0 \quad \forall \mathbb{P} \in \mathcal{F}, \tag{4.5}$$

也就是约束 (4.3) 需在其左端项通过均值来度量的情况下满足。

事实上，目标函数参数不确定性亦可纳入约束讨论，因为通过引入辅助决策变量 $t \in \mathbb{R}$，(4.1) 等价于

$$\begin{aligned} \min_{t,\boldsymbol{x}} \quad & t, \\ \text{s.t.} \quad & \boldsymbol{a}_0^\top \boldsymbol{x} \le t, \\ & \boldsymbol{a}_m^\top \boldsymbol{x} \le b_m \quad \forall m \in [M]. \end{aligned}$$

进而将目标函数中不确定参数 $\boldsymbol{a}_0$ 置于约束中。

约束式 (4.4) 和 (4.5) 中，随机变量 $\tilde{\boldsymbol{\varepsilon}}$ 服从联合概率分布 $\mathbb{P}$，而 $\mathbb{P}$ 本身也不确定，属于模糊集 $\mathcal{F}$；第4.1节将介绍模糊集相关内容。分布鲁棒优化采取保守策略，令这两个约束条件对模糊集中所有概率分布皆满足，它也是因此得名--- “鲁棒”的内涵是考虑最坏情况，而“分布”表明最坏情况的主体是环境参数的分布函数。望本章内容能抛砖引玉，启发读者研究和处理更复杂的形式，并用于解决实际问题。

# 4.1 模糊集（Ambiguity set）

问题环境中随机参数的分布函数往往难以从现实世界直接获取。鉴于此，分布鲁棒优化方法假设其分布函数并不明确，而是处于一个**模糊集** (ambiguity set) 中。模糊集通过随机变量的不完全分布信息构建而成。特别地，它还需保证相应分布鲁棒优化模型在计算上可处理 (tractable)，也就是现实规模问题可在允许时间范围内求解。

从数学上说，分布鲁棒优化囊括随机规划 (stochastic programming) 和传统鲁棒优化 (robust optimization) 为特殊形式，因此分布鲁棒优化更具一般性。当环境变量 $\tilde{\boldsymbol{\varepsilon}}$ 的分布函数 $\mathbb{P}_0$ 可获知时，可令模糊集为单元素集 $\mathcal{F}_S \triangleq \{\mathbb{P}_0\}$，则分布鲁棒优化退化为随机规划；当仅知环境变量的不确定集 $\Xi$ 时，可令模糊集为 $\mathcal{F}_R \triangleq \mathcal{P}_0(\Xi)$，即支撑集为 $\Xi$ 的所有概率分布函数之集合，则分布鲁棒优化退化为经典鲁棒优化。

按照描述分布函数的信息种类划分，目前相关研究主要提出了两类模糊集。

## 4.1.1 基于广义矩信息（generalized moment information）的模糊集

在统计学中，矩 (moment) 表征随机变量的分布。对于随机变量 $\tilde{\varepsilon}$，其 $n$ 阶矩被定义为 $\mathbb{E}_{\mathbb{P}}[\tilde{\varepsilon}^n]$，$n \ge 1$。因此，随机变量一阶矩为均值，表征其位置 (location)，二阶矩与方差有关，表征其散度 (dispersion)，三阶矩表征其偏斜度，等等。

更广义地，还可利用其它形式表征随机变量的位置、散度、偏斜度等特性。例如，绝对离差均值 $\mathbb{E}_{\mathbb{P}}[|\tilde{\varepsilon}-\mu|]$ 可表征 $\tilde{\varepsilon}$ 的散度，其中 $\mu$ 为其均值。再如，半绝对离差均值 $\mathbb{E}_{\mathbb{P}}[(\tilde{\varepsilon}-\mu)^{+}]$ 和 $\mathbb{E}_{\mathbb{P}}[(\mu-\tilde{\varepsilon})^{+}]$ 可从某种程度刻画 $\tilde{\varepsilon}$ 的偏斜度，其中 $(x)^{+} \triangleq \max\{x, 0\}$。

