arXiv:2301.01287v2 [math.ST] 4 Jan 2023

估计成本下的经验最优传输：分布极限与统计应用

Shayan Hundrieser[†]、Gilles Mordant[†]、Christoph A. Weitkamp[†]和Axel Munk[†‡] *

[†]哥廷根大学数学随机学研究所，Goldschmidtstraße 7，37077 Göttingen
[‡]马克斯·普朗克多学科科学研究所，Am Faßberg 11，37077 Göttingen
*卓越集群“多尺度生物成像：从分子机器到可兴奋细胞网络”（MBExC），大学医学中心，Robert-Koch-Straße 40，37075 Göttingen

2023年1月5日

摘要

基于最优传输（OT）的数据分析常常面临这样的问题：底层代价函数部分或完全未知。本文旨在推导当代价函数和测度均由数据估计时，经验最优传输值的分布极限。无论是出于统计推断的目的，还是从稳定性分析的角度来看，理解此类量的波动都至关重要。我们的结果可直接应用于群族的拟合优度检验、涉及不变传输成本的机器学习应用、估计分布混合之间距离的问题，以及经验切片最优传输量的分析。

所建立的分布极限要求满足以下条件之一：成本过程在一致范数下弱收敛，或者代价由固定参数空间上的最优传输值优化问题确定。对于第一种情形，我们结合 Skorokhod 表示，依据测度和代价对最优传输值给出细致的下界和上界。第二种情形则基于参数空间上最优传输值过程的泛函δ方法。这些证明技术本身也可能具有独立的研究价值。

1 引言

近年来，依托最优传输（OT）理论（例如，Rachev 和 Rüschendorf（1998）、Villani（2008）以及 Santambrogio（2015））进行数据分析的统计可靠方法备受瞩目。例如，这些方法可用于拟合生成对抗网络（Arjovsky、Chintala 和 Bottou，2017）、提出多元分位数（Chernozhukov 等，2017；Hallin 等，2021）和依赖关系（Nies、Staudt 和 Munk，2021；Mordant 和 Segers，2022；Wiesel，2022）的新概念，以及开发因果推断工具（Torous、Gunsilius 和 Rigollet，2021）。

回顾一下，对于波兰空间 $\mathcal{X}$ 和 $\mathcal{Y}$ 以及连续成本函数 c: $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$，两个（博雷尔）概率测度 $\mu \in \mathcal{P}(\mathcal{X})$ 和 $\nu \in \mathcal{P}(\mathcal{Y})$ 之间的最优传输值定义为

$$OT(\mu, \nu, c) := \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, \mathrm{d}\pi(x, y), \tag{1.1}$$

其中，$\Pi(\mu, \nu)$表示μ与 ν的耦合集合。在温和的假设下，（1.1）也有对偶形式（参见 Santambrogio 2015），

$$OT(\mu, \nu, c) = \sup_{f \in C(\mathcal{X})} \int_{\mathcal{X}} f^{cc}(x)\, \mathrm{d}\mu(x) + \int_{\mathcal{Y}} f^{c}(y)\, \mathrm{d}\nu(y), \tag{1.2}$$

关键词和短语。最优传输、中心极限定理、稳定性分析、维数灾难、经验过程、自助法。2020年数学主题分类。主要分类：60B12、60F05、60G15、62E20、62F40；次要分类：90C08、90C31。

其中，$C(\mathcal{X})$表示定义在$\mathcal{X}$上的实值连续函数集合。此外，$f_c(y) := \inf_{x\in\mathcal{X}} c(x, y)–f(x)$和$f^{cc}(x) := \inf_{y\in\mathcal{Y}} c(x, y) – f_c(y)$分别表示f及$f^c$在代价c下的代价变换；它们也常称为c-变换。

若 $\mathcal{X} = \mathcal{Y}$，且代价函数 $c = d_{\mathcal{X}}^p$ 是度量 $d_{\mathcal{X}}$ 在 $\mathcal{X}$ 上的 p 次幂（$p \geq 1$），则最优传输值可导出 p-Wasserstein距离

$$W_p(\mu, \nu) := (OT(\mu, \nu, d_{\mathcal{X}}^p))^{1/p},$$

它在具有 p 阶矩的概率测度空间上定义了一种度量（Villani，2008，第 6 章）。由于这种度量能够感知 $\mathcal{X}$的“内在几何”，因此对许多数据分析任务特别有用。例如，若将（归一化）图像，更确切地说，将相应的像素位置和强度视为概率测度，那么有观点认为，最优传输诱导的距离符合人眼对远近的自然判断（Rubner、Tomasi 和 Guibas，2000）。与此同时，现实中有大量案例表明，基于最优传输的距离（及其相关传输计划）在诸多应用中都很有用，例如细胞生物学（Tameling 等，2021）、遗传学（Evans 和 Matsen，2012；Schiebinger 等，2019）、蛋白质结构分析（Gellert 等，2019；Weitkamp 等，2022）或指纹分析（Sommerfeld 和 Munk，2018），这里只举几个例子。在这些研究中，代价函数是给定且已知的量，由具体应用决定；例如，Evans 和 Matsen（2012）采用了系统发育树空间上的树距离。

然而，尽管上述内容提及了多种成功应用，但在某些情况下，底层代价自然依赖于测度。在某些问题中，例如基于 Wasserstein 距离、针对群族的拟合优度检验（Hallin、Mordant 和 Segers，2021），或 Wasserstein Procrustes分析（Grave、Joulin 和 Berthet，2019），底层最优传输问题相对于某些变换具有不变性至关重要。这只能通过依赖测度的成本来实现。此外，对于切片最优传输（Bonneel 等，2015），需要计算测度多个一维投影之间的 Wasserstein 距离。对所有方向取最大值便得到最大切片 Wasserstein 距离（Deshpande 等，2019）；由于使距离达到最大的方向由底层测度决定，因此该距离可纳入采用依赖测度的成本的最优传输框架。受这些考量启发，本文为采用依赖底层测度的成本的经验最优传输问题的统计分析提供了一个通用框架。

采用这一统计学视角，我们假设无法获取测度 μ 和 ν，只能获得独立样本 $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ 和 $\{Y_i\}_{i=1}^m \sim \nu^{\otimes m}$，其中 $n、m \in \mathbb{N}$。定义经验测度 $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ 和 $\nu_m := \frac{1}{m}\sum_{i=1}^m \delta_{Y_i}$，并给定随机代价函数[1] $c_{n,m}$，使得 $OT(\mu_n, \nu_m, c_{n,m})$ 可以估计量 $OT(\mu, \nu, c)$，我们的主要关注点是刻画当 $n、m \to \infty$ 且 $m/(n+m) \to \lambda \in (0, 1)$ 时的极限分布

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_{n,m}) - OT(\mu, \nu, c)\Big). \tag{1.3}$$

这对于基于最优传输值、检验 μ 与 ν 之间关系且 c 未知的渐近检验尤为重要。此外，这也使得推导置信区间成为可能。由于在实践中更具相关性，我们主要关注 μ 和 ν 两个测度均未知的情形。不过，我们强调，我们的理论也给出了单样本情形下的分布极限，即仅根据数据估计 μ，而假设 ν 已知（见注 2.4 和注 2.9）。此外，尽管我们主要关注用经验测度估计底层测度，但只要能够确定相应的分布极限，我们的理论也能推导出其他测度估计量的分布极限。

[1]这里，$c_{n,m}$ 要么是 c 的直接估计量，要么是通过在参数类上求解与最优传输相关的优化问题来选取。

对于固定代价函数，即 $c_{n,m} \equiv c$，其中 $c \in C(\mathcal{X} \times \mathcal{Y})$，已有多项研究推导出式 (1.3) 中经验最优传输量的极限分布结果。一个特殊情形是定义在 $\mathbb{R}$ 上、满足 $c_p(x, y) = |x - y|^p$ 的概率测度，其中 $p \geq 1$（Munk 和 Czado，1998；del Barrio、Giné 和 Matrán，1999；del Barrio、Giné 和 Utzet，2005；Mason，2016；del Barrio、Gordaliza 和 Loubes，2019）；在这种情形下，最优传输计划可通过分位数耦合表示。对于这一情形，研究者利用分位数过程理论（Csörgö 和 Horváth，1993），结合关于底层密度的可积性条件，推导出了分布极限。

此外，对于一般的欧几里得空间 $\mathbb{R}^d$，其维数为 $d \geq 1$，且采用 p 阶幂代价 $c_p(x, y) = \|x - y\|^p$，其中 $p > 1$，del Barrio 和 Loubes（2019）以及 del Barrio、González-Sanz 和 Loubes（2021）已证明，对于具有连通支撑且有限 $2p$ 阶矩的概率测度 μ，ν 当 n、$m \to \infty$ 满足 $m/(n+m) \to \lambda \in (0, 1)$ 时，

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_p) - \mathbb{E}\left[OT(\mu_n, \nu_m, c_p)\right]\Big) \rightsquigarrow \mathcal{N}(0, \sigma^2_{\mu,\nu}), \tag{1.4}$$

其中 $\sigma^2_{\mu,\nu} > 0$当且仅当 $\mu \neq \nu$。本文中，“ ”均表示 Hoffman-Jørgensen 意义下的弱收敛（参见 van der Vaart 和 Wellner 1996年，第 1.3 章）。他们的证明基于最优传输值的 L^2 线性化技术，并依赖 Efron-Stein不等式。一般而言，式（1.4）中的中心化量 $\mathbb{E}[OT(\mu_n, \nu_m, c_p)]$ 不能替换为其总体量 $OT(\mu, \nu, c_p)$，这妨碍了进一步开展统计推断。事实上，对于相同的绝对连续概率测度 μ，$= \nu$ 定义在 $\mathbb{R}^d$ 上且具有足够多矩的情形，当 $d > 2p$ 时，根据 Fournier 和 Guillin（2015）以及 Weed 和 Bach（2019）的结果可知，

$$\mathbb{E}\left[OT(\mu_n, \nu_m, c_p)\right] \asymp \min(n, m)^{-p/d}.$$

此外，对于绝对连续且具有次威布尔性质的不同测度 $\mu \neq \nu$ on $\mathbb{R}^d$，Manole 和 Niles-Weed（2021）已证明，当 $d \geq 5$时，

$$\mathbb{E}\left[OT(\mu_n, \nu_m, c_p)\right] - OT(\mu, \nu, c_p) \asymp \min(n, m)^{-\min(p,2)/d}.$$

对于适当的相同测度 $\mu = \nu$ 集合（Singh 和 Póczos，2018）以及不同测度 $\mu \neq \nu$ 集合（Manole 和 Niles-Weed，2021），这些收敛速率也达到了极小极大最优（至多相差对数因子）。具体而言，这表明最优传输值的估计会受到维数灾难的影响，也说明通常在 $d \geq 5$时，由于偏差占主导地位，无法在（1.4）中用 $OT(\mu, \nu, c_p)$ 替换 $\mathbb{E}[OT(\mu_n, \nu_m, c_p)]$。

不过，根据最近发现的经验最优传输低复杂度适应原理（Hundrieser, Staudt, and Munk，2022），只要总体测度 μ 或 ν 中有一个支撑在足够低维的定义域上，仍可获得快速收敛速率。基于这一观察，Hundrieser 等人（2022）利用泛函δ方法（Römisch, 2006）证明，对于紧致支撑的 μ, ν on $\mathbb{R}^d$，其中 $\tilde{\mu}$ 支撑在一个有限集上，或支撑在一个维数为 $d < 2\min(p, 2)$ 的光滑子流形上，

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_p) - OT(\mu, \nu, c_p)\Big) \rightsquigarrow \sup_{f \in S_{c_p}(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^{\mu}(f^{c_p c_p}) + \sqrt{1-\lambda}\mathbb{G}^{\nu}(f^{c_p}), \tag{1.5}$$

其中，$S_{c_p}(\mu, \nu)$ 是 (1.2) 的最优解集合，而 $\mathbb{G}^{\mu}, \mathbb{G}^{\nu}$ 表示 μ-、ν-布朗桥，即协方差结构由下式刻画的中心高斯过程

$$\mathrm{Cov}[\mathbb{G}^{\mu}(f^{cc}), \mathbb{G}^{\mu}(g^{cc})] = \int f^{cc} g^{cc} \mathrm{d}\mu - \int f^{cc} \mathrm{d}\mu \int g^{cc} \mathrm{d}\mu \quad \text{for } f, g \in C(\mathcal{X}) \tag{1.6}$$

对于 $\mathbb{G}^{\nu}$，情况也同样如此。(1.5) 中阐述的渐近理论还为离散总体测度下经验最优传输值的分布极限提供了统一框架（Sommerfeld

和 Munk，2018；Tameling、Sommerfeld 和 Munk，2019），以及半离散情形（del Barrio、González-Sanz 和 Loubes，2022）。

本文的核心贡献是将此类分布极限从 (1.5) 扩展到代价函数并非固定、且可能依赖于底层测度的情形。我们重点研究以下两个特殊情形。

(A) 代价估计量 $c_{n,m}$ 减去其总体对应量 c 后，经过适当缩放，在 $C(\mathcal{X}\times\mathcal{Y})$ 中弱收敛到一个紧极限，即 $\sqrt{nm/(n+m)}(c_{n,m}-c) \rightsquigarrow \mathbb{G}^c$ in $C(\mathcal{X}\times\mathcal{Y})$。

(B) 存在一组代价 $\{c_\theta\}_{\theta\in\Theta}$，使得对于任意 $\mu\in\mathcal{P}(\mathcal{X}), \nu\in\mathcal{P}(\mathcal{Y})$，相应的代价函数 $c_{\mu,\nu} := c_\theta$ 均根据关于 Θ 上最优传输值的优化问题选定，即 $\theta\in\arg\max_{\theta\in\Theta} OT(\mu,\nu,c_\theta)$ 或 $\theta\in\arg\min_{\theta\in\Theta} OT(\mu,\nu,c_\theta)$。

这两种设定都很自然，能够处理范围广泛的问题。此外，它们之间有很强的关联。值得注意的是，可以通过估计最优 θ，在设定 (A) 的框架下处理设定 (B)。然而，这种方法要求总体代价函数存在且唯一，并要求成本过程作为 $C(\mathcal{X}\times\mathcal{Y})$ 中的随机元素弱收敛。由于我们只关注经验下确界或上确界最优传输值，因此采用另一种方法更为自然；该方法既不要求总体代价函数具有唯一性，也不要求成本过程弱收敛。

对于设定 (A)，我们允许根据给定数据估计代价函数，从而捕捉代价估计量与经验测度之间的渐近依赖关系。特别是，这使我们能够分析由代入估计量（例如最大似然程序）参数化代价估计量时的经验最优传输代价。值得注意的是，设定 (A) 也允许利用独立数据估计代价函数。总的来说，这种设定涵盖了许多涉及“外在估计成本”的情形。相关示例见第 4.1 节和第 4.3 节。设定 (B) 的动机则略有不同。在这里，所选代价函数取决于最优传输问题本身，并且往往使最优传输问题具有关于由Θ参数化的某个变换类的不变性。这可以描述为具有“内在估计成本”的最优传输。第 4.2 节和第 4.4 节给出了此类设定的示例。

在适当的假设下，我们在定理 2.2 中证明，对于情形 (A)，

$$\sqrt{\frac{nm}{n+m}}\big(OT(\mu_n,\nu_m,c_{n,m})-OT(\mu,\nu,c)\big)\rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\mathbb{G}^c)+\sup_{f\in S_c(\mu,\nu)}\sqrt{\lambda}\mathbb{G}^\mu(f^{cc})+\sqrt{1-\lambda}\mathbb{G}^\nu(f^c),$$

其中，$\Pi_c^\star(\mu,\nu)$ 表示针对 μ,ν、代价为 c 且 $\pi(\mathbb{G}^c) := \mathrm{R}\,\mathbb{G}^c\mathrm{d}\pi$ 的 (1.1) 的最优解集合。对于情形 (B)，我们仅陈述上确界代价下的分布极限；下确界代价下也有类似的分布极限（定理 2.6）。定义最大化解集合 $S_+(\Theta,\mu,\nu) = \arg\max_{\theta\in\Theta} OT(\mu,\nu,c_\theta)$ 后，我们在定理 2.7 中证明，

$$\sqrt{\frac{nm}{n+m}}\left(\sup_{\theta\in\Theta}OT(\mu_n,\nu_m,c_\theta)-\sup_{\theta\in\Theta}OT(\mu,\nu,c_\theta)\right)\rightsquigarrow \sup_{\theta\in S_+(\Theta,\mu,\nu)}\sup_{f_\theta\in S_{c_\theta}(\mu,\nu)}\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})+\sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}).$$

除了这些分布极限之外，我们还证明了自助法原理在情形 (A) 和 (B) 下均具有一致性。这在实践中十分重要，因为相应分布极限的分位数难以显式表示，这是由于它们依赖于总体测度和代价对应的原始与对偶优化器集合。

我们在设定 (A) 下证明分布极限所采用的证明技术不同于以往的方法，本身或许也颇具价值。更具体地说，由于代价函数需要估计，我们无法依赖上述参考文献中的任何技术。相反，我们推导出最优传输值的若干下界和上界，这些界满足适当的（半）连续性

性质。结合经验过程与成本过程的斯科罗霍德表示，我们得以证明，采用估计成本的经验最优传输值的分布，在渐近意义下分别被所断言的极限分布从上方和下方随机支配。

对于设定 (B) 的分析，我们证明，在成本族 $\{c_\theta\}_{\theta\in\Theta}$ 和底层概率测度满足适当假设时，经验最优传输过程 $\sqrt{n}(OT(\mu_n, \nu_m, c_\theta) - OT(\mu, \nu, c_\theta))_{\theta\in\Theta}$ 在 $C(\Theta)$ 中弱收敛到一个紧随机变量。我们通过结合泛函δ方法与一项关于极值型泛函在紧参数空间上统一满足 Hadamard 方向可微性的普遍结果来证明这一结论（见附录 A）。后者可视为 Fang 和 Santos（2019，引理 S.4.9）对 Θ 上过程的推广，并依赖于 Dini定理（Toma 1997，推论 1）。这一可微性结果的关键在于：不同参数对应的最大化元素集合之间满足某种连续性条件。对于最优传输过程，若对每个 $\theta \in \Theta$，对偶优化器集合 $S_{c_\theta}(\mu, \nu)$ 唯一（相差一个常数平移），则该条件成立。Xi 和 Niles-Weed（2022）在证明经验切片最优传输过程弱收敛时也采用了类似假设；该过程可视为我们关于一般最优传输过程的结果的一个特例，见第 4.4 节。随后，再次应用泛函δ方法即可得到 $\theta \in \Theta$ 上经验下确界和上确界最优传输值的分布极限。

内容概述 我们首先在第 2.1 节推导最优传输值的一种适当对偶形式，这对后续讨论很有用。随后，我们介绍主要贡献：第 2.2 节给出弱收敛代价下经验最优传输值的分布极限，第 2.3 节则给出极值型代价函数下经验最优传输值的分布极限。这些渐近结果还辅以第 2.4 节关于自助法重抽样方案一致性的结果。我们在第 3 节讨论分布极限和自助法原理所需的假设，并给出其成立的充分条件。第 4 节介绍本理论的统计应用，并推导出测度与代价函数联合扰动下最优传输代价的确定性（一阶）稳定性结果。第 5 节明确构造若干泛函，使代价估计量的正则性得以“提升”至其总体对应对象的正则性。我们在主要结果的证明中运用这些泛函，而主要结果陈述于第 6 节。其余证明以及辅助结果和引理均置于附录中。

记号与概率空间 给定集合 T，记 $l^\infty(T)$ 为 T 上有界泛函所构成的巴拿赫空间，并赋予一致范数 $\|\varphi\| := \sup_{t\in T} |\varphi(t)|$。此外，若 T 配备了拓扑 τ，则记 $C(T)$ 为 T 上实值、有界连续函数所构成的巴拿赫空间，并赋予一致范数。若 d_T 表示 T 上的度量，则定义 $C_u(T, d_T)$；当度量 d_T 可由上下文明确时，也记作 $C_u(T)$，表示 (T, d) 上实值、有界、一致连续函数构成的空间。赋予一致范数后，它也是一个巴拿赫空间。实值函数类 $\mathcal{F}$ 在 $\mathcal{X}$ 上总是赋予一致范数。这定义了巴拿赫空间 $C_u(\mathcal{F})$，它是巴拿赫空间 $l^\infty(\mathcal{F})$ 的闭子集。此外，对于 $\varepsilon > 0$，覆盖数 $\mathcal{N}(\varepsilon, T, d)$ 表示以直径为 2ε 的集合覆盖 T 所需的最少集合数；若存在常数 $C > 0$，使得 $x \le Cy$，则记作 $x \lesssim y$。

对于拓扑空间 $\mathcal{X}$，集合 $\mathcal{P}(\mathcal{X})$表示定义在 $\mathcal{X}$上的博雷尔概率测度的集合。实值博雷尔可测函数 $f: \mathcal{X} \to \mathbb{R}$ 关于 $\mu \in \mathcal{P}(\mathcal{X})$ 的积分 $\int f \, d\mu$ 简记为 $\mu(f)$ 或 μf。此外，我们用 $f_{\#}\mu$ 表示 μ 在 f 下的推前测度。我们在同一个概率空间 $(\Omega, \mathcal{A}, P)$ 上定义所有随机变量。为定义样本和自助法的随机权重，我们进一步假设该空间具有乘积结构，即 $\Omega = \Omega_0 \times \Omega_1 \times \ldots$ 且 $P = P_0 \otimes P_1 \times \ldots$，因此样本仅取决于 (Ω_0, P_0)，第一个自助法重复样本的权重取决于 (Ω_1, P_1)，依此类推。随机变量 X 的分布记为 $\mathcal{L}(X)$。最后，我们假设存在无穷多个可测映射序列 $X_1, X_2, \ldots$，其定义域为

(Ω_0, P_0)，值域分别为 $\mathcal{X}$，并且基数为 n 的样本是通过投影该无穷序列的前 n 个坐标得到的。外概率测度记为 P^*（参见 van der Vaart 和 Wellner，1996年，第 1.2 章）。记 BL_1 为度量空间 (T, d_T) 上所有一致范数不超过 1 且满足 $|f(x) - f(y)| \le d_T(x, y)$ 的实值函数的集合，其中 $x, y \in T$。我们将两个概率测度 μ、ν 之间的有界 Lipschitz 度量定义为 $d_{BL}(\mu, \nu) := \sup_{f \in BL_1} |\mu(f) - \nu(f)|$。对于集合 A 和函数 f，记 $f(A) := \{f(a) \mid a \in A\}$。对于向量空间的两个子集 A、B，定义 $A + B := \{a + b \mid a \in A, b \in B\}$。

2 主要结果

2.1 预备知识

为研究估计代价函数下经验最优传输值的分布极限理论，我们始终考虑紧波兰空间 $\mathcal{X}$和 $\mathcal{Y}$。给定连续代价函数 $c \in C(\mathcal{X} \times \mathcal{Y})$ 和概率测度 $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，原问题和对偶问题总是都存在最优解（Villani，2008，定理 4.1 和 5.10）。

根据 Villani（2003，注释 1.13），总可以从以下函数类中选取对偶优化解

$$\mathcal{H}_c := \left\{ h : \mathcal{X} \to \mathbb{R} \,\middle|\, \exists g \colon \mathcal{Y} \to [-\|c\|_\infty, \|c\|_\infty], h(\cdot) = \inf_{y \in \mathcal{Y}} c(\cdot, y) - g(y) \right\}, \tag{2.1}$$

由此，对于任意 $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，可得到最优传输值的另一种对偶表示 e,

$$OT(\mu, \nu, c) = \sup_{h \in \mathcal{H}_c} \mu(h^{cc}) + \nu(h^c). \tag{2.2}$$

函数类 $\mathcal{H}_c$ 一致有界，且其中每个元素都具有与 c 相同的连续性模数，因此根据 Arzelà-Ascoli 定理，它在 $C(\mathcal{X})$ 中是紧致的。Hundrieser 等人（2022）利用公式（2.2）研究了固定代价函数下经验最优传输值的分布极限。

就我们的目的而言，我们需要在一个固定函数类上建立对偶形式，且该形式对不止一个代价函数成立。为避免潜在的可测性问题，我们希望该函数类在 $C(\mathcal{X})$ 中是紧致的（参见引理 6.4）。为此，令 $B > 0$，并考虑凹连续模 $w : \mathbb{R}_+ \to \mathbb{R}_+$。那么，对于 $\mathcal{X}$ 上的连续度量 $d_{\mathcal{X}}$，我们定义紧函数类 $\mathcal{F}(B, w) \subseteq C(\mathcal{X})$,

$$\mathcal{F}(B, w) := \left\{ f \colon \mathcal{X} \to \mathbb{R} \,\middle|\, \|f\|_\infty \le 2B,\ |f(x) - f(x')| \le w(d_{\mathcal{X}}(x, x')) \ \text{ for all } x, x' \in \mathcal{X} \right\}, \tag{2.3}$$

该函数类将在适当的代价条件下用于最优传输值的对偶表示。

引理 2.1（对偶形式）。设 $c \in C(\mathcal{X} \times \mathcal{Y})$，其中 $\|c\|_\infty \le B$，且对所有 $x、x' \in \mathcal{X}、y \in \mathcal{Y}$，有 $|c(x, y) - c(x', y)| \le w(d_{\mathcal{X}}(x, x'))$。那么，对于 $\mathcal{F} := \mathcal{F}(B, w)$，以下包含关系成立

$$\mathcal{H}_c \subseteq \mathcal{F}^{cc} \subseteq \mathcal{H}_c + [-2B, 2B] \quad \textit{and} \quad \mathcal{H}_c^c \subseteq \mathcal{F}^c \subseteq \mathcal{H}_c^c + [-2B, 2B].$$

此外，对于任意概率测度 $\mu \in \mathcal{P}(\mathcal{X})$ 和 $\nu \in \mathcal{P}(\mathcal{Y})$，可得

$$OT(\mu, \nu, c) = \sup_{f \in \mathcal{F}} \mu(f^{cc}) + \nu(f^c) \tag{2.4}$$

以及 t （2.4）的对偶优化解集 $S_c(\mu, \nu)$，称为康托罗维奇势，是非经验的 *ty.*

引理 2.1 的证明推迟至第6.1节。总体而言，引理 2.1 证明了将函数类 $\mathcal{F} = \mathcal{F}(B, w)$ 用于对偶最优传输表述的合理性，并使我们能够用 $\mathcal{F}$ 而非可能变化的函数集合来表述分布极限的条件。

2.2 弱收敛代价下的分布极限

对于以下所有陈述中的分布极限，我们考虑在引言中提出的概率空间上定义的独立同分布随机变量 $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$ 和相互独立的 $\{Y_i\}_{i=1}^m \sim \nu^{\otimes m}$。基于这些样本，我们定义经验测度 $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ 和 $\nu_m := \frac{1}{m}\sum_{i=1}^m \delta_{Y_i}$。所有后续渐近结果均针对 $n, m \to \infty$ 且 $m/(n+m) \to \lambda \in (0,1)$ 的情形；为节省篇幅，我们不再每次重复说明。

对于双样本情形，我们的主要结果如下，给出了弱收敛代价下经验最优传输值的极限分布。单样本情形见备注 2.4(ii)。

定理 2.2（弱收敛代价下的最优传输）。设 $c \in C(\mathcal{X}\times\mathcal{Y})$，并考虑 c 的一个估计量 $c_{n,m} \in C(\mathcal{X}\times\mathcal{Y})$，满足对每个 $(x,y) \in \mathcal{X}\times\mathcal{Y}$，$c_{n,m}(x,y)$ 都是可测的。设 $w: \mathbb{R}_+ \to \mathbb{R}_+$ 是 c 的凹连续模，且对 $\delta > 0$ 有 $w(\delta) > 0$，使得对所有 $x, x' \in \mathcal{X}, y \in \mathcal{Y}$，$|c(x,y) - c(x',y)| \le w(d_{\mathcal{X}}(x,x'))$。假设对于 $\mu \in \mathcal{P}(\mathcal{X})$、$\nu \in \mathcal{P}(\mathcal{Y})$，满足以下条件。

(JW) 对于 (2.3) 中的函数类 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$，存在联合弱收敛，

$$\sqrt{\frac{nm}{n+m}}\begin{pmatrix}\mu_n - \mu\\ \nu_m - \nu\\ c_{n,m} - c\end{pmatrix} \rightsquigarrow \begin{pmatrix}\sqrt{\lambda}\,\mathbb{G}^\mu\\ \sqrt{1-\lambda}\,\mathbb{G}^\nu\\ \mathbb{G}^c\end{pmatrix} \quad \text{in } \ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X}\times\mathcal{Y}),$$

其中 $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ 是一个紧随机变量，且 $\mathbb{G}^\mu, \mathbb{G}^\nu$ 具有 (1.6) 中所示的协方差结构。

此外，假设以下两个假设之一成立。

(OP) 对于代价函数 c，μ 与 ν 之间存在唯一的最优传输计划 $\pi \in \Pi_c^\star(\mu,\nu)$。

(Sup) *The empirical processes* $\mathbb{G}_n^\mu := \sqrt{n}(\mu_n - \mu)$ *and* $\mathbb{G}_m^\nu := \sqrt{m}(\nu_m - \nu)$ *fulfill the convergence* $\sup_{f\in\mathcal{F}} \mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}} - f^{cc}) \xrightarrow{P^*} 0$ *and* $\sup_{f\in\mathcal{F}} \mathbb{G}_m^\nu(f^{c_{n,m}} - f^c) \xrightarrow{P^*} 0$.

于是可得

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)\Big) \rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\mathbb{G}^\nu(f^c).$$

定理 2.2 的一个关键见解是，估计的最优传输值的极限分布可以分解为两项：在一组最优传输计划处取值的成本估计量的波动，以及在经验过程极限处取值的康托罗维奇势。在总体最优传输问题的原始与对偶优化器唯一的条件下，我们得到以下结果。

推论 2.3（弱收敛代价下的最优传输与唯一性）。在定理 2.2 的设定下，假设满足 (JW) 和 (OP)，并且对于 μ、ν，代价函数 c 对应的康托罗维奇势函数集合 $S_c(\mu,\nu)$ 唯一（相差一个常数平移）[2]。那么，对于 $\pi \in \Pi_c^\star(\mu,\nu)$ 以及 $f \in S_c(\mu,\nu)$，可得

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)\Big) \rightsquigarrow \pi(\mathbb{G}^c) + \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c). \tag{2.5}$$

特别地，若 $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ 是一个取值于 $l^\infty(\mathcal{F}cc) \times l^\infty(\mathcal{F}c) \times C(\mathcal{X}\times\mathcal{Y})$ 的联合中心化高斯过程，则 (2.5) 中的弱极限为中心正态。

[2]这里的意思是，对于任意 $f、g \in S_c(\mu,\nu)$，差值 $f - g$ 在 μ 的支撑上为常数。

定理 2.2 的证明推迟至第 6.2.1 节，并依赖于对偶形式 (2.4) 和原问题 (1.1) 所对应的经验最优传输值的精细下界和上界，以及经验过程理论中的论证。其中一个关键论证是将引理 2.1 应用于 $c_{n,m}$ 和 c。值得注意的是，我们并不要求代价估计量 $c_{n,m}$ 有适当的界，或具有与 c 本身相似的连续模。相反，我们通过推论 5.4 构造另一个代价估计量 $c_{n,m}$，使得条件 $\|c_{n,m}\|_\infty \le 2\|c\|_\infty + 1$ 以及 $|c_{n,m}(x,y) - c_{n,m}(x',y)| \le 2w(d_{\mathcal{X}}(x,x'))$ 对所有 $x, x' \in \mathcal{X}$、$y \in \mathcal{Y}$ 都确定性地成立，并且 $\sqrt{nm/(n+m)}\|c_{n,m} - c_{n,m}\|_\infty \xrightarrow{P} 0$。由引理 6.2 可知，后者意味着

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n,\nu_m,\bar{c}_{n,m}) - OT(\mu_n,\nu_m,c_{n,m})\Big) \le \sqrt{\frac{nm}{n+m}}\|\bar{c}_{n,m} - c_{n,m}\|_\infty \xrightarrow{P} 0.$$

因此，只需证明关于 $c_{n,m}$ 的断言，此时代价类 $\mathcal{F}(2\|c\|_\infty + 1, 2w)$ 所涉及的引理 2.1 中的对偶形式是可用的。我们称 $c_{n,m}$ 为 $c_{n,m}$ 的正则性提升；关于不同类型正则性提升的细节见第 5 节。正则性提升这一概念也有助于证明条件 (Sup) 成立，详见第 3.3 节。

注 2.4. 我们想对所推导出的分布极限的若干方面作些说明。

(i) 定理 2.2 的假设及其成立的充分条件将在第 3.1 节至第 3.3 节讨论。实际上，(JW) 将该理论限定在低维情形。在这类情形下，只要总体代价足够正则，条件 (Sup) 通常也成立。

(ii) 我们用于定理 2.2 和推论 2.3 的证明技术也适用于单样本情形的分布极限，即 μ 由 μ_n 估计且 ν 已知时。对于这种情形，(JW) 可简化为条件

$$\sqrt{n}\begin{pmatrix}\mu_n - \mu\\ c_n - c\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{G}^\mu\\ \mathbb{G}^c\end{pmatrix} \quad \text{in } \ell^\infty(\mathcal{F}^{cc}) \times C(\mathcal{X}\times\mathcal{Y}).$$

此外，在 (Sup) 中，我们只要求 $\sup_{f\in\mathcal{F}} \mathbb{G}_n^\mu(f_{c_n c_n} - f_{cc}) \xrightarrow{P^*} 0$。于是，

$$\sqrt{n}\Big(OT(\mu_n,\nu,c_n) - OT(\mu,\nu,c)\Big) \rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^\mu(f^{cc}).$$

(iii) 对于固定代价函数的情形，即选择 $c_n = c$ 时，定理 2.2 的条件可简化为 $\mathcal{F}_{cc}$ 是 μ-Donsker 类，且 $\mathcal{F}^c$ 是 ν-Donsker 类。此外，根据引理 2.1，这等价于 $\mathcal{H}_c$ 和 $\mathcal{H}_c^c$ 分别对 μ 和 ν 是 Donsker 类（van der Vaart 和 Wellner，1996，定理 2.10.1 和例 2.10.7）；这与 Hundrieser 等人（2022）定理 2.1 的条件 (C) 和 (S2) 相符，它们意味着

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n,\nu_m,c) - OT(\mu,\nu,c)\Big) \rightsquigarrow \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c).$$

(iv) 我们的证明技术还可以推导出估计的最优传输值的分布极限：将经验测度 μ_n and ν_m 替换为分别满足 $\tilde{\mu}_n \rightsquigarrow \mu$ 和 $\tilde{\nu}_m \rightsquigarrow \nu$（以概率收敛）的可测估计量 $\tilde{\mu}_n \in \mathcal{P}(\mathcal{X})$、$\tilde{\nu}_m \in \mathcal{P}(\mathcal{Y})$。这意味着在假设 (JW) 和 (Sup) 中，分别用 $\tilde{\mu}_n$ 和 $\tilde{\nu}_m$ 替换经验测度 μ_n and ν_m。此外，我们的证明技术理论还允许采用~~不同于缩放~~速率 $\sqrt{nm/(n+m)}$ 的缩放速率 $a_{n,m}$，该速率在 $n, m \to \infty$ 时趋于无穷。

(v) 在命题 4.9 中，我们证明，对于 $(\Delta^\mu, \Delta^\nu, \Delta^c) \in (\mathcal{P}(\mathcal{X}) - \mu) \times (\mathcal{P}(\mathcal{Y}) - \nu) \times C(\mathcal{X} \times \mathcal{Y})$ 中的可容许方向，最优传输值关于 (μ, ν, c) 盖托可微，其导数为，

$$(\Delta^\mu, \Delta^\nu, \Delta^c) \mapsto \inf_{\pi \in \Pi_c^\star(\mu,\nu)} \pi(\Delta^c) + \sup_{f \in S_c(\mu,\nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c).$$

因此，定理 2.2 中描述的渐近分布也可解释为在极限过程处，最优传输值关于三元组 (μ, ν, c) 的导数。若通过应用泛函 δ 方法来证明定理 2.2，就需要证明最优传输值具有 Hadamard 方向可微性（Römisch，2006）。然而，若不对测度估计量和代价估计量施加额外假设，这一问题颇具挑战性，参见注 4.10。

(vi) 若 (2.5) 中的极限为中心正态分布，则极限方差为

$$\begin{aligned}&\operatorname{Var}(\pi(\mathbb{G}^c)) + \lambda \operatorname{Var}_{X\sim\mu}(f^{cc}(X)) + (1-\lambda)\operatorname{Var}_{Y\sim\mu}(f^c(Y))\\&+ 2\sqrt{\lambda}\operatorname{Cov}(\pi(\mathbb{G}^c), \mathbb{G}^\mu(f^{cc})) + 2\sqrt{1-\lambda}\operatorname{Cov}(\pi(\mathbb{G}^c), \mathbb{G}^\nu(f^c)),\end{aligned}$$