早期研究往往假设随机参数的概率分布无法准确获取，但其部分广义矩信息（和支撑集）可获取或估计，于是根据这些信息构建模糊集。例如，通过 $\tilde{\boldsymbol{\varepsilon}}$ 的均值 $\boldsymbol{\mu}$ 和协方差矩阵 $\boldsymbol{\Sigma}$ 构成的模糊集 (Ghaoui et al., 2003; Popescu, 2007; Chen and Sim, 2009) 为

$$\mathcal{F}_{MV}=\left\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^J)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{\varepsilon}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})']=\boldsymbol{\Sigma}\end{array}\right\}. \tag{4.6}$$

如果进一步考虑支撑集 $\Xi$，则模糊集为 $\mathcal{F}_{MVS}=\mathcal{F}_{MV}\cap\mathcal{P}_0(\Xi)$. 但研究表明，基于 $\mathcal{F}_{MVS}$ 的分布鲁棒优化模型一般不可处理 (Bertsimas and Popescu, 2005; Natarajan et al., 2011)。而如果给定的 $\boldsymbol{\Sigma}$ 不是准确协方差而是协方差的上界时，则模糊集为

$$\mathcal{F}_{M}=\left\{\mathbb{P}\in\mathcal{P}_0(\Xi)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{\varepsilon}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})']\preceq\boldsymbol{\Sigma}\end{array}\right\}, \tag{4.7}$$

其中“$\preceq$”为半正定锥空间意义上的小于等于，也就是 $\boldsymbol{X}\preceq\boldsymbol{Y}$ 意味着 $\boldsymbol{Y}-\boldsymbol{X}$ 为半正定矩阵。有趣的是，基于 $\mathcal{F}_M$ 的分布鲁棒线性优化模型却可处理 (Wiesemann et al., 2014; Hanasusanto et al., 2015)。

Delage and Ye (2010) 研究了 $\mathcal{F}_M$ 的一个变种

$$\mathcal{F}_{DY}=\left\{\mathbb{P}\in\mathcal{P}_0(\Xi)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{\varepsilon}}\sim\mathbb{P}\\ (\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]-\boldsymbol{\mu})^{\top}\boldsymbol{\Sigma}^{-1}(\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]-\boldsymbol{\mu})\le\gamma_1\\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})']\preceq\gamma_2\boldsymbol{\Sigma}\end{array}\right\}, \tag{4.8}$$

其中第一个约束指 $\tilde{\boldsymbol{\varepsilon}}$ 的均值处于一个以 $\boldsymbol{\mu}$ 为球心的椭球中，$\gamma_1\ge 0$ 和 $\gamma_2\ge 1$ 为两个参数。从数据驱动的视角看，假设 $\tilde{\boldsymbol{\varepsilon}}$ 客观上服从概率分布 $\mathbb{P}_0$，但无法观测该分布，而仅能观测其 $N$ 组样本/历史数据/观测值 $(\hat{\boldsymbol{\varepsilon}}_\omega)_{\omega\in[N]}$，令

$$\boldsymbol{\mu}\triangleq\frac{1}{N}\sum_{\omega\in[N]}\hat{\boldsymbol{\varepsilon}}_\omega,\qquad \boldsymbol{\Sigma}\triangleq\frac{1}{N}\sum_{\omega\in[N]}(\hat{\boldsymbol{\varepsilon}}_\omega-\boldsymbol{\mu})(\hat{\boldsymbol{\varepsilon}}_\omega-\boldsymbol{\mu})^{\top},$$

且 $\gamma_1$ 和 $\gamma_2$ 通过与样本量 $N$ 和参数 $\delta>0$ 有关的某函数给定时（随着 $N\to\infty$，有 $\gamma_1\to 0$ 和 $\gamma_2\to 1$），则 Delage and Ye (2010) 证明了在统计学上 $\mathbb{P}_0\in\mathcal{F}_{DY}$ 的置信度大于等于 $1-\delta$。

并非任意基于广义矩信息（和支撑集）的模糊集都能保证相应分布鲁棒优化模型可处理。Wisemann，Kuhn 和 Sim 提出了一种具有一般性的模糊集表达形式，能囊括 $\mathcal{F}_M$ 和 $\mathcal{F}_{DY}$ 为其特殊形式，能建模许多其它的广义矩信息，如绝对离差、半方差、高阶矩等，

且（在一些技术性假设条件下）对于分布鲁棒线性优化在计算上可处理 (Wiesemann et al., 2014; Hanasusanto et al., 2015)。其形式为

$$\mathcal{F}_{WKS}=\left\{\mathbb{P}\in\mathcal{P}_0(\mathbb{R}^J\times\mathbb{R}^L)\;\middle|\;\begin{array}{l}(\tilde{\boldsymbol{\varepsilon}},\tilde{\boldsymbol{u}})\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\boldsymbol{A}\tilde{\boldsymbol{\varepsilon}}+\boldsymbol{B}\tilde{\boldsymbol{u}}]=\boldsymbol{b}\\ \mathbb{P}[(\tilde{\boldsymbol{\varepsilon}},\tilde{\boldsymbol{u}})\in\Xi_k]\in[\underline{p}_k,\overline{p}_k]\;\forall k\in[K]\end{array}\right\},\tag{4.9}$$

其中 $\mathbb{P}$ 为 $\tilde{\boldsymbol{\varepsilon}}$ 和辅助随机变量 $\tilde{\boldsymbol{u}}$ 的联合概率分布，$\boldsymbol{A}\in\mathbb{R}^{J\times Q}$，$\boldsymbol{B}\in\mathbb{R}^{L\times Q}$，$\boldsymbol{b}\in\mathbb{R}^Q$；置信集合 $\Xi_k$ 给定为

$$\Xi_k=\{(\boldsymbol{\varepsilon},\boldsymbol{u})\in\mathbb{R}^J\times\mathbb{R}^L|\boldsymbol{C}_k\boldsymbol{\varepsilon}+\boldsymbol{D}_k\boldsymbol{u}\preceq_{\mathcal{K}_k}\boldsymbol{c}_k\},\tag{4.10}$$

其中 $\boldsymbol{C}_k\in\mathbb{R}^{J\times R}$，$\boldsymbol{D}_k\in\mathbb{R}^{L\times R}$，$\boldsymbol{c}\in\mathbb{R}^R$，而 $\mathcal{K}_k$ 代表某一真锥 (proper cone)，如非负象限、二阶锥、半正定锥等。在此，“$\preceq_{\mathcal{K}_k}$” 是在该锥空间意义上的小于等于；关于锥和相应的锥规划 (conic programming) 问题相关介绍，可参见 Ben-Tal and Nemirovski (2001)。对于表示概率界的 $\underline{\boldsymbol{p}},\overline{\boldsymbol{p}}\in[0,1]^K$，有 $\underline{\boldsymbol{p}}\leq\overline{\boldsymbol{p}}$。

模糊集 $\mathcal{F}_{WKS}$ 巧妙之处在于引入了辅助随机变量 $\tilde{\boldsymbol{u}}$，这为计算上的可处理性提供了一种有效途径，如下例所示。

**示例 4.1:** 通过 $\tilde{\boldsymbol{\varepsilon}}$ 的均值 $\boldsymbol{\mu}$ 、绝对离差均值上界 $\boldsymbol{\sigma}$、半绝对离差均值上界 $\boldsymbol{h}$ 及满足 (4.10) 形式的支撑集 $\Xi$ 构成的模糊集为

$$\mathcal{F}_{MAD}=\left\{\mathbb{P}\in\mathcal{P}_0(\Xi)\;\middle|\;\begin{array}{l}\tilde{\boldsymbol{\varepsilon}}\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[|\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu}|]\leq\boldsymbol{\sigma}\\ \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{\varepsilon}}-\boldsymbol{\mu})^+]\leq\boldsymbol{h}\end{array}\right\},\tag{4.11}$$