其中我们利用了随机变量 $X_1, \ldots, X_n$ 和 $Y_1, \ldots, Y_n$ 相互独立这一事实。特别地，如果两个康托罗维奇势 (f^{cc}, f^c) 均几乎处处为 (μ, ν)-常数，且 $c_{n,m}$ 以快于 $(nm/(n+m))^{-1/2}$ 的速率收敛至 c，并且该收敛在最优传输计划 π 的支撑上是一致的，那么极限分布会退化。关于几乎处处为常数的康托罗维奇势何时出现的精确刻画，可参见 Hundrieser 等人（2022）的第 4 节。作者指出，对于大多数实际关注的代价函数，如果底层测度不同，几乎处处为常数的情形通常不会出现。

2.3 极值型代价函数下的分布极限

如引言中所述，也可以利用前文介绍的框架，分析在一组固定的代价函数上取下确界或上确界的经验最优传输值。然而，采用这一方法需要存在一个单一的总体代价函数，并且代价估计量弱收敛。为拓展本理论的适用范围，本小节采用另一种途径来推导极限分布，而不要求满足上述条件。具体而言，我们首先证明由这组代价函数索引的经验最优传输过程的一致分布极限，然后利用 δ 方法刻画相应下确界统计量和上确界统计量的分布极限。

对于后续结论，我们再次遵循第 2.2 节开头给出的抽样约定。单样本情形见注 2.9(iii)。

定理 2.5（在紧致 Θ 上一致的 OT 过程）。设 Θ 是一个紧 Polish 空间，并考虑连续映射 $c: \Theta \to C(\mathcal{X} \times \mathcal{Y}), \theta \mapsto c_\theta$。设 $w: \mathbb{R}_+ \to \mathbb{R}_+$ 为一个连续性模数，使得对所有 $x, x' \in \mathcal{X}, y \in \mathcal{Y}$，均有 $\sup_{\theta\in\Theta} |c_\theta(x, y) - c_\theta(x', y)| \le w(d_{\mathcal{X}}(x, x'))$。假设对于 $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，满足以下条件。

（Don）对于（2.3）中的函数类 $\mathcal{F} = \mathcal{F}(\sup_{\theta\in\Theta} \|c_\theta\|_\infty, w)$，集合 $S_{\theta\in\Theta} \mathcal{F}_{c_\theta c^\theta}$ 是 μ-Donsker，且 $S_{\theta\in\Theta} \mathcal{F}^{c_\theta}$ 是 ν-Donsker。

（KP）对于任意 $\theta \in \Theta$，最优传输问题中 μ 与 ν 之间、代价为 c_θ 的 Kantorovich 势函数集合 $S_{c_\theta}(\mu, \nu) \subseteq \mathcal{F}$ 唯一（相差一个常数平移）。

于是，对任意 $\theta \in \Theta$ 选取 $f_\theta \in S_{c_\theta}(\mu, \nu)$ 后，可得

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu_m, c_\theta) - OT(\mu, \nu, c_\theta)\Big)_{\theta\in\Theta} \rightsquigarrow \Big(\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta})\Big)_{\theta\in\Theta} \quad \text{in } C(\Theta).$$

定理 2.5 的证明基于 OT 代价过程的 Hadamard 方向可微性，这一性质源于对紧参数空间上极值型函数的一般敏感性分析（附录 A）。经验 OT 过程的结论则可通过应用泛函 δ方法（Römisch，2006）得到；证明留待第 6.3.1 节。

根据上述结果，对于任意在函数 $OT(\mu, \nu, c(\cdot)) \in C(\Theta)$ 处 Hadamard 方向可微的泛函 $\Phi: C(\Theta) \to \mathbb{R}$，定理 2.5 通过再次应用泛函δ方法，得出以下分布极限

$$\sqrt{\frac{nm}{n+m}}\Big(\Phi((OT(\mu_n, \nu_m, c_\theta))_{\theta\in\Theta}) - \Phi((OT(\mu, \nu, c_\theta))_{\theta\in\Theta})\Big) \rightsquigarrow D^H_{OT(\mu,\nu,c(\cdot))}\Phi\Big((\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}))_{\theta\in\Theta}\Big).$$

这里，$D^H_{OT(\mu,\nu,c(\cdot))}\Phi$表示 Φ的方向 Hadamard 导数。这使得我们能够使用 Fang 和 Santos（2019，引理 S.4.9）推导下确界映射的极限分布（另见 Cárcamo、Cuevas 和 Rodríguez 2020，推论 2.3）。

定理 2.6（紧致 Θ上的最优传输下确界）。考虑定理 2.5 的设定。于是，对任意 $\theta \in \Theta$，选取 $f_\theta \in S_{c_\theta}(\mu, \nu)$，可得

$$\sqrt{\frac{nm}{n+m}}\Big(\inf_{\theta\in\Theta} OT(\mu_n, \nu_m, c_\theta) - \inf_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\Big) \rightsquigarrow \inf_{\theta\in S_-(\Theta,\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}),$$

其中，$S_-(\Theta, \mu, \nu) = \arg\min_{\theta\in\Theta} OT(\mu, \nu, c_\theta)$ 表示在 Θ上使 $OT(\mu, \nu, c_\theta)$ 达到极小值的极小化元集合。

即使仅满足（Don），仍可推得经验上确界最优传输值的极限分布。

定理 2.7（紧致 Θ 上的最优传输上确界）。考虑定理 2.5 的设定，并且仅假设（Don）。那么可得

$$\sqrt{\frac{nm}{n+m}}\Big(\sup_{\theta\in\Theta} OT(\mu_n, \nu_m, c_\theta) - \sup_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\Big) \rightsquigarrow \sup_{\substack{\theta\in S_+(\Theta,\mu,\nu)\\ f_\theta\in S_{c_\theta}(\mu,\nu)}} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}),$$

其中，$S_+(\Theta, \mu, \nu) = \arg\max_{\theta\in\Theta} OT(\mu, \nu, c_\theta)$ 表示在 Θ 上使 $OT(\mu, \nu, c_\theta)$ 取得最大值的最大化解集合。

定理 2.6 和定理 2.7 的证明分别见第 6.3.2 节和第 6.3.3 节。此外，在某些情境下，关于 Θ 的紧性假设可能过于严格。以下结果将其推广到非紧空间 Θ，并聚焦于下确界统计量；上确界统计量也有类似的结论。其证明推迟至第 6.3.4 节。

命题 2.8（一般 Θ上的最优传输下确界）。设 Θ是一个波兰空间，并考虑连续映射 $c: \Theta \to C(\mathcal{X} \times \mathcal{Y})$。令 $\mu \in \mathcal{P}(\mathcal{X})$，$\nu \in \mathcal{P}(\mathcal{Y})$，并假设存在一个紧集 $K \subseteq \Theta$，使得 $S_-(\Theta, \mu, \nu) \subseteq K$；存在极小化元序列 $\theta_{n,m} \in S_-(\Theta, \mu_n, \nu_m)$，满足 $\lim_{n,m\to\infty} P^*(\theta_{n,m} \notin K) = 0$；并且定理 2.6 的假设在将 Θ替换为 K 后成立。那么，定理 2.6 关于 Θ上的经验下确界最优传输值的结论仍然成立。

备注 2.9。关于经验最优传输代价过程的弱极限，以及相应的下确界统计量和上确界统计量，下面作几点说明。

(i) 在定理 2.5 的设定下，参数空间 Θ 是紧致的，且 $c: \Theta \to C(\mathcal{X} \times \mathcal{Y})$ 连续，因此值域 $c(\Theta)$ 在 $C(\mathcal{X} \times \mathcal{Y})$ 中也紧致。特别地，根据 Arzelà–Ascoli 定理，我们得出 $\sup_{\theta\in\Theta} \|c_\theta\|_\infty < \infty$，并且存在一个合适的连续模，使所有代价函数在 Θ 上一致满足该连续模。

(ii) 定理 2.5 的两个假设及其充分条件在第 3.4 节和第 3.5 节中进行了讨论。假设（Don）看来自然适用于控制经验最优传输过程在 Θ 上的一致性，而 (KP) 则用于确保极限过程支撑于 $C(\Theta)$ 且保持紧。我们的证明技术表明，(KP) 可以略微放宽，但幅度不大。例如，可以要求：在导数中达到上确界的康托罗维奇势 $S^{c_\theta}(\mu, \nu)$，能够由 θ 紧邻区域内 θ' 的康托罗维奇势 $S^{c_{\theta'}}(\mu, \nu)$ 逼近，正如引理 A.3(i) 所要求的那样。特别是，如果 $\Theta := \{\theta_1, \ldots, \theta_K\}$ 是配备离散拓扑的有限集，那么可以省略 (KP)。

（iii）这些结果同样适用于单样本情形，即当 μ 由 μ_n 估计且假设 ν 已知时。对于定理 2.5 的单样本版本，只需在（Don）中假设函数类 $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c^\theta}$ 是 μ-Donsker，并结合假设（KP）即可。选取 $f_\theta \in S^{c_\theta}(\mu, \nu)$（任意 $\theta \in \Theta$）后，对于 $n \to \infty$，极限分布为

$$\sqrt{n}\Big(OT(\mu_n, \nu, c_\theta) - OT(\mu, \nu, c_\theta)\Big)_{\theta\in\Theta} \rightsquigarrow \Big(\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})\Big)_{\theta\in\Theta} \quad \text{in } C(\Theta).$$

在相同假设下，定理 2.6 的单样本对应结果同样成立。对于定理 2.7 中单样本结果的有效性，只需满足 $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c^\theta}$ 是 μ-Donsker。

（iv）所得的弱极限凸显了极限分布与康托罗维奇势函数集合之间的密切依赖关系。在定理 2.5 中，由于假设（KP），极限过程是中心高斯过程。对于固定的 $\theta \in \Theta$，如果相应的康托罗维奇势函数是 (μ, ν)-几乎处处为常数，则极限随机变量退化为集中于零点的狄拉克测度。此外，如果对于 μ、ν 和 c_θ，康托罗维奇势函数（$f^{c_\theta c^\theta}$，f^{c^θ}）在 $\mathrm{supp}(\mu) \times \mathrm{supp}(\nu)$ 上（相差一个常数平移）一致，且对任意 $\theta \in S_-(\Theta, \mu, \nu)$，则定理 2.6 中的极限分布同样为中心正态。在关于 $\theta \in S_+(\Theta, \mu, \nu)$ 的类似假设下，定理 2.7 中的极限分布为中心正态。特别地，在假设（KP）下，如果 $S_-(\Theta, \mu, \nu)$ 或 $S_+(\Theta, \mu, \nu)$ 是单元素集，则该条件得到满足。当康托罗维奇势函数是 (μ, ν)-几乎处处为常数时，所得极限分布退化。关于几乎处处为常数的势函数的精确刻画，详见 Hundrieser 等人（2022）的第 4 节。

2.4 最优传输代价的自助法原理

由于定理 2.2、2.6 和 2.7 中的极限分布涉及康托罗维奇势函数集（以及最优传输计划），当最优解不唯一时，几乎不可能显式写出这些分布的分位数的闭式表达，而这对于进一步的实际应用是必需的。为解决这一问题，我们建议采用 n 中取 k 自助法，其中 $k = o(n)$；本小节将证明其一致性。

为简单起见，后续结果均针对样本量相等的情形，即 $n = m$，且自助样本的样本量也相等，为 $k = o(n)$。若样本量不同，即 n、m 不同，则应选择样本量分别为 $k = o(n)$、$l = o(m)$ 的自助样本，使得 $l/(l + k) \approx m/(n + m)$。下文始终采用以下介绍的同一种自助法。对于两列独立同分布随机变量 $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}$, $\{Y_i\}_{i=1}^n \sim \nu^{\otimes n}$，其相应的经验测度为 μ_n, ν_n，考虑另一列独立同分布的自助随机变量 $\{X_i^b\}_{i=1}^k \sim \mu_n^{\otimes k}$, $\{Y_i^b\}_{i=1}^k \sim \nu_n^{\otimes k}$，并定义自助经验测度 $\mu_{n,k}^b := \frac{1}{k}\sum_{i=1}^k \delta_{X_i^b}$ 和 $\nu_{n,k}^b := \frac{1}{k}\sum_{i=1}^k \delta_{Y_i^b}$。此外，在下文的陈述中，用 c_n 表示代价估计量，用 $c_{n,k}^b$ 表示自助代价估计量。

命题 2.10（弱收敛代价下最优传输的自助法）。在定理 2.2 的设定下，假设 (JW) 以及 (OP) 或 (Sup)。令 $c_{n,k}^b \in C(\mathcal{X} \times \mathcal{Y})$ 为自助代价估计量，使得对所有 $(x, y) \in \mathcal{X} \times \mathcal{Y}$，$c_{n,k}^b(x, y)$ 均可测。此外，假设以下条件成立。

(JW)* 自助经验过程在给定 $X_1, \ldots, X_n, Y_1, \ldots Y_n$ 的条件下，在空间 $l^\infty(\mathcal{F}_{cc}) \times l^\infty(\mathcal{F}_c) \times C(\mathcal{X} \times \mathcal{Y})$ 中相合，其中 $n, k \to \infty$ 且 $k = o(n)$，即，

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k}\begin{pmatrix}\mu_{n,k}^b - \mu_n\\ \nu_{n,k}^b - \nu_n\\ c_{n,k}^b - c_n\end{pmatrix}\middle| X_1, \ldots, X_n, Y_1, \ldots Y_n\right), \mathcal{L}\left(\sqrt{n}\begin{pmatrix}\mu_n - \mu\\ \nu_n - \nu\\ c_n - c\end{pmatrix}\right)\right) \xrightarrow{P^*} 0.$$

在采用（Sup）设定的情况下，还需作如下假设。

(Sup)* *The unconditional bootstrap empirical processes* $\mathbb{G}_{n,k}^\mu := \sqrt{k}(\mu_{n,k}^b - \mu)$ *and* $\mathbb{G}_{n,k}^\nu := \sqrt{k}(\nu_{n,k}^b - \nu)$ *fulfill* $\sup_{f\in\mathcal{F}} \mathbb{G}_{n,k}^\mu(f^{c_{n,k}^b c_{n,k}^b} - f^{cc}) \xrightarrow{P^*} 0$, $\sup_{f\in\mathcal{F}} \mathbb{G}_{n,k}^\nu(f^{c_{n,k}^b} - f^c) \xrightarrow{P^*} 0$ *for* $n, k \to \infty$, $k = o(n)$.

于是，对于满足 $k = o(n)$ 的 $n, k \to \infty$，有

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{k}\left(OT(\mu_{n,k}^b, \nu_{n,k}^b, c_{n,k}^b) - OT(\mu_n, \nu_n, c_n)\right)\Big| X_1, \ldots, X_n, Y_1, \ldots, Y_n\Big), \mathcal{L}\Big(\sqrt{n}\left(OT(\mu_n, \nu_n, c_n) - OT(\mu, \nu, c)\right)\Big)\Big) \xrightarrow{P^*} 0.$$

尽管在推导弱收敛代价下经验最优传输值的极限分布时，并未依赖泛函δ方法，但通过采用自助法一致性的一种等价表述（Bücher 和 Kojadinovic，2019），并结合使用斯科罗霍德表示，我们得到了与 Dümbgen（1993，命题 2）类似的自助法原理。完整证明见第 6.2.2 节。

注 2.11. 使用泛函δ方法时，如果 Hadamard方向导数不是线性的，则 n中取n自助法不一致（Dümbgen，1993，命题 1）。尽管命题 2.10 并非基于可微性结果，但我们将在第 4.5 节证明，若原始或对偶最优解不唯一，则 OT 泛函具有 Gateaux方向可微性，且其导数非线性。由于 Gateaux方向可微性由 Hadamard 方向可微性蕴含，这表明在最优解不唯一的情形下，对于弱收敛代价下的经验最优传输成本，朴素的 n中取n自助法并不一致。

在定理 2.5–2.7 的设定下，验证自助法一致性十分直接。这直接源于当 $k = o(n)$ 时，n中取k自助法经验过程的一致性（van der Vaart 和 Wellner，1996，定理3.6.13），以及针对自助法的泛函δ方法（Dümbgen，1993，命题 2）。因此，我们省略下面命题的证明。

命题 2.12（OT 过程、上确界与下确界的自助法）。设 $\mathcal{X}, \mathcal{Y}$ 为紧波兰空间，Θ 为紧拓扑空间。考虑连续映射 $c: \Theta \to C(\mathcal{X} \times \mathcal{Y})$, $\theta \mapsto c_\theta$，令 $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，并假设 (Don)。

(i)（$C(\Theta)$ 中的 OT 过程）那么，在 (KP) 条件下，对于满足 $k \le n$ 的 $n,k \to \infty$，有

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{k}\left(OT(\mu_{n,k}^b, \nu_{n,k}^b, c_\theta) - OT(\mu_n, \nu_n, c_\theta)\right)_{\theta\in\Theta}\Big| X_1, \ldots, X_n, Y_1, \ldots, Y_n\Big), \mathcal{L}\Big(\sqrt{n}\left(OT(\mu_n, \nu_n, c_\theta) - OT(\mu, \nu, c_\theta)\right)\Big)_{\theta\in\Theta}\Big) \xrightarrow{P^*} 0.$$

(ii)（Θ上的最优传输下确界）那么，在（KP）条件下，对于满足 $k \le o(n)$ 的 $n,k \to \infty$，有

$$d_{BL}\Big(\mathcal{L}\Big(\sqrt{k}\left(\inf_{\theta\in\Theta} OT(\mu_{n,k}^b, \nu_{n,k}^b, c_\theta) - \inf_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta)\right)\Big| X_1, \ldots, X_n, Y_1, \ldots, Y_n\Big), \mathcal{L}\Big(\sqrt{n}\left(\inf_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta) - \inf_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\right)\Big)\Big) \xrightarrow{P^*} 0.$$

(iii)（Θ上的最优传输上确界）那么，对于满足 k = o(n) 的 n,k → ∞，有

$$d_{BL}\Bigg(\mathcal{L}\Bigg(\sqrt{k}\Bigg(\sup_{\theta\in\Theta} OT(\mu_{n,k}^{b}, \nu_{n,k}^{b}, c_\theta) - \sup_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta)\Bigg)\Bigg| X_1, \ldots, X_n, Y_1, \ldots, Y_n\Bigg),$$
$$\mathcal{L}\Bigg(\sqrt{n}\Bigg(\sup_{\theta\in\Theta} OT(\mu_n, \nu_n, c_\theta) - \sup_{\theta\in\Theta} OT(\mu, \nu, c_\theta)\Bigg)\Bigg)\Bigg) \xrightarrow{\mathrm{P}^*} 0.$$

值得注意的是，对于情形 (i)，我们还得到了 n中取n自助法的一致性，因为 (KP) 蕴含 Hadamard方向导数的线性性。

3 假设讨论

本节讨论分布极限和自助法一致性所需的假设，并给出其成立的充分条件。所有证明均置于附录B。

3.1 假设 (JW) 和 (JW)*：联合弱收敛

对于采用估计成本的经验最优传输值，我们在 (JW) 和 (JW)* 中要求经验过程在 $l^\infty(\mathcal{F}_{cc})$ 和 $l^\infty(\mathcal{F}_c)$ 中弱收敛，其中 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ 的选取方式与定理 2.2 中相同。这要求 $\mathcal{F}^{cc}$ 和 $\mathcal{F}_c$ 分别为 μ-和 ν-Donsker。此外，我们要求估计成本函数在 $C(\mathcal{X}\times\mathcal{Y})$ 中弱收敛，以确保 μ_n, ν_m 和 $c_{n,m}$ 的任意一列最优传输计划都趋于 $\Pi_c^\star(\mu, \nu)$ 中的一个最优传输计划。最后，我们强调 (JW) 和 (JW)* 中联合弱收敛的必要性，因为极限分布由随机变量 $(\mathbb{G}^\mu, \mathbb{G}^\nu, \mathbb{G}^c)$ 决定，因此取决于它们之间的依赖关系。

尽管这些条件表面上似乎不可避免，但它们在某种程度上具有限制性，将理论局限于低维情形。这是意料之中的，因为在总体代价下估计最优传输值会遭遇维数灾难（Manole和Niles-Weed，2021），当两个总体测度 μ、ν 都具有高维支撑时，会导致较慢的收敛速率。然而，鉴于最近发现的低复杂度适应原理（Hundrieser、Staudt和Munk，2022），只需测度 μ 或 ν 中的一个由低维空间支撑即可。以下命题给出了 $\mathcal{F}^c$ 和 $\mathcal{F}^{cc}$ 在一致范数下的覆盖数界（定义见符号说明部分），由此可为两个函数类建立普适 Donsker 性质。

命题3.1（普适 Donsker 性质）。设 $c \in C(\mathcal{X}\times\mathcal{Y})$ 是一个连续代价函数，且$\|c\|_\infty \le 1$。假设满足以下三种情形之一。

(i) $\mathcal{X} = \{x_1, \ldots, x_N\}$ 是一个有限空间（且对 c 没有额外假设）。

(ii) 存在一个伪度量[3] $\tilde{d}_\mathcal{X}$ 定义在 $\mathcal{X}$ 上，使得当 $\varepsilon > 0$ 足够˜小时，$\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_\mathcal{X}) \lesssim \varepsilon^{-\beta}$，且对于某个 $\beta \in (0, 2)$，对所有 $y \in \mathcal{Y}$，$c(\cdot, y)$ 关于 $d_\mathcal{X}$ 是 1-Lipschitz。

(iii) 对于具有非空内部的紧凸子集 I 和映射，若对每个 i……，函数 c（），y) 在上是 (，1)-赫尔德的，其中某个 d/，对所有 y 均成立。

[3]集合 $\mathcal{M}$ 上的非负函数 $d: \mathcal{M}\times\mathcal{M} \to \mathbb{R}_+$ 若满足以下三个条件，则称为伪度量：对于任意 x、y、$z \in \mathcal{M}$，d(x, x) = 0，d(x, y) = d(y, x)，且 d(x, y) ≤ d(x, z) + d(z, y}。[4]定义在具有非空内部的凸集 $\mathcal{U} \subseteq \mathbb{R}^d$ 上的函数 $f: \mathcal{U} \to \mathbb{R}$，若 $\|f\|_\infty < \Lambda$ 且对于任意 x、$y \in \mathcal{U}$，当 $\|x - y\|$ 时，$|f(x) - f(y)| \le \Lambda\|$，则称其为模数为 $\Lambda \ge 0$ 且$\gamma \in (0, 1]$ 的(γ, Λ)-赫尔德函数。此外，若 f 的每个偏导数都是$(\gamma - 1, \Lambda)$-赫尔德函数，则称 f 为(γ, Λ)-赫尔德函数，其中 $\gamma \in (1, 2]$。如果 $\mathcal{U}$ 不是开集，我们假设存在 f 的一个扩张 ˜f，将其定义在包含 $\mathcal{U}$ 的开凸集上，且 ˜f 在该集合上是(γ, Λ)-赫尔德函数；参见 Hundrieser, Staudt, and Munk (2022)。

设 B，并考虑相对于上的度量 d 的连续模 w：。那么，对于每种情形，都存在某个，使得当足够小时，

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^c, \|\cdot\|_\infty) = \log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad for\, \mathcal{F} = \mathcal{F}(B, w),$$

其中，隐藏常数在 (i) 中取决于自然数集，在 (ii) 中取决于 $\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\mathcal{X}})$，在 (iii) 中取决于 $(\zeta_i, \mathcal{U}_i)_{i=1}^I$。特别地，函数类 $\mathcal{F}^c$ 和 $\mathcal{F}^{cc}$ 是普适 Donsker。

上述命题中关于覆盖数的界对于经验过程 $\sqrt{n}(\mu_n - \mu)$ 和 $\sqrt{m}(\nu_m - \nu)$ 的弱收敛至关重要，也是验证 (JW) 的重要工具。为阐明命题3.1的假设，我们通过一个简单的例子来展示这些假设。有关更多示例，还可参见 Hundrieser, Staudt, and Munk（2022，第 3 节）以及 Hundrieser 等人（2022，第 5 节）。

例 3.2. 假设 $\mathcal{X}$ 和 $\mathcal{Y}$ 是 $\mathbb{R}^3$ 的紧子集，且 $c: \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ 二阶连续可微。将 $\mathcal{X}$ 扩展为一个紧致凸集，并且由于可以对 c 进行缩放，使得 c(·,y) 在 $\mathcal{X}$上满足 (2,1)-赫尔德条件，因此命题3.1适用于这一情形。

为了给出满足 (JW) 和 (JW)* 的充分条件，我们假设总体代价以及经验估计量和自助法估计量，均由底层测度通过一个 Hadamard 方向可微泛函确定。为简化表述，在下述命题中，我们考虑样本量相同、均为 n 的随机变量 $\{X_i\}_{i=1}^n \sim \mu^{\otimes n}, \{Y_i\}_{i=1}^n \sim \nu^{\otimes n}$，其经验测度为 μ_n, ν_n；以及大小为 $k = k(n) = o(n)$ 的自助样本 $\{X_i^b\}_{i=1}^k \sim \mu_n^{\otimes k}, \{Y_i^b\}_{i=1}^k \sim \nu_n^{\otimes k}$，相应的自助经验测度为 $\mu_{n,k}^b, \nu_{n,k}^b$。

命题 3.3（联合弱收敛）。设 $\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{Y}}$ 分别是定义在 $\mathcal{X}$ 和 $\mathcal{Y}$ 上的有界函数类，并假设存在泛函 $\Phi_c: \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \subseteq l^\infty(\mathcal{F}_{\mathcal{X}}) \times l^\infty(\mathcal{F}_{\mathcal{Y}}) \to C(\mathcal{X} \times \mathcal{Y})$，使得对所有 $n,k \in \mathbb{N}$，

$$c = \Phi_c(\mu, \nu), \quad c_n = \Phi_c(\mu_n, \nu_n), \quad and \quad c_{n,k}^b = \Phi_c(\mu_{n,k}^b, \nu_{n,k}^b).$$

如果 Φ_c 在 (μ, ν) 处沿切于 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 的方向 Hadamard 方向可微，且 $\mathcal{F}_{\mathcal{X}} \cup \mathcal{F}^{cc}$ 是 μ-Donsker，而 $\mathcal{F}_{\mathcal{Y}} \cup \mathcal{F}^c$ 是 ν-Donsker，则 (JW) 和 (JW)* 均得到满足。

注 3.4。我们要指出的是，如果泛函 Φ_c 相对于 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 上由弱收敛诱导的拓扑还具有连续性，则可推出，对于每个 $(x, y) \in \mathcal{X} \times \mathcal{Y}$，$c_n(x, y)$ 和 $c_{n,k}^b(x, y)$ 都是可测的；此外，由于 $\mathcal{X}$ 和 $\mathcal{Y}$ 的紧致性，它们在 $C(\mathcal{X} \times \mathcal{Y})$ 中也是可测的。

3.2 假设 (OP)：最优传输计划的唯一性

关于概率测度之间、给定代价函数下最优传输计划唯一性的问题，长期以来一直受到关注，并已由多位作者加以研究。连续情形下的一般条件由 Gangbo 和 McCann（1996年）以及 Levin （ 1999 ）在前人工作的基础上提出。此后，该主题在参考教材 Villani （ 2008 ）第9章和第10章中得到了深入论述；自那以后又取得了进一步进展。

为保证最优传输计划的唯一性，许多研究采用所谓的扭转条件：对于可微代价函数，该条件要求对所有 $x \in \mathcal{X}$，映射 $y \to \nabla_x c(x, y)$ 都是单射。以下命题将基于这一条件的唯一性判据形式化，应能满足读者在许多实际应用中的需求。该结果可由 Villani（2008）中的定理10.28和注释10.33推导得出。

命题3.5。假设 $\mathcal{X}$、$\mathcal{Y}$ 是紧波兰空间，其中 $\mathcal{X} \subseteq \mathbb{R}^d$ 是具有非空内部的欧几里得子集，且 μ 关于勒贝格测度绝对连续。此外，假设 c 在 $\mathcal{X} \times \mathcal{Y}$ 上局部利普希茨连续，对每个 $y \in \mathcal{Y}$，$c(\cdot, y)$ 在 $\mathrm{int}(\mathcal{X})$ 上可微，并且对每个 $x \in \mathcal{X}$，映射 $y \mapsto \nabla_x c(x, y)$ 是单射。那么，最优传输计划 $\pi^\star \in \Pi(\mu, \nu)$ 是唯一的。

尽管在某些情况下，较弱的条件也能保证唯一性（Ahmad、Kim 和 McCann，2011年），但这些更一般的条件通常难以验证得多。尽管如此，除非代价函数具有某种对称性，或在某个区域内为常数，否则通常可以预期最优传输计划具有唯一性。事实上，对于固定的测度，存在一个代价函数集合的剩余集，使得对其中任意代价函数，最优传输计划都是唯一的（McCann 和 Rifford，2016）。

在有限离散情形下，也就是说，当两个底层测度的支撑均为有限个点时，关于最优传输计划唯一性的结果大多基于有限维线性规划理论；详细论述可参见 Klatt、Munk 和 Zemel（2022，第 6 节）。他们等人还给出了若干最优传输计划唯一性的充分条件，这些条件仅取决于代价函数和支撑点，而与测度的权重无关。对于基于欧氏距离的代价，几乎所有支撑点的排列（勒贝格意义下）都满足其条件；然而，若支撑点遵循某种规则或重复模式，则该条件不成立。

3.3 假设 (Sup) 和 (Sup)*：控制经验过程上确界

我们关于经验过程上确界的假设确保，由代价函数估计所导致的可行对偶势函数集合上的波动在渐近意义下可忽略。还需指出，根据引理 6.4 和 6.5，(Sup) 和 (Sup)* 中的上确界是（Borel）可测的。这意味着外概率意义下的收敛实际上是概率收敛。事实上，沿着引理 6.5 的证明思路，并由 $c_{n,m}$ 的可测性可知，对于固定的 $f \in \mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$，映射 $\omega \mapsto \mathbb{G}_n^\mu(f^{cc})$ 和 $\omega \mapsto \mathbb{G}_n^\mu(f^{c_{n,m}, c_{n,m}})$ 均可测。结合 $\mathbb{G}_n^\mu((\cdot)^{cc})$、$\mathbb{G}_n^\mu((\cdot)^{c_{n,m}, c_{n,m}}) \in C_u(\mathcal{F}, \|\cdot\|_\infty)$，由引理 6.4 以及 $(\mathcal{F}, \|\cdot\|_\infty)$ 的紧致性，可推出 $\mathbb{G}_n^\mu((\cdot)^{cc} - (\cdot)^{c_{n,m}, c_{n,m}})$ 及其上确界的可测性。

下文中，我们将推导使假设 (Sup)（以及假设 (Sup)*）成立的充分条件。基于经验过程理论，为了适当地控制上确界

$$\sup_{f \in \mathcal{F}} \mathbb{G}_n^\mu(f^{c_{n,m}, c_{n,m}} - f^{cc}) \quad \text{and} \quad \sup_{f \in \mathcal{F}} \mathbb{G}_n^\nu(f^{c_{n,m}} - f^{c})$$

一种规范做法是对 $\mathcal{F}_{c_{n,m}, c_{n,m}} \cup \mathcal{F}^{cc}$ 和 $\mathcal{F}_{c_{n,m}} \cup \mathcal{F}^{c}$ 施加度量熵界。然而，这类界会对代价估计量 $c_{n,m}$ 施加一定的正则性要求。因此，为了不缩小我们关于代价估计量的研究范围，我们采用与第 2.2 节相同的思想，并用更具正则性的代价估计量 $\tilde{c}_{n,m}$ 对代价估计量 $c_{n,m}$ 进行近似。下面的结果针对我们的情形将这些考虑形式化。其证明依赖于 van der Vaart 和 Wellner (2007) 针对以估计函数类为指标的经验过程所发展的技术。

P命题 3.6。设 $\mathcal{X}$ 和 $\mathcal{Y}$ 为紧波兰空间，并考虑一个连续代价函数 c.