其中 $|\tilde{\boldsymbol{\varepsilon}}|$ 代表对向量 $\tilde{\boldsymbol{\varepsilon}}$ 的每个元素分别取绝对值构成的向量，取正符 $(\tilde{\boldsymbol{\varepsilon}})^+$ 亦然。目前虽无法**直接**处理基于 $\mathcal{F}_{MAD}$ 的分布鲁棒优化模型，但如引入辅助变量，$\mathcal{F}_{MAD}$ 则变成

$$\mathcal{F}_{MADL}=\left\{\mathbb{P}\in\mathcal{P}_0(\bar{\Xi})\;\middle|\;\begin{array}{l}(\tilde{\boldsymbol{\varepsilon}},\tilde{\boldsymbol{u}},\tilde{\boldsymbol{v}})\sim\mathbb{P}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}]=\boldsymbol{\mu}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{u}}]=\boldsymbol{\sigma}\\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]=\boldsymbol{h}\end{array}\right\},\tag{4.12}$$

其中，扩展的支撑集为

$$\bar{\Xi}=\{(\boldsymbol{\varepsilon},\boldsymbol{u},\boldsymbol{v})|\boldsymbol{\varepsilon}\in\Xi,\boldsymbol{u}\geq\boldsymbol{\varepsilon}-\boldsymbol{\mu},\boldsymbol{u}\geq\boldsymbol{\mu}-\boldsymbol{\varepsilon},\boldsymbol{v}\geq\boldsymbol{\varepsilon}-\boldsymbol{\mu},\boldsymbol{v}\geq\boldsymbol{0}\}.$$

在此，$\mathcal{F}_{MADL}$ 即为 $\mathcal{F}_{WKS}$ 的一个特例，因此变得可处理；对于具体处理方法，可参见 Wiesemann et al. (2014); Hanasusanto et al. (2015)。类似地，前文提到的 $\mathcal{F}_{DY}$ 和 $\mathcal{F}_M$ 以及文献中更多有趣的形式 (例如， Bertsimas et al., 2019) 也可借助辅助变量化为 $\mathcal{F}_{WKS}$ 的形式。

### 4.1.2 基于统计距离 (statistical distance) 的模糊集

在大数据时代背景下，问题环境中随机参数的历史数据越来越容易获取。为了获得随机参数概率分布，一种自然的想法是通过历史数据对应的经验分布来近似描述真实概率分布。经验分布是建立在 $N$ 条历史数据点上的离散均匀分布，它视每一条历史数据 $\hat{\boldsymbol{\varepsilon}}_\omega$ 为随机变量的一个支撑点，出现概率为 $1/N$，即

$$\hat{\mathbb{P}}[\tilde{\boldsymbol{\varepsilon}}^\dagger = \hat{\boldsymbol{\varepsilon}}_\omega] = \frac{1}{N} \quad \forall \omega \in [N].$$

其中 $\tilde{\boldsymbol{\varepsilon}}^\dagger$ 表示 $\tilde{\boldsymbol{\varepsilon}}$ 对应的经验随机变量。

但经验分布并不等同于真正概率分布，分布鲁棒优化的思想是假设真正概率分布与经验分布在概率空间中的统计距离不超过某一阈值，并以此构建模糊集。基于此思想，如何定义两个概率分布间的统计距离成为了关键，它不仅需要具有良好的统计学意义，而且需要保证相应的分布鲁棒优化模型可处理。在此例举两种典型形式。

第一是基于 $\phi$-散度 ($\phi$-divergence) 的模糊集，定义为

$$\mathcal{F}_\phi = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^J) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{\varepsilon}} \sim \mathbb{P}, \tilde{\boldsymbol{\varepsilon}}^\dagger \sim \hat{\mathbb{P}}, \\ D_\phi(\mathbb{P}||\hat{\mathbb{P}}) \le \theta \end{array} \right\}, \tag{4.13}$$