(i) 对随机元素 $c_{n,m} \in C(\mathcal{X} \times \mathcal{Y})$ 假设 (JW)，并取随机元素 $\tilde{c}_{n,m} \in C(\mathcal{X} \times \mathcal{Y})$，满足 $\sqrt{nm/(n+m)}\|c_{n,m} - \tilde{c}_{n,m}\|_\infty \xrightarrow{P} 0$，其中 $n,m = m(n) \to \infty$，且 $m/(n+m) \to \lambda \in (0,1)$，使得对于 $\varepsilon > 0$ 足够小，

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) + \sup_{n \in \mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \textit{with } \alpha < 2. \tag{3.1}$$

那么假设 (Sup) 成立。

(ii) 假设 (JW)，并且对随机元素 $c_{n,k}^b \in C(\mathcal{X} \times \mathcal{Y})$，令 $\tilde{c}_{n,k}^b \in C(\mathcal{X} \times \mathcal{Y})$ 为满足以下条件的随机元素：对于 $n,k = k(n) \to \infty$ 且 $k = o(n)$，有 $\sqrt{k}\|c_{n,k}^b - \tilde{c}_{n,k}^b\|_\infty \xrightarrow{P} 0$，且对足够小的 $\varepsilon > 0$sufficiently small,

$$\log \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) + \sup_{n \in \mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,k}^b \tilde{c}_{n,k}^b}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \textit{with } \alpha < 2. \tag{3.2}$$

那么，假设 (Sup)* 成立。

作为命题 3.6 的直接推论，我们发现，如果代价估计量 $c_{n,\,m}$ 和 $c^{b}_{n,\,k}$ 满足某些确定性正则条件，那么当 n、m、k 足够大时，(Sup) 和 (Sup)* 成立。在大样本情形下，我们进而选择 $\tilde{c}_{n,\,m} := c_{n,\,m}$ 和 $\tilde{c}^{b}_{n,\,k} := c^{b}_{n,\,k}$。

推论 3.7. 设 $\mathcal{X}, \mathcal{Y}$ 为紧波兰空间，考虑连续成本函数 c。假设 c_n 满足 (JW)（且 $c^{b}_{n,k}$ 满足 (JW)* ），并且 c、c_n （以及 $c^{b}_{n,k}$）各自满足命题3.1中针对 $n \geq$ 自然数集、$k \geq K$ 的三个条件之一，其中 $N,K \in \mathbb{N}$ 为随机变量。则 (Sup)（以及 (Sup)*）成立。

因此，若总体代价 c 及其估计量 $c_n, c_{n,k}$ 由某个参数$\theta \in \Theta$ 和估计量 $\theta_n, \theta_{n,k}$确定，且命题3.1的正则性在 θ 的某个开邻域内一致满足，并且这些估计量相合，则推论 3.7 表明，假设 (Sup) 和 (Sup)* 成立。

此外，在对空间 $\mathcal{X}$ 和成本函数 c 作出温和的附加假设后，我们可以构造一个泛函 $\Psi: C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y})$，使得 $\tilde{c}_n := \Psi(c_n)$ 满足熵界 (3.1)，同时满足 $\sqrt{n}\left\|\tilde{c}_{n,m} - c_{n,m}\right\|_\infty \xrightarrow{P} 0$，其中 $n \to \infty$。我们称这样的泛函为正则性提升泛函，因为它提升了代价估计量的正则性阶数。关于正则性提升的细节留待第 5 节讨论。

推论 3.8。设 $\mathcal{X}, \mathcal{Y}$ 为紧波兰空间，并考虑一个连续代价。假设 (JW)（以及 (JW)*）。设 c 满足命题3.1 的三个条件之一。在 (ii) 或 (iii) 情形下，还分别假设满足后续条件 (ii)' 或 (iii)'。

(ii)' 弱极限 $\mathbb{G}^c$ 几乎必然关于 $(\mathcal{X}, \tilde{d}_{\mathcal{X}}) \times \mathcal{Y}$ 连续。

(iii)' 对每个 $i \in \{1, \ldots I\}$，集合 $\mathcal{U}_i \subseteq \mathbb{R}^{d_i}$ 是凸且紧致的，映射 $\zeta_i: \mathcal{U}_i \to \zeta_i(\mathcal{U}_i)$ 是一个同胚，并且 $c_i := c(\zeta_i(\cdot), \cdot): \mathcal{U}_i \times \mathcal{Y} \to \mathbb{R}$ 关于 u 在 $\mathcal{U}_i \times \mathcal{Y}$ 上连续可微，即导数 $\nabla_u c_i: \mathrm{int}(\mathcal{U}_i) \times \mathcal{Y} \to \mathbb{R}^d$ 可以连续延拓至 $\mathcal{U}_i \times \mathcal{Y}$。此外，存在定义在 $\mathcal{X}$ 上的连续单位分解[5] $\{\eta_i\}_{i=1}^{I}$，其 $\mathrm{supp}(\eta_i) \subseteq \zeta_i(\mathcal{U}_i)$。

那么，假设 (Sup)（以及 (Sup)*）成立。

3.4 假设(Don)：在 Θ 上一致的 Donsker 性质

关于 Hundrieser 等人（2022）对固定代价函数 c 下经验最优传输值的分布极限，作者实际上假设函数类 $\mathcal{F}_{cc}$ 和 $\mathcal{F}^c$ 分别是 μ-Donsker 和 ν-Donsker（注 2.4(iii)）。因此，对于定理 2.5 中的一致收敛结果，自然需要并集 $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c^\theta}$ 和 $\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}$ 在$\mathcal{F} = \mathcal{F}(\sup_{\theta\in\Theta}\|c\|_\infty, w)$ 下具有 μ-Donsker 和 ν-Donsker 性质。通过对定义域 $\mathcal{X}$ 的复杂度作出假设，并结合对代价函数施加的正则性条件，可以确保这一条件成立。

命题 3.9（在 Θ 上的普适 Donsker 性质）。设 $\mathcal{X}, \mathcal{Y}$ 是紧波兰空间，并设 (Θ, d_Θ) 是一个度量空间，满足 $\log \mathcal{N}(\varepsilon, \Theta, d_\Theta) \lesssim \varepsilon^{-\alpha}$ for $\alpha < 2$。假设 $c: (\Theta, d_\Theta) \to C(\mathcal{X} \times \mathcal{Y}), \theta \mapsto c_\theta$ 是 1-Lipschitz，并且 $\sup_{\theta\in\Theta}\|c_\theta\|_\infty \leq 1$。考虑以下三种情形之一。

(i) $\mathcal{X} = \{x_1, \ldots, x_N\}$ 是一个有限空间（且对 c 没有额外假设）。

(ii) 对任意 $\theta \in \Theta$，存在一个伪度量 $\tilde{d}_{\theta,\,\mathcal{X}}$ 定义在 $\mathcal{X}$ 上，使得 $\sup_{\theta\in\Theta}\mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\theta,\,\mathcal{X}}) \lesssim \varepsilon^{-\beta}$ 对~$\beta < 2$ 成立，并且对所有 $y \in \mathcal{Y}$，$c_\theta(\cdot, y)$ 在度量 $d_{\theta,\,\mathcal{X}}$ 下是 1-Lipschitz。

(iii) 对于 I，$\mathcal{X} = \bigcup_{i=1}^{I} \zeta_i(\mathcal{U}_i)$，其中 $I \in \mathbb{N}$ 是紧凸子集 $\mathcal{U}_i \subseteq \mathbb{R}^{d_i}$，$d_i \leq 3$ 具有非空内部，且映射 $\zeta_i: \mathcal{U}_i \to \mathcal{X}$ 满足：对每个 $i \in \{1, \ldots..., I\}$ 上的函数 $c_\theta(\zeta_i(\cdot), y)$ 是 $(\gamma_i, 1)$-赫尔德函数（参见脚注 4），其中某个 $\gamma_i \in (d_i/2, 2]$ 对所有 y 均成立 $\in \mathcal{Y}, \theta \in \Theta$。

[5]如果一个集合 $\{\eta_i\}_{i=1}^{I}$ 满足 $\eta_i \in C(\mathcal{X})$，$\eta_i \geq 0$ 对每个 i 均成立，且 $\sum_{i=1}^{I}\eta_i \equiv 1$ 定义在 $\mathcal{X}$上，则称其为连续单位分解。

那么，对于每种情形，都存在某个 $\alpha < 2$，使得

$$\log \mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha} \quad \textit{and} \quad \log \mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}, \|\cdot\|_\infty) \lesssim \varepsilon^{-\alpha}.$$

特别地，$\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c^\theta}, \cup_{\theta\in\Theta}\mathcal{F}^{c^\theta}$ 是普适 Donsker 的，且满足假设（Don）。

命题 3.9 的证明可直接由命题3.1与下文的引理推出，该引理的证明见附录 B.6。

引理 3.10. 设 $\mathcal{X}, \mathcal{Y}$ 为紧波兰空间，且 (Θ, d_Θ) 为度量空间。假设 $c : (\Theta, d_\Theta) \to C(\mathcal{X}\times\mathcal{Y})$ 是 1-Lipschitz 的。那么，对任意 $\varepsilon > 0$，有

$$\max\left(\mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}, \|\cdot\|_\infty), \mathcal{N}(\varepsilon, \cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty)\right) \le \mathcal{N}\left(\frac{\varepsilon}{4}, \Theta, d_\Theta\right) \sup_{\theta\in\Theta} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty\right).$$

3.5 假设（KP）：康托罗维奇势的唯一性

定理 2.5 中经验最优传输过程的一致弱极限表明，它与康托罗维奇势的集合密切相关。特别地，若要使该极限以 $C(\Theta)$ 为支撑，就要求康托罗维奇势 $S_{c_\theta}(\mu, \nu)$ 关于 θ 满足某种连续性性质。关于康托罗维奇势唯一性的假设（KP）是确保这一性质成立的充分条件。

Staudt、Hundrieser 和 Munk（2022）的近期研究全面分析了康托罗维奇势唯一性这一主题，并指出人们通常预期其具有唯一性。更确切地说，对于可微代价函数，若假设某一概率测度支撑在光滑流形上某个连通开集的闭包中，则康托罗维奇势是唯一的。正如其研究中的例 3 所示，当一个测度是离散的而另一个测度具有连通支撑时，即使代价函数连续，也会出现唯一性。如果两个测度都具有不连通支撑，那么只要受限最优传输子问题中的势函数具有唯一性，并且存在一个用 Staudt、Hundrieser 和 Munk 的术语来说的非退化最优传输计划，仍然可以保证唯一性；这意味着两个测度的所有连通分支都通过该最优传输计划相互连接。在底层测度满足温和条件时，可以保证此类最优传输计划的存在（见（3.3））；直观而言，这要求最优传输问题不能被划分为彼此独立的不同子问题。

以下陈述是 Staudt、Hundrieser 和 Munk（2022）理论的一个简单推论，列于此处以便查阅。

命题 3.11. 设 $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ 是一个可微代价函数。考虑具有紧支撑的概率测度 $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$，并假设 $\operatorname{supp}(\mu) = \bigcup_{i\in I} \mathcal{X}_i$ 且 $\operatorname{supp}(\nu) = \bigcup_{j\in J} \mathcal{Y}_j$，其中涉及有限个互不相交的集合。假设每个集合 $\mathcal{X}_i$ 要么是 (i) 连通开集的闭包，(ii) $\mathbb{R}^d$的光滑紧子流形中某个连通开集的闭包，要么是 (iii) 单点。此外，若 $\min(|I|, |J|) \ge 2$，则对所有非空真子集 $I' \subset I$ 和 $J' \subset J$，假设

$$\sum_{i\in I'} \mu(X_i) \neq \sum_{j\in J'} \nu(Y_j), \tag{3.3}$$

那么，μ、ν 和 c 的康托罗维奇势函数唯一（相差一个常数平移）。

该命题可通过验证 Staudt、Hundrieser 和 Munk（2022）中定理 1 的条件得出。具体而言，由于紧致性假设和代价函数的连续性，康托罗维奇势在 $\operatorname{supp}(\nu)$ 上连续；根据代价和集合 $\mathcal{X}_i$ 的假设，子问题上的 Kantorovich 势函数唯一（Staudt、Hundrieser 和 Munk，2022，推论 2）；此外，由于条件 (3.3)，根据（Staudt、Hundrieser 和 Munk，2022，引理 6）可知非退化计划存在。

4 应用

本节将运用第 2 节的理论，深入探讨最优传输相关主题。 本节所有证明均见附录 C。

4.1 基于最优传输的单样本拟合优度检验

Hallin、Mordant 和 Segers（2021）提出，在群作用下使用样本测度与参考测度之间的 Wasserstein 距离进行拟合优度检验。下文简要回顾紧支撑测度的设定。设 $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$ 为紧支撑测度，定义 $\mathcal{Y}$ 为 $\mathrm{supp}(\nu_0)$ 的凸包，并设 $\mathrm{G}_\Theta = \{\mathrm{g}_\vartheta : \vartheta \in \Theta\}$ 为一群可测变换 $\mathrm{g}_\vartheta : \mathbb{R}^d \to \mathbb{R}^d$，其由 $\vartheta \in \Theta \subseteq \mathbb{R}^k$ 参数化，其中 $\mathrm{k} \in \mathbb{N}$。此外，假设对每个 $\vartheta \in \Theta$，映射 $\mathrm{x} \mapsto \mathrm{g}_\vartheta(\mathrm{x})$ 连续，并且映射 $\vartheta \mapsto \mathrm{g}_\vartheta$ 和 $\mathrm{g}_\vartheta \mapsto (\mathrm{g}_\vartheta)_{\#}\nu_0$ 均为双射（这意味着模型参数具有可识别性）。Hallin、Mordant 和 Segers（2021）考虑了以下检验问题：

> 设 G_Θ 为一个群，并定义 $\mathcal{M} = \{\mathrm{g}_{\vartheta\#}\nu_0 : \mathrm{g}_\vartheta \in \mathrm{G}_\Theta\}$。给定来自某个未知分布 $\mu \in \mathcal{P}(\mathcal{X})$ 的独立同分布样本 $\{\mathrm{X}_i\}_{i=1}^n$，其中 $\mathcal{X} \subset \mathbb{R}^d$ 紧致，目标是检验
>
> $$\mathcal{H}_0^* : \mu \in \mathcal{M} \quad against \quad \mathcal{H}_1^* : \mu \notin \mathcal{M}. \tag{4.1}$$

注意，在 $\mathcal{H}_0$ 下的参数 ϑ^*（使得 $(\mathrm{g}_{\vartheta^*})_{\#}\nu_0 = \mu$）是未知的。为检验上述假设——例如，在位置尺度族的分析中，这一检验尤其值得关注——作者建议采用（2-）Wasserstein 距离，也就是说，提出了一种基于其经验版本的检验

$$OT\left(\mu, \nu_0, \left\|g_{\vartheta^*}^{-1}(\cdot) - \cdot\right\|^2\right) = \inf_{\pi \in \Pi(\mu,\nu_0)} \int \left\|g_{\vartheta^*}^{-1}(x) - y\right\|^2 \, \mathrm{d}\pi(x, y).$$

为此，用 μ_n 替代未知测度 μ，并用 $\mathrm{c}_n(\mathrm{x}, \mathrm{y}) = \|\mathrm{g}_{\vartheta_n}^{-1}(\mathrm{x}) - \mathrm{y}\|^2$ 替代代价函数 $\mathrm{c}(\mathrm{x}, \mathrm{y}) = \|\mathrm{g}_{\vartheta^*}^{-1}(\mathrm{x}) - \mathrm{y}\|^2$，其中 $\vartheta_n \in \Theta$ 表示 ϑ^*的一个合适估计量。因此，所提出的检验统计量定义为

$$OT\left(\mu_n, \nu_0, \left\|g_{\vartheta_n}^{-1}(\cdot) - \cdot\right\|^2\right) = \inf_{\pi \in \Pi(\mu_n,\nu_0)} \int \left\|g_{\vartheta_n}^{-1}(x) - y\right\|^2 \, \mathrm{d}\pi(x, y), \tag{4.2}$$

这相当于求解一个具有估计代价函数的最优传输问题。因此，我们可以应用本文的理论推导

$$\sqrt{n}\left(OT\left(\mu_n, \nu_0, \left\|g_{\vartheta_n}^{-1}(\cdot) - \cdot\right\|^2\right) - OT\left(\mu, \nu_0, \left\|g_{\vartheta^*}^{-1}(\cdot) - \cdot\right\|^2\right)\right) \tag{4.3}$$

在 (4.1) 中的原假设 H_0^* 下的极限分布（相关讨论见注 4.4）。此外，我们还能够将其扩展到检验 H_0^* 是否近似成立；在实践中，这通常更为可取（例如，参见 Munk 和 Czado，1998；Dette 和 Munk，1998；Dette 和 Wu，2019）。为此，我们固定一个针对 ϑ^* 的估计程序，即选择一个特定的估计量 ϑ_n（取值于 Θ）来估计 ϑ^*，并将其总体量记为 $\vartheta^o \in \Theta$（在 $\mathcal{H}_0^*$ 下，我们假设 $\vartheta^* = \vartheta^o$）。接下来，我们考虑以下检验问题：

> 设 G_Θ 为一个群。给定来自某个未知分布 $\mu \in \mathcal{P}(\mathcal{X})$ 的独立同分布样本 $\{\mathrm{X}_i\}_{i=1}^n$，其中 $\mathcal{X} \subset \mathbb{R}^d$ 紧致，目标是检验某个预先指定的 $\Delta > 0$ 假设
>
> $$\mathcal{H}_0 : W_2(\mu, (g_{\vartheta^o})_{\#}\nu_0) \le \Delta \quad versus \quad \mathcal{H}_1 : W_2(\mu, (g_{\vartheta^o})_{\#}\nu_0) > \Delta. \tag{4.4}$$

为了针对上述问题构造检验，我们必须推导 $\mu \notin \mathcal{M}$ 假设下 (4.3) 的分布极限。为此，我们采用第2节和第3节中的理论。推导 (4.3) 分布极限的第一步，是证明由 C_Θ 诱导的代价具有赫尔德正则性（回顾脚注 4）： $\Theta \to C(\mathcal{X} \times \mathcal{Y})$， $\vartheta \mapsto ((x, y) \mapsto \|g_\vartheta^{-1}(x) - y\|^2)$ 在 ϑ^o 附近。

引理 4.1。设 $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ 为紧致集，并用 $C(\mathcal{X}, \mathbb{R}^d)$ 表示从 $\mathcal{X}$ 到 $\mathbb{R}^d$ 的连续函数空间。假设 $K_\Theta : \Theta \subseteq \mathbb{R}^k \to C(\mathcal{X}, \mathbb{R}^d)$， $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ 在 ϑ^o 附近连续。那么，存在 ϑ^o 的一个开邻域 $U \subseteq \Theta$（相对于相对拓扑而言）以及某个 $\Lambda \geq 0$，使得对任意 $x \in \mathcal{X}$ 和 $\vartheta \in U$，代价函数 $C_\Theta(\vartheta)(x, \cdot) := \|g_\vartheta^{-1}(x) - \cdot\|^2$ 在 $\mathcal{Y}$ 上是 $(2, \Lambda)$-赫尔德的。

接下来，我们验证 K_Θ 在 ϑ^o 处的 Hadamard 可微性蕴含代价参数化映射 $C_\Theta : \Theta \to C(\mathcal{X} \times \mathcal{Y})$， $\vartheta \mapsto ((x, y) \mapsto \|g_\vartheta^{-1}(x) - y\|^2)$ 在 ϑ^o 处的 Hadamard 可微性。为此，我们还需作出以下假设。

(G) 对于 ϑ^o，存在 $m_{\vartheta^o} > 0$，使得对于某个 ϑ^o 的邻域中的所有 $\vartheta' \in \Theta$，

$$\sup_{x \in \mathbb{R}^d} \frac{\left\| g_{\vartheta'}^{-1}(x) - g_{\vartheta^o}^{-1}(x) \right\|}{1 + \left\| g_{\vartheta^o}^{-1}(x) \right\|} \leq m_{\vartheta^o} \left\| \vartheta' - \vartheta^o \right\|.$$

例如，位置尺度族和仿射变换满足这一条件。Hallin、Mordant 和 Segers（2021）使用了上述假设的全局版本，即要求该条件对任意 ϑ 都成立，而不仅仅是 ϑ^0，以确保其上述拟合优度检验的一致性。

引理 4.2. 假设函数 $K_\Theta : \Theta \to C(\mathcal{X}, \mathbb{R}^d)$， $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ 在 ϑ^o 处沿 Θ Hadamard 可微；也就是说，对于任意序列 $(\vartheta^o + t_n h_n)_{n \in \mathbb{N}} \subseteq \Theta$，若 $t_n \searrow 0$ 且 $h_n \to h \in \mathbb{R}^k$，当 $n \to \infty$ 时，

$$\lim_{n \to \infty} \left\| \frac{K_\Theta(\vartheta^o + t_n h_n) - K_\Theta(\vartheta^o)}{t_n} - D^H_{|\vartheta^o} K_\Theta(h) \right\|_\infty = 0,$$

其中 $D^H_{|\vartheta^o} K_\Theta(h) : \mathcal{X} \to \mathbb{R}^d$ 是连续函数。那么，若假设 (G) 成立，C_Θ 在 ϑ^o 处沿 Θ Hadamard 可微，其导数为 $D^H_{|\vartheta^o} C_\Theta(h) \in C(\mathcal{X} \times \mathcal{Y})$，由下式给出

$$D^H_{|\vartheta^o} C_\Theta(h) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}, \quad (x, y) \mapsto 2\left\langle D^H_{\vartheta^o} K_\Theta(h)(x), g_{\vartheta^o}^{-1}(x) - y \right\rangle.$$

此外，如果对于 $n \to \infty$， $\sqrt{n}(\vartheta_n - \vartheta^o) \Rightarrow \mathbb{G}^\vartheta$，则对于 $c_n := C_\Theta(\vartheta_n)$ 和 $c := C_\Theta(\vartheta)$，有

$$\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c := \left(2\left\langle D^H_{\vartheta^o} K_\Theta(\mathbb{G}^\vartheta)(x), g_{\vartheta^o}^{-1}(x) - y \right\rangle \right)_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \quad \text{属于 } C(\mathcal{X} \times \mathcal{Y})。 \tag{4.5}$$

在上述命题的条件下，定理 2.2 给出的主要结果表明，在假设 $\mu \notin \mathcal{M}$ 的情况下，统计量 $OT(\mu_n, \nu, c_n)$ 具有（通常）非退化的极限分布。特别地，这使我们能够针对式 (4.4) 中给出的原假设构造一个渐近水平为 α 的检验（精确构造见 Munk 和 Czado，1998）。

命题4.3. 设 $\nu_0 \in \mathcal{P}(\mathbb{R}^d)$ 对于 $d \leq 3$具有紧致支撑，并定义 $\mathcal{Y}$ 为 $\operatorname{supp}(\nu_0)$ 的凸包，且设 $\mathcal{X} \subseteq \mathbb{R}^d$ 为紧致集。假设 (G) 成立，并设 $G_\Theta : \Theta \to C(\mathcal{X}, \mathbb{R}^d)$， $\vartheta \mapsto (x \mapsto g_\vartheta^{-1}(x))$ 在 ϑ^o 附近连续，且在 ϑ^o 处 Hadamard 可微。对引理 4.2 中的$\Lambda \geq 0$，定义函数类

$$\mathcal{F} := \left\{ f : \mathcal{Y} \to \mathbb{R} \;\middle|\; \|f\|_\infty \leq \Lambda + 1, |f(y) - f(y')| \leq 2\Lambda \left\| y - y' \right\| \text{ for all } y, y' \in \mathcal{Y} \right\}.$$

那么，函数类 $\mathcal{F}^{C_\Theta(\vartheta^o)}$ 在 $\mathcal{X}$ 上是普适 Donsker。此外，对于独立同分布随机变量{X $_i\}_{i=1}^n \sim \mu^{\otimes n}$，考虑一个可测估计量 ϑ_n，并假设当 n $\to \infty$ 时联合弱收敛，

$$\sqrt{n}\begin{pmatrix}\mu_n - \mu \\ \vartheta_n - \vartheta^o\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathbb{G}^\mu \\ \mathbb{G}^\vartheta\end{pmatrix} \quad in\ \ell^\infty(\mathcal{F}^{C_\Theta(\vartheta^o)}) \times \mathbb{R}^k. \tag{4.6}$$

那么，对于 $c_n := C_\Theta(\vartheta_n)$ 和 $c := C_\Theta(\vartheta)$，并将 (4.5) 中的极限记为 $\mathbb{G}^c$，可得

$$\sqrt{n}\,(OT\,(\mu_n, \nu_0, c_n) - OT\,(\mu, \nu_0, c)) \rightsquigarrow \inf_{\pi \in \Pi_c^\star(\mu,\nu_0)} \pi(\mathbb{G}^c) + \sup_{f \in S_c(\mu,\nu_0)} \mathbb{G}^\mu(f^c),$$

其中，$S_c(\mu, \nu_0)$ 表示 sup f$_{\in\mathcal{F}}\mu($ f c) $+ \nu_0($ f cc) 的最优解集合。

备注 4.4. 关于分布极限，有几点需要说明。

(i) 鉴于函数类 $\mathcal{F}^{C_\Theta(\vartheta^o)}$ 是普适 Donsker 类，因而也是 μ-Donsker 类，并假设 $\sqrt{n}(\vartheta_n - \vartheta^o)$ 在分布上收敛，则 (4.6) 所要求的联合收敛条件相当宽松。事实上，如果 $\sqrt{n}(\vartheta_n - \vartheta^o)$ 渐近地可表示为经验过程的某个适当线性泛函，即存在渐近影响函数 $\psi \in L^2(\mu)$（参见 van der Vaart 1998，第 58 页），则由于并集$\mathcal{F}^{cc} \cup \{\psi\}$ 是 μ-Donsker 类，联合收敛便随之成立。

(ii) 我们想指出，即使 $\mu \in \mathcal{M}$，命题4.3 仍然成立。然而，在这一假设下可得 $(g_{\vartheta^o}^{-1})_\#\mu = \nu_0$，这意味着 μ 与 ν_0 之间对应的最优传输计划为 $\pi = (\mathrm{Id}, g_{\vartheta^o}^{-1}(\cdot))_\#\mu$。因此，根据 (4.5)，随机过程 $\mathbb{G}^c$ 在 π 的支撑上消失，极限中的第一项退化。此外，如果 ν_0 的支撑是连通的，由于此时康托罗维奇势在相差一个常数平移的意义下唯一（Staudt、Hundrieser 和 Munk，2022，推论 2），且几乎必然为常数（Hundrieser 等人，2022，推论 4.6(i)），因此在这种情形下，相应的极限分布是退化的。相反，如果 ν_0 的支撑不连通，则存在非恒定的康托罗维奇势（Staudt、Hundrieser 和 Munk，2022，引理 11），从而得到非退化的极限。

（iii）针对单样本情形介绍的内容也可以推广到两种经验测度分别或联合进行变换的情形。例如，可以选择马氏距离 $(x - y)^\top \Sigma^{-1}(x - y)$ 作为代价函数，其中 Σ^{-1}需要估计，且例如可以是一个对角矩阵。由于最优传输值对仿射变换不具有不变性，对变量进行重新缩放可确保没有任何一个分量相较于其他分量对代价函数产生过大的影响。

4.2 带有嵌入不变性的最优传输

与上一节的思路类似，另一支文献（Alvarez-Melis、Jegelka 和 Jaakkola，2019；Grave、J oulin 和 Berthet，2019）旨在通过考虑以下形式，使最优传输对一类变换 $\mathcal{T}$ 具有不变性，其中 τ：对于每个 $\tau \in \mathcal{T}$，$\mathbb{R}^d \to \mathbb{R}^d$ 连续可微

$$\inf_{\tau \in \mathcal{T}} OT\left(\mu, \nu, \|\cdot - \tau(\cdot)\|^2\right) = \inf_{\tau \in \mathcal{T}} \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x - \tau(y)\|^2 \mathrm{d}\pi(x, y). \tag{4.7}$$

这种距离在许多情境下都很有用，例如词嵌入问题或蛋白质比对。如果所考虑的变换类是旋转集合，那么基于该距离的分析称为瓦瑟斯坦–普罗克鲁斯特分析（Grave、Joulin 和 Berthet，2019；Jin、Liu 和 Xia，2021）。定理 2.6 为优化问题 (4.7) 的经验版本提供了统计推断所需的工具。

命题 4.5。考虑一个变换集合 $(\mathcal{T}, \mathrm{d}_{\mathcal{T}})$，它是一个紧致度量空间，且 $\log\ \mathcal{N}(\varepsilon, \mathcal{T}, \mathrm{d}_{\mathcal{T}}) \lesssim \varepsilon^{-\alpha}$ for $\alpha < 2$。令 $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^d$ 为紧子集，并假设泛函 $\mathrm{c} : (\mathcal{T}, \mathrm{d}_{\mathcal{T}}) \to \mathrm{C}(\mathcal{X} \times \mathcal{Y})$, $\tau \mapsto \mathrm{c}_\tau$，其中 $\mathrm{c}_\tau(\mathrm{x}, \mathrm{y}) = \| \mathrm{x} - \tau(\mathrm{y})\|^2$，对某个 $\mathrm{L} \geq 0$是 L-利普希茨的。此外，假设对于 $\mathcal{X}$ 和 $\{\mathrm{c}_t\}_{t\in\mathcal{T}}$，可取命题 3.9 中的任一设定，并令 $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，使得 μ 或 ν 的支撑是 $\mathbb{R}^d$中的一个连通开集的闭包。那么，对于 $\{\mathrm{X}_i\}_{i=1}^n \sim \mu^{\otimes n}$ 和$\{\mathrm{Y}_i\}_{i=1}^m \sim \nu^{\otimes m}$，分别取 $\mathrm{n}, \mathrm{m} \to \infty$，使得 $\mathrm{m}/(\mathrm{n} + \mathrm{m}) \to \lambda \in (0, 1)$，有

$$\sqrt{\frac{nm}{n+m}}\Big(\inf_{\tau\in\mathcal{T}} OT(\mu_n, \nu_m, c_\tau) - \inf_{\tau\in\mathcal{T}} OT(\mu, \nu, c_\tau)\Big) \rightsquigarrow \inf_{\tau\in S_-(\mathcal{T},\mu,\nu)} \sqrt{\lambda}\, \mathbb{G}^\mu(f_\tau^{c_\tau c_\tau}) + \sqrt{1-\lambda}\, \mathbb{G}^\nu(f_\tau^{c_\tau}),$$

其中，$\mathrm{f}_\tau \in \mathrm{S}_{\mathrm{c}_\tau}(\mu, \nu)$表示$\mu$与 ν 之间、代价为c_τ的 $\tau \in \mathrm{S}_-(\mathcal{T}, \mu, \nu)$的康托罗维奇势。

如前所述，可以放宽 $\mathcal{T}$ 为紧致的要求，改为假设估计得到的最优变换序列 τ_n 以趋于 1 的概率包含在某个紧集内（命题 2.8）。在 $\mathcal{T}$ 由微分同胚构成的情形下，根据 Nies、Staudt 和 Munk 的引理 1（2021），我们有

$$\inf_{\tau\in\mathcal{T}} OT\left(\mu, \nu, \|\cdot - \tau(\cdot)\|^2\right) = \inf_{\tau\in\mathcal{T}} OT\left(\mu, (\tau^{-1})_\# \nu, \|\cdot - \cdot\|^2\right) = \inf_{\tau\in\mathcal{T}} W_2^2\left(\mu, (\tau^{-1})_\# \nu\right),$$

利用 Bernton 等人（2019）的结果，可以在各种情形下验证经验极小化解 τ_n 的收敛。

注 4.6（瓦瑟斯坦–普罗克鲁斯特）。在对测度作出温和正则性假设的条件下，上述命题可应用于特殊正交群 $\mathcal{T} := \mathrm{SO(d)}$，其中 $\mathrm{d} \leq 3$。事实上，取 $\mathcal{X}$、$\mathcal{Y}$ 为该设定中的紧凸集，即可满足命题 3.9 的情形 (iii)，从而断言 (Don)。此外，若 μ 或 ν 的支撑是 $\mathbb{R}^d$中的某个连通开集的闭包，则 (KP) 成立，定理 2.6 的分布极限也随之成立。

4.3 混合分布的草图瓦瑟斯坦距离

最近，Delon 和 Desolneux（2020）以及 Bing、Bunea 和 Niles-Weed（2022）研究了（高斯）混合分布之间的一种距离。这些分布在统计学和机器学习中无处不在，参见 McLachlan、Lee 和 Rathnayake（2019）及其中的参考文献。理解这种距离的一种方式是从离散测度之间的 Wasserstein 距离出发，但不再使用点与点之间的代价函数，而是用分布替代这些点，因此必须选择分布之间的代价。在正式定义这一概念之前，回顾一下：对于一个由 $\mathcal{A} := (\mathrm{A}_1, \ldots, \mathrm{A}_K)$ 构成的有限基数为 K 的分布集合，混合分布 r 是由 $\mathcal{A}$ 中的分量构成的凸组合，可由向量 $\alpha \in \Delta_K$ 给出，即 $\mathrm{r} = \mathrm{P}_{i=1}^K \alpha_i \mathrm{A}_i$，其中 Δ_K 是 $\mathbb{R}^K$ 中的概率单纯形。给定 $\mathcal{A}$ 的混合分量之间的距离 $\mathrm{d}: \mathcal{A} \times \mathcal{A} \to \mathbb{R}_+$，上述作者将权重为 α 和 β 的两个混合分布之间的草图瓦瑟斯坦距离定义为

$$W(\alpha, \beta, d) := \inf_{\pi\in\Pi(\alpha,\beta)} \sum_{k,\ell=1}^K \pi_{k,\ell} d(\mathrm{A}_k, \mathrm{A}_\ell),$$

其中，下确界取遍耦合集合中的元素

$$\Pi(\alpha, \beta) = \left\{ \pi \in \Delta_{K\times K} \;\middle|\; \begin{array}{ll} \sum_{\ell=1}^K \pi_{k,\ell} = \alpha_k, & \text{for all } k \in \{1, \ldots, K\} \\ \sum_{k=1}^K \pi_{k,\ell} = \beta_\ell, & \text{for all } \ell \in \{1, \ldots, K\} \end{array} \right\}.$$

利用本文提出的理论，可以理解该距离估计量的波动。以下命题对此进行了形式化表述。

命题 4.7. 设 $(\alpha_n, \beta_n, \mathrm{d}_n) \in \Delta_K \times \Delta_K \times \mathbb{R}_+^{K^2}$ 分别是 $\alpha, \beta, \mathrm{d}$ 的可测估计量。此外，对于正序列 $(\mathrm{a}_n)_{n\in\mathbb{N}}$，若 $\lim_{n\to\infty} \mathrm{a}_n = \infty$，则假设对 $\mathrm{n} \to \infty$ 有

$$a_n \begin{pmatrix} \alpha_n - \alpha \\ \beta_n - \beta \\ d_n - d \end{pmatrix} = a_n \begin{pmatrix} (\alpha_{n,k} - \alpha_k)_{k=1}^K \\ (\beta_{n,k} - \beta_k)_{k=1}^K \\ (d_n(\mathrm{A}_k, \mathrm{A}_\ell) - d(\mathrm{A}_k, \mathrm{A}_\ell))_{l,k=1}^K \end{pmatrix} \rightsquigarrow \begin{pmatrix} \mathbb{G}^\alpha \\ \mathbb{G}^\beta \\ \mathbb{G}^d \end{pmatrix} \quad \textit{in } \mathbb{R}^{2K+K^2}, \tag{4.8}$$

where $(\mathbb{G}^\alpha, \mathbb{G}^\beta, \mathbb{G}^d)$ 表示 $\mathbb{R}^{2K+K^2}$ 上一个紧（可能是非高斯的）随机变量。于是

$$a_n\Big(W(\alpha_n, \beta_n, d_n) - W(\alpha, \beta, d)\Big) \rightsquigarrow \inf_{\pi \in \Pi_d^\star(\alpha,\beta)} \langle \pi, \mathbb{G}^d \rangle + \sup_{f \in S_d(\alpha,\beta)} \langle f^{dd}, \mathbb{G}^\alpha \rangle + \langle f^d, \mathbb{G}^\beta \rangle.$$

证明沿用证明定理 2.2 的相同思路，因此从略，参见注 2.4 (iv)。在此背景下，测度估计量 (α_n, β_n) (α, β) 的弱收敛要求在概率意义下成立，这是由假设 (4.8) 推出的，因为总体测度及其估计量均支撑在有限个点上。

在 Bing、Bunea 和 Niles-Weed（2022）中，他们针对代价的渐近波动相较于估计测度可忽略的情形，得到了分布极限。命题 4.7 推广了他们的结果，同时涵盖代价基于同一批数据估计且以相同速率收敛的情形。最后，我们强调，高斯混合这一在应用中尤为重要的情形也涵盖在我们的理论之中。尽管如此，该情形下仍缺乏关于混合参数估计量的分布结果，这阻碍了进一步发展，值得后续研究。

4.4 切片最优传输

第 2.3 节的理论也使我们能够分析切片最优传输量，并补充或扩展现有文献中的结果（Goldfeld 等，2022；Manole、Balakrishnan 和 Wasserman，2022；Xi and Niles-Weed，2022；Xu 和 Huang，2022）。下文将对此进行形式化表述。对于两个博雷尔概率测度 $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$，$1 \le \mathrm{p} < \infty$ 阶的平均切片和最大切片 Wasserstein 距离分别定义为

$$\underline{W}_p(\mu, \nu) := \left(\int_{\mathbb{S}^{d-1}} OT(\mathfrak{p}_\#^\theta \mu, \mathfrak{p}_\#^\theta \nu, |\cdot - \cdot|^p) \, \mathrm{d}\sigma(\theta) \right)^{\frac{1}{p}} \text{ and } \overline{W}_p(\mu, \nu) := \max_{\theta \in \mathbb{S}^{d-1}} \left(OT(\mathfrak{p}_\#^\theta \mu, \mathfrak{p}_\#^\theta \nu, |\cdot - \cdot|^p) \right)^{\frac{1}{p}},$$

其中，$\mathfrak{p}^\theta \colon \mathbb{R}^d \to \mathbb{R}$ 是投影映射 $\mathrm{x} \mapsto \theta^T \mathrm{x}$，而 σ 表示单位球面 $\mathbb{S}^{d-1}$ 上的均匀分布。根据 Nies、Staudt 和 Munk（2021）中的引理 1，注意到对任意 $\theta \in \mathbb{S}^{d-1}$，都有

$$OT(\mathfrak{p}_\#^\theta \mu, \mathfrak{p}_\#^\theta \nu, |\cdot - \cdot|^p) = OT(\mu, \nu, |\mathfrak{p}^\theta(\cdot) - \mathfrak{p}^\theta(\cdot)|^p),$$

这使我们能够在第 2.3 节的框架下考察切片 Wasserstein 量，并由定理 2.5–2.7 得到以下结果。

命题 4.8. 设 $\mathrm{p} \ge 1$、$\mathrm{d} \ge 2$，并对 $\theta \in \mathbb{S}^{d-1}$ 定义代价 $\mathrm{c}_\theta \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, (\mathrm{x}, \mathrm{y}) \mapsto |\mathfrak{p}^\theta(\cdot) - \mathfrak{p}^\theta(\cdot)|^p$。此外，取紧支撑概率测度 μ、$\nu \in \mathcal{P}(\mathbb{R}^d)$，其经验测度分别为 μ_n, ν_m。对于所有断言，令 $\mathrm{n}, \mathrm{m} \to \infty$，且 $\mathrm{m}/(\mathrm{n} + \mathrm{m}) \to \lambda \in (0, 1)$。

（i）假设对于任意 $\theta \in \mathbb{S}^{d-1}$，康托罗维奇势集合 $S_{c_\theta}(\mu, \nu)$ 在常数平移意义下是唯一的。那么，对于任意 $\theta \in \mathbb{S}^{d-1}$ 选取 $\mathrm{f}_\theta \in S_{c_\theta}(\mu, \nu)$ 后，可得

$$\sqrt{\frac{nm}{n+m}}\Big(OT(\mu_n, \nu, c_\theta) - OT(\mu, \nu, c_\theta)\Big)_{\theta \in \mathbb{S}^{d-1}} \rightsquigarrow \Big(\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}) \Big)_{\theta \in \mathbb{S}^{d-1}} \quad \textit{in } C(\mathbb{S}^{d-1})$$

（ii）假设与（i）中的相同。那么，可得

$$\sqrt{\frac{nm}{n+m}}\left(\underline{W}_p^p(\mu_n,\nu_m)-\underline{W}_p^p(\mu,\nu)\right)\rightsquigarrow\int_{\mathbb{S}^{d-1}}\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})+\sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta})\,\mathrm{d}\theta.$$

(iii) 无需假设康托罗维奇势函数的唯一性，即可得出 th *at*

$$\sqrt{\frac{nm}{n+m}}\left(\overline{W}_p^p(\mu_n,\nu_m)-\overline{W}_p^p(\mu,\nu)\right)\rightsquigarrow\sup_{\substack{\theta\in S_+(\mathbb{S}^{d-1},\mu,\nu)\\ f_\theta\in S_{c_\theta}(\mu,\nu)}}\sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta})+\sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}).$$

将命题 4.8 与文献中针对 $\mathrm{p}>1$ 的结果进行比较，可以发现，在略弱的假设下即可得到 Goldfeld 等人（2022）以及 Xi 和 Niles-Weed（2022）的结果。对于两类经验切片 Wasserstein 距离的分析，Goldfeld 等人（2022）要求底层测度具有紧凸支撑。此外，在 Xi 和 Niles-Weed （ 2022 ）关于切片最优传输过程的一致中心极限定理中，他们假设对每个 $\mathrm{u}\in\mathbb{S}^{d-1}$ ，投影测度之一具有紧致连通支撑。这些条件足以保证康托罗维奇势函数的唯一性，但对于支撑不连通的测度，也可以保证唯一性（参见命题 3.11，以及更一般的 Staudt、Hundrieser 和 Munk 2022）。命题 4.8(ii) 也补充了 Manole、Balakrishnan 和 Wasserman（2022）关于修剪切片 Wasserstein 距离的结果，因为我们不要求存在密度，只要求底层测度具有紧支撑。

对于特殊情形 $\mathrm{p}=1$，与我们的结果不同，Goldfeld 等人（2022）以及 Xu 和 Huang（ 2022）关于平均切片 Wasserstein 距离和最大切片 Wasserstein 距离的分布极限并不要求康托罗维奇势函数的唯一性。此外，通过施加适当的矩条件，他们的理论仍适用于非紧支撑测度。他们的方法关键在于将 1-Wasserstein 距离刻画为关于 Lipschitz 函数的积分概率度量这一特殊性质（Villani，2008，备注 6.5）；而在我们的一般理论中，并未利用这一性质。不过，在康托罗维奇势函数具有唯一性的情况下——例如，当一个测度是离散的，而另一个测度具有连通支撑且绝对连续时（Staudt、Hundrieser 和 Munk，2022，例 3）——命题 4.8(i) 断言切片最优传输过程在 $\mathrm{C}(\mathbb{S}^{d-1})$ 中弱收敛。

4.5 最优传输的稳定性分析

除统计应用外，我们关于代价弱收敛时经验最优传输值的理论，还能对代价和测度联合扰动下的最优传输问题 (1.1) 进行确定性稳定性分析，这一点也可能具有独立的研究价值，例如从优化的角度来看。更准确地说，我们将在下文证明，对于所有可容许方向，最优传输值在 $(\mu,\nu,\mathrm{c})\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times\mathrm{C}(\mathcal{X}\times\mathcal{Y})$ 上 Gâteaux 可微。这拓展了有限维线性规划的经典稳定性结果（Gal 和 Greenberg，1997，定理 3.1）；该结果涵盖了支撑在有限多个点上的概率测度对应的最优传输问题。

命题 4.9. 设 $\mu\in\mathcal{P}(\mathcal{X})$、$\nu\in\mathcal{P}(\mathcal{Y})$ 和 $\mathrm{c}\in\mathrm{C}(\mathcal{X}\times\mathcal{Y})$ 固定。对充分小的 $\mathrm{t}>0$，定义量 $\mu_t=\mu+\mathrm{t}\Delta^\mu$ 和 $\nu_t=\nu+\mathrm{t}\Delta^\nu$，其中 $\Delta^\mu\in(\mathcal{P}(\mathcal{X})-\mu)$ 和 $\Delta^\nu\in(\mathcal{P}(\mathcal{Y})-\nu)$ 分别成立。此外，设 $\mathrm{c}_t=\mathrm{c}+\mathrm{t}\Delta^c$，其中某个 $\Delta^c\in\mathrm{C}(\mathcal{X}\times\mathcal{X})$。于是可得

$$\lim_{t\searrow 0}\frac{1}{t}\left(OT(\mu_t,\nu_t,c_t)-OT(\mu,\nu,c)\right)=\inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\Delta^c)+\sup_{f\in S_c(\mu,\nu)}\Delta^\mu(f^{cc})+\Delta^\nu(f^c).$$

注 4.10（关于 Hadamard 方向可微性）。由于可容许方向的集合 $(\mathcal{P}(\mathcal{X})-\mu)\times(\mathcal{P}(\mathcal{Y})-\nu)\times\mathrm{C}(\mathcal{X}\times\mathcal{Y})$ 不是赋范向量空间，一般而言，我们无法据此推断

Hadamard 方向可微性；要得到这一结论，还需证明最优传输问题关于测度 μ、ν 及代价函数 c 具有利普希茨连续性。

采用与命题 4.9 相同的证明策略，需要我们证明对于任意序列 $(\mu_n, \nu_n, c_n) = (\mu + t_n\Delta_n^\mu, \nu + t_n\Delta_n^\nu, c + t_n\Delta_n^c) \in \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \times C(\mathcal{X}\times\mathcal{Y})$，其中 $t_n \searrow 0$ and $(\Delta_n^\mu, \Delta_n^\nu, \Delta_n^c) \to (\Delta^\mu, \Delta^\nu, \Delta^c)$ 属于 $l^\infty(\mathcal{F}_{cc}) \times l^\infty(\mathcal{F}_c) \times C(\mathcal{X}\times\mathcal{Y})$,

$$\sup_{f\in\mathcal{F}} \left|\Delta_n^\mu(f^{c_nc_n} - f^{cc}) + \Delta_n^\nu(f^{c_n} - f^c)\right| \to 0. \tag{4.9}$$

证明这一点仍然是一个挑战；若能做到，我们便可在定理2.2和定理2.10的表述中分别省去条件 (Sup) 和 $(\text{Sup})^*$。此类尝试面临的另一个挑战是，任意这样的序列 (μ_n, ν_n) 未必会在 $n \to \infty$ 时弱收敛到 (μ, ν)；这与我们的证明相关，因为由 $l^\infty(\mathcal{F}^{cc}) \times l^\infty(\mathcal{F}_c)$ 诱导的拓扑可能过弱。

虽然可能可以通过选择一种足够强、能够度量测度弱收敛的范数，证明最优传输问题关于测度和代价联合具有 Hadamard 方向可微性，但泛函δ方法必然要求经验过程在该范数下弱收敛，并施加额外条件。文献中已知，范数的选择存在类似的权衡（参见 Dudley 1990, p.76；Jourdain 和 Tse，2021）。

5 正则性提升泛函

本节中，我们构造正则性提升映射，即连续映射 $\Psi: C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$，使得对于满足 $\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c$ 的可测估计量 c_n，当 $n \to \infty$ 时，有

$$(i)\ \sqrt{n}(c_n - \Psi(c_n)) \xrightarrow{P} 0 \quad \text{and} \quad (ii)\ \Psi(c_n) \text{ fulfills certain regularity properties.} \tag{5.1}$$

基于最优传输值关于代价函数的利普希茨连续性（引理 6.2），条件 (i) 使我们能够用具有特定正则性的代价估计量替代原代价估计量，从而有效地“提升”其正则性。这类映射在我们的研究中有两个特别有用的应用。首先，它使我们能够在定理 2.2 的证明中假设代价估计量有适当的界，且具有与总体成本函数相同的连续模（参见推论 5.4）。这是运用引理 2.1 的重要一步。此外，正则性提升这一概念也有助于证明推论 3.8；为此，我们运用命题 3.6，并针对一个合适的正则性提升映射设定 $\tilde{c}_n := \Psi(c_n)$。因此，这些映射是进行分布极限理论分析的有效工具。

下面的结果给出第一组条件，以确保满足 (5.1) 的条件 (i)。本节后续所有结果的证明，以及本结果的证明，详见附录 D。

命题 5.1. 设 $\mathcal{X}, \mathcal{Y}$ 为紧波兰空间，且 $c_n \in C(\mathcal{X}\times\mathcal{Y})$ 是一个（博雷尔可测的）随机序列，使得 $a_n(c_n - c)$ $\mathcal{L}$ 属于 $C(\mathcal{X}\times\mathcal{Y})$，其中 $c \in C(\mathcal{X}\times\mathcal{Y})$，并且 $(a_n)_{n\in\mathbb{N}}$ 满足 $a_n \to \infty$，对 $n \to \infty$ 成立。设 $U \subseteq C(\mathcal{X}\times\mathcal{Y})$ 是一个线性子空间，且 $\mathcal{L}$ 几乎必然属于 U。那么，若 $\Psi: C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$ 在 c 附近连续，在 f 处 Hadamard 方向可微，且其导数满足 $D_c^H\Psi|_U = \mathrm{Id}_U$ 以及 $\Psi(c) = c$，则对 $n \to \infty$ 有

$$a_n(c_n - \Psi(c_n)) \xrightarrow{P} 0 \quad \textit{for } n \to \infty.$$

值得注意的是，在 Ψ 是 Hadamard 可微且 $D_f^H\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$ 的情况下，可以选择 $U = C(\mathcal{X}\times\mathcal{Y})$，此时关于极限 $\mathcal{L}$ 的条件便不再需要。

为得出多种所需的正则性（如 (5.1) 的 (ii) 所要求），我们将在以下小节中定义若干映射，使其满足命题 5.1 的条件。此外，我们还给出适当的度量熵界，这些界对 $\mathcal{F}^\Psi(\tilde{c})\Psi(\tilde{c})$ 而言不依赖于 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$。

5.1 正则性提升至确定性有界性

考虑紧波兰空间 $\mathcal{X}, \mathcal{Y}$，并令 $c \in C(\mathcal{X} \times \mathcal{Y})$ 为连续成本函数，满足$\|c\|_\infty \le 1$。我们将有界性的正则性提升泛函定义为

$$\Psi_{\mathrm{bdd}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}), \quad \tilde{c} \mapsto \Big((x, y) \mapsto \max(\min(\tilde{c}(x, y), 2), -2)\Big).$$

命题 5.2。在上述设定下，$\Psi = \Psi_{\mathrm{bdd}}$ 满足 $\Psi(c) = c$，具有连续性，并且在 c 处 Hadamard 可微，且 $\mathrm{D}^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。特别地，如果 $\mathcal{X}$ 是一个有限空间，则对于 $\mathcal{Y}$ 上的任意一致有界函数类 $\mathcal{G}$，我们得到

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim |\log(\varepsilon)|.$$

因此，在分析具有估计代价函数的经验最优传输值时，我们可以不失一般性地假设，代价估计量由一个确定性常数有界，而该常数取决于总体代价。下面，我们将对紧（伪）度量空间上代价估计量的连续模证明类似的结论。

5.2 正则性提升至凹连续模与利普希茨连续性

考虑紧波兰空间 $\mathcal{X}$、$\mathcal{Y}$，并令 $d_{\mathcal{X}}$ 为 $\mathcal{X}$上的连续（伪）度量。记 $\tilde{\mathcal{X}}$ 为赋予由 $d_{\mathcal{X}}$ 诱导的拓扑的空间 $\mathcal{X}$，该空间也是紧的（引理 F.2），但可能不满足 Hausdorff 性。令 $c \in C(\mathcal{X} \times \mathcal{Y})$ 为一个代价函数，使得$\|c\|_\infty \le 1$，并考虑一个凹连续模 $w\colon \mathbb{R}_+ \to \mathbb{R}_+$，其中对于满足以下条件的 $\delta > 0$，$w(\delta) > 0$

$$|c(x, y) - c(x', y)| \le w(\tilde{d}_{\mathcal{X}}(x, x')) \quad \text{for any } x, x' \in \tilde{\mathcal{X}}, y \in \mathcal{Y}. \tag{5.2}$$

若 $c(\cdot, y)$ 在 $d_{\mathcal{X}}$ 下是 1-Lipschitz，则取 $w(t) := t$；若 $c(\cdot, y)$ 对于 $\gamma \in (0, 1]$（回顾脚注 4）是 $(\gamma, 1)$-赫尔德连续，则取 $w(t) := t^\gamma$。此时，$w \circ d_{\mathcal{X}}$ 对应的正则性提升泛函为

$$\Psi^{w\circ\tilde{d}_{\mathcal{X}}}_{\mathrm{mod}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\tilde{\mathcal{X}} \times \mathcal{Y}), \quad \tilde{c} \mapsto \left((x, y) \mapsto \inf_{x' \in \mathcal{X}} \tilde{c}(x', y) + 2w(\tilde{d}_{\mathcal{X}}(x, x'))\right)$$

命题 5.3。对于上述设定，$\Psi = \Psi^{w\circ d_{\mathcal{X}}}_{\text{模}} \circ \Psi_{\mathrm{bdd}}$ 满足 $\Psi(c) = c$，在 c 附近连续，并且在 c 处 Hadamard 方向可微，其导数为 $\mathrm{D}^H_{|c}\Psi|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$。此外，对于 $\mathcal{Y}$ 上任意一致有界函数类 $\mathcal{G}$，有

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim \mathcal{N}(\varepsilon/8, \mathcal{X}, w \circ \tilde{d}_{\mathcal{X}}) |\log(\varepsilon)|.$$

上述考量的一个重要推论是，我们可以根据 $c_{n,m}$ 构造正则性提升估计量 $\tilde{c}_{n,m}$，使得对于 (2.3) 中定义的 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$，$\mathcal{H}_{\tilde{c}_{n,m}} \subseteq \mathcal{F}_{\tilde{c}_{n,m}, \tilde{c}_{n,m}}$ 确定性地成立。

推论 5.4。设 $c \in C(\mathcal{X} \times \mathcal{Y})$，令 $B := \|c\|_\infty + 1/2$，并令 $w\colon \mathbb{R}_+ \to \mathbb{R}_+$ 为凹模数，满足 $w(\delta) > 0$，其中 $\delta > 0$ 使得 (5.2) 对 $\mathcal{X}$上的度量 $d_{\mathcal{X}}$ 成立。假设对于一个随机序列 $c_n \in C(\mathcal{X} \times \mathcal{Y})$，$a_n(c_n - c)$ $\mathbb{G}^c$ 属于 $C(\mathcal{X} \times \mathcal{Y})$，且 $a_n \to \infty$。那么，该随机序列

$$\bar{c}_n := B \cdot \Psi^{w\circ d_{\mathcal{X}}/B}_{\mathrm{mod}} \circ \Psi_{\mathrm{bdd}}(c_n/B) \in C(\mathcal{X} \times \mathcal{Y})$$

满足 $a_n \|c_n - c_n\|_\infty \xrightarrow{P} 0$，对 $n \to \infty$ 成立，并确定性地满足 $\|c_n\|_\infty \le 2B = 2\|c\|_\infty + 1$、关系 (5.2) 以及包含关系 $\mathcal{H}_{\bar{c}_n} \subseteq \mathcal{F}_{c_n c_n}(2\|c\|_\infty + 1, 2w)$。

5.3 提升至阶数为 $\gamma \in (1,2]$ 的赫尔德函数的正则性

由于我们能够利用底层代价函数的正则性（最高达赫尔德阶数 $\gamma \le 2$）来获得经验最优传输值的收敛速率（回顾命题 3.1(i)），因此本小节给出相应的正则性提升映射。由于 $\gamma \le 1$ 的情形可以利用上一小节的理论处理，我们仅关注 $\gamma \in (1,2]$ 的情形。

考虑一个具有非空内部的凸紧集 $\mathcal{X} \subseteq \mathbb{R}^d$。令 $c \in C(\mathcal{X} \times \mathcal{Y})$ 为一个代价函数，满足 $\|c\|_\infty \le 1$，并假设 c 关于 x 连续可微，即假设 $\nabla_x c : \mathrm{int}(\mathcal{X}) \times \mathcal{Y} \to \mathbb{R}^d$ 可连续延拓至 $\mathcal{X} \times \mathcal{Y}$。此外，假设对于每个 $y \in \mathcal{Y}$（其中 $\gamma \in (1,2]$），函数 $c(\cdot, y)$ 都是 $(\gamma, 1)$-赫尔德函数。我们将阶数为 $\gamma \in (1,2]$ 的赫尔德函数的正则性提升映射定义为

$$\Psi_{\mathrm{Hol}}^{c,\gamma} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}), \tilde{c} \mapsto \left((x,y) \mapsto \inf_{x' \in \mathcal{X}} \tilde{c}(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + 2\sqrt{d} \left\| x - x' \right\|^\gamma \right)$$

值得注意的是，标量积项涉及相应（总体）代价函数 c 的偏导数，这一点至关重要。此外，我们想指出，经由 $\Psi^{c,\gamma}$ Hol 得到的像未必是 $(\gamma,1)$-赫尔德函数，但仍能保证适当的度量熵界。

命题 5.5. 对于上述设定，若 $\mathcal{X} \subseteq \mathbb{R}^d$ 凸且紧致，$\Psi = \Psi^{c,\ \gamma\ \mathrm{Hol}} \circ \Psi_{\mathrm{bdd}}$ 满足 $\Psi(c) = c$，则其在 c 附近连续，且在 c 处 Hadamard 可微，导数为 $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。此外，对于 $\mathcal{Y}$ 上任意一致有界函数类 $\mathcal{G}$，我们有

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \lesssim \varepsilon^{-d/\gamma}.$$

5.4 正则性提升的组合

最后，我们还概述一种构造性方法，用于组合定义在不同空间上的正则性提升映射。这一点很重要，因为它使我们能够利用定义域不同区域的总体成本函数的正则性。

因此，令 $\mathcal{X}$、$\mathcal{Y}$ 为紧波兰空间，并假设存在一组同胚 $\zeta_i \colon \mathcal{U}_i \to \zeta_i(\mathcal{U}_i)$，其中 $1 \le i \le I$，且满足 $\mathcal{X} = \bigcup_{i=1}^I \zeta_i(\mathcal{U}_i)$。此外，假设存在 $\mathcal{X}$ 上的单位分解 $\{\eta_i\}_{i=1}^I$，满足 $\mathrm{supp}(\eta_i) \subseteq \zeta_i(\mathcal{U}_i)$。考虑连续代价函数 $c : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$，并令 $c_i \colon \mathcal{U}_i \times \mathcal{Y} \to \mathbb{R}$，$(u, y) \mapsto c(\zeta_i(u), y)$。假设存在映射 $\Psi_i \colon C(\mathcal{U}_i \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y})$，使得 $\Psi_i(c_i) = c_i$，且其中 Ψ_i 在 c_i 附近连续，并在 c_i 处 Hadamard 可微，导数为 $D^H_{|c_i}\Psi_i = \mathrm{Id}$。利用这些映射，我们将正则性提升组合定义为

$$\Psi_{\mathrm{com}} : C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}), \quad \tilde{c} \mapsto \left((x,y) \mapsto \sum_{i=1}^I \eta_i(x) \Psi_i\Big(\tilde{c}(\zeta_i(\cdot),\cdot)\Big)(\zeta_i^{-1}(x), y) \right).$$

事实上，由于单位分解 η_i 的连续性，以及泛函 Ψ_i 和 ζ_i，对于每个 $i \in \{1, \ldots\ldots, I\}$，可知该泛函的值域确实包含于 $C(\mathcal{X} \times \mathcal{Y})$。

命题 5.6. 对于上述设定，$\Psi = \Psi_{\mathrm{com}}$ 满足 $\Psi(c) = c$，在 c 附近连续，并且在 c 处 Hadamard 可微，且 $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。此外，对于任意定义在 $\mathcal{Y}$ 上的一致有界函数类 $\mathcal{G}$，我们有

$$\sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \le \sum_{i=1}^I \sup_{\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c}(\zeta_i(\cdot),\cdot))}, \|\cdot\|_\infty).$$

6 主要结果的证明

本节给出引理 2.1（关于最优传输值的对偶表示）、定理 2.2 和命题 2.10（关于代价弱收敛时经验最优传输的分布极限），以及定理 2.5–2.7 和命题 2.8（关于具有极值型代价的经验最优传输）的完整证明。本小节所有辅助结果的证明均见附录 E。

6.1 引理 2.1 的证明：最优传输值的对偶表示

下面的辅助引理证明了成本变换的一个重要性质，该性质在本节中至关重要。

˜引理 6.1（成本变换的利普希茨性质）。对于任意函数 f、f: $\mathcal{X} \to \mathbb{R}$ 以及代价函数 c，˜c: $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}$，可得 $\|f^c - \tilde{f}^{\tilde{c}}\|_\infty \le \|f - \tilde{f}\|_\infty + \|c - \tilde{c}\|_\infty$。特别地，˜选取常值函数 f，˜c ≡ 0后，可得 $\|f^c\|_\infty \le \|f\|_\infty + \|c\|_\infty$。

*Proof*引理 2.1 的证明。对于任意 $h \in \mathcal{H}_c$，都存在 $g: \mathcal{Y} \to [-\|c\|_\infty, \|c\|_\infty]$，使得 $h = g^c$，因此

$$-\|c\|_\infty - \sup_{y\in\mathcal{Y}} g(y) \le h(x) = \inf_{y\in\mathcal{Y}} c(x,y) - g(y) \le \|c\|_\infty - \sup_{y\in\mathcal{Y}} g(y).$$

由此可得 $\|h\|_\infty \le 2\|c\|_\infty \le 2B$。此外，对于任意 x，$x' \in \mathcal{X}$ 和 $\varepsilon > 0$，考虑满足 $h(x') \ge c(x', y') - g(y') - \varepsilon$ 的 $y' \in \mathcal{Y}$。于是可得

$$\begin{aligned} h(x) - h(x') &= \left[\inf_{y\in\mathcal{Y}} c(x,y) - g(y)\right] - \left[\inf_{y\in\mathcal{Y}} c(x',y) - g(y)\right] \\ &\le c(x,y') - g(y') - c(x',y') + g(y') + \varepsilon \\ &\le w(d_{\mathcal{X}}(x,x')) + \varepsilon. \end{aligned}$$

由于 $\varepsilon > 0$ 可以任意小，我们得到 $|h(x) - h(x')| \le w(d_{\mathcal{X}}(x, x'))$。这就推出 $\mathcal{H}_c \subseteq \mathcal{F}$，因此 $\mathcal{H}_c^c \subseteq \mathcal{F}_c$。此外，由 Santambrogio（2015，命题 1.34）可推得$\mathcal{H}_c = \mathcal{H}_c^{cc} \subseteq \mathcal{F}_{cc}$。为证明其余包含关系，注意对于 $f \in \mathcal{F}$，有

$$-\|c\|_\infty - \sup_{x\in\mathcal{X}} f(x) \le f^c \le \|c\|_\infty - \sup_{x\in\mathcal{X}} f(x).$$

因此，函数 $g := f^c + \sup_{x\in\mathcal{X}} f(x)$ 满足 $\|g\|_\infty \le \|c\|_\infty$，又由于 $\|f\|_\infty \le 2B$，我们得到

$$f^{cc}(x) = (f^c)^c = (g)^c + \sup_{x\in\mathcal{X}} f(x) \in \mathcal{H}_c + [-2B, 2B],$$

由此可得 $\mathcal{F}cc \subseteq \mathcal{H}_c + [-2B, 2B]$ 以及 $\mathcal{F}c = \mathcal{F}ccc \subseteq \mathcal{H}_c^c + [-2B, 2B]$。为证明表示式 (2.4)，我们结合包含关系 $\mathcal{H}_c \subseteq \mathcal{F} \subseteq C(\mathcal{X})$ 与替代的对偶表示 (1.2) 和 (2.2)。对于最后一个结论，取 (2.4) 的一个最大化序列 $\{f_n\}_{n\in\mathbb{N}}$；由 $\mathcal{F}$ 的紧性，该序列存在一个收敛子序列 $\{f_{n_k}\}_{k\in\mathbb{N}}$，其一致极限为 $f \in \mathcal{F}$。由引理 6.1 可知，$\{f^{cn_k}\}_{k\in\mathbb{N}}$ 和 $\{f^{ccn_k}\}_{k\in\mathbb{N}}$ 也分别一致收敛于 f^c 和 f^{cc}。因此，我们得到 $\mu(f^{cc}) + \nu(f^c) = \lim_{k\to\infty} \mu(f^{ccn_k}) + \nu(f^{cn_k}) =$最优传输 (μ, ν, c)，这表明 $f \in \mathcal{F}$是一个最大化元素，因此 (2.4) 的最优解集合 $S_c(\mu, \nu)$ 非空。□

6.2 弱收敛代价下分布极限的证明

6.2.1 定理 2.2 的证明

证明定理 2.2 的关键在于以下辅助结果。首先，我们给出代价和概率测度变化时最优传输值之差的下界和上界；这是因为最优传输问题既可以表示为对可行耦合取下确界，也可以表示为对可行势函数取上确界。

引理 6.2（下界和上界）。对于 B $>$ 0 和凹连续模 w: $\mathbb{R}_+ \to \mathbb{R}_+$，定义集合

$$C(B,w) := \{\bar{c} \in C(\mathcal{X}\times\mathcal{Y}) \,\big|\, \|\bar{c}\|_\infty \le B, |\bar{c}(x,y)-\bar{c}(x',y)| \le w(d_{\mathcal{X}}(x,x')) \text{ for all } x,x' \in \mathcal{X}, y \in \mathcal{Y}\}.$$

那么，对于代价 c，˜c $\in$ C(B, w) 以及概率测度 μ，˜μ $\in \mathcal{P}(\mathcal{X})$，$\nu, \tilde{\nu} \in \mathcal{P}(\mathcal{Y})$，有 *at*

$$\begin{aligned}
&\inf_{\pi\in\Pi_{\tilde{c}}^\star(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c}-c) + \sup_{f\in S_c(\mu,\nu)} (\tilde{\mu}-\mu)f^{cc} + (\tilde{\nu}-\nu)f^c \\
&\le\ OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\mu,\nu,c) \\
&\le \min\Big(\inf_{\pi\in\Pi_{\tilde{c}}^\star(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c}-c) + \sup_{f\in S_c(\tilde{\mu},\tilde{\nu})} (\tilde{\mu}-\mu)f^{cc} + (\tilde{\nu}-\nu)f^c, \\
&\quad \inf_{\pi\in\Pi_{\tilde{c}}^\star(\mu,\nu)} \pi(\tilde{c}-c) + \sup_{f\in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu}-\mu)f^{cc} + (\tilde{\nu}-\nu)f^c + \sup_{f\in\mathcal{F}} (\tilde{\mu}-\mu)(f^{\tilde{c}\tilde{c}}-f^{cc}) + (\tilde{\nu}-\nu)(f^{\tilde{c}}-f^c).\Big)
\end{aligned}$$

特别地，对于固定的测度或固定的代价，可得

$$|OT(\mu,\nu,\tilde{c}) - OT(\mu,\nu,c)| \le \|\tilde{c}-c\|_\infty, \quad |OT(\tilde{\mu},\tilde{\nu},c) - OT(\mu,\nu,c)| \le \sup_{f\in\mathcal{F}^{cc}} \big|(\tilde{\mu}-\mu)f\big| + \sup_{f\in\mathcal{F}^c} \big|(\tilde{\nu}-\nu)f\big|.$$

为在定理 2.2 的证明中运用引理 6.2 的下界和上界，我们还需要若干连续性和可测性性质，以下引理对这些性质作了概括。值得注意的是，我们赋予 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 有界 Lipschitz 范数，使其成为波兰度量空间，并使该范数能够刻画测度的弱收敛。

引理 6.3（连续性与可测性）。设 μ $\in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$，且 c $\in$ C($\mathcal{X}\times\mathcal{Y}$)。取一个关于 c 的凹连续模 w : $\mathbb{R}_+ \to \mathbb{R}_+$，并令 C := C(2 ‖c‖$_\infty$ + 1, 2w)（关于 C(2 ‖c‖$_\infty$ + 1, 2w) 的定义，见引理 6.2）。此外，回顾式 (2.3) 中引入的函数类 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$，并定义函数

$$\begin{aligned}
&T_1 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C \to \mathbb{R}, && (\mu',\nu',c') \mapsto OT(\mu',\nu',c'), \\
&T_2 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C\times C(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}, && (\mu',\nu',c',h_c) \mapsto \inf_{\pi\in\Pi_{c'}^\star(\mu',\nu')} \pi(h_c), \\
&T_3 : \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})\times C\times C_u(\mathcal{F})^2 \to \mathbb{R}, && (\mu',\nu',c',h_\mu,h_\nu) \mapsto \sup_{f\in S_{c'}(\mu',\nu')} h_\mu(f) + h_\nu(f), \\
&T_4 : C_u(\mathcal{F})^4 \to \mathbb{R}, && (h_\mu,\tilde{h}_\mu,h_\nu,\tilde{h}_\nu) \mapsto \sup_{f\in\mathcal{F}} h_\mu(f) - \tilde{h}_\mu(f) + h_\nu(f) - \tilde{h}_\nu(f).
\end{aligned}$$

那么，T_1 和 T_4 连续，T_2 下半连续，而 T_3 上半连续。若$\Pi_{c'}^\star(\mu',\nu')$唯一，则 T_2 在 (μ′, ν′, c′, h_c) 处连续。此外，对于固定的 (μ′, ν′, c′)，映射 T_2 关于 h_c 连续，而 T_3 关于 (h_μ, h_ν) 连续。特别地，对于 $1 \le i \le 4$，每个函数 T_i 均为博雷尔可测。

前两个断言完整地处理了关于最优传输泛函的确定性陈述，以及由相应界所产生的相关项。以下两个结果提供了控制随机性方面的相关工具。更具体地，在证明定理 2.2 时，我们考虑 (JW) 中详述的随机序列的斯科罗霍德表示，该表示还满足如下性质：μ_n 和 ν_n 分别弱收敛于 μ 和 ν。为此，我们陈述以下可测性断言和联合弱收敛结论。

引理 6.4（经验过程的可测性）。对于波兰空间 $\mathcal{X}$，考虑在一致范数下全有界的函数类 $\mathcal{G} \subseteq C(\mathcal{X})$。那么，以下断言成立。

(i) 任意概率测度 $\mu \in \mathcal{P}(\mathcal{X})$ 都通过求值定义了 $\mathcal{G}$ 上的一致连续泛函，即 $\mu \in C_u(\mathcal{G})$。

(ii) 映射 $\omega \mapsto \mu(\omega) \in \mathcal{P}(\mathcal{X}) \subseteq C_u(\mathcal{G})$ 是博雷尔可测的，当且仅当对任意 $g \in \mathcal{G}$，求值映射 $\omega \mapsto \mu(\omega)(g)$ 是博雷尔可测的。

(iii) 经验过程 $\sqrt{n}(\mu_n - \mu)$ 和自助法经验过程 $\sqrt{k}(\mu_{n,k}^b - \mu_n)$ 都是 $C_u(\mathcal{G})$ 中的博雷尔可测随机变量。

引理 6.5（联合弱收敛）。对于定理 2.2 的设定，假设 (JW) 成立。那么，对于 $n, m \to \infty$，在波兰空间 $C_u(\mathcal{F})^2 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 中，弱收敛至一个紧极限成立

$$\left(\left(\mathbb{G}_n^\mu(f^{cc}), \mathbb{G}_m^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,m}^c, \mu_n, \nu_m\right) \rightsquigarrow \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu\right). \tag{6.1}$$

如果定理 2.2 的 (Sup) 也成立，那么对于 $n, m \to \infty$，在波兰空间 $C_u(\mathcal{F})^4 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 中，有

$$\begin{aligned}&\left(\left(\mathbb{G}_n^\mu(f^{cc}), \mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}}), \mathbb{G}_m^\nu(f^c), \mathbb{G}_m^\nu(f^{c_{n,m}})\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,m}^c, \mu_n, \nu_n,\right)\\ &\quad\rightsquigarrow \left(\left(\mathbb{G}^\mu(f^{cc}), \mathbb{G}^\mu(f^{cc}), \mathbb{G}^\nu(f^c), \mathbb{G}^\nu(f^c)\right)_{f\in\mathcal{F}}, \mathbb{G}^c, \mu, \nu,\right),\end{aligned} \tag{6.2}$$

(6.1) 和 (6.2) 中的每个序列元素以及弱极限都是博雷尔可测的。

注 6.6（斯科罗霍德表示）。在非可分空间中处理经验过程的弱收敛时，由于可能存在可测性问题，需要格外谨慎。不过，由于我们关注的不同映射定义在波兰空间之间且均可测，因此可以避开此类问题。特别是，引理 6.5 中的随机变量弱收敛到一个支撑可分的紧极限，因此满足 Billingsley（1999，定理 6.7）的条件，存在一个（可测的）斯科罗霍德表示。

有了这些工具，我们现在开始证明定理 2.2。

定理 2.2 的证明。根据推论 5.4，由于 $\sqrt{nm/(n+m)}(c_{n,m} - c) =: \mathbb{G}_{n,m}^c \rightsquigarrow \mathbb{G}^c$ 在空间 $C(\mathcal{X}\times\mathcal{Y})$ 中收敛，因此存在 $c_{n,m}$，使得包含关系 $\mathcal{H}_{c_{n,m}} \subseteq \mathcal{F}_{c_{n,m}, \bar{c}_{n,m}}$（回忆第 2.1 节中的函数类）对于 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ 和 $\sqrt{n}(c_{n,m} - c_{n,m}) \xrightarrow{P} 0$ 确定性地成立。由引理 6.2，后者推出

$$\sqrt{\frac{nm}{n+m}}\left(OT(\mu_n, \nu_m, \bar{c}_{n,m}) - OT(\mu_n, \nu_m, c_{n,m})\right) \xrightarrow{P} 0.$$

因此，根据斯卢茨基引理，只需证明

$$\sqrt{\frac{nm}{n+m}}\left(OT(\mu_n, \nu_m, \bar{c}_{n,m}) - OT(\mu, \nu, c)\right) \rightsquigarrow \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f^{cc}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f^c). \tag{6.3}$$

因此，不失一般性，我们可以假设 $c_{n,m} = c_{n,m}$。此外，令 $\lambda_n := m/(n+m)$。于是，

根据引理 6.2，可得以下下界和上界，

$$\begin{aligned}
&\inf_{\pi\in\Pi^\star_{c_{n,m}}(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}^\nu_m(f^c) \\
&\le \sqrt{\frac{nm}{n+m}}(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)) \qquad (6.4)\\
&\le \min\Bigg(\inf_{\pi\in\Pi^\star_c(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}^\nu_m(f^c), \\
&\qquad \inf_{\pi\in\Pi^\star_c(\mu,\nu)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_{c_{n,m}}(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}^\nu_m(f^c) \\
&\qquad + \sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\mathbb{G}^\mu_n(f^{c_{n,m}c_{n,m}}) - \mathbb{G}^\mu_n(f^{cc})\right) + \sqrt{1-\lambda_n}\,\left(\mathbb{G}^\nu_m(f^{c_{n,m}}) - \mathbb{G}^\nu_m(f^c)\right)\Bigg).
\end{aligned}$$

对于每种情形 (OP) 和 (Sup)，我们证明上界和下界在分布上渐近收敛于 (6.3) 中的极限，进而表明经验最优传输值也趋于该极限。为此，我们在概率空间 $(\Omega, \mathcal{A}, \mathrm{P})$ 上，为引理 6.5 中的随机变量取一个斯科罗霍德表示（Billingsley，1999，第 70 页）；根据注 6.6，该表示定义良好。更确切地说，在 (OP) 下，我们取斯科罗霍德表示使得

$$\left(\left(\tilde{\mathbb{G}}^\mu_n(f^{cc}), \tilde{\mathbb{G}}^\nu_m(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c_{n,m}, \tilde{\mu}_n, \tilde{\nu}_m\right) \xrightarrow{\text{a.s.}} \left(\left(\tilde{\mathbb{G}}^\mu(f^{cc}), \tilde{\mathbb{G}}^\nu(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c, \mu, \nu\right) \qquad (6.5)$$

在 $\mathrm{C}_u(\mathcal{F})^2 \times \mathrm{C}(\mathcal{X}\times\mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 中，而在 (Sup) 下，我们选择其使得

$$\begin{aligned}
&\left(\left(\tilde{\mathbb{G}}^\mu_n(f^{cc}), \tilde{\mathbb{G}}^\mu_n(f^{\tilde{c}_{n,m}\tilde{c}_{n,m}}), \tilde{\mathbb{G}}^\nu_m(f^c), \tilde{\mathbb{G}}^\nu_m(f^{\tilde{c}_n})\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c_{n,m}, \tilde{\mu}_n, \tilde{\nu}_m\right) \\
&\qquad \xrightarrow{\text{a.s.}} \left(\left(\tilde{\mathbb{G}}^\mu(f^{cc}), \tilde{\mathbb{G}}^\mu(f^{cc}), \tilde{\mathbb{G}}^\nu(f^c), \tilde{\mathbb{G}}^\nu(f^c)\right)_{f\in\mathcal{F}}, \tilde{\mathbb{G}}^c, \mu, \nu\right) \qquad (6.6)
\end{aligned}$$

在 $\mathrm{C}_u(\mathcal{F})^4 \times \mathrm{C}(\mathcal{X}\times\mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 中。我们还令 $\tilde{c}_{n,m} := c + \tilde{\mathbb{G}}^c_{n,m}/\sqrt{nm/(n+m)}$，其几乎必然收敛于 c。