其中 $D_\phi(\mathbb{P}||\hat{\mathbb{P}})$ 表示所认为的真正分布 $\mathbb{P}$ 对于经验分布 $\hat{\mathbb{P}}$ 的 **$\phi$-散度**（“距离”），定义如 (4.14)；而 $\theta > 0$ 为给定的“距离”上界。

$$D_\phi(\mathbb{P}||\hat{\mathbb{P}}) = \sum_{\omega \in [N]} \hat{\mathbb{P}}(\omega) \phi \left( \frac{\mathbb{P}(\omega)}{\hat{\mathbb{P}}(\omega)} \right) \tag{4.14}$$

在 (4.14) 中，$\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ 为满足以下条件的凸函数：$\phi(1) = 0$，对于 $x > 0$ 有 $0\phi(x/0) \triangleq x \lim_{t \to +\infty} \phi(t)/t$，且 $0\phi(0/0) \triangleq 0$；$\mathbb{P}(\omega)$ 表示概率分布 $\mathbb{P}$ 中第 $\omega \in [N]$ 个观测值发生的概率。可见，该模糊集要求真正的概率分布函数支撑集与经验分布支撑集相同，也就无法考量到历史数据以外的点/场景，而仅仅是将历史数据发生的概率从经验分布的各 $1/N$ 变成了更具“鲁棒性”的值。

**示例 4.2:** 在此例举一种研究相对较多的 $\phi$ 函数形式：当 $\phi(x) = x \log x - x + 1$ 时，$\phi$ 散度具体化为 (4.15)，名为 **Kullback-Leibler 散度** (KL divergence)，又名**相对熵** (relative entropy)。

$$D_{KL}(\mathbb{P}||\hat{\mathbb{P}}) = \sum_{\omega \in [N]} \mathbb{P}(\omega) \log \left( \frac{\mathbb{P}(\omega)}{\hat{\mathbb{P}}(\omega)} \right) \tag{4.15}$$

为后文阐述方便，将其模糊集记为 $\mathcal{F}_{KL}$，也就是 (4.13) 中的将 $D_\phi$ 具体化为 $D_{KL}$。对于更多的 $\phi$ 函数形式，可参见 Ben-Tal et al. (2013); Jiang and Guan (2016)。

第二是基于 Wasserstein 距离的模糊集，定义为

$$\mathcal{F}_W = \left\{ \mathbb{P} \in \mathcal{P}_0(\Xi) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{\varepsilon}} \sim \mathbb{P}, \tilde{\boldsymbol{\varepsilon}}^\dagger \sim \hat{\mathbb{P}}, \\ d_W(\mathbb{P}, \hat{\mathbb{P}}) \le \theta, \end{array} \right\}. \tag{4.16}$$

此模糊集囊括了概率空间中以 Wasserstein 距离为度量标准，以经验分布 $\hat{\mathbb{P}}$ 为球心，以 $\theta \in \mathbb{R}_+$ 为半径的球中所有的概率分布。Esfahani and Kuhn (2018) 的研究表明，记真实但未知的概率分布为 $\mathbb{P}_0$，则当 $\theta$ 通过与样本量 $N$ 和参数 $\beta \in (0,1)$ 有关的某函数取值时（随着 $N \to \infty$，有 $\theta \to 0$），从统计学上可证明 $\mathbb{P} \in \mathcal{F}_W$ 的置信度大于等于 $1-\beta$。**Wasserstein 距离** $d_W : \mathcal{P}_0(\Xi) \times \mathcal{P}_0(\Xi) \mapsto [0,+\infty)$ 表示所考虑的分布与经验分布在概率空间中的一种距离，定义为