对于后续论证，请回顾引理 6.3 中的函数 T_1, T_2, T_3, T_4 及其（半）连续性性质。此外，注意到，将引理 6.1 与引理 6.4 (i) 的证明中的论证相结合，可得映射 $\mathcal{F} \to \mathbb{R}$，

$$f \mapsto \mathbb{G}^\mu_n(f^{cc}), \quad f \mapsto \mathbb{G}^\mu_n(f^{c_{n,m}c_{n,m}}), \quad f \mapsto \mathbb{G}^\nu_m(f^c), \quad f \mapsto \mathbb{G}^\nu_m(f^{c_{n,m}})$$

是一致连续的，即属于 $\mathrm{C}_u(\mathcal{F})$。对于 (OP) 和 (Sup) 两种情形，由于映射 T_1, T_2, T_3 对每个 $n, m \in \mathbb{N}$ 均可测，可得

$$\begin{aligned}
&\sqrt{\frac{nm}{n+m}}(OT(\mu_n,\nu_m,c_{n,m}) - OT(\mu,\nu,c)) \overset{d}{=} \sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c)) \\
&\inf_{\pi\in\Pi^\star_{c_{n,m}}(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}^\nu_m(f^c) \\
&\qquad \overset{d}{=} \inf_{\pi\in\Pi^\star_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}^c_{n,m}) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}^\nu_m(f^c).
\end{aligned}$$

在 (OP) 条件下，我们还注意到

$$\begin{aligned}
&\inf_{\pi\in\Pi^\star_c(\mu_n,\nu_m)} \pi(\mathbb{G}^c_{n,m}) + \sup_{f\in S_c(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}^\nu_m(f^c) \\
&\qquad \overset{d}{=} \inf_{\pi\in\Pi^\star_c(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}^c_{n,m}) + \sup_{f\in S_c(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}^\mu_n(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}^\nu_m(f^c),
\end{aligned}$$

而在 (Sup) 条件下，我们还利用 T_4 的可测性推得，对于每个 $n \in \mathbb{N}$，

$$\begin{aligned}
&\inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}_{n,m}^c) + \sup_{f\in S_{c_{n,m}}(\mu_n,\nu_m)} \sqrt{\lambda_n}\,\mathbb{G}_n^\mu(f^{cc}) + \sqrt{1-\lambda_n}\,\mathbb{G}_m^\nu(f^c) \\
&+ \sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\mathbb{G}_n^\mu(f^{cc}) - \mathbb{G}_n^\mu(f^{c_{n,m}c_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\mathbb{G}_m^\nu(f^c) - \mathbb{G}_m^\nu(f^{c_{n,m}})\right) \\
&\qquad \overset{d}{=} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^\mu(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^\nu(f^c) \\
&\qquad + \sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\tilde{\mathbb{G}}_n^\mu(f^{cc}) - \tilde{\mathbb{G}}_n^\mu(f^{\tilde{c}_{n,m}\tilde{c}_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\tilde{\mathbb{G}}_m^\nu(f^c) - \tilde{\mathbb{G}}_m^\nu(f^{\tilde{c}_{n,m}})\right).
\end{aligned}$$

因此，只需利用斯科罗霍德表示得到经验最优传输值的弱极限。调用引理 6.2，即可得出，与式（6.4）相同的关于目标量 $\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c))$ 的下界和上界。

为得到 $\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c))$ 下极限的合适上界，对于 (OP) 和 (Sup)，分别取一个全测度可测集 $A \in \mathcal{A}$，使得式（6.5）和（6.6）中的收敛分别在该集合上成立。然后，对于每个 $\omega \in A$，由 T_2 的下半连续性以及在固定 (μ,ν,c) 下 T_3 的连续性可知

$$\begin{aligned}
&\liminf_{n,m\to\infty} \inf_{\pi\in\Pi_{\tilde{c}_{n,m}}^\star(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^\mu(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^\nu(f^c) \\
&\qquad \geq \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c).
\end{aligned}$$

在（OP）条件下，根据（OP）以及 T_3 的上半连续性，由 T_2 在 $(\mu,\nu,c,\mathbb{G}_{n,m}^c)$ 处的连续性可知，对每个 $\omega \in A$,

$$\begin{aligned}
&\limsup_{n,m\to\infty} \inf_{\pi\in\Pi_c^\star(\tilde{\mu}_n,\tilde{\nu}_m)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_c(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^\mu(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^\nu(f^c) \\
&\qquad \leq \pi^\star(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c) \\
&\qquad = \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c).
\end{aligned}$$

在（Sup）条件下，对于每个 $\omega \in A$，注意到：当 (μ,ν,c) 固定时，T_2 关于 h_c 连续，且 T_3 上半连续、T_4 连续，因此

$$\begin{aligned}
&\limsup_{n,m\to\infty} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}_{n,m}^c) + \sup_{f\in S_{\tilde{c}_{n,m}}(\tilde{\mu}_n,\tilde{\nu}_m)} \sqrt{\lambda_n}\,\tilde{\mathbb{G}}_n^\mu(f^{cc}) + \sqrt{1-\lambda_n}\,\tilde{\mathbb{G}}_m^\nu(f^c) \\
&+ \sup_{f\in\mathcal{F}} \sqrt{\lambda_n}\,\left(\tilde{\mathbb{G}}_n^\mu(f^{cc}) - \tilde{\mathbb{G}}_n^\mu(f^{c_{n,m}c_{n,m}})\right) + \sqrt{1-\lambda_n}\,\left(\tilde{\mathbb{G}}_m^\nu(f^c) - \tilde{\mathbb{G}}_m^\nu(f^{c_{n,m}})\right) \\
&\qquad \leq \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c) \\
&\qquad + \sup_{f\in\mathcal{F}} \sqrt{\lambda}\,\left(\tilde{\mathbb{G}}^\mu(f^{cc}) - \tilde{\mathbb{G}}^\mu(f^{cc})\right) + \sqrt{1-\lambda}\,\left(\tilde{\mathbb{G}}^\nu(f^c) - \tilde{\mathbb{G}}^\nu(f^c)\right) \\
&\qquad = \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c).
\end{aligned}$$

由于对于所有 $\omega \in A$，$\sqrt{nm/(n+m)}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c))$ 的下界和上界在渐近意义下相匹配，因此在 (OP) 和 (Sup) 两种情况下均有

$$\lim_{n,m\to\infty} \sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m}) - OT(\mu,\nu,c)) = \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^c) + \sup_{f\in S_c(\mu,\nu)} \sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc}) + \sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c).$$

由于集合 A 具有全测度，我们得到

$$\sqrt{\frac{nm}{n+m}}(OT(\tilde{\mu}_n,\tilde{\nu}_m,\tilde{c}_{n,m})-OT(\mu,\nu,c))\xrightarrow{\text{a.s.}}\inf_{\pi\in\Pi_c^\star(\mu,\nu)}\pi(\tilde{\mathbb{G}}^c)+\sup_{f\in S_c(\mu,\nu)}\sqrt{\lambda}\,\tilde{\mathbb{G}}^\mu(f^{cc})+\sqrt{1-\lambda}\,\tilde{\mathbb{G}}^\nu(f^c),$$

由于 T_2 和 T_3 的可测性，该极限与断言中的极限具有相同的博雷尔分布，由此完成证明。□

6.2.2 命题 2.10 的证明

在开始证明自助法一致性（即命题 2.10）之前，我们先介绍一个关于自助经验测度收敛的重要结果。

引理 6.7. 设 $\mathcal{X}$ 为波兰空间，$\in\mathcal{P}(\mathcal{X})$ 为其上的博雷尔分布。考虑独立同分布随机变量 $\{X_i\}_{i=1}^n\sim$，其分布为 $\mu^{\otimes n}$，并据此定义经验测度 $\mu_n=n^{-1}\sum_{i=1}^n\delta_{X_i}$。此外，考虑 k(n) 个独立同分布随机变量$\{X_k^b\}_{i=1}^n\sim$，其分布为 $\mu_n^{\otimes k(n)}$，并据此定义自助经验测度 $\mu_{n,k}^b:=\frac{1}{k(n)}\sum_{j=1}^{k(n)}\delta_{X_i^b}$。那么，只要 $k(n)\to\infty$，当 $n\to\infty$时，在 $n\to\infty$ 的条件下，$\mu_{n,k}^b$ 依概率弱收敛于 μ。

上述引理是 (Beran, Le Cam, 和 Millar, 1987) 中定理 2 的推论，加入该引理是为了便于后续引用。现在我们可以证明在弱收敛代价下自助法一致性的命题 2.10。

命题 2.10 的证明。由假设 (JW) 和 (JW)*，以及经验过程和自助法经验过程的可测性（引理 6.4），利用 Bücher 和 Kojadinovic (2019) 中针对两个自助法版本 ($\mu^{(1)}n,k$, $\nu^{(1)}n,k$, $c^{(1)}n,k$)、($\mu^{(2)}n,k$, $\nu^{(2)}n,k$, $c^{(2)}n,k$) 的引理 2.2(c) $\Rightarrow$ (a)，它们基于相互独立的自助样本 $\{X^{(1)}i\}_{i=1}^k,\{X^{(2)}i\}_{i=1}^k\sim\mu_n^{\otimes k}$和 $\{Y^{(1)}i\}_{i=1}^k,\{Y^{(2)}i\}_{i=1}^k\sim\nu_n^{\otimes k}$，对于 $n,k\to\infty$，$k=o(n)$，可推得：

$$\begin{pmatrix}\sqrt{n}\begin{pmatrix}\mu_n-\mu\\ \nu_n-\nu\\ c_n-c\end{pmatrix}\\ \sqrt{k}\begin{pmatrix}\mu_{n,k}^{(i)}-\mu_n\\ \nu_{n,k}^{(i)}-\nu_n\\ c_{n,k}^{(i)}-c_n\end{pmatrix}_{i=1,2}\end{pmatrix}\rightsquigarrow\begin{pmatrix}\begin{pmatrix}\mathbb{G}^\mu\\ \mathbb{G}^\nu\\ \mathbb{G}^c\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}^{\mu,(i)}\\ \mathbb{G}^{\nu,(i)}\\ \mathbb{G}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}$$

在 $\left(C_u(\mathcal{F}^{cc})\times C_u(\mathcal{F}_c)\times C(\mathcal{X}\times\mathcal{Y})\right)^3$ 中。由于 $k=o(n)$，由斯卢茨基引理还可得

$$\begin{pmatrix}\sqrt{n}\begin{pmatrix}\mu_n-\mu\\ \nu_n-\nu\\ c_n-c\end{pmatrix}\\ \sqrt{k}\begin{pmatrix}\mu_{n,k}^{(i)}-\mu\\ \nu_{n,k}^{(i)}-\nu\\ c_{n,k}^{(i)}-c\end{pmatrix}_{i=1,2}\end{pmatrix}=:\begin{pmatrix}\begin{pmatrix}\mathbb{G}_n^\mu\\ \mathbb{G}_n^\nu\\ \mathbb{G}_n^c\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}_{n,k}^{\mu,(i)}\\ \mathbb{G}_{n,k}^{\nu,(i)}\\ \mathbb{G}_{n,k}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}\rightsquigarrow\begin{pmatrix}\begin{pmatrix}\mathbb{G}^\mu\\ \mathbb{G}^\nu\\ \mathbb{G}^c\end{pmatrix}\\ \begin{pmatrix}\mathbb{G}^{\mu,(i)}\\ \mathbb{G}^{\nu,(i)}\\ \mathbb{G}^{c,(i)}\end{pmatrix}_{i=1,2}\end{pmatrix}.$$

在此，三元组（$\mathbb{G}^\mu$、$\mathbb{G}^\nu$、$\mathbb{G}^c$）、（$\mathbb{G}^\mu$、$^{(1)}$、$\mathbb{G}^\nu$、$^{(1)}$、$\mathbb{G}^c$、$^{(1)}$）和（$\mathbb{G}^\mu$、$^{(2)}$、$\mathbb{G}^\nu$、$^{(2)}$、$\mathbb{G}^c$、$^{(2)}$）相互独立且同分布。值得注意的是，利用推论 5.4，我们不失一般性地假设，对于 $i\in\{1,2\}$，经验代价函数 c_n 和自助法代价函数 $c^{(i)}n,k$ 确定性地满足关系 $\mathcal{F}_{\bar{c}}\subseteq\mathcal{F}^{\bar{}}c\text{-}c$，$^{\bar{}}c\in\{c_n,c^{(i)}n,k\}$。此外，根据 Varadarajan（1958），我们知道当 $n\to\infty$ 时，$\mu_n\rightsquigarrow\mu,\nu_n\rightsquigarrow\nu$ 几乎必然成立；根据引理 6.7，我们推得对于 $i\in\{1,2\}$，$\mu^{(i)}n,k\,\mu,\nu^{(i)}n,k\ \nu$ 依概率收敛，条件为

n, k $\to \infty$, k = o(n)。因此，斯卢茨基引理断言

$$\begin{pmatrix} \left(\mathbb{G}_n^{\mu}, \mathbb{G}_n^{\nu}, \mathbb{G}_n^{c}, \mu_n, \nu_n\right)^T \\ \left(\mathbb{G}_{n,k}^{\mu,(i)}, \mathbb{G}_{n,k}^{\nu,(i)}, \mathbb{G}_{n,k}^{c,(i)}, \mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}\right)_{i=1,2}^T \end{pmatrix} \rightsquigarrow \begin{pmatrix} \left(\mathbb{G}^{\mu}, \mathbb{G}^{\nu}, \mathbb{G}^{c}, \mu, \nu\right)^T \\ \left(\mathbb{G}^{\mu,(i)}, \mathbb{G}^{\nu,(i)}, \mathbb{G}^{c,(i)}, \mu, \nu\right)_{i=1,2}^T \end{pmatrix} \tag{6.7}$$

在 $(C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^{c}) \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}))^3$ 中。此外，采用与引理 6.5 证明类似的论证，我们得出

$$\begin{pmatrix} \left(\left(\mathbb{G}_n^{\mu}(f^{cc}), \mathbb{G}_n^{\nu}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}_n^{c}, \mu_n, \nu_n\right)^T \\ \left(\left(\mathbb{G}_{n,k}^{\mu,(i)}(f^{cc}), \mathbb{G}_{n,k}^{\nu,(i)}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,k}^{c,(i)}, \mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}\right)_{i=1,2}^T \end{pmatrix} \rightsquigarrow \begin{pmatrix} \left(\left(\mathbb{G}^{\mu}(f^{cc}), \mathbb{G}^{\nu}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}^{c}, \mu, \nu\right)^T \\ \left(\left(\mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\nu,(i)}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}^{c,(i)}, \mu, \nu\right)_{i=1,2}^T \end{pmatrix} \tag{6.8}$$

在波兰空间 $(C_u(\mathcal{F})^2 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}))^3$ 中，并且在假设 (OP) 下，我们对 (6.8) 中的随机过程使用斯科罗霍德表示。

在假设 (Sup) 和 $(\text{Sup})^*$ 下，由于 c_n 和 $c_{n,k}$ 作为到 $C(\mathcal{X} \times \mathcal{Y})$ 的映射具有可测性，且 c-变换具有利普希茨连续性（引理 6.1），再结合斯卢茨基引理，我们得出随机变量的弱收敛

$$\begin{aligned} &\begin{pmatrix} \left(\left(\mathbb{G}_n^{\mu}(f^{cc}), \mathbb{G}_n^{\mu}(f^{c_n c_n}), \mathbb{G}_n^{\nu}(f^{c}), \mathbb{G}_n^{\nu}(f^{c_n})\right)_{f\in\mathcal{F}}, \mathbb{G}_n^{c}, \mu_n, \nu_n\right)^T \\ \left(\left(\mathbb{G}_{n,k}^{\mu,(i)}(f^{cc}), \mathbb{G}_{n,k}^{\mu,(i)}(f^{c_{n,k}^{(i)}}), \mathbb{G}_{n,k}^{\nu,(i)}(f^{c}), \mathbb{G}_{n,k}^{\nu,(i)}(f^{c_{n,k}^{(i)}})\right)_{f\in\mathcal{F}}, \mathbb{G}_{n,k}^{c,(i)}, \mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}\right)_{i=1,2}^T \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} \left(\left(\mathbb{G}^{\mu}(f^{cc}), \mathbb{G}^{\mu}(f^{cc}), \mathbb{G}^{\nu}(f^{c}), \mathbb{G}^{\nu}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}^{c}, \mu, \nu\right)^T \\ \left(\left(\mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\mu,(i)}(f^{cc}), \mathbb{G}^{\nu,(i)}(f^{c}), \mathbb{G}^{\nu,(i)}(f^{c})\right)_{f\in\mathcal{F}}, \mathbb{G}^{c,(i)}, \mu, \nu\right)_{i=1,2}^T \end{pmatrix} \end{aligned} \tag{6.9}$$

在波兰空间 $(C_u(\mathcal{F})^4 \times C(\mathcal{X} \times \mathcal{Y}) \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}))^3$ 中。对于 (6.9) 中的随机变量，我们现在采用斯科罗霍德表示。

为表示斯科罗霍德表示中的随机元素，我们在相应的随机变量上加波浪号，例如，将 $\tilde{}\mu_n$ 记作 μ_n 的表示。按照定理 2.2 中相同的证明技术，结合引理 6.2 和引理 6.3，我们由此得到

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}\left(OT(\tilde{\mu}_n, \tilde{\nu}_n, \tilde{c}_n) - OT(\mu, \nu, c)\right) \\ \sqrt{k}\left(OT(\tilde{\mu}_{n,k}^{(i)}, \tilde{\nu}_{n,k}^{(i)}, \tilde{c}_{n,k}^{(i)}) - OT(\mu, \nu, c)\right)_{i=1,2} \end{pmatrix} \\ &\xrightarrow{\text{a.s.}} \begin{pmatrix} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^{c}) + \sup_{f\in S_c(\mu,\nu)} \tilde{\mathbb{G}}^{\mu}(f^{cc}) + \tilde{\mathbb{G}}^{\nu}(f^{c}) \\ \left(\inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\tilde{\mathbb{G}}^{c,(i)}) + \sup_{f\in S_c(\mu,\nu)} \tilde{\mathbb{G}}^{\mu,(i)}(f^{cc}) + \tilde{\mathbb{G}}^{\nu,(i)}(f^{c})\right)_{i=1,2} \end{pmatrix}. \end{aligned}$$

因此，对于原始随机变量，并利用 k = o(n)，我们推得

$$\begin{aligned} &\begin{pmatrix} \sqrt{n}\left(OT(\mu_n, \nu_n, c_n) - OT(\mu, \nu, c)\right) \\ \sqrt{k}\left(OT(\mu_{n,k}^{(i)}, \nu_{n,k}^{(i)}, c_{n,k}^{(i)}) - OT(\mu_n, \nu_n, c_n)\right)_{i=1,2} \end{pmatrix} \\ &\rightsquigarrow \begin{pmatrix} \inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^{c}) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^{\mu}(f^{cc}) + \mathbb{G}^{\nu}(f^{c}) \\ \left(\inf_{\pi\in\Pi_c^\star(\mu,\nu)} \pi(\mathbb{G}^{c,(i)}) + \sup_{f\in S_c(\mu,\nu)} \mathbb{G}^{\mu,(i)}(f^{cc}) + \mathbb{G}^{\nu,(i)}(f^{c})\right)_{i=1,2} \end{pmatrix}. \end{aligned}$$

由于极限中的三个分量具有相同的分布且相互独立，结论可由 Bücher 和 Kojadinovic（2019，引理 2.2 (a) $\Rightarrow$ (c)）立即得出。□

6.3 极值型代价函数下分布极限的证明

在继续证明第 2.3 节中依赖于应用泛函δ方法的结果之前，我们先给出一个关于极限随机过程支撑的简单结果。其证明推迟至附录 E.6。

引理 6.8. 对于波兰空间 $\mathcal{X}$，令 $\mu \in \mathcal{P}(\mathcal{X})$，并考虑定义在 $\mathcal{X}$ 上的有界可测函数类 $\mathcal{F}$。则以下结论成立。

(i) $\mathcal{P}(\mathcal{X})$ 在 μ 处的切锥为 $T_\mu\mathcal{P}(\mathcal{X}) = \mathrm{Cl}\{t(\mu' - \mu) \mid t > 0, \mu' \in \mathcal{P}(\mathcal{X})\} \subseteq l^\infty(\tilde{\mathcal{F}})$。

(ii) 对任意 $\Delta \in T_\mu\mathcal{P}(\mathcal{X})$ 及 f，若 $f' \in \mathcal{F}$ 且 $f - f' \equiv \kappa$ 对某个 $\kappa \in \mathbb{R}$ 成立，则有 $\Delta(f) = \Delta(f')$。

(iii) 若 $\tilde{\mathcal{F}}$ 是 μ-Donsker，则经验过程 $\sqrt{n}(\mu_n - \mu)$ 在 $l^\infty(\tilde{\mathcal{F}})$ 中的紧极限 $\mathbb{G}^\mu$ 几乎必然属于 $T_\mu\mathcal{P}(\mathcal{X})$。

6.3.1 定理 2.5 的证明

该结果可由泛函δ方法（Römisch，2006）推出。不失一般性，假设 $\mathcal{X} = \mu$的支撑和 $\mathcal{Y} = \nu$。这保证康托罗维奇势函数根据 (KP) 在整个定义域 $\mathcal{X}$ 和 $\mathcal{Y}$ 上唯一。假设（Don）结合底层随机变量与 μ 及 ν 的独立性，由 van der Vaart 和 Wellner（1996，例 1.4.6）可知，联合随机过程 $\sqrt{nm/m+n}(\mu_n - \mu, \nu_m - \nu)$ 在 $l^\infty(\cup_{\theta\in\Theta}\mathcal{F}_{c_\theta c^\theta}) \times l^\infty(\cup_{\theta\in\Theta}\mathcal{F}_{c^\theta})$ 中弱收敛。此外，由引理 6.8 可知，该极限几乎必然属于 $T_\mu\mathcal{P}(\mathcal{X}) \times T_\nu\mathcal{P}(\mathcal{Y})$。接下来只需证明映射

$$(OT(\cdot,\cdot,c_\theta))_{\theta\in\Theta} : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \subseteq \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}) \times \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}) \to C(\Theta),$$

$$(\mu,\nu) \mapsto \left(\theta \mapsto \sup_{f\in\mathcal{F}} \mu(f^{c_\theta,c_\theta}) + \nu(f^{c_\theta})\right)$$

在 (μ,ν) 处沿 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ Hadamard 方向可微。按照定理 A.2 的表述，取 $\mathcal{F}$ 和 Θ，并令

$$V := \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta}) \times \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}), \quad U := \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}), \quad E((\mu,\nu),f,\theta) := \mu(f^{c_\theta c_\theta}) + \nu(f^{c_\theta}).$$

那么，假设（EC）由引理 6.1 得出，而（Lip）和（Lin）根据 V 和 E 的定义很容易验证。此外，根据假设（KP），引理 A.3 中第 (ii) 点的条件成立，因为 E 对 f 的取值在 $(\Delta^\mu, \Delta^\nu) \in T_\mu\mathcal{P}(\mathcal{X}) \times T_\nu\mathcal{P}(\mathcal{Y})$ 下不受常数平移影响（引理 6.8），且康托罗维奇势在 $\mathcal{X}$ 和 $\mathcal{Y}$ 上除相差一个常数平移外唯一。这就证明了（DC），证毕。□

6.3.2 定理 2.6 的证明

由于 Θ 是一个紧 Polish 空间，由 Fang 和 Santos（2019，引理 S.4.9）（另见 Cárcamo、Cuevas 和 Rodríguez 2020年，推论 2.3）可知，下确界映射，

$$I : C(\Theta) \to \mathbb{R}, \quad h \mapsto \inf_{\theta\in\Theta} h(\theta),$$

在 $OT(\mu, \nu, c(\cdot)) \in C(\Theta)$ 处是 Hadamard 方向可微的，其导数为

$$D^H_{OT(\mu,\nu,c(\cdot))} I : C(\Theta) \to \mathbb{R}, \quad \Delta^h \mapsto \inf_{\theta\in S_-(\Theta,\mu,\nu)} \Delta^h(\theta).$$

因此，对定理 2.5 中经验最优传输过程的均匀弱极限应用下确界映射 I 的泛函δ方法（Römisch，2006），即可得出所述结论。□

6.3.3 定理 2.7 的证明

由对偶形式 (2.4)， Θ 上的上确界最优传输值为

$$\sup_{\theta\in\Theta} OT(\cdot,\cdot,c_\theta)\colon \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}) \subseteq \ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta c_\theta})\times\ell^\infty(\cup_{\theta\in\Theta}\mathcal{F}^{c_\theta}) \to \mathbb{R},$$
$$(\mu,\nu)\mapsto \sup_{(f,\theta)\in\mathcal{F}\times\Theta} \mu(f^{c_\theta c_\theta}) + \nu(f^{c_\theta}).$$

附录 A 的结果可直接应用，其中 V、U 和 E 的选取与定理 2.5 的证明相同；唯一的区别是，上确界取自 $\mathcal{F}\times\Theta$，而非 $\mathcal{F}$。特别地，（EC）、（Lip）和（Lin）均成立，而（DC）现在显然满足。总之，定理 A.2 断言，$\sup_{\theta\in\Theta} \mathrm{OT}(\cdot,\cdot,c_\theta)$ 在 $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ 上切向 Hadamard 方向可微，其导数为

$$D^H_{|(\mu,\nu)} \sup_{\theta\in\Theta} OT(\cdot,\cdot,c_\theta)\colon T_\mu\mathcal{P}(\mathcal{X})\times T_\nu\mathcal{P}(\mathcal{Y})\to\mathbb{R},\quad (\Delta^\mu,\Delta^\nu)\mapsto \sup_{\substack{\theta\in S_+(\Theta,\mu,\nu)\\ f_\theta\in S_{c_\theta}(\mu,\nu)}} \Delta^\mu(f_\theta^{c_\theta c_\theta}) + \Delta^\nu(f_\theta^{c_\theta}).$$

结合 (Don) 所给出的 $\sqrt{\mathrm{nm/n}+\mathrm{m}}(\overline{\mu_n-\mu,\ \nu_m}-\nu)$ 在 $\mathrm{l}^\infty(\cup_{\theta\in\Theta}\mathcal{F}_{c_\theta c^\theta})\times\mathrm{l}^\infty(\cup_{\theta\in\Theta}\mathcal{F}_{c^\theta})$ 中的弱收敛、底层样本的独立性（van der Vaart 和 Wellner，1996年，例 1.4.6），以及根据引理 6.8，极限属于 $\mathrm{T}_\mu\mathcal{P}(\mathcal{X})\times\mathrm{T}_\nu\mathcal{P}(\mathcal{Y})$，泛函 δ 方法（ Römisch，2006 年）即推出所述结论。□

注 6.9. 除了上述证明之外，也可以采用与 Fang 和 Santos（2019年，引理 S.4.9）或 Cárcamo、Cuevas 和 Rodríguez（2020年，推论 2.3）证明中类似的论证来证明定理 2.7。然而，他们的结论仅为在切空间 $\mathrm{C}(\cup_{\theta\in\Theta}\mathcal{F}_{c_\theta c^\theta})\times\mathrm{C}(\cup_{\theta\in\Theta}\mathcal{F}_{c^\theta})$ 上的 Hadamard 方向可微性提供充分条件，而上确界最优传输值仅定义在严格子集 $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ 上。

6.3.4 命题 2.8 的证明

定义 $\Delta(\mu_n,\nu_m,\mathrm{K}) := \inf_{\theta\in\Theta}$最优传输 $(\mu_n,\nu_m,\mathrm{c}_\theta) - \inf_{\theta\in K}$最优传输$(\mu_n,\nu_m,\mathrm{c}_\theta)$。那么，

$$\mathrm{P}^*(\Delta(\mu_n,\nu_m,K)\neq 0) \le \mathrm{P}^*(\{\Delta(\mu_n,\nu_m,K)\neq 0\}\cap\{\theta_{n,m}\in K\}) + \mathrm{P}^*(\theta_{n,m}\notin K),$$

并且，由于上面表达式中的第一项为零，而 $\lim_{n\to\infty}\mathrm{P}^*(\theta_{n,\ m}\notin \mathrm{K}) = 0$，右侧收敛到零。因此，应用斯卢茨基引理（van der Vaart 和 Wellner，1996年，例 1.4.7），由定理 2.6 可得

$$\begin{aligned}
&\sqrt{\frac{nm}{n+m}}\left(\inf_{\theta\in\Theta} OT(\mu_n,\nu_m,c_\theta) - \inf_{\theta\in\Theta} OT(\mu,\nu,c_\theta)\right)\\
&= \sqrt{\frac{nm}{n+m}}\Delta(\mu_n,\nu_m,K) + \sqrt{\frac{nm}{n+m}}\left(\inf_{\theta\in K} OT(\mu_n,\nu_m,c_\theta) - \inf_{\theta\in K} OT(\mu,\nu,c_\theta)\right)\\
&\rightsquigarrow 0 + \inf_{\theta\in S_-(K,\mu,\nu)} \sqrt{\lambda}\mathbb{G}^\mu(f_\theta^{c_\theta c_\theta}) + \sqrt{1-\lambda}\mathbb{G}^\nu(f_\theta^{c_\theta}).
\end{aligned}$$

观察到 $\mathrm{S}_-(\mathrm{K},\mu,\nu) = \mathrm{S}_-(\Theta,\mu,\nu)$ 后，结论立即成立。 □

致谢：S. Hundrieser 和 C.A. Weitkamp 衷心感谢德国研究基金会研究培训组 2088“发现复杂数据中的结构：统计学与优化及逆问题相遇”提供的支持。G. Mordant 衷心感谢德国研究基金会协同研究中心 1456“实验数学”A04 提供的支持；A. Munk 衷心感谢德国研究基金会协同研究中心 1456 的 A04、C06 项目，以及卓越集群 2067 MBExC“多尺度生物成像——从分子机器到可兴奋细胞网络”提供的支持。

参考文献

Ahmad, Najma、Hwa Kil Kim 和 Robert J. McCann（2011）。“最优输运、拓扑与唯一性”。数学科学通报 1.1，第 13–32 页。Albano, Paolo（2002）。“具有一般模数的半凹函数的一些性质”。数学分析与应用杂志 271.1，第 217–231 页。Alvarez-Melis, David、Stefanie Jegelka 和 Tommi S. Jaakkola（2019）。“迈向具有全局不变性的最优传输”。第 22 届人工智能与统计国际会议。PMLR，第 1870–1879 页。Arjovsky, Martin、Soumith Chintala 和 Léon Bottou（2017）。“Wasserstein生成对抗网络”。国际机器学习会议。PMLR，第 214–223 页。Aubin, J.P. 和 H. Frankowska（1990）。《集值分析》。现代Birkhäuser经典丛书。Springer。Beran, Rudolf J.、Lucien Le Cam 和 P. Warwick Millar（1987）。“随机经验测度的收敛”。《多元分析杂志》23.1，第 159–168 页。Bernton, Espen、Pierre E. Jacob、Mathieu Gerber 和 Christian P. Robert（2019）。“基于 Wasserstein 距离的参数估计”。Information and Inference: A Journal of the IMA 8.4，第 657–676 页。Billingsley, Patrick（1999）。《概率测度的收敛》。纽约：Wiley。Bing, Xin、Florentina Bunea 和 Jonathan Niles-Weed（2022）。“混合分布的草图瓦瑟斯坦距离”。预印本 arXiv:2206.12768。Bolley, François（2008）。“Wasserstein 距离的可分性与完备性”。《概率研讨会XLI》。Springer，第 371–377 页。Bonneel, Nicolas、Julien Rabin、Gabriel Peyré 和 Hanspeter Pfister（2015）。“测度的切片与 Radon Wasserstein 重心”。Journal of Mathematical Imaging and Vision 51.1，第 22–45 页。Bronshtein, Efim Mikhailovich（1976）。“凸集与函数的 ε-熵”。Siberian Mathematical Journal 17.3，第 393–398 页。Bücher, Axel 和 Ivan Kojadinovic（2019）。“关于自助法一致性结果中的条件弱收敛与联合无条件弱收敛的说明”。理论概率杂志 32.3，第 1145–1165 页。Cárcamo, Javier、Antonio Cuevas 和 Luis-Alberto Rodríguez（2020）。“上确界型泛函的方向可微性：统计应用”。《伯努利》26.3，第 2143–2175 页。Chernozhukov, Victor、Alfred Galichon、Marc Hallin 和 Marc Henry（2017）。“Monge–Kantorovich深度、分位数、秩与符号”。统计年刊 45.1，第 223–256 页。Csörgö, Miklós 和 Lajos Horváth（1993）。《概率与统计中的加权逼近》。约翰·威利父子公司。del Barrio, Eustasio、Evarist Giné 和 Carlos Matrán（1999）。“经验分布与真实分布之间 Wasserstein 距离的中心极限定理”。The Annals of Probability 27.2，第 1009–1071 页。del Barrio, Eustasio、Evarist Giné 和 Frederic Utzet（2005）。“经验分位数过程的 L_2 泛函的渐近性质及其在基于加权Wasserstein距离的拟合优度检验中的应用”。《伯努利》11.1，第 131–189 页。del Barrio, Eustasio、Alberto González-Sanz 和 Jean-Michel Loubes（2021）。“一般输运代价的中心极限定理”。预印本 arXiv:2102.06379。—（2022）。“半离散Wasserstein距离的中心极限定理”。预印本 arXiv:2202.0638

0.

del Barrio, Eustasio, Paula Gordaliza 和 Jean-Michel Loubes（2019）。“实数轴上 L^p 传输代价的中心极限定理及其在机器学习公平性评估中的应用”。Information and Inference: A Journal of the IMA 8.4，第 817–849 页。

del Barrio, Eustasio 和 Jean-Michel Loubes（2019）。“一般维度下经验传输代价的中心极限定理”。《概率年刊》47.2，第 926–951 页。Delon, Julie 和 Agnes Desolneux（2020）。“高斯混合模型空间中的一种 Wasserstein 型距离”。SIAM《成像科学期刊》13.2，第 936–970 页。Deshpande, Ishan 等（2019）。“最大切片 Wasserstein 距离及其在生成对抗网络中的应用”。IEEE/CVF 计算机视觉与模式识别会议论文集，第 10648–10656 页。Dette, Holger 和 Axel Munk（1998）。“线性回归模型的验证”。《统计年刊》26.2，第 778–800 页。Dette, Holger 和 Weichi Wu（2019）。“检测非平稳过程均值的相关变化——一种质量超额方法”。《统计年刊》47.6，第 3578–3608 页。Dudley, Richard M（1990）。“经验测度的非线性泛函与自助法”。《巴拿赫空间中的概率》7。Springer，第 63–82 页。Dümbgen, Lutz（1993）。“关于不可微函数与自助法”。《概率论及相关领域》95.1，第 125–140 页。Evans, Steven N 和 Frederick A Matsen（2012）。“环境序列样本的系统发育 Kantorovich–Rubinstein 度量”。《皇家统计学会杂志：B辑（统计方法学）》74.3，第 569–592 页。Fang, Zheng 和 Andres Santos（2019）。“方向可微函数的推断”。《经济研究评论》86.1，第 377–412 页。Fournier, Nicolas 和 Arnaud Guillin（2015）。“经验测度在 Wasserstein 距离下的收敛速率”。《概率论及相关领域》162.3，第 707–738 页。Gal, Tomas 和 Harvey J. Greenberg（1997）。《敏感性分析与参数规划进展》。第 1 卷。Springer Science & Business Media。Gangbo, Wilfrid 和 Robert J. McCann（1996）。“最优传输的几何学”。《数学学报》177.2，第 113–161 页。Gellert, Manuela 等（2019）。“硫氧还蛋白和谷氧还蛋白的底物特异性——迈向功能分类”。Heliyon 5.12，e02943。Giné, Evarist 和 Richard Nickl（2016）。《无限维统计模型的数学基础》。剑桥大学出版社。Goldfeld, Ziv、Kengo Kato、Gabriel Rioux 和 Ritwik Sadhu（2022）。“正则化最优传输的统计推断”。预印本 arXiv:2205.04283。Grave, Edouard、Armand Joulin 和 Quentin Berthet（2019）。“使用 Wasserstein Procrustes 对嵌入进行无监督对齐”。第 22 届人工智能与统计国际会议。PMLR，第 1880–1890 页。Guntuboyina, Adityanand 和 Bodhisattva Sen（2013）。“凸函数的覆盖数”。IEEE《信息论汇刊》59.4，第 1957–1965 页。Hallin, Marc、Eustasio del Barrio、Juan Cuesta-Albertos 和 Carlos Matrán （ 2021 ）。“d 维中的分布函数与分位数函数、秩和符号：一种测度传输方法”。《统计年刊》49.2，第 1139–1165 页。Hallin, Marc、Gilles Mordant 和 J ohan Segers（2021）。“基于 Wasserstein 距离的多变量拟合优度检验”。《电子统计学杂志》15.1，第 1328–1371 页。Hundrieser, Shayan、Marcel Klatt、Thomas Staudt 和 Axel Munk（2022）。“经验最优传输分布极限的统一方法”。预印本 arXiv:2202.12790。Hundrieser, Shayan、Thomas Staudt 和 Axel Munk（2022）。“不同测度之间的经验最优传输可适应较低复杂度”。预印本 arXiv:2202.10434。Jin, Kun、Chaoyue Liu 和 Cathy Xia（2021）。“双侧 Wasserstein Procrustes 分析”。IJCAI，第 3515–3521 页。

Jourdain, Benjamin 和 Alvin Tse（2021）。“经验测度非线性泛函的中心极限定理及其在相互作用扩散过程平均场涨落中的应用”。《概率电子期刊》26，页码 1–34。Klatt, Marcel、Axel Munk 和 Yoav Zemel（2022）。“随机线性规划中经验最优解的极限定律”。《运筹学年鉴》315，页码 251–278。Kolmogorov, Andrei N. 和 Vladimir M. Tikhomirov（1961）。“函数空间中集合的 ϱ-熵与 ϱ-容量”。《代数与实函数论文十二篇》。编者：S.N. Cernikov、N.V. Cernikova、A.N. Kolmogorov、A.I. Mal’cev 和 B.I. Plotkin。《美国数学会译丛》第二辑。美国数学会，页码 277–364。Levin, Vladimir（1999）。“一般 Monge–Kantorovich 问题的抽象循环单调性与蒙日解”。《集值分析》7.1，页码 7–32。Manole, Tudor、Sivaraman Balakrishnan 和 Larry Wasserman（2022）。“切片 Wasserstein 距离的极小极大置信区间”。《电子统计学杂志》16.1，页码 2252–2345。Manole, Tudor 和 Jonathan Niles-Weed（2021）。“光滑代价函数下经验最优传输的精确收敛速率”。预印本 arXiv:2106.13181v2。Mason, David M.（2016）。“研究经验 Wasserstein 距离的加权逼近方法”。《高维概率 VII》。Springer，页码 137–154。McCann, Robert J. 和 Ludovic Rifford（2016）。“最优传输的内禀动力学”。《巴黎综合理工学院数学杂志》3，页码 67–98。McLachlan, Geoffrey J.、Sharon X. Lee 和 Suren I. Rathnayake（2019）。“有限混合模型”。《统计学及其应用年度评论》6.1，页码 355–378。Mordant, Gilles 和 Johan Segers （ 2022 ）。“ 通过最优传输度量随机向量之间的依赖性”。《多元分析杂志》189，页码 104912。Munk, Axel 和 Claudia Czado（1998）。“相似分布的非参数验证与拟合优度评估”。《皇家统计学会杂志：B辑（统计方法学）》60.1，页码 223–241。Nies, Thomas Giacomo、Thomas Staudt 和 Axel Munk（2021）。“传输依赖性：基于最优传输的依赖性度量”。预印本 arXiv:2105.02073。Rachev, Svetlozar T. 和 Ludger Rüschendorf（1998）。《质量传输问题：第一卷：理论》。第 1 卷。Springer Science & Business Media。Römisch, Werner（2006）。“δ 方法，无限维”。《统计科学百科全书》。第 16 卷。纽约：Wiley，页码 1575–1583。Rubner, Yossi、Carlo Tomasi 和 Leonidas J. Guibas（2000）。“作为图像检索度量的推土机距离”。《国际计算机视觉杂志》40.2，页码 99–121。Santambrogio, Filippo（2015）。《应用数学家的最优传输》。第 87 卷。《非线性微分方程及其应用进展》。Birkhäuser/Springer，Cham。Schiebinger, Geoffrey 等（2019）。“单细胞基因表达的最优传输分析揭示重编程中的发育轨迹”。《细胞》176.4，页码 928–943。Singh, Shashank 和 Barnabás Póczos（2018）。“Wasserstein 距离下的极小极大分布估计”。预印本 arXiv:1802.08855。Sommerfeld, Max 和 Axel Munk（2018）。“有限空间上经验 Wasserstein 距离的推断”。《皇家统计学会杂志：B辑（统计方法学）》80.1，页码 219–238。Staudt, Thomas、Shayan Hundrieser 和 Axel Munk（2022）。“关于康托罗维奇势函数的唯一性”。预印本 arXiv:2201.08316。Tameling, Carla、Max Sommerfeld 和 Axel Munk（2019）。“可数度量空间上的经验最优传输：分布极限与统计应用”。《应用概率年刊》29.5，页码 2744–2781。

Tameling, Carla 等（2021）。“通过最优传输实现超分辨率显微术中的共定位”。《自然·计算科学》1.3，页码 199–211。Toma, Vladimír（1997）。“非一致空间中的强收敛与 Dini 定理”。《布莱兹·帕斯卡数学年刊》。第 4 卷，第 2 期，页码 97–102。Torous, William、Florian Gunsilius 和 Philippe Rigollet（2021）。“因果推断的最优传输方法”。预印本 arXiv:2108.05858。van der Vaart, Aad W.（1998）。《渐近统计学》。剑桥统计与概率数学丛书。剑桥大学出版社。van der Vaart, Aad W. 和 Jon A. Wellner（1996）。《弱收敛与经验过程：及其在统计学中的应用》。施普林格统计学丛书。Springer。—（2007）。“由估计函数索引的经验过程”。《渐近学：粒子、过程与逆问题》。数理统计学会，页码 234–252。Varadarajan, Veeravalli S.（1958）。“关于样本概率分布的收敛”。Sankhy¯a：《印度统计学杂志》（1933–1960）19.1/2，页码 23–26。Villani, Cédric（2003）。《最优输运专题》。第 58 卷。美国数学会。—（2008）。《最优传输：旧与新》。第 338 卷。Springer Science & Business Media。Wainwright, Martin J.（2019）。《高维统计学：非渐近视角》。第 48 卷。剑桥统计与概率数学丛书。剑桥大学出版社。Weed, Jonathan 和 Francis Bach（2019）。“经验测度在 Wasserstein 距离下的精确渐近收敛速率与有限样本收敛速率”。《伯努利》25.4A，页码 2620–2648。Weitkamp, Christoph Alexander、Katharina Proksch、Carla Tameling 和 Axel Munk（2022）。“基于距离分布的对象匹配：渐近推断”。《美国统计学会杂志》，即将发表。Wiesel, Johannes C.W.（2022）。“用 Wasserstein 距离度量关联”。《伯努利》28.4，页码 2816–2832。Xi, Jiaqi 和 Jonathan Niles-Weed（2022）。“切片 Wasserstein 过程的分布收敛”。预印本 arXiv:2206.00156。Xu, Xianliang 和 Zhongyi Huang（2022）。“切片 1-Wasserstein 距离与最大切片 1-Wasserstein 距离的中心极限定理”。预印本 arXiv:2205.14624。

极值型泛函的一致 Hadamard 方向可微性

本文中的许多结果都依赖于 Hadamard 方向可微性和泛函δ方法的概念。更具体地说，第 2.3 节关于经验最优传输过程弱收敛的结果，以及第 5 节对正则性提升泛函的表述，都采用了这一方法。尽管这两项发现从概念上看并无太大关联，但其证明技术都基于一个更一般的见解，我们将在本节中加以阐述。

设 $(\mathrm{V}, \|\cdot\|_V)$ 为一个赋范向量空间，并考虑集合 $\mathcal{F}$ 和 Θ。此外，考虑一个实值函数 $\mathrm{E} : \mathrm{V} \times \mathcal{F} \times \Theta \to \mathbb{R}$，它将每个三元组 $(\mathrm{v}, \mathrm{f}, \theta)$ 映射为一个目标值 $\mathrm{E}(\mathrm{v}, \mathrm{f}, \theta)$。我们关注极值型泛函的敏感性结果

$$\Psi(v) := \left(\sup_{f \in \mathcal{F}} E(v, f, \theta) \right)_{\theta \in \Theta} \qquad \text{and} \qquad \tilde{\Psi}(v) := \left(\inf_{f \in \mathcal{F}} E(v, f, \theta) \right)_{\theta \in \Theta}.$$

其中，Θ表示影响该优化问题的可行参数集合，而 $\mathcal{F}$ 表示可行解集合。空间 V 表示另一组决定该优化问题的参数，并具有向量空间结构。总之，这些优化问题刻画了以 Θ为指标的随机过程的一般结构；这类过程逐点定义为某个集合 $\mathcal{F}$ 上的上确界或下确界，并依赖于 V 中具有加法结构的某个参数。

为分析 v 的扰动下的敏感性，我们只需关注 Ψ，因为

$$\inf_{f \in \mathcal{F}} E(v, f, \theta) = - \sup_{f \in \mathcal{F}} -E(v, f, \theta) \quad \text{for any } (v, \theta) \in V \times \Theta.$$

接下来，我们首先给出关于 E 的充分条件，以保证 Ψ 及其底层最优解集合的连续性性质。

引理 A.1（连续性）。设 $(\mathrm{V}, \|\cdot\|_V)$ 为一个赋范向量空间，考虑拓扑分别由（伪）度量 $\mathrm{d}_{\mathcal{F}}$ 和 d_Θ 生成的紧拓扑空间 $\mathcal{F}$ 和 Θ，并假设 $\mathrm{E} : \mathrm{V} \times \mathcal{F} \times \Theta \to \mathbb{R}$ 满足以下条件。

（EC）对于任意 $\mathrm{v} \in \mathrm{V}$，泛函 $\mathrm{E}(\mathrm{v}, \cdot, \cdot) : \mathcal{F} \times \Theta \to \mathbb{R}$ 连续。

（Lip）存在某个 $\mathrm{L} \geq 0$，使得对于任意 $(\mathrm{f}, \theta) \in \mathcal{F} \times \Theta$，泛函 $\mathrm{E}(\cdot, \mathrm{f}, \theta): \mathrm{V} \to \mathbb{R}$ 关于 $\|\cdot\|_V$ 是 L-利普希茨的。

则 $\mathrm{Range}(\Psi) \subseteq \mathrm{C}(\Theta)$，且泛函 $\Psi: \mathrm{V} \to \mathrm{C}(\Theta)$ 是 L-利普希茨的。此外，对于任意 $(\mathrm{v}, \theta) \in \mathrm{V} \times \Theta$，最优解集合 $\mathrm{S}(\mathrm{v}, \theta) := \{\mathrm{f} \in \mathcal{F} | \sup \mathrm{f}' \in \mathcal{F} \mathrm{E}(\mathrm{v}, \mathrm{f}', \theta) = \mathrm{E}(\mathrm{v}, \mathrm{f}, \theta)\}$ 非空；对于固定的 $\mathrm{v} \in \mathrm{V}$，集合值映射

$$(\theta, t) \in \Theta \times \mathbb{R}_+ \mapsto S(v, \theta; t) := \left\{ f \in \mathcal{F} \;\middle|\; \sup_{f' \in \mathcal{F}} E(v, f', \theta) \leq E(v, f, \theta) + t \right\}$$

关于包含关系是上半连续的，即对于 $\theta_n \to \theta$ and $\mathrm{t}_n \to \mathrm{t}$，任意序列 $\mathrm{f}_n \in \mathrm{S}(\mathrm{v}, \theta_n; \mathrm{t}_n)$ 都存在一个收敛子序列 $(\mathrm{f}_{n_k})_{k \in \mathbb{N}}$，其在 $\mathcal{F}$ 中的极限 f 满足 $\mathrm{f} \in \mathrm{S}(\mathrm{v}, \theta; \mathrm{t})$。

引理 A.1 的证明。由假设（EC）以及 $\Theta \times \mathcal{F}$ 的紧致性可知，对于任意 $\mathrm{v} \in \mathrm{V}$，$\mathrm{E}(\mathrm{v}, \cdot, \cdot)$ 一致连续，因此函数

$$w_{E,v} : \mathbb{R}_+ \to \mathbb{R}_+, \quad t \mapsto \sup_{\substack{d_\Theta(\theta, \theta') \leq t \\ d_{\mathcal{F}}(f, f') \leq t}} |E(v, f, \theta) - E(v, f', \theta')|$$

对所有 $t \geq 0$ 都是有限的，且满足 $\lim_{t \searrow 0} w_{E,v}(t) = 0$。因此，对于 $\theta, \theta' \in \Theta$，我们有

$$\left| \sup_{f \in \mathcal{F}} E(v, f, \theta) - \sup_{f \in \mathcal{F}} E(v, f, \theta') \right| \leq \sup_{f \in \mathcal{F}} |E(v, f, \theta) - E(v, f, \theta')| \leq w_{E,v}(d_\Theta(\theta, \theta')),$$

这意味着对于 $v \in V$，有 $\Psi(v) \in C(\Theta)$，因此 $\mathrm{Range}(\Psi) \subseteq C(\Theta)$。关于 Ψ 的利普希茨连续性，根据假设（Lip），对于任意 $v, v' \in V$，有

$$\begin{aligned} \left\| \Psi(v) - \Psi(v') \right\|_{C(\Theta)} &= \sup_{\theta \in \Theta} \left| \sup_{f \in \mathcal{F}} E(v, f, \theta) - \sup_{f \in \mathcal{F}} E(v', f, \theta) \right| \\ &\leq \sup_{\substack{\theta \in \Theta \\ f \in \mathcal{F}}} |E(v, f, \theta) - E(v', f, \theta)| \leq L \left\| v - v' \right\|_V . \end{aligned}$$

为证明 $S(v, \theta) \neq \emptyset$，注意到函数 $E(v, \cdot, \theta): \mathcal{F} \to \mathbb{R}$ 对任意 $(v, \theta) \in V \times \Theta$ 都是连续的；因此，根据 $\mathcal{F}$ 的紧致性，关于 $\mathcal{F}$ 的上确界可以取到。

下面仍需证明关于上半连续性的断言。考虑收敛序列 $t_n \to t \geq 0$ 和 $\theta_n \to \theta \in \Theta$，并取序列 $f_n \in S(v, \theta_n; t_n)$。由 $\mathcal{F}$ 的紧致性，存在一个收敛子序列 $(f_{n_k})_{k \in \mathbb{N}}$，其极限为 $f \in \mathcal{F}$。因此，根据假设（EC），并且由于 $\sup_{f \in \mathcal{F}} E(v, f, \cdot) = \Psi(v)(\cdot) \in C(\Theta)$，我们得到 $f \in S(v, \theta; t)$，因为

$$E(v, f, \theta) + t = \lim_{k \to \infty} E(v, f_{n_k}, \theta_{n_k}) + t_{n_k} \geq \lim_{k \to \infty} \sup_{f \in \mathcal{F}} E(v, f, \theta_{n_k}) = \sup_{f \in \mathcal{F}} E(v, f, \theta). \qquad \square$$

有了这些工具，我们可以陈述一般的敏感性结果。

定理 A.2（可微性）。假设处于引理 A.1 的设定中，且满足条件（EC）和（Lip）。令 $v \in V$，并考虑一个凸集 $U \subset V$。将 $T_v U := \mathrm{Cl}\{t(\bar{v} - v) \mid t > 0, \bar{v} \in U\} \subseteq V$ 在 v 处的切锥记为它。此外，假设如下：

（Lin）对于任意 $(f, \theta) \in \mathcal{F} \times \Theta$，函数 $\Delta_{\bar{v}} E(\cdot, f, \theta): V \to \mathbb{R}, v \mapsto E(v + \bar{v}, f, \theta) - E(v, f, \theta)$ 是线性的。

（DC）对于任意 $h \in T_v U$，函数 $\theta \in \Theta \mapsto \sup_{f \in S} \Delta_{\bar{v}} E(h, f, \theta)$ 是下半连续的。

那么，该泛函

$$\Psi: V \to C(\Theta), \quad v \mapsto \left(\sup_{f \in \mathcal{F}} E(v, f, \theta) \right)_{\theta \in \Theta}$$

在 v 处沿 U Hadamard 方向可微，其导数为

$$D^H_{\bar{v}} \Psi: T_v U \to C(\Theta), \quad h \mapsto \left(\sup_{f \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(h, f, \theta) \right)_{\theta \in \Theta} .$$

定理 A.2 可视为 Fang 和 Santos（2019，引理 S.4.9）的扩展，将其推广为参数空间 Θ 上的一致扰动结果。此外，我们的结果不要求整个定义域 V 满足正则性，只要求凸集 U 满足正则性。这一优点颇具吸引力，我们在分析 OT 过程（其中选择 $U = \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$）以及正则性提升时都利用了这一点（见命题 5.3 的证明）。

假设（EC）、（Lip）和（Lin）相当直观，通常也很容易验证。前两个条件似乎也是推断 $\mathrm{Range}(\Psi) \subseteq C(\Theta)$ 以及 $\Psi: V \to C(\Theta)$ 的利普希茨连续性所必需的。假设（DC）则更具技术性，需要对最优解集合 $S(v, \theta)$ 有所了解。正如定理 A.2 的证明所揭示的，在引理 A.1 的假设下，泛函 $\theta \in \Theta \mapsto \sup_{f \in S(v, \theta)} E(h, f, \theta)$ 总是上半连续的。因此，（DC）的唯一作用是确保 $\mathrm{Range}(D^H_v \Psi) \subseteq C(\Theta)$。其成立的充分条件如下所述。

引理 A.3. 设定与引理 A.1 和定理 A.2 相同。若满足以下任一条件，则定理 A.2 中的假设（DC）成立。

(i) 对于任意 $\theta \in \Theta$ 和 $h \in T_{\bar{v}}U$，存在 $f \in S(v, \theta)$，使得 $\sup_{f' \in S(v,\theta)} \Delta_{\bar{v}} E(h, f', \theta) = \Delta_{\bar{v}} E(h, f, \theta)$，从而任意收敛序列 $\theta_n \to \theta$ 都存在一个子序列 (θ_{n_k}) 以及一个收敛序列 $f_{n_k} \in S(v, \theta_{n_k})$，并且 $f_{n_k} \to f$ 属于 $\mathcal{F}$。

(*ii*) 对于任意 $\theta \in \Theta$ 和 $h \in T_{\bar{v}}U$，有 $\Delta_{\bar{v}} E(h, f, \theta) = \Delta_{\bar{v}} E(h, f', \theta)$，其中任意 $f, f' \in S(v, \theta)$.

引理 A.3 的证明。设 $\theta_n \to \theta$，并考虑元素 $f \in S(v, \theta)$，使得 $\Delta_{\bar{v}} E(h, f, \theta) = \sup_{f' \in S(v,\theta)} \Delta_{\bar{v}} E(h, f, \theta)$。对于情形 (i)，取任意子序列 θ_{n_k}，并再取一个子序列 $\theta_{n_{k_l}}$，使得 $f^{n_{k_l}} \in S(v, \theta_{n_{k_l}})$ 收敛到 f（对 $l \to \infty$）。于是，根据 (EC),

$$\lim_{l \to \infty} \Delta_{\bar{v}} E(h, f_{n_{k_l}}, \theta_{n_{k_l}}) = \Delta_{\bar{v}} E(h, f, \theta) = \sup_{f' \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(h, f', \theta).$$

由下极限的单调性和引理 F.1 可知,

$$\liminf_{n \to \infty} \sup_{f' \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(h, f', \theta_n) \geq \liminf_{n \to \infty} \Delta_{\bar{v}} E(h, f_n, \theta_n) \geq \sup_{f' \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(h, f', \theta),$$

这证明了定理 A.2 中假设（DC）的有效性。对于情形 (ii)，取 $f_n \in S(v, \theta_n)$，并根据引理 A.1 考虑一个收敛子序列 f_{n_k}，其极限为 $f \in S(v, \theta)$。因此，有 $\Delta_{\bar{v}} E(h, f, \theta) = \overline{\sup}_{f' \in S(v,\theta)} \Delta_{\bar{v}} E(h, f', \theta)$，结论由 (i) 得出。□

定理 A.2 的证明。证明策略受到 Römisch（2006）的启发，他针对 Θ 为单元素集的情形进行了敏感性分析。为了将结论推广到紧拓扑空间 Θ，我们采用以下版本的 Dini定理。

引理 A.4（Dini定理，Toma 1997，推论 1）。设 Θ 为紧拓扑空间，并考虑一个递减的 $f_n: \Theta \to \mathbb{R}$ 序列（即，对于所有 $n \in \mathbb{N}$，都有 $f_n \geq f_{n+1}$），其成员均为上半连续函数。此外，假设 f_n 逐点收敛到一个（下半）连续函数 $f: \Theta \to \mathbb{R}$。那么，f_n 在 Θ 上一致收敛到 f。

取一个正零序列 $t_n \searrow 0$，使得对所有 $n \in \mathbb{N}$ 都有 $t_n > 0$，并令 $h \in T_{\bar{v}}U$。此外，取一个序列 $h_n \in V$，使得对所有 $n \in \mathbb{N}$ 都有 $v_n := v + t_n h_n \in U$，且 $h_n \to h$ 属于 V。对于任意 $\theta \in \Theta$，由 (Lin) 和 (Lip) 可知，对任意 $n \in \mathbb{N}$，下界为

$$\begin{aligned}
\Psi(v_n)(\theta) - \Psi(\bar{v})(\theta) &= \sup_{f \in \mathcal{F}} E(v_n, f, \theta) - \sup_{f \in \mathcal{F}} E(\bar{v}, f, \theta) \\
&\geq \sup_{f \in S(\bar{v}, \theta)} E(v_n, f, \theta) - E(\bar{v}, f, \theta) \\
&\geq \sup_{f \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(t_n h_n, f, \theta) \\
&\geq t_n \sup_{f \in S(\bar{v}, \theta)} \Delta_{\bar{v}} E(h, f, \theta) - 2 t_n L \|h - h_n\|_V . \qquad \text{(A.1)}
\end{aligned}$$

类似地，我们得到上界

$$\begin{aligned}
\Psi(v_n)(\theta) - \Psi(\bar{v})(\theta) &= \sup_{f \in \mathcal{F}} E(v_n, f, \theta) - \sup_{f \in \mathcal{F}} E(\bar{v}, f, \theta) \\
&\leq \sup_{f \in S(v_n, \theta)} E(v_n, f, \theta) - E(\bar{v}, f, \theta) \\
&\leq t_n \sup_{f \in S(v_n, \theta)} \Delta_{\bar{v}} E(h, f, \theta) + 2 t_n L \|h - h_n\|_V . \qquad \text{(A.2)}
\end{aligned}$$

注意，$S(v_n,\theta) \subseteq S(v,\theta;2L\|v_n-v\|_V)$，因为任意 $f^* \in S(v_n,\theta)$ 根据（Lip）都满足以下界

$$\begin{aligned} E(v,f^*,\theta) &\geq E(v_n,f^*,\theta) - L\|v_n-v\|_V \\ &= \sup_{f\in\mathcal{F}} E(v_n,f,\theta) - L\|v_n-v\|_V \\ &\geq \sup_{f\in\mathcal{F}} E(v,f,\theta) - 2L\|v_n-v\|_V . \end{aligned}$$

因此，根据（A.2），定义 $\varepsilon_n := \sup_{k\geq n} 2L\|v_k-v\|_V$ 后可得

$$\begin{aligned} \Psi(v_n)(\theta)-\Psi(v)(\theta) &\leq t_n \sup_{f\in S(v,\theta;2L\|v_n-v\|_V)} \Delta_{|v}E(h,f,\theta) + 2t_nL\|h-h_n\|_V \\ &\leq t_n \sup_{f\in S(v,\theta;\varepsilon_n)} \Delta_{|v}E(h,f,\theta) + 2t_nL\|h-h_n\|_V . \end{aligned} \tag{A.3}$$

结合（A.1）和（A.3），我们因此对任意 $\theta\in\Theta$ 得到

$$\sup_{f\in S(v,\theta)} \Delta_{|v}E(h,f,\theta) - 2L\|h-h_n\|_V \leq \frac{\Psi(v_n)(\theta)-\Psi(v)(\theta)}{t_n} \leq \sup_{f\in S(v,\theta;\varepsilon_n)} \Delta_{|v}E(h,f,\theta) + 2L\|h-h_n\|_V .$$

为完成该论断，我们证明下界和上界在 Θ 上一致收敛（当 $n\to\infty$ 时）到 $D^H_{|v}\Psi$。由于 $\|h_n-h\|_V\to\infty$，只需证明对于以下函数

$$\Phi := D^H_{|v}\Psi : \Theta\to\mathbb{R},\ \theta\mapsto \sup_{f\in S(v,\theta)} \Delta_{|v}E(h,f,\theta), \qquad \Phi_n : \Theta\to\mathbb{R},\ \theta\mapsto \sup_{f\in S(v,\theta,\varepsilon_n)} \Delta_{|v}E(h,f,\theta),$$

满足 $\lim_{n\to\infty}\|\Phi-\Phi_n\|_{C(\Theta)}=0$。为此，我们使用 Dini定理（引理 A.4）。

在此需要注意，由于 $(\varepsilon_n)_{n\in\mathbb{N}}$ 是一个递减的零序列，对于所有 $n\in\mathbb{N}$ 以及任意 $\theta\in\Theta$，有 $S(v,\theta)\subseteq \bar{S}(v,\theta;\bar{\varepsilon}_{n+1})\subseteq S(v,\theta;\bar{\varepsilon}_n)$，因此

$$\Phi(\theta)\leq\Phi_{n+1}(\theta)\leq\Phi_n(\theta)\leq 2\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}} E(h,f,\theta)<\infty, \tag{A.4}$$

其中，上界由于假设（EC）以及 $\mathcal{F}\times\Theta$ 的紧致性而为有限值。

进一步，我们证明对任意 $\theta\in\Theta$，有 $\lim_{n\to\infty}\Phi_n(\theta)=\Phi(\theta)$。取一个序列 $f_n\in S(v,\theta;\bar{\varepsilon}_n)$ 使得 $\Phi_n(\theta)\leq \Delta_{|v}E(h,f_n,\theta)+1/n$。考虑一个收敛子序列 $(f_{n_k})_{k\in\mathbb{N}}$，其极限为 $f_\infty\in S(v,\theta)$。于是，由（EC）可得

$$\limsup_{k\to\infty}\Phi_{n_k}(\theta)\leq\lim_{k\to\infty}\Delta_{|v}E(h,f_{n_k},\theta)+1/n_k=\Delta_{|v}E(h,f_\infty,\theta)\leq\sup_{f\in S(v,\theta)}\Delta_{|v}E(v,f,\theta)=\Phi(\theta).$$

回顾（A.4），由此可得 $\lim_{n\to\infty}\Phi_n(\theta)=\Phi(\theta)$。

要借助 Dini定理得出该结论，还需证明 Φ_n 和 Φ 上半连续；回顾假设（DC），可知 Φ 已经是下半连续的。为此，取 $\varepsilon\geq 0$，并考虑一个收敛序列 $\theta_n\to\theta$。选取 $f_n\in S(v,\theta_n,\varepsilon)$，使得 $\sup_{f\in S(v,\theta_n,\varepsilon)}\Delta_{|v}E(h,f,\theta_n)\leq\Delta_{|v}E(h,f_n,\theta_n)+1/n$。取一个子序列 f_{n_k}，并根据引理 A.1 再选取另一个收敛子序列 $f^{n_{k_l}}$，其极限为 $f_\infty\in S(v,\theta;\varepsilon)$。利用假设（EC），由此可得

$$\lim_{l\to\infty}\Delta_{|v}E(h,f_{n_{k_l}},\theta_{n_{k_l}})+1/n_{k_l}=\Delta_{|v}E(h,f_\infty,\theta)\leq\sup_{f\in S(v,\theta;\varepsilon)}\Delta_{|v}E(h,f,\theta)$$

利用上极限的单调性和引理 F.1，我们由此得到

$$\limsup_{n\to\infty}\sup_{f\in S(v,\theta_n;\varepsilon)}\Delta_{|v}E(h,f,\theta)\leq\limsup_{l\to\infty}\Delta_{|v}E(h,f_n,\theta_n)+1/n\leq\sup_{f\in S(v,\theta;\varepsilon)}\Delta_{|v}E(h,f,\theta),$$

因此，由引理 F.1 可得，Φ_n 是上半连续的，而 Φ 是连续的。由此，Dini定理（引理 A.4）推出 $\lim_{n\to\infty}\|\Phi-\Phi_n\|_\infty=0$，这表明 Ψ 在 v 处沿 U 方向具有 Hadamard 方向可微性。最后注意，$D^H{}_v\Psi$ 的值域确实包含于 $C(\Theta)$ 中。□

B 第 3 节证明：假设的充分判据

B.1 命题 3.1 的证明

由引理 2.1 可知，$\mathcal{F}_c \subseteq \mathcal{H}_c^c + [-2B, 2B]$ 且 $\mathcal{F}_{cc} \subseteq \mathcal{H}_c + [-2B, 2B]$，其中 $\mathcal{H}_c$ 由式 (2.1) 定义。根据 Hundrieser, Staudt, and Munk（2022，引理 2.1）和 Santambrogio（2015，命题 1.34），可得对任意 $\varepsilon > 0$，

$$\mathcal{N}(\varepsilon, \mathcal{F}^c, \|\cdot\|_\infty) = \mathcal{N}(\varepsilon, \mathcal{F}^{cc}, \|\cdot\|_\infty) \leq \left\lceil \frac{2B}{\varepsilon} \right\rceil \mathcal{N}(\varepsilon/2, \mathcal{H}_c^c, \|\cdot\|_\infty) = \left\lceil \frac{2B}{\varepsilon} \right\rceil \mathcal{N}(\varepsilon/2, \mathcal{H}_c, \|\cdot\|_\infty).$$

对于函数类 $\mathcal{H}_c$，所述一致度量熵界见 Hundrieser, Staudt, and Munk (2022) 的第3.1节和附录 A。注意，由于代价函数一致有界，$\mathcal{H}_c$ 和 $\mathcal{H}_c^c$ 均一致有界。关于普适 Donsker 性质的结论由 van der Vaart 和 Wellner (1996, 定理 2.5.6) 推出。□

B.2 命题 3.3 的证明

根据假设，泛函

$$\overline{\Phi}_c : \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \ell^\infty(\mathcal{F}^{cc}) \times \ell^\infty(\mathcal{F}^c) \times C(\mathcal{X} \times \mathcal{Y})$$
$$(\mu, \nu) \mapsto (\mu, \nu, \Phi_c(\mu, \nu)),$$

其中定义域被视为 $\mathrm{l}^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}\mathrm{cc}) \times \mathrm{l}^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}\mathrm{c})$ 的子集，在 (μ, ν) 处沿着 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 切向 Hadamard 可微。此外，由于 $\mathcal{F}_\mathcal{X} \cup \mathcal{F}\mathrm{cc}$ 是 μ-Donsker，可得$\sqrt{\mathrm{n}/2}(\mu_n - \mu)$ $\mathbb{G}^\mu$ 属于 $\mathrm{l}^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}\mathrm{cc})$。同理，由于 $\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c$ 是 ν-Donsker，可得$\sqrt{\mathrm{n}/2}(\nu_n - \nu)$ $\mathbb{G}^\nu$ 属于 $\mathrm{l}^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}^c)$。此外，由于随机变量 $\{\mathrm{X}_i\}_{i=1}^n$和 $\{\mathrm{Y}_i\}_{i=1}^n$ 相互独立，根据 van der Vaart 和 Wellner, 1996, 定理 1.4.6，联合经验过程 $\sqrt{\mathrm{n}/2}(\mu_n - \mu, \nu_n - \nu)$ 在 $\mathrm{l}^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}\mathrm{cc}) \times \mathrm{l}^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}\mathrm{c})$ 中弱收敛至 $(\mathbb{G}^\mu, \mathbb{G}^\nu)$；根据引理 6.8，该极限属于 $\mathrm{T}_\mu\mathcal{P}(\mathcal{X}) \times \mathrm{T}_\nu\mathcal{P}(\mathcal{Y})$。因此，由泛函 δ 方法（Römisch，2006），对于 Φ_c，(JW) 得以满足。

此外，由于 Donsker 性质和随机变量的独立性，根据 van der Vaart 和 Wellner（1996 年，定理3.6.13），我们还可以在空间 $\mathrm{l}^\infty(\mathcal{F}_\mathcal{X} \cup \mathcal{F}^{cc}) \times \mathrm{l}^\infty(\mathcal{F}_\mathcal{Y} \cup \mathcal{F}_c)$ 中推得

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k} \begin{pmatrix} \mu_{n,k}^b - \mu_n \\ \nu_{n,k}^b - \nu_n \end{pmatrix} \middle| X_1, \ldots, X_n, Y_1, \ldots Y_n \right), \mathcal{L}\begin{pmatrix} \mathbb{G}^\mu \\ \mathbb{G}^\nu \end{pmatrix} \right) \xrightarrow{\mathrm{P}^*} 0.$$

因此，对于条件弱收敛随机变量的泛函δ方法（Düm- bgen，1993年），我们可针对 Ψ_c w 推得

$$d_{BL}\left(\mathcal{L}\left(\sqrt{k} \begin{pmatrix} \mu_{n,k}^b - \mu_n \\ \nu_{n,k}^b - \nu_n \\ c_{n,k}^b - c_n \end{pmatrix} \middle| X_1, \ldots, X_n, Y_1, \ldots Y_n \right), \mathcal{L}\left(\sqrt{n} \begin{pmatrix} \mu_n - \mu \\ \nu_n - \nu \\ c_n - c \end{pmatrix} \right) \right) \xrightarrow{\mathrm{P}^*} 0. \qquad \square$$

B.3 命题 3.6 的证明

在开始证明命题 3.6 之前，我们先建立一个辅助引理。

引理B.1。设 $\mathcal{X}$ 和 $\mathcal{Y}$ 为紧波兰空间，并考虑 $\mathrm{c} \in \mathrm{C}(\mathcal{X} \times \mathcal{Y})$。

(i) 对任意函数 $\mathrm{g} : \mathcal{X} \to \mathbb{R}$ 和任意常数 κ，均有 $(\mathrm{g} + \kappa)^c = \mathrm{g}^c - \kappa$。

(ii) 令 $\mathrm{B} > 0$。那么，对于任意 $\mathrm{g} : \mathcal{X} \to \mathbb{R}$ and $\Delta^c \in \mathrm{C}(\mathcal{X} \times \mathcal{Y})$，若 $\|\mathrm{g}\|_\infty + 2\|\mathrm{c} + \Delta^c\|_\infty \leq \mathrm{B}$，则有 $\mathrm{g}^{(c+\Delta^c)(\mathrm{c}+\Delta^c)\mathrm{cc}} \in \mathcal{H}_c + [-\mathrm{B}, \mathrm{B}]$。

上述引理的证明见附录 E.7。

命题 3.6 的证明。证明主要借鉴 van der Vaart 和 Wellner，2007，定理 2.3，并采用标准的经验过程论证。为简化记号，我们仅考虑 $n = m$ 的情形，并用 c_n 代替 $c_{n,n}$。注意，n, m 时的结论可通过类似论证得到。

为证明 (i)，首先根据三角不等式并利用引理 6.1，可得

$$\begin{aligned}\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{c_nc_n}-f^{cc})| &\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{c_nc_n}-f^{\tilde{c}_n\tilde{c}_n})| + \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})| \\ &\le 4\sqrt{n}\,\|c_n-\tilde{c}_n\|_{\infty} + \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})|. \end{aligned} \tag{B.1}$$

根据假设，第一项在 $n \to \infty$ 时依概率收敛于零。对于后一项，由 (JW) 以及关于 $\tilde{c}_n$ 的假设可知 $\sqrt{n/2}(\tilde{c}_n - c) \rightsquigarrow \mathbb{G}^c$。由 $\mathbb{G}^c$ 的分布的紧性可知，对任意 $\varepsilon > 0$，存在 $C(\mathcal{X}\times\mathcal{Y})$ 中的紧集 $K \subseteq$，使得 $P(\mathbb{G}^c \in K) > 1-\varepsilon$；因此，对任意 $\delta > 0$，$C(\mathcal{X}\times\mathcal{Y})$ 中与 K 的距离小于 $\delta > 0$ 的元素所构成的集合 K^{δ} 满足

$$\liminf_{n\to\infty} P\left(\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta}\right) \ge P(\mathbb{G}^c \in K^{\delta}) > 1-\varepsilon. \tag{B.2}$$

由 K 的紧致性，存在一个有限的 $\delta/2$-覆盖 $\{h_1, \ldots, h_p\}$，这意味着 $K^{\delta/2} \subseteq \bigcup_{i=1}^{p} B(h_i, \delta)$，其中 $B(h, \delta)$ 表示空间 $C(\mathcal{X}\times\mathcal{Y})$ 中以 h 为中心、半径为 δ 的开球。由此我们得到

$$\left\{\sqrt{n/2}(\tilde{c}_n - c) \in K^{\delta/2}\right\} \subset \bigcup_{i=1}^{p}\left\{\tilde{c}_n \in B(c + 2n^{-1/2}h_i, \delta)\right\}.$$