$$
\begin{aligned}
d_W(\mathbb{P},\hat{\mathbb{P}}) = \inf \quad & \mathbb{E}_{\bar{\mathbb{P}}}\left[\|\tilde{\boldsymbol{\varepsilon}} - \tilde{\boldsymbol{\varepsilon}}^{\dagger}\|\right] \\
\text{s.t.} \quad & \left(\tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\varepsilon}}^{\dagger}\right) \sim \bar{\mathbb{P}}, \\
& \tilde{\boldsymbol{\varepsilon}} \sim \mathbb{P}, \\
& \tilde{\boldsymbol{\varepsilon}}^{\dagger} \sim \hat{\mathbb{P}}, \\
& \bar{\mathbb{P}}\left[\left(\tilde{\boldsymbol{\varepsilon}}, \tilde{\boldsymbol{\varepsilon}}^{\dagger}\right) \in \Xi \times \Xi\right] = 1,
\end{aligned} \tag{4.17}
$$

其中的 $\bar{\mathbb{P}}$ 表示 $\tilde{\boldsymbol{\varepsilon}}$ 和 $\tilde{\boldsymbol{\varepsilon}}^{\dagger}$ 的联合概率分布，$\|\cdot\|$ 表示范数。根据定义，可直观地将 Wasserstein 距离视为从真实分布 $\mathbb{P}$ 向经验分布 $\hat{\mathbb{P}}$ 移动概率质量 (probability mass) 的最小费用。上述定义准确说是 **1 型 Wasserstein 距离**，对于更一般的 Wasserstein 距离定义及更详细深入的介绍可参见 Esfahani and Kuhn (2018); Gao and Kleywegt (2016); Zhao and Guan (2018)。

## 4.2 机会约束问题（Chance constraint）

机会约束规划是指当优化问题环境参数为随机变量时，在以一定概率满足约束条件的情况下进行优化。自从 Charnes and Cooper (1959) 提出来以来，该框架在管理科学等各领域得到了广泛研究与应用。作为抛砖引玉，本节讨论如何处理分布鲁棒独立机会约束 (4.4)，读者可进而自行研究更一般也更难处理的联合机会约束 (4.18)：

$$
\mathbb{P}[y_m^0(\boldsymbol{x}) + \boldsymbol{y}_m(\boldsymbol{x})^{\top}\tilde{\boldsymbol{\varepsilon}} \leq 0 \quad \forall m \in [M]] \geq 1-\epsilon \quad \forall \mathbb{P} \in \mathcal{F}, \tag{4.18}
$$

其中的 $M$ 个约束同时成立的概率不小于 $1-\epsilon$。

从计算角度看，约束式 (4.4) 左端项可等价表示为

$$
\mathbb{E}_{\mathbb{P}}\left[\mathbb{1}\{y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^{\top}\tilde{\boldsymbol{\varepsilon}} \leq 0\}\right], \tag{4.19}
$$

其中，$\mathbb{1}\{\cdot\}$ 为指示函数，当其事件发生取值为 1，否则为 0。指示函数为非凸函数，导致 (4.19) 对 $\boldsymbol{x}$ 或 $\tilde{\boldsymbol{\varepsilon}}$ 而言皆为非凸函数，除个别特殊情况（如 $\tilde{\boldsymbol{\varepsilon}}$ 服从联合正态分布）外难以处理。事实上，即便给定决策变量 $\boldsymbol{x}$ 和概率分布 $\mathbb{P}$，计算 (4.4) 左端项的概率值一般而言已是 NP 难问题，更何况还要基于此对 $\boldsymbol{x}$ 进行优化 (Nemirovski and Shapiro, 2006)。而有趣的是，在给定某些模糊集 $\mathcal{F}$ 的情况下，(4.4) 却可处理。

接下来讲述如何在分布鲁棒优化框架下对机会约束式 (4.4) 进行处理。易知，约束式 (4.4) 等价于

$$
\inf_{\mathbb{P} \in \mathcal{F}} \mathbb{P}[y^0(\boldsymbol{x}) + \boldsymbol{y}(\boldsymbol{x})^{\top}\tilde{\boldsymbol{\varepsilon}} \leq 0] \geq 1-\epsilon. \tag{4.20}
$$