此外，由 Santambrogio（2015，命题 1.34）可知，对于任意 $f \in \mathcal{F}$ and $\bar{c} \in C(\mathcal{X}\times\mathcal{Y})$，均有 $f^{cc} = f^{cccc}$。因此，由三角不等式可得，

$$\begin{aligned}\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})| &= \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n\tilde{c}_n\tilde{c}_n}-f^{cccc})| \\ &\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n\tilde{c}_n\tilde{c}_n}-f^{\tilde{c}_n\tilde{c}_ncc})| + \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_ncc}-f^{cccc})| \\ &\le \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})| + \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_ncc}-f^{cccc})|. \end{aligned} \tag{B.3}$$

假设 $\sqrt{n/2}(\bar{c}_n - c) \in K^{\delta/2}$，则对于 (B.3) 中的第一项，

$$\begin{aligned}\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})| &\le \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}\max_{i=1,\ldots,p}\sup_{\|h-h_i\|_{\infty}<\delta}|\mathbb{G}_n^{\mu}(f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})}-f^{cc})| \\ &\le \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}\max_{i=1,\ldots,p}\sup_{\|h-h_i\|_{\infty}<\delta}|\mathbb{G}_n^{\mu}(f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})}-f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})})| \\ &+ \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}\max_{i=1,\ldots,p}|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}-f^{cc})| \\ &\le 8\delta + \sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}\max_{i=1,\ldots,p}|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}-f^{cc})|. \end{aligned} \tag{B.4}$$

这里，在最后一个不等式中，我们使用引理 6.1 推得

$$\left\|f^{(c+2h/\sqrt{n})(c+2h/\sqrt{n})} - f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}\right\|_{\infty} \le 4\,\|h_i - h\|_{\infty}/\sqrt{n} \le 4\delta/\sqrt{n}$$

结合 $\mathbb{G}_n^{\mu}(g) = \sqrt{n}(\mu_n - \mu)(g) \leq 2\sqrt{n}\,\|g\|_{\infty}$，对 $\mathcal{X}$ 上的任意可测函数 g 均成立。现在，对 $1 \leq i \leq p$，定义函数类

$$\tilde{\mathcal{G}}_n^i := \tilde{\mathcal{G}}_n^i(h_i) := \left\{ f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc} \middle| f \in \mathcal{F}^{\tilde{c}_n \tilde{c}_n} \right\}.$$

对于每个 $1 \leq i \leq p$ 以及任意 $\varepsilon > 0$，我们接下来观察到

$$\begin{aligned} \log N(\varepsilon, \tilde{\mathcal{G}}_n^i, \|\cdot\|_{\infty}) &\leq \log N(\varepsilon, \mathcal{F}^{\tilde{c}_n \tilde{c}_n (c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}, \|\cdot\|_{\infty}) + \log N(\varepsilon, \tilde{\mathcal{F}}^{\tilde{c}_n \tilde{c}_n cc}, \|\cdot\|_{\infty}) \\ &\leq 2 \log N(\varepsilon, \mathcal{F}^{\tilde{c}_n \tilde{c}_n}, \|\cdot\|_{\infty}), \end{aligned}$$

其中最后一步由 Hundrieser、Staudt和Munk，2022 中的引理2.1得出。因此，根据 Dudley 熵积分（例如，参见 Wainwright，2019，第5章），可得

$$\begin{aligned} \mathbb{E}\left[\sup_{f \in \mathcal{F}^{\tilde{c}_n \tilde{c}_n}} \max_{i=1,\ldots,p} \left| \mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc}) \right| \right] &\leq \sum_{i=1}^{p} \mathbb{E}\left[\sup_{g \in \tilde{\mathcal{G}}_n^i} \left| \mathbb{G}_n^{\mu}(g) \right| \right] \\ &\lesssim \sum_{i=1}^{p} \int_0^{4\|h_i\|_{\infty}/\sqrt{n}} \sqrt{\log\left(N(\varepsilon, \mathcal{F}^{\tilde{c}_n \tilde{c}_n}, \|\cdot\|_{\infty})\right)} d\varepsilon \\ &\lesssim \sum_{i=1}^{p} \int_0^{4\|h_i\|_{\infty}/\sqrt{n}} \varepsilon^{-\alpha/2} d\varepsilon \\ &\lesssim \sum_{i=1}^{p} (\|h_i\|_{\infty} / \sqrt{n})^{1-\alpha/2}, \end{aligned}$$

其中，根据假设，隐藏常数不依赖于 n。因此，在事件 $\sqrt{n}/2(\tilde{c}_n - c) \in K^{\delta/2}$ for $n \to \infty$ 上，我们可以条件性地推断出

$$\sup_{f \in \mathcal{F}^{\tilde{c}_n \tilde{c}_n}} \max_{i=1,\ldots,p} \mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})} - f^{cc}) \xrightarrow{P} 0. \tag{B.5}$$

对于 (B.3) 中的第二项，我们假设 $\sqrt{n}/2(\tilde{c}_n - c) \in K^{\delta/2}$，并通过类似的论证得到：

$$\sup_{f \in \mathcal{F}} |\mathbb{G}_n^{\mu}(f^{\tilde{c}_n \tilde{c}_n cc} - f^{cccc})| \leq 8\delta + \sup_{f \in \mathcal{F}} \max_{i=1,\ldots,p} |\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc})|. \tag{B.6}$$

定义函数类

$$\tilde{\mathcal{G}}_n := \tilde{\mathcal{G}}_n(h_1, \ldots, h_p) := \left\{ f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} - f^{cccc} \middle| f \in \mathcal{F} \right\} \tag{B.7}$$

我们再次注意到，根据引理 6.1，任意 $g \in \mathcal{G}_n$ 都满足 $\|g\|_{\infty} \leq \max_{i=1,\ldots,p} 4\|h_i\|_{\infty}/\sqrt{n}$。此外，当 n 充分大时，存在常数 $B > 0$，使得引理B.1适用于任意 $f \in \mathcal{F}$，从而得到

$$f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc} \in \mathcal{H}_c + [-B, B].$$

因此，当 n 足够大时，由引理 2.1 可知，对于任意 $\varepsilon > 0$，

$$N(\varepsilon, \tilde{\mathcal{G}}_n(h_1, \ldots, h_p), \|\cdot\|_{\infty}) \leq \left(N(\varepsilon, \mathcal{F}^{cc} + [-B, B], \|\cdot\|_{\infty})\right)^2.$$

再次利用 Dudley 熵积分，可断言对于这样的 n,

$$\begin{aligned}
\mathbb{E}\left[\sup_{f\in\mathcal{F}}\max_{i=1,\dots,p}\left|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc}-f^{cccc})\right|\right] &= \mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_n}\left|\mathbb{G}_n^{\mu}(\tilde{f})\right|\right] \\
&\lesssim \int_0^{\max_{i=1,\dots,p}4\|h_i\|_\infty/\sqrt{n}}\sqrt{\log\left(\mathcal{N}(\varepsilon,\tilde{\mathcal{G}}_n,\|\cdot\|_\infty)\right)}d\varepsilon \\
&\le \int_0^{\max_{i=1,\dots,p}4\|h_i\|_\infty/\sqrt{n}}\sqrt{\log\left(\mathcal{N}(\varepsilon,\mathcal{F}^{cc}+[-B,B],\|\cdot\|_\infty)\right)}d\varepsilon \\
&\lesssim \int_0^{\max_{i=1,\dots,p}4\|h_i\|_\infty/\sqrt{n}}\varepsilon^{-\alpha/2}d\varepsilon \\
&\lesssim \left(\max_{i=1,\dots,p}\|h_i\|_\infty/\sqrt{n}\right)^{1-\alpha/2}.
\end{aligned}$$

这意味着，在事件 $\sqrt{n}/2(\tilde{c}_n - c) \in K^{\delta/2}$ for $n \to \infty$ 条件下，

$$\sup_{f\in\mathcal{F}}\max_{i=1,\dots,p}|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc}-f^{cccc})| \xrightarrow{P} 0. \tag{B.8}$$

综上，对于任意 $\varepsilon > 0$，由 (B.2)–(B.8) 可得 $\delta := \varepsilon/32 > 0$

$$\begin{aligned}
&\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})|>\varepsilon\right) \\
\le\ &\limsup_{n\to\infty}\left(\mathbb{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})|>\varepsilon,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right)+\mathbb{P}\left(\sqrt{n/2}(\tilde{c}_n-c)\notin K^{\delta/2}\right)\right) \\
\le\ &\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})|>\varepsilon,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right)+\varepsilon \\
\le\ &\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n}-f^{cc})|>\varepsilon/2,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right) \\
&+\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}_n cc}-f^{cccc})|>\varepsilon/2,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right)+\varepsilon \\
\le\ &\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}^{\tilde{c}_n\tilde{c}_n}}\max_{i=1,\dots,p}|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})}-f^{cc})|>\varepsilon/4,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right) \\
&+\limsup_{n\to\infty}\mathbb{P}\left(\sup_{f\in\mathcal{F}}\max_{i=1,\dots,p}|\mathbb{G}_n^{\mu}(f^{(c+2h_i/\sqrt{n})(c+2h_i/\sqrt{n})cc}-f^{cccc})|>\varepsilon/4,\ \sqrt{n/2}(\tilde{c}_n-c)\in K^{\delta/2}\right)+\varepsilon=\varepsilon,
\end{aligned}$$

这表明 $\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\mu}(f^{\tilde{c}_n\tilde{c}^n}-f^{cc})|$ 依概率收敛于零。因此，我们得出 (B.1) 中两项均依概率收敛。类似的论证可得，当 $n \to \infty$ 时，$\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{\nu}(f^{c^n}-f^{c})| \xrightarrow{P} 0$ 收敛；此处我们应用 Hundrieser、Staudt和Munk，2022 的引理 2.1，得到

$$\begin{aligned}
\sup_{n\in\mathbb{N}}\log\mathcal{N}(\varepsilon,\mathcal{F}^{\tilde{c}_n}\cup\mathcal{F}^{c},\|\cdot\|_\infty) &\le \sup_{n\in\mathbb{N}}\left(\log\mathcal{N}(\varepsilon,\mathcal{F}^{\tilde{c}_n},\|\cdot\|_\infty)+\log\mathcal{N}(\varepsilon,\mathcal{F}^{c},\|\cdot\|_\infty)\right) \\
&= \sup_{n\in\mathbb{N}}\left(\log\mathcal{N}(\varepsilon,\mathcal{F}^{\tilde{c}_n\tilde{c}_n},\|\cdot\|_\infty)+\log\mathcal{N}(\varepsilon,\mathcal{F}^{cc},\|\cdot\|_\infty)\right) \\
&\le \sup_{n\in\mathbb{N}}2\log\mathcal{N}(\varepsilon,\mathcal{F}^{\tilde{c}_n\tilde{c}_n}\cup\mathcal{F}^{cc},\|\cdot\|_\infty)\lesssim\varepsilon^{-\alpha}\quad\text{for }\alpha<2,
\end{aligned}$$

这总体上验证了定理 2.2 的 (Sup)。

对于 (ii)，由 Bücher 和 Kojadinovic (2019)，并且由于当 $n \to \infty$ 时 $k = k(n) = o(n)$，可知

$$\sqrt{k}(\tilde{c}^b_{n,k} - c) = \sqrt{k}(\tilde{c}^b_{n,k} - c^b_{n,k}) + \sqrt{k}(c^b_{n,k} - c_n) + \sqrt{\frac{k}{n}}\sqrt{n}(c_n - c) \rightsquigarrow \mathbb{G}^c.$$

同样，对于 $n \to \infty$，可得 $\sqrt{k}(\mu^b_{n,k} - \mu) \rightsquigarrow \mathbb{G}^\mu$ 在 $l^\infty(\mathcal{F}^{cc})$ 中，且 $\sqrt{k}(\nu^b_{n,k} - \nu) \rightsquigarrow \mathbb{G}^\nu$ 在 $l^\infty(\mathcal{F}_c)$ 中。这意味着我们可以采用与 (i) 类似的证明策略。定义 $\mathbb{G}^\mu_{n,k} := \sqrt{k}(\mu^b_{n,k} - \mu)$ 和 $\mathbb{G}^\nu_{n,k} := \sqrt{k}(\nu^b_{n,k} - \nu)$。于是，由引理 6.1 可推得

$$\begin{aligned}\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{c^b_{n,k}c^b_{n,k}} - f^{cc})| &\le \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{c^b_{n,k}c^b_{n,k}} - f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}})| + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \\ &\le 4\sqrt{k}\left\|c^b_{n,k} - \tilde{c}^b_{n,k}\right\|_\infty + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})|, \end{aligned} \tag{B.9}$$

其中第一项当 $n \to \infty$ 时依概率收敛于零。由 Santambrogio (2015, 命题 1.34)，可得

$$\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \le \sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| + \sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}cc} - f^{cccc})|.$$

此外，采用与 (i) 类似的论证可得，当 n 充分大时，以至少 $1-\varepsilon$ 的概率有

$$\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}} - f^{cc})| \le 8\delta + \sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})| \tag{B.10}$$

以及

$$\sup_{f\in\mathcal{F}} |\mathbb{G}^\mu_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}cc} - f^{cccc})| \le 8\delta + \sup_{f\in\mathcal{F}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})cc} - f^{cccc})|. \tag{B.11}$$

接下来，我们验证 (B.10) 和 (B.11) 右侧的上确界趋于零（相对于 μ_n 无条件而言，但在该集合上以至少 $1-\varepsilon$ 的概率条件而言）。根据自助法经验过程 $\sqrt{k}(\mu^b_{n,k} - \mu_n)$ 和经验过程 $\sqrt{n}(\mu_n - \mu)$ 的 Dudley 熵积分，并结合我们之前的讨论，可知

$$\begin{aligned}&\mathbb{E}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} \max_{i=1,\dots,p} |\mathbb{G}^\mu_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})|\right] \\ &= \sum_{i=1}^p \mathbb{E}_{\mu_n}\left[\mathbb{E}_{\mu^b_{n,k}}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} \left|\sqrt{k}(\mu^b_{n,k} - \mu_n)(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})\right| \,\middle|\, \mu_n\right]\right] \\ &\quad + \sqrt{\frac{k}{n}}\mathbb{E}\left[\sup_{f\in\mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}} |\mathbb{G}^\mu_n(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})} - f^{cc})|\right] \\ &\lesssim \sum_{i=1}^p \mathbb{E}_{\mu_n}\int_0^{4\|h_i\|_\infty/\sqrt{k}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}, \|\cdot\|_\infty)\right)}d\varepsilon + \sqrt{\frac{k}{n}}\int_0^{4\|h_i\|_\infty/\sqrt{k}} \sqrt{\log\left(\mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}, \|\cdot\|_\infty)\right)}d\varepsilon \\ &\lesssim \sum_{i=1}^p \left(1 + \sqrt{\frac{k}{n}}\right)\int_0^{4\|h_i\|_\infty/\sqrt{k}} \varepsilon^{-\alpha/2}d\varepsilon \lesssim \sum_{i=1}^p \left(1 + \sqrt{\frac{k}{n}}\right)\left(\|h_i\|_\infty / \sqrt{k}\right)^{1-\alpha/2},\end{aligned}$$

由于隐藏常数与 $\tilde{}$n,k 无关，当 n $\to \infty$ 且 k = k(n) = o(n) 时，该式趋于零。回顾将 (B.7) 中 n 替换为 k 后的函数类 $\mathcal{G}_k$ 的定义，可得

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\max_{i=1,\dots,p}|\mathbb{G}^{\mu}_{n,k}(f^{(c+2h_i/\sqrt{k})(c+2h_i/\sqrt{k})cc}-f^{cccc})|\right]=\mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k}\left|\mathbb{G}^{\mu}_{n,k}(f)\right|\right].$$

因此，结合此前的分析与 Dudley 熵积分可得

$$\begin{aligned}\mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k}\left|\mathbb{G}^{\mu}_{n,k}(f)\right|\right]&\le\mathbb{E}_{\mu_n}\left[\mathbb{E}_{\mu^b_{n,k}}\left[\sup_{f\in\tilde{\mathcal{G}}_k}|\sqrt{k}(\mu^b_{n,k}-\mu_n)(f)|\,\middle|\,\mu_n\right]\right]+\sqrt{\frac{k}{n}}\mathbb{E}\left[\sup_{f\in\tilde{\mathcal{G}}_k}|\mathbb{G}^{\mu}_{n}(f)|\right]\\&\lesssim\left(1+\sqrt{\frac{k}{n}}\right)\int_0^{\max_{i=1,\dots,p}4\|h_i\|_\infty/\sqrt{k}}\sqrt{\log\left(\mathcal{N}(\varepsilon,\mathcal{F}^{cc}+[-B,B],\|\cdot\|_\infty)\right)}d\varepsilon\\&\lesssim\left(1+\sqrt{\frac{k}{n}}\right)\left(\max_{i=1,\dots,p}\|h_i\|_\infty/\sqrt{k}\right)^{1-\alpha/2},\end{aligned}$$

当 n、k(n) $\to \infty$ 且 k(n) = o(n) 时，该式趋于零（隐藏常数与 n、k 无关）。运用与 (i) 中相同的论证，我们得出

$$\sup_{f\in\mathcal{F}}|\mathbb{G}^{\mu}_{n,k}(f^{\tilde{c}^b_{n,k}\tilde{c}^b_{n,k}}-f^{cc})|\xrightarrow{\mathrm{P}}0.$$

最后，类似的论证可得 $\sup_{f\in\mathcal{F}}|\mathbb{G}^{\nu}_{n,k}(f_{\tilde{c}^b_{n,k}}-f_c)|\xrightarrow{\mathrm{P}}0$，从而证明 (ii)。 □

B.4 推论3.7的证明

定义随机变量

$$\tilde{c}_n:=\begin{cases}c & \text{if } n<N,\\ c_n & \text{if } n\ge N,\end{cases}\quad\text{and}\quad\tilde{c}^b_{n,k}:=\begin{cases}c & \text{if } n<N \text{ or } k<K,\\ c^b_{n,k} & \text{if } n\ge N \text{ and } k\ge K.\end{cases}$$

根据命题3.1，代价估计量 $\tilde{c}_n$ 和 $\tilde{c}^b_{n,k}$ 满足 (3.1) 和 (3.2) 中的熵界。自然数集 N 和 K 的紧性意味着 $\sqrt{n}\,\|\tilde{c}_n-c_n\|_\infty\xrightarrow{\mathrm{P}}0$ 且 $\sqrt{k}\|\tilde{c}^b_{n,k}-c^b_{n,k}\|_\infty\xrightarrow{\mathrm{P}}0$ 对 n、k $\to\infty$成立，由命题 3.6 可得所述结论。□

B.5 推论3.8的证明

由假设（JW）可知 $\sqrt{nm/(n+m)}(c_{n,m}-c)\rightsquigarrow\mathbb{G}^c$；而在（JW）下，根据 Bücher 和 Kojadinovic（2019）以及 k = o(n)，我们可无条件地推出 $\sqrt{k}(c^b_{n,k}-c)\rightsquigarrow\mathbb{G}^c$。下文给出（Sup）的论证；对于（Sup）*，只需将经验过程替换为自助法成本过程，即可采用类似的证明策略。

首先，不失一般性地假设总体成本函数满足 $\|c\|_\infty\le 1$。那么，对于命题3.1的三种情形，均可推出 $\log\mathcal{N}(\varepsilon,\mathcal{F}^{cc},\|\cdot\|_\infty)\lesssim\varepsilon^{-\alpha}$，其中 $\alpha<2$。

对于情形(i)，我们令 $\tilde{c}_{n,m}:=\Psi_{\mathrm{bdd}}(c_{n,m})$，其中 Ψ_{bdd} 在第 5.1 节中定义。由于 $\|\tilde{c}_{n,m}\|_\infty\le 2$且 $\mathcal{F}=\mathcal{F}(2\|c\|_\infty+1,2w)$ 一致有界于 6，因此我们得到 $\mathcal{F}^{\tilde{c}_{n,m}}$ 一致有界于 8。根据命题 5.1 和 5.2，命题 3.6(i) 的两个条件均得到满足，从而可断定（Sup）成立。

对于设定 (ii)，我们取 $\tilde{c}_{n,m}:=\Psi_{d_\chi}(c_{n,m})$，其中 Ψ_{d_χ} 由第 5.2 节中定义。于是，$\|\tilde{c}_{n,m}\|_\infty\le 2$，且 $\mathcal{F}^{\tilde{c}_{n,m}}$ 一致有界于 8。此外，由假设 (ii)’，结合命题 5.1 和 5.3，可得 $\sqrt{nm/(n+m)}\|\tilde{c}_{n,m}-c_{n,m}\|_\infty\xrightarrow{\mathrm{P}}0$，并且

$$\sup_{n\in\mathbb{N}}\log\mathcal{N}(\varepsilon,\mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}},\|\cdot\|_\infty)\lesssim\mathcal{N}(\varepsilon/8,\mathcal{X},\tilde{d}_\chi)|\log(\varepsilon)|\lesssim\varepsilon^{-\beta}|\log(\varepsilon)|\lesssim\varepsilon^{-2+(2-\beta)/2},$$

这里我们使用了关于 $\mathcal{X}$ 的覆盖数假设。根据命题 3.6(i)，即可得到 (Sup)。

对于情形 (iii)，定义 $c_i \in C(\mathcal{U}_i \times \mathcal{Y})$ 为 $c_i(u, y) := c(\zeta_i(u), y)$。我们考虑 $\tilde{c}_{n,m} := \Psi_{com}(c_{n,m})$，其中 Ψ_{com} 5 4 Ψ_i $\mathcal{U}_i \times$ 表示由第 . 节中的 mod（当 $\mathcal{Y} \to \mathcal{U}_i \times \mathcal{Y}$ $\Psi_i = \Psi^{\|\cdot\|^{\gamma_i}} \circ \Psi_{bdd}$ 5 2 $\gamma_i \in (0, 1]$, $\Psi_i = \Psi^{c_i\,\gamma_i} \circ \Psi_{bdd}$（且 .）） 以及第 5.3 节中的 Hol（当 $\gamma_i \in (1, 2]$, 时，我们将 $\mathcal{X}$ 替换为 $\mathcal{U}_i$）定义的正则性提升泛函 : C( ) C( ) 的组合（第 . 节）。于是，根据命题 5.3 、 5.5 和 5.6，

泛函 Ψ 满足命题 5.1 的假设，因此 $\sqrt{nm/(n+m)}\|\tilde{c}_{n,m} - c_{n,m}\|_\infty$ P $\to 0$。此外，由于对于任意 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$，存在一个仅依赖于函数 c_i 和空间 $\mathcal{U}_i$的确定性常数 $C \geq 0$，使得 $\|\Psi(\tilde{c})\|_\infty < C$，因此 $\mathcal{F}^{\Psi(\tilde{c})}$ 由 $C + 6$ 一致有界，从而

$$\sup_{n\in\mathbb{N}} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\tilde{c}_{n,m}}, \|\cdot\|_\infty) \lesssim \sum_{i=1}^{I} \sup_{\tilde{c}_i \in C(\mathcal{U}_i \times \mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{F}^{\tilde{c}_{n,m}\Psi_i(\tilde{c}_i)}, \|\cdot\|_\infty) \lesssim \max_{i=1,\dots,I} \varepsilon^{-d_i/\gamma_i},$$

其中，第一处不等式用到了命题 5.6；第二处则对 $0 < \gamma_i \leq 1$ 使用了命题 5.3 中关于 $\mathcal{N}(\varepsilon, \mathcal{U}_i, \|\cdot\|^{\gamma_i}) \lesssim \varepsilon^{-d_i/\gamma_i}$ 的界，并对 $1 < \gamma_i \leq 2$使用了命题 5.5。由命题 3.6(i) 即可得到所述结论。□

B.6 引理 3.10 的证明

对于 $\varepsilon > 0$，不妨设右端有限，否则结论显然成立。令 $k = \mathcal{N}(\varepsilon/4, \Theta, d_\Theta)$，并令 $\{\theta_1, \dots, \theta_k\}$ 为 Θ的一个极小 $\varepsilon/4$-覆盖。此外，对每个 $i = 1, \dots, k$，令$\{f_1^i, \dots, f_{k_i}^i\}$为 $\mathcal{F}^{c_{\theta_i} c_{\theta_i}}$的一个极小 $\varepsilon/2$-覆盖，即 $k_i = \mathcal{N}(\varepsilon/2, \mathcal{F}^{c_{\theta_i} c_{\theta_i}}, \|\cdot\|_\infty)$。一旦证明$\mathcal{F}_\mathcal{X}(\varepsilon) := \bigcup_{i=1}^k \{f_{i1}, \dots, f_{ik_i}\}$是 $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta c_\theta}$的一个 ε-覆盖，且 $\mathcal{F}_\mathcal{Y}(\varepsilon) := \bigcup_{i=1}^k \{(f_{i1})^{c_{\theta_i}}, \dots, (f_{ik_i})^{c_{\theta_i}}\}$是 $\bigcup_{\theta\in\Theta} \mathcal{F}^{c_\theta}$ 的一个 ε-覆盖，结论便成立，因为

$$|\mathcal{F}_\mathcal{Y}(\varepsilon)| \leq |\mathcal{F}_\mathcal{X}(\varepsilon)| = \sum_{i=1}^{k} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_{\theta_i} c_{\theta_i}}, \|\cdot\|_\infty\right) \leq \mathcal{N}\left(\frac{\varepsilon}{4}, \Theta, d_\Theta\right) \sup_{\theta\in\Theta} \mathcal{N}\left(\frac{\varepsilon}{2}, \mathcal{F}^{c_\theta c_\theta}, \|\cdot\|_\infty\right).$$

~ ~因此，令 $\theta \in \Theta$ 且 $f \in \mathcal{F}^{c_\theta c_\theta}$，并选取满足 $f = f^{c_\theta c_\theta}$ 的 $f \in \mathcal{F}$。选取满足 $d_\Theta(\theta, \theta_i) \leq \varepsilon/4$ 的 θ_i，并选择 $f_{il_i} \in \mathcal{F}_\mathcal{X}(\varepsilon)$，使得 $\left\| f_{il_i} - \tilde{f}^{c_{\theta_i} c_{\theta_i}} \right\|_\infty \leq \varepsilon/2$。现在，根据代价在 θ 上的利普希茨连续性和引理 6.1，我们推得 $\left\| \tilde{f}^{c_{\theta_i} c_{\theta_i}} - \tilde{f}^{c_\theta c_\theta} \right\|_\infty \leq 2 d_\Theta(\theta, \theta_i) \leq \varepsilon/2$，从而

$$\left\| f_{l_i}^i - f \right\|_\infty = \left\| f_{l_i}^i - \tilde{f}^{c_\theta c_\theta} \right\|_\infty \leq \left\| f_{l_i}^i - \tilde{f}^{c_{\theta_i} c_{\theta_i}} \right\|_\infty + \left\| \tilde{f}^{c_{\theta_i} c_{\theta_i}} - \tilde{f}^{c_\theta c_\theta} \right\|_\infty \leq \varepsilon,$$

这验证了 $\mathcal{F}_\mathcal{X}(\varepsilon)$ 是 $\cup_{\theta\in\Theta} \mathcal{F}^{c_\theta c^\theta}$ 的 ε-覆盖。

~ ~此外，对于任意满足 $f \in \mathcal{F}^{c^\theta}$ 的 f， $\in \mathcal{F}$都存在一个满足 $f = f^{c_\theta}$ 的 f。由 Santambrogio（2015，命题 1.34）可知 $\tilde{f}^{c^\theta} = \tilde{f}^{c_\theta c_\theta c^\theta}$。因此，按照上述方式选取 $f_{il_i} \in \mathcal{F}_\mathcal{X}(\varepsilon)$ 后，由引理 6.1 可得

$$\left\| (f_{l_i}^i)^{c_{\theta_i}} - \tilde{f}^{c_{\theta_i}} \right\|_\infty = \left\| (f_{l_i}^i)^{c_{\theta_i}} - \tilde{f}^{c_{\theta_i} c_{\theta_i} c_{\theta_i}} \right\|_\infty \leq \left\| f_{l_i}^i - \tilde{f}^{c_{\theta_i} c_{\theta_i}} \right\|_\infty \leq \varepsilon/2.$$

再次应用引理 6.1 可得 $\left\| \tilde{f}^{c^{\theta_i}} - \tilde{f}^{c^\theta} \right\|_\infty \leq d(\theta, \theta_i) \leq \varepsilon/4$。因此，我们得到

$$\left\| (f_{l_i}^i)^{c_{\theta_i}} - f \right\|_\infty = \left\| (f_{l_i}^i)^{c_{\theta_i}} - \tilde{f}^{c_\theta} \right\|_\infty \leq \left\| (f_{l_i}^i)^{c_{\theta_i}} - \tilde{f}^{c_{\theta_i}} \right\|_\infty + \left\| \tilde{f}^{c_{\theta_i}} - \tilde{f}^{c_\theta} \right\|_\infty \leq \frac{3\varepsilon}{4} \leq \varepsilon,$$

这证明了 $\mathcal{F}_\mathcal{Y}(\varepsilon)$ 是 ε 的 $\cup_{\theta\in\Theta} \mathcal{F}^{c_\theta}$ 覆盖，并完成了证明。 □

C 第 4 节应用的证明

C.1 引理 4.1 的证明

选取 U 为 $\{\tilde{g} \in C(\mathcal{X}, \mathbb{R}^d) : \|\tilde{g} - g_{\vartheta^o}^{-1}\|_\infty < 1\}$ 在 K_Θ 下的原像。由于连续性，它在 Θ 中是开集（相对开集）。因此，根据 $\mathcal{X}$ 的紧致性，集合 $\{g_\vartheta^{-1}\}_{\vartheta \in U}$ 在 $\mathcal{X}$ 上一致有界。利用柯西-施瓦茨不等式，并根据 $\mathcal{Y}$ 的紧致性，我们推知 $\{C_\Theta(\vartheta)(x, \cdot)\}_{\vartheta \in U}$ 在 $\mathcal{Y}$ 上也一致有界。此外，由于对 $y \in \mathrm{int}(\mathcal{Y})$，有 $\nabla_y C_\Theta(\vartheta)(x, y) = 2(g_\vartheta^{-1}(x) - y)$，集合 $\{\nabla_y C_\Theta(\vartheta)(x, \cdot)\}$ 在 $\mathcal{Y}$ 上有界，且对 $\vartheta \in U, x \in \mathcal{X}$ 一致成立。最后，注意 $\mathrm{Hess}_y C_\Theta(\vartheta)(x, y) = -2\mathrm{Id}$，且与 $\vartheta \in U, x \in \mathcal{X}$ 无关。综上可知，存在 $\Lambda \geq 0$，使得 $(2, \Lambda)$-赫尔德正则性成立。□

C.2 引理 4.2 的证明

为证明 C_Θ 在 ϑ^o 处的 Hadamard 可微性，注意到

$$
\begin{aligned}
&\left\| \frac{C_\Theta(\vartheta^o + t_n h_n) - C_\Theta(\vartheta^o)}{t_n} - D^H_{|\vartheta^o} C_\Theta(h) \right\|_\infty \\
&= \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left| \frac{1}{t_n} \left\langle g^{-1}_{\vartheta^o + t_n h_n}(x) - g^{-1}_{\vartheta^o}(x), g^{-1}_{\vartheta^o + t_n h_n}(x) + g^{-1}_{\vartheta^o}(x) - 2y \right\rangle - 2 \left\langle D^H_{\vartheta^o} K_\Theta(h)(x), g^{-1}_{\vartheta^o}(x) - y \right\rangle \right| \\
&\leq \sup_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \left(\left| 2 \left\langle \frac{1}{t_n} \left(g^{-1}_{\vartheta^o + t_n h_n}(x) - g^{-1}_{\vartheta^o}(x) \right) - D^H_{\vartheta^o} K_\Theta(h)(x), g^{-1}_{\vartheta^o}(x) - y \right\rangle \right| + \frac{1}{t_n} \left\| g^{-1}_{\vartheta^o + t_n h_n}(x) - g^{-1}_{\vartheta^o}(x) \right\|^2 \right).
\end{aligned}
$$

对于 $n \to \infty$，第一项由 K_Θ 的 Hadamard 可微性趋于零，而第二项由 (G) 趋于零。因此，C_Θ 在 ϑ^o 处 Hadamard 可微。第二个断言由 Hadamard 可微泛函的泛函δ方法（Römisch，2006年）得出。□

C.3 命题 4.3 的证明

首先注意，与第2节和第3节相比，$\mathcal{X}$ 和 $\mathcal{Y}$ 的作用互换了。由命题3.1(iii)可知，$\mathcal{F}^{C_\Theta(\vartheta^o)}$具有普适 Donsker 性质，因为 $d \leq 3$，且 $C_\Theta(\vartheta^o)(x, \cdot)$ 对某个 $\Lambda \geq 0$ 在关于 $x \in \mathcal{X}$ 的意义下是一致的 $(2, \Lambda)$-赫尔德函数（引理 4.1）。此外，注意到由 ϑ_n 的可测性以及 C_Θ 在 ϑ^o 附近的连续性可知，c_n 也是可测的。由联合弱收敛 (4.6)，利用 C_Θ 在 ϑ^o 处的 Hadamard 可微性（引理 4.2）以及泛函δ方法，我们推得

满足 (JW) 的单样本版本（回忆注 2.4(ii)）。进一步，由于当 n 趋于无穷时 $\vartheta_n \xrightarrow{P} \vartheta^o$，由推论 3.8 和引理 4.1 可知，(Sup) 的单样本版本也成立。结论随即由定理 2.2 得出。□

C.4 命题4.5的证明

注意，根据假设，$(\mathcal{T}, d_\mathcal{T})$ 和 c 满足定理 2.6 的要求。此外，注意到假设（Don）可由命题 3.9 得出，而假设（ KP ）则由关于 μ 和 ν 的支撑的假设推出（ Staudt 、 Hundrieser 和 Munk，2022，推论 2）。因此，结论由定理 2.6 得出。□

C.5 命题4.8的证明

选取 $\mathcal{X} \subseteq \mathbb{R}^d$ 作为一个紧集，使其包含 μ 和 ν 的支撑。注意，$\mathbb{S}^{d-1}$ 是一个紧致波兰空间，并考虑利普希茨映射 $c_{\mathbb{S}^{d-1}} : (\mathbb{S}^{d-1}, \|\cdot\|) \to C(\mathcal{X} \times \mathcal{X})$，$\theta \mapsto c_\theta|_{\mathcal{X} \times \mathcal{X}}$，其模数取决于 $\mathcal{X}$ 和 p。由 $\mathcal{X}$ 和 $\mathbb{S}^{d-1}$ 的紧致性，根据 Arzelà–Ascoli 定理可知，$\{c_\theta|_{\mathcal{X} \times \mathcal{X}}\}_{\theta \in \mathbb{S}^{d-1}}$ 一致有界且等度连续，且具有一致的模数。

因此，选择定理 2.5 中的函数类 $\mathcal{F}$ 后，只要验证假设（Don）成立，即可由定理 2.5 得出断言 (i)。为此，注意到 $\log N(\varepsilon, \mathbb{S}^{d-1}, \|\cdot\|) \lesssim |\log(\varepsilon)|$。此外，对 $\theta \in \mathbb{S}^{d-1}$ 定义伪度量 $\tilde{d}_{\theta, \mathcal{X}}(x, y) = |\theta^T x - \theta^T y|$，它定义在 $\mathcal{X}$ 上并满足 $\sup_{\theta \in \mathbb{S}^{d-1}} \mathcal{N}(\varepsilon, \mathcal{X}, \tilde{d}_{\theta, \mathcal{X}}) \lesssim \varepsilon^{-1}$；且对于任意 $x, x', y \in \mathcal{X}$,

$$\left|c_\theta(x, y) - c_\theta(x', y)\right| \leq p \operatorname{diam}(\mathfrak{p}_\theta(\mathcal{X}))^{p-1} |\theta^T x - \theta^T x'| \leq p \operatorname{diam}(\mathcal{X})^{p-1} \tilde{d}_{\theta, \mathcal{X}}(x, x').$$

由于利普希茨模数的上界不依赖于 θ，命题 3.9(ii) 适用，因此我们得出 $S_{\theta \in \mathbb{S}^{d-1}} \mathcal{F}^{c_\theta c^\theta}$ 和 $S_{\theta \in \mathbb{S}^{d-1}} \mathcal{F}^{c_\theta}$ 是普适 Donsker。对 $\mathbb{S}^{d-1}$ 上的积分算子应用连续映射定理（van der Vaart 和 Wellner，1996年，定理 1.11.1），即可得到断言 (ii)。最后，断言 (iii) 由定理 2.7 得出。□

C.6 命题4.9的证明

由于 $\mathcal{X} \times \mathcal{Y}$ 是紧致的，且 c 和 Δ^c 连续，因此存在一个共同的连续模 w，适用于 $\{c_t(\cdot, y)\}_{y \in \mathcal{Y},\ t \in [0,1]}$。因此，对于任意 $t \in [0, 1]$，我们有 $c, c_t \in C(\|c\|_\infty + \|\Delta^c\|_\infty + 1, w)$（$C(\cdot, \cdot)$ 的定义见引理 6.2）。由此，根据引理 6.2，我们得到以下不等式,

$$\begin{aligned}
&\frac{1}{t}\left(\inf_{\pi \in \Pi^\star_{c_t}(\mu_t, \nu_t)} \pi(t\Delta^c) + \sup_{f \in S_c(\mu, \nu)} t\Delta^\mu(f^{cc}) + t\Delta^\nu(f^c)\right) \\
\leq &\frac{1}{t}(OT(\mu_t, \nu_t, c_t) - OT(\mu, \nu, c)) \\
\leq &\frac{1}{t}\left(\inf_{\pi \in \Pi^\star_c(\mu, \nu)} \pi(t\Delta^c) + \sup_{f \in S_{c_t}(\mu_t, \nu_t)} t\Delta^\mu(f^{cc}) + t\Delta^\nu(f^c) + \sup_{f \in \mathcal{F}} t\Delta^\mu(f^{c_t c_t} - f^{cc}) + t\Delta^\nu(f^{c_t} - f^c)\right).
\end{aligned}$$

接下来，我们注意到，对于某个 $\tilde{\mu} \in \mathcal{P}(\mathcal{X})$，有 $\Delta^\mu = \tilde{\mu} - \mu$。根据相对于代价函数的成本变换下的利普希茨性（引理 6.1），可得

$$\sup_{f \in \mathcal{F}} \left|\Delta^\mu(f^{c_t c_t} - f^{cc})\right| = \sup_{f \in \mathcal{F}} \left|(\tilde{\mu} - \mu)(f^{c_t c_t} - f^{cc})\right| \leq 4 \|c_t - c\|_\infty = 4t \left\|\Delta^c\right\|_\infty \xrightarrow{t \to 0} 0.$$

同样可得，对于 $t \to 0$，有 $|\sup_{f \in \mathcal{F}} \Delta^\nu(f^{c_t} - f^c)| \to 0$。最后，由于当 $t \searrow 0$ 时，(μ_t, ν_t) 弱收敛到 (μ, ν)，由引理 6.3 可得

$$\begin{aligned}
&\liminf_{t \searrow 0} \inf_{\pi \in \Pi^\star_{c_t}(\mu_t, \nu_t)} \pi(\Delta^c) + \sup_{f \in S_c(\mu, \nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c) \\
\geq &\inf_{\pi \in \Pi^\star_c(\mu, \nu)} \pi(\Delta^c) + \sup_{f \in S_c(\mu, \nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c)
\end{aligned}$$

以及

$$\begin{aligned}
&\limsup_{t \searrow 0} \inf_{\pi \in \Pi^\star_c(\mu, \nu)} \pi(\Delta^c) + \sup_{f \in S_{c_t}(\mu_t, \nu_t)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c) \\
\leq &\inf_{\pi \in \Pi^\star_c(\mu, \nu)} \pi(\Delta^c) + \sup_{f \in S_c(\mu, \nu)} \Delta^\mu(f^{cc}) + \Delta^\nu(f^c),
\end{aligned}$$

由此得出该结论。 □

D 第 5 节的证明：正则性提升泛函

D.1 命题 5.1 的证明

根据泛函δ方法（Römisch，2006年）以及关于 Ψ 和 $\mathcal{L}$ 的假设，可得

$$a_n\begin{pmatrix}(f_n - f)\\(\Psi(f_n) - f)\end{pmatrix} \rightsquigarrow \begin{pmatrix}\mathcal{L}\\D_f^H\Psi(\mathcal{L})\end{pmatrix} \overset{d}{=} \begin{pmatrix}\mathcal{L}\\\mathcal{L}\end{pmatrix} \quad \text{for } n \to \infty.$$

连续映射定理（van der Vaart 和 Wellner，1996年，定理 1.11.1）结合随机元素 f_n 和 $\Psi(f_n)$ 的可测性（由于 f 附近的连续性 Ψ），由此断言

$$a_n(\Psi(f_n) - f_n) \xrightarrow{P} 0 \quad \text{for } n \to \infty. \qquad \square$$

D.2 命题 5.2 的证明

首先注意，对于任意 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$，作为连续函数的复合，并在 $\|\tilde{c}\|_\infty < 2$ 满足 $\Psi(\tilde{c}) = \tilde{c}$ 的条件下，有 $\Psi(\tilde{c}) \in C(\mathcal{X}\times\mathcal{Y})$，从而得到 $\Psi(c) = c$。特别地，这表明 $\Psi: C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$ 在 c 附近连续。为考察 c 处的 Hadamard 可微性，取一个正序列 $t_n \searrow 0$，并取一个收敛序列 $(h_n)_{n\in\mathbb{N}} \subseteq C(\mathcal{X}\times\mathcal{Y})$，其极限为 h。由于 h 有界且$\|c\|_\infty \le 1$，当 n 充分大时，有 $\|c + t_n h_n\|_\infty < 2$，因此 $\Psi(c + t_n h_n) = c + t_n h_n$。于是我们得到

$$\left\|\frac{\Psi(c + t_n h_n) - \Psi(c)}{t_n} - h\right\|_\infty = \|h_n - h\|_\infty \to 0.$$

最后，由于对任意 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$，都有 $\|g^{\Psi(\tilde{c})}\|_\infty \le B + 2$，其中 $B := \sup_{g\in\mathcal{G}}\|g\|_\infty$，因此对于有限空间 $\mathcal{X}$，我们得到

$$\sup_{\tilde{c}\in C(\mathcal{X}\times\mathcal{Y})} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_\infty) \le |\mathcal{X}|(\log(B+2) + |\log(\varepsilon)|) \lesssim |\log(\varepsilon)|. \qquad \square$$

D.3 命题5.3的证明

由条件 (5.2) 可知，对于满足 $d_{\mathcal{X}}(x, x') = 0$ 的 $x, x' \in \mathcal{X}$，有 $c(x, y) = c(x', y)$；而在满足$\tilde{d}_{\mathcal{X}}(x, x') > 0$ 的情况下，由 $w(\delta) > 0$ 可知，对于 $\delta > 0$，有

$$c(x, y) \le c(x', y) + w(\tilde{d}_{\mathcal{X}}(x, x')) < c(x', y) + 2w(\tilde{d}_{\mathcal{X}}(x, x')).$$

这表明对任意 $(x, y) \in \mathcal{X}\times\mathcal{Y}$，都有

$$S(c, (x, y)) := \arg\min_{x'\in\mathcal{X}} c(x', y) + 2w\left(\tilde{d}_{\mathcal{X}}(x, x')\right) = \{x'' \in \mathcal{X} \mid \tilde{d}(x, x'') = 0\},$$

并且总体上由 $\|c\|_\infty \le 1$可得 $\Psi(c) = c$。

对于第二个和第三个结论，由命题 5.2 可知，$\Psi_{bdd}: C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y})$ 在 c 附近连续，且在 c 处 Hadamard 可微，其导数为 $\mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。因此，只需验证 $\Psi^{w\circ d_{\mathcal{X}}}$mod 在 c 附近连续且 Hadamard 方向可微，满足 $D_{|c}^H\Psi|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$；为此我们借助引理 A.1 和定理 A.2。定义空间 $V := C(\mathcal{X}\times\mathcal{Y}), \mathcal{F} = \mathcal{X}, \Theta = \mathcal{X}\times\mathcal{Y}$ 以及泛函

$$E^{w\circ\tilde{d}_{\mathcal{X}}}: V\times\mathcal{F}\times\Theta = C(\mathcal{X}\times\mathcal{Y})\times\mathcal{X}\times(\tilde{\mathcal{X}}\times\mathcal{Y}) \mapsto \mathbb{R}, \quad (\tilde{c}, x', (x, y)) \mapsto -\tilde{c}(x', y) - 2w(\tilde{d}_{\mathcal{X}}(x, x')).$$

对于任意 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$，函数 $E^{w\circ d_{\mathcal{X}}}(\tilde{c},\cdot,\cdot): \mathcal{X}\times(\tilde{\mathcal{X}}\times\mathcal{Y}) \to \mathbb{R}$ 作为连续函数之和是连续的。此外，对于任意 $(x', (x, y)) \in \mathcal{X}\times(\tilde{\mathcal{X}}\times\mathcal{Y})$，注意函数 $E^{w\circ d_{\mathcal{X}}}(\cdot, x', (x, y)) : C(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}$ 在一致范数下是 1-Lipschitz 的，并且

$$\Delta_c E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}, x', (x, y)) := E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}+c, x', (x, y)) - E^{w\circ\tilde{d}_{\mathcal{X}}}(c, x', (x, y)) = -\tilde{c}(x', y)$$

关于 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$ 是线性的。因此，由引理 A.1 可得该泛函的连续性

$$\Psi_{\mathrm{mod}}^{w\circ\tilde{d}_{\mathcal{X}}} : C(\mathcal{X}\times\mathcal{Y}) \to C(\tilde{\mathcal{X}}\times\mathcal{Y}),$$
$$\tilde{c} \mapsto \left((x, y) \mapsto \inf_{x'\in\mathcal{X}} \tilde{c}(x', y) + 2w(\tilde{d}_{\mathcal{X}}(x, x')) = -\sup_{x'\in\mathcal{X}} E^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c}, x', (x, y))\right).$$

考虑闭子向量空间 $U := C(\mathcal{X}\times\mathcal{Y}) \subseteq C(\mathcal{X}\times\mathcal{Y})$，参见引理 F.2。接下来只需证明定理 A.2 的假设（DC）。为此，注意对于 $h \in C(\mathcal{X}\times\mathcal{Y})$，有

$$h(\overline{x}, y) + 2w(\tilde{d}_{\mathcal{X}}(\overline{x}, x')) = h(\overline{x}', y) + 2w(\tilde{d}_{\mathcal{X}}(\overline{x}', x')) \quad \text{for any } \overline{x}, \overline{x}' \in S(c, (x, y))$$

因为 $\tilde{d}_{\overline{\mathcal{X}}}(\overline{x}, x') = 0$。由引理 A.3 可知，这意味着（DC）成立。因此，定理 A.2 断言 $\Psi^{w\circ d_{\mathcal{X}}}\mathrm{mod}$ 在 c 处 Hadamard 方向可微，其导数为

$$D^H_{|c}\Psi_{\mathrm{mod}}^{w\circ\tilde{d}_{\mathcal{X}}} : C(\mathcal{X}\times\mathcal{Y}) \to C(\tilde{\mathcal{X}}\times\mathcal{Y}),$$
$$h \mapsto \left((x, y) \mapsto \inf_{x' :\, \tilde{d}_{\mathcal{X}}(x', x)=0} h(x', y) = -\sup_{x' :\, \tilde{d}_{\mathcal{X}}(x', x)=0} -\Delta_c E^{w\circ\tilde{d}_{\mathcal{X}}}(h, x', (x, y))\right).$$

因此，若 $h \in C(\tilde{\mathcal{X}}\times\mathcal{Y})$，则 $D^H_{|c}\Psi^{w\circ d_{\mathcal{X}}}\mathrm{mod}(h) = h$，从而得到 $D^H_{|c}\Psi^{w\circ d_{\mathcal{X}}}\mathrm{mod}\,|_{C(\tilde{\mathcal{X}}\times\mathcal{Y})} = \mathrm{Id}_{C(\tilde{\mathcal{X}}\times\mathcal{Y})}$。

对于最后一个结论，注意任意 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$ 对于 $(x, y) \in \mathcal{X}\times\mathcal{Y}$ 都满足

$$-\|\tilde{c}\|_\infty \le \Psi_{\mathrm{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x, y) \le \tilde{c}(x, y) \le \|\tilde{c}\|_\infty,$$

因此 $\|\Psi(\tilde{c})\|_\infty = \|\Psi^{w\circ d_{\mathcal{X}}}\mathrm{mod}\circ\Psi_{\mathrm{bdd}}(\tilde{c})\|_\infty \le 2$。此外，对于任意 x、$x' \in \mathcal{X}$、$y \in \mathcal{Y}$，我们有

$$\begin{aligned}\Psi_{\mathrm{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x, y) - \Psi_{\mathrm{mod}}^{w\circ\tilde{d}_{\mathcal{X}}}(\tilde{c})(x', y) &\le \inf_{x''\in\mathcal{X}} c(x'', y) + 2w(\tilde{d}_{\mathcal{X}}(x'', x)) - c(x'', y) - 2w(\tilde{d}_{\mathcal{X}}(x'', x')) \\ &\le 2w(\tilde{d}_{\mathcal{X}}(x, x')), \end{aligned} \tag{D.1}$$

这里我们使用了反三角不等式，因为 $w\circ d_{\mathcal{X}}$ 在 $\mathcal{X}$ 上定义了一个（伪）度量。因此，对于任意 $\tilde{c} \in C(\mathcal{X}\times\mathcal{Y})$ 以及满足 $B := \sup_{g\in\mathcal{G}}\|g\|_\infty < \infty$ 的有界函数类 $\mathcal{G}$（其中 $\|\Psi(\tilde{c})\|_\infty \le 2$ 且满足 (D.1)），我们得出 $\mathcal{G}^{\Psi(\tilde{c})}$ 中的元素均由 $B+2$ 界定；由于其是这类 2-利普希茨函数的下确界，这些元素在 $w\circ\tilde{d}_{\mathcal{X}}$ 下均为 2-利普希茨函数。因此，$\mathcal{G}^{\Psi(\tilde{c})} \subseteq \mathrm{BL}_{(B+2),\,2}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}})$；对于后一个函数类，Kolmogorov 和 Tikhomirov（1961，第 9 节）给出了一致度量熵界，断言对任意 $\varepsilon > 0$

$$\begin{aligned}\mathcal{N}(\varepsilon, \mathrm{BL}_{(B+2),2}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}}), \|\cdot\|_\infty) &= \mathcal{N}(\varepsilon/2, \mathrm{BL}_{(B+2)/2,1}(\mathcal{X}, w\circ\tilde{d}_{\mathcal{X}}), \|\cdot\|_\infty) \\ &\lesssim \mathcal{N}(\varepsilon/8, \mathcal{X}, w\circ\tilde{d}_{\mathcal{X}})|\log(\varepsilon)|. \end{aligned}$$

□

D.4 推论 5.4 的证明

由命题 5.3 可知，$\Psi := B\cdot\Psi^{w\circ d_{\mathcal{X}}/B}\mathrm{mod}\circ\Psi_{\mathrm{bdd}}(\cdot/B)$ 在 c 附近连续，且在 c 处 Hadamard 可微，其导数为 $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。因此，根据命题 5.1 可得，对于 $n \to \infty$，$a_n(c_n - c_n)$ $P \to 0$。此外，根据 Ψ_{bdd} 和 $\Psi^{w/B,\,d_{\mathcal{X}}}\mathrm{mod}$ 的定义可知，$\|c_n\|_\infty \le 2B$，且 c_n 满足将 w 替换为 $2w$ 后的 (5.2)。由引理 2.1 可立即得出该包含关系。□

D.5 命题 5.5 的证明

由于 c 是 $(\gamma, 1)$-赫尔德的，根据凸性，由 Hundrieser, Staudt, and Munk (2022) 的引理 A.4 可知，对于任意 x、$x' \in \mathcal{X}$ 和 $y \in \mathcal{Y}$，

$$c(x,y) = c(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + R_{x'}(x) \quad \text{with} \quad |R_{x'}(x)| \le \sqrt{d} \left\| x - x' \right\|^{\gamma},$$

因此，对于 x、x'，我们得到

$$c(x,y) < c(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + 2\sqrt{d} \left\| x - x' \right\|^{\gamma}.$$

这表明，对于任意 $(x, y) \in \mathcal{X} \times \mathcal{Y}$，

$$S(c,(x,y)) := \operatorname*{arg\,min}_{x' \in \mathcal{X}} c(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + 2\sqrt{d} \left\| x - x' \right\|^{\gamma} = \{x\}$$

并由 $\|c\|_\infty \le 1$ 得出 $\Psi(c) = c$。

为了证明关于连续性和 Hadamard 可微性的结论，只需验证 Ψ_{Hol} 在 c 附近连续，且在 c 处 Hadamard 可微，其导数为 $D^H_{|c}\Psi_{\mathrm{Hol}} = \mathrm{Id}_{C(\mathcal{X}\times\mathcal{Y})}$。为此，我们将借助引理 A.1 和定理 A.2。令 $V := C(\mathcal{X} \times \mathcal{Y})$，$\mathcal{F} := \mathcal{X}$ 和 $\Theta := \mathcal{X} \times \mathcal{Y}$，并定义泛函

$$\begin{aligned} E_{\mathrm{Hol}} \colon V \times \mathcal{F} \times \Theta = C(\mathcal{X}\times\mathcal{Y}) \times \mathcal{X} \times (\mathcal{X}\times\mathcal{Y}) &\to \mathbb{R}, \\ (\tilde{c}, x', (x,y)) &\mapsto -\left(\tilde{c}(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + 2\sqrt{d} \left\| x - x' \right\|^{\gamma} \right). \end{aligned}$$

对于任意 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$，泛函 $E_{\mathrm{Hol}}(\tilde{c}, \cdot, \cdot) \colon \mathcal{X} \times (\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ 由 $\nabla_x c(\cdot,\cdot)$ 的连续性可知是连续的；对于任意 $(x', (x, y)) \in \mathcal{X} \times (\mathcal{X} \times \mathcal{Y})$，泛函 $E_{\mathrm{Hol}}(\cdot, x', (x, y)) : C(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}$ 在一致范数下是 1-Lipschitz，而

$$\Delta_c E_{\mathrm{Hol}}(\tilde{c}, x', (x,y)) = E_{\mathrm{Hol}}(\tilde{c} + c, x', (x,y)) - E_{\mathrm{Hol}}(c, x', (x,y)) = -\tilde{c}(x',y)$$

关于 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$ 是线性的。最后，由于 $S(c, (x, y)) = \{x\}$ 是单元素集，条件（DC）由引理 A.3 得出。因此，根据引理 A.1 和定理 A.2，该泛函

$$\Psi_{\mathrm{Hol}} \colon C(\mathcal{X}\times\mathcal{Y}) \to C(\mathcal{X}\times\mathcal{Y}),\ \tilde{c} \mapsto \left((x,y) \mapsto -\sup_{x' \in \mathcal{X}} E_{\mathrm{Hol}}(\tilde{c}, x', (x,y)) \right)$$

在 c 附近连续，且在 c 处 Hadamard 可微，其导数为

$$D^H_{|c}\Psi_{\mathrm{Hol}} \colon C(\mathcal{X}\times\mathcal{Y}) \to C(\tilde{\mathcal{X}}\times\mathcal{Y}), \quad h \mapsto \Big((x,y) \mapsto h(x,y) = -\Delta_c E_{\mathrm{Hol}}(h, x', (x,y)) \Big).$$

关于一致度量熵界的论断，令 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$，并假设（应用 Ψ_{bdd} 后）$\|\tilde{c}\|_\infty \le 2$。定义函数集合 $(E_{x',\,y})_{x' \in \mathcal{X},\ y \in \mathcal{Y}}$，其中

$$\tilde{E}_{x',y} \colon \mathcal{X} \to \mathbb{R}, \quad x \mapsto \tilde{c}(x',y) + \langle \nabla_x c(x',y), x - x' \rangle + 2\sqrt{d} \left\| x - x' \right\|^{\gamma},$$

它在 $\mathcal{X}$ 上是 $(\gamma,2)$-赫尔德连续的。因此，根据 Hundrieser, Staudt, and Munk（2022，引理 A.5），存在另一个函数集合 $(E^{\sigma}_{x',\,y})_{x' \in \mathcal{X},\ y \in \mathcal{Y},\ \sigma \in (0,1]}$，其由定义在 $\mathcal{X}$ 上的光滑函数构成，使得

$$\sup_{\substack{x' \in \mathcal{X} \\ y \in \mathcal{Y}}} \left\| \tilde{E}_{x',y} - \tilde{E}^{\sigma}_{x',y} \right\|_\infty \le K\sigma^{\gamma} \quad \text{and} \quad \sup_{\substack{x' \in \mathcal{X} \\ y \in \mathcal{Y}}} \left\| \tilde{E}^{\sigma}_{x',y} \right\|_{C^2(\mathcal{X})} \le K\sigma^{\gamma-2}, \tag{D.2}$$

对于所有 $\sigma > 0$ 以及某个独立的 $K > 0$。这里，二次连续可微函数 $g \colon \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ 的 $C^2(\mathcal{X})$-范数定义为

$$\|g\|_{C^2(\mathcal{X})} := \max_{|\beta| \le 2} \left\| D^{\beta} g \right\|_\infty, \quad \text{where} \quad D^{\beta} g = \partial^{|\beta|} g / \partial x_1^{\beta_1} \cdots x_d^{\beta_d} \ \text{ for } \beta \in \mathbb{N}_0^d.$$

注意，若函数满足 $\|g\|_{C^2(\mathcal{X})} \le \Gamma$，对 $\Gamma > 0$ 成立，则它绝对有界，界为 Γ；它是 Γ-利普希茨的，并且是 $d\Gamma$-半凹的（形式化定义见 Albano，2002 或 Hundrieser、Staudt和 Munk，2022），因为其海森矩阵的特征值被 $d \cdot \Gamma$ 所界定。定义 $c(x, y) := \Psi(\tilde{c})(x, y) = \inf_{x' \in \mathcal{X}} \tilde{E}_{x',\, y}(x)$ 和 $c^{\sigma}(x, y) := \inf_{x' \in \mathcal{X}} \tilde{E}^{\sigma}_{x',\, y}(x)$ 后，由 (D.2) 可得 $\|c - c^{\sigma}\|_{\infty} \le K\sigma^{\gamma}$，且 c^{σ} 是阶数为 $\Gamma(\sigma) := dK\sigma^{\gamma-2}$ 的半凹函数。因此，沿用 Hundrieser、Staudt和Munk，2022 中引理 A.4 的证明思路，对于任意 $\varepsilon > 0$ 和 $\sigma(\varepsilon) := (\varepsilon/4K)^{1/\gamma}$，我们得到

$$\mathcal{N}(\varepsilon, \mathcal{G}^{\bar{c}}, \|\cdot\|_{\infty}) \le \mathcal{N}(\varepsilon/2, \mathcal{G}^{\bar{c}^{\sigma(\varepsilon)}}, \|\cdot\|_{\infty}) = \mathcal{N}\left(\frac{\varepsilon}{2\Gamma(\sigma(\varepsilon))}, \frac{\mathcal{G}^{\bar{c}^{\sigma(\varepsilon)}}}{2\Gamma(\sigma(\varepsilon))}, \|\cdot\|_{\infty}\right) \lesssim \left(\frac{\varepsilon}{2\Gamma(\sigma(\varepsilon))}\right)^{-d/2} \lesssim \varepsilon^{-d/\gamma}.$$

这里，在第二个不等式中我们用到了：$\mathcal{G}^{\bar{c}^{\sigma(\varepsilon)}}/2\Gamma(\sigma(\varepsilon))$ 属于定义在 $\mathcal{X}$ 上的函数集合，其中函数绝对有界于 $B \ge 0$，利普希茨模数为 $L \ge 0$，且为 1-半凹函数；B 取决于 $\mathcal{G}$，L 取决于 $\mathcal{X}$。此外，我们还结合了 Bronshtein（1976）以及 Guntuboyina 和 Sen（2013）关于凸函数的一致度量熵界。特别地，由于隐藏常数不依赖于 $\tilde{c}$，因此结论成立。□

D.6 命题5.6的证明

对于第一个结论，注意到对每个 $i \in \{1, \dots, I\}$，$\Psi_i(c_i) = c_i$，因此对 $x \in \zeta_i(\mathcal{U}_i)$，有 $\Psi_i(c_i)(\zeta_i^{-1}(x), y) = c(x, y)$。于是，由于 $\sum_{i=1}^{I} \eta_i(x) \equiv 1$，可得 $\Psi(c) = c$。

由泛函 $\Psi_i \colon C(\mathcal{U}_i \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y})$ 在 c_i 附近的连续性（对每个 $1 \le i \le I$）可得，Ψ 在 c 附近连续。

F关于 Ψ 的 Hadamard 可微性，对每个 $i \in \{1, \dots, I\}$，定义泛函

$$\Psi^1_{\mathrm{com},i} \colon C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y}), \qquad \tilde{c} \mapsto ((u, y) \mapsto \tilde{c}(\zeta_i(u), y)),$$
$$\Psi^2_{\mathrm{com},i} \colon C(\mathcal{U}_i \times \mathcal{Y}) \to C(\zeta_i(\mathcal{U}_i) \times \mathcal{Y}), \qquad \tilde{c} \mapsto \left((x, y) \mapsto \tilde{c}(\zeta_i^{-1}(x), y)\right),$$

由于 ζ_i^{-1} 和 ζ_i 均连续，这两个映射分别映射到相应的连续函数空间。此外，注意到对任意 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$，都有 $\Psi(\tilde{c}) = \sum_{i=1}^{I} \eta_i \cdot \Psi^2_{\mathrm{com},i} \circ \Psi_i \circ \Psi^1_{\mathrm{com},i}(\tilde{c})$。两个泛函 $\Psi^1_{\mathrm{com},i}$ 和 $\Psi^2_{\mathrm{com},i}$ 在 c_i 处均 Hadamard 可微，其导数为

$$D^H_{|c_i}\Psi^1_{\mathrm{com},i} \colon C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{U}_i \times \mathcal{Y}), \quad h \mapsto ((u, y) \mapsto h(\zeta_i(u), y)),$$
$$D^H_{|c_i}\Psi^2_{\mathrm{com},i} \colon C(\mathcal{U}_i \times \mathcal{Y}) \to C(\zeta_i(\mathcal{U}_i) \times \mathcal{Y}), \quad h \mapsto \left((x, y) \mapsto h(\zeta_i^{-1}(x), y)\right).$$

根据关于 Ψ_i 的假设及链式法则，我们可推知 Ψ 在 c 处 Hadamard 可微，其导数为

$$D^H_c\Psi \colon C(\mathcal{X} \times \mathcal{Y}) \to C(\mathcal{X} \times \mathcal{Y}),$$
$$h \mapsto \left((x, y) \mapsto \sum_{i=1}^{I} \eta_i(x) h(\zeta_i^{-1}(\zeta_i(x)), y) = \sum_{i=1}^{I} \eta_i(x) h(x, y) = h(x, y)\right)$$

并得出 $D^H_{|c}\Psi = \mathrm{Id}_{C(\mathcal{X} \times \mathcal{Y})}$。

最后，覆盖数的界是 Hundrieser、Staudt和Munk，2022 中的引理3.1和引理 A.1 的推论，因为它们断言，对于任意的 $\tilde{c} \in C(\mathcal{X} \times \mathcal{Y})$，

$$\begin{aligned}
\log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}, \|\cdot\|_{\infty}) &\le \sum_{i=1}^{I} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi(\tilde{c})}|_{\zeta_i(\mathcal{U}_i)}, \|\cdot\|_{\infty}) \\
&\le \sum_{i=1}^{I} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c})} \circ \zeta_i, \|\cdot\|_{\infty}) \\
&= \sum_{i=1}^{I} \log \mathcal{N}(\varepsilon, \mathcal{G}^{\Psi_i(\tilde{c}(\zeta_i(\cdot), \cdot))}, \|\cdot\|_{\infty}).
\end{aligned}$$

□

E 第 6 节证明：分布极限引理

E.1 引理 6.1 的证明

~ ~假设 $\|f - f\|_\infty + \|c - c\|_\infty < \infty$，否则该结论显然成立。对于 f 和 ~c，存在 for~ ~$y \in \mathcal{Y}$ and $\varepsilon > 0$ some $x' \in \mathcal{X}$，使得 $f^c(y) \geq \tilde{c}(x', y) - f(x') - \varepsilon$。因此，

$$\begin{aligned} f^c(y) - \tilde{f}^{\tilde{c}}(y) &= \left[\inf_{x\in\mathcal{X}} c(x,y) - f(x)\right] - \left[\inf_{x\in\mathcal{X}} \tilde{c}(x,y) - \tilde{f}(x)\right] \\ &\leq c(x',y) - f(x') - \tilde{c}(x',y) + \tilde{f}(x') + \varepsilon \\ &\leq \left\|f - \tilde{f}\right\|_\infty + \|c - \tilde{c}\|_\infty + \varepsilon. \end{aligned}$$

由于 $\varepsilon > 0$ 可以任意小，对于任意 $y \in \mathcal{Y}$，我们得到不等式

$$f^c(y) - \tilde{f}^{\tilde{c}}(y) \leq \left\|f - \tilde{f}\right\|_\infty + \|c - \tilde{c}\|_\infty .$$

对 f 和 c 重复上述论证，可得反向不等式，从而证明该结论。 □

E.2 引理 6.2 的证明

我们先从两种不同的方式对问题进行拆分，

$$\begin{aligned} OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\mu,\nu,c) &= (OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\tilde{\mu},\tilde{\nu},c)) + (OT(\tilde{\mu},\tilde{\nu},c) - OT(\mu,\nu,c)) \\ &= (OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\mu,\nu,\tilde{c})) + (OT(\mu,\nu,\tilde{c}) - OT(\mu,\nu,c)). \end{aligned}$$

由于 $c, \tilde{c} \in C(2\|c\|_\infty + 1, 2w)$，我们可以在 $\mathcal{F} = \mathcal{F}(2\|c\|_\infty + 1, 2w)$ 下，利用引理 2.1 中 OT 值的对偶表示。因此，对于上面展示式中的每个括号，都可以选择代入原始形式中的一个可行方案，或对偶形式 $\mathcal{F}$ 中的一个势函数，从而得到上界和下界。这样，我们得到

$$\begin{aligned} \inf_{\pi\in\Pi_{\tilde{c}}^\star(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c} - c) \leq OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\tilde{\mu},\tilde{\nu},c) &\leq \inf_{\pi\in\Pi_c^\star(\tilde{\mu},\tilde{\nu})} \pi(\tilde{c} - c), \\ \sup_{f\in S_c(\mu,\nu)} (\tilde{\mu} - \mu)f^{cc} + (\tilde{\nu} - \nu)f^c \leq OT(\tilde{\mu},\tilde{\nu},c) - OT(\mu,\nu,c) &\leq \sup_{f\in S_c(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu)f^{cc} + (\tilde{\nu} - \nu)f^c, \\ OT(\mu,\nu,\tilde{c}) - OT(\mu,\nu,c) &\leq \inf_{\pi\in\Pi_{\tilde{c}}^\star(\mu,\nu)} \pi(\tilde{c} - c), \\ OT(\tilde{\mu},\tilde{\nu},\tilde{c}) - OT(\mu,\nu,\tilde{c}) &\leq \sup_{f\in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu} - \mu)f^{\tilde{c}\tilde{c}} + (\tilde{\nu} - \nu)f^{\tilde{c}}. \end{aligned}$$

特别是，对于最后一个上界，我们还注意到

$$\sup_{f\in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu}-\mu)f^{\tilde{c}\tilde{c}} + (\tilde{\nu}-\nu)f^{\tilde{c}} \leq \sup_{f\in S_{\tilde{c}}(\tilde{\mu},\tilde{\nu})} (\tilde{\mu}-\mu)f^{cc} + (\tilde{\nu}-\nu)f^c + \sup_{f\in\mathcal{F}} (\tilde{\mu}-\mu)(f^{\tilde{c}\tilde{c}} - f^{cc}) + (\tilde{\nu}-\nu)(f^{\tilde{c}} - f^c),$$

综合起来，这给出了测度和代价变化时 OT 代价的下界和上界。

最后，对于固定测度 μ, ν，由霍尔德不等式可知，对任意 $\pi \in \Pi(\mu, \nu)$，$|\pi(\tilde{c} - c)| \leq \|\tilde{c} - c\|_\infty$；而对于固定代价函数 c，我们有

$$\begin{aligned} \sup_{f\in S_c(\tilde{\mu},\tilde{\nu})\cup S_c(\mu,\nu)} \left|(\tilde{\mu} - \mu)f^{cc} + (\tilde{\nu} - \nu)f^c\right| &\leq \sup_{f\in\mathcal{F}} \left|(\tilde{\mu} - \mu)f^{cc}\right| + \sup_{f\in\mathcal{F}} \left|(\tilde{\nu} - \nu)f^c\right| \\ &= \sup_{f\in\mathcal{F}^{cc}} \left|(\tilde{\mu} - \mu)f\right| + \sup_{f\in\mathcal{F}^c} \left|(\tilde{\nu} - \nu)f\right|. \end{aligned}$$ □

E.3 引理 6.3 的证明

T_1 的连续性是 Villani（2008，定理 5.20）的推论。事实上，任意一个收敛序列 $(\mu_n, \nu_n, \mathrm{c}_n)$，其极限为 $(\mu_\infty, \nu_\infty, \mathrm{c}_\infty)$，都存在一个最优传输计划序列 $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$，该序列沿某个子序列弱收敛，记作 $(\pi_{n_k})_{k\in\mathbb{N}}$，并收敛到一个最优传输计划 $\pi_\infty \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$。因此，

$$\begin{aligned}\limsup_{k\to\infty} |T_1(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - T_1(\mu_\infty, \nu_\infty, c_\infty)| &= \limsup_{k\to\infty} |OT(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - OT(\mu_\infty, \nu_\infty, c_\infty)| \\ &= \limsup_{k\to\infty} |\pi_{n_k}(c_{n_k}) - \pi_\infty(c_\infty)| \\ &\le \limsup_{k\to\infty} |(\pi_{n_k} - \pi_\infty)(c_\infty)| + \left\|c_{n_k} - c_\infty\right\|_\infty = 0.\end{aligned}$$

由于这对任意收敛的最优传输计划序列都成立，因此由引理 F.1 可知，T_1 的连续性得证。

为证明 T_2 的下半连续性，取一个序列 $(\mathrm{h}_{c,n})_{n\in\mathbb{N}}$，其极限为 $\mathrm{h}_{c,\infty}$，并考虑最优传输计划 $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$，使得

$$\inf_{\pi\in\Pi^\star_{c_n}(\mu_n,\nu_n)} \pi(h_{c,n}) \ge \pi_n(h_{c,n}) - 1/n.$$

于是，根据 Villani（2008，定理 5.20），存在一个收敛子序列 $(\pi_{n_k})_{k\in\mathbb{N}}$，其极限为 $\pi_\infty \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$，由此可得

$$\begin{aligned}\liminf_{k\to\infty} T_2(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{c,n_k}) &= \liminf_{k\to\infty} \inf_{\pi\in\Pi^\star_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} \pi(h_{c,n_k}) \\ &\ge \liminf_{k\to\infty} \pi_{n_k}(h_{c,n_k}) - 1/n_k \\ &\ge \liminf_{k\to\infty} \pi_{n_k}(h_{c,\infty}) - \left\|h_{c,\infty} - h_{c,n_k}\right\|_\infty - 1/n_k \\ &= \pi_\infty(h_{c,\infty}) \ge T_2(\mu_\infty, \nu_\infty, c_\infty, h_{c,\infty}).\end{aligned}$$

因此，由引理 F.1 可知，T_2 具有下半连续性。为推得 T_2, 的上半连续性，进而证明其在 $(\mu_\infty, \nu_\infty, \mathrm{c}_\infty, \mathrm{h}_{c,\infty})$ 处的连续性，需假设存在唯一的最优传输计划 $\pi^\star \in \Pi^\star_{c_\infty}(\mu_\infty, \nu_\infty)$。根据 Villani（2008，定理 5.20），注意到对任意最优传输计划序列 $\pi_n \in \Pi^\star_{c_n}(\mu_n, \nu_n)$，都存在一个弱收敛子序列 π_{n_k}，其极限为 $\pi^\star$，当 $\mathrm{k} \to \infty$ 时成立。因此，我们得出

$$\begin{aligned}\limsup_{k\to\infty} T_2(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{c,n_k}) &= \limsup_{k\to\infty} \inf_{\pi\in\Pi^\star_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} \pi(h_{c,n_k}) \\ &\le \limsup_{k\to\infty} \pi_{n_k}(h_{c,n_k}) \\ &\le \limsup_{k\to\infty} \pi_{n_k}(h_{c,\infty}) - \left\|h_{c,\infty} - h_{c,n_k}\right\|_\infty \\ &= \pi_\infty(h_{c,\infty}) = T_2(\mu_\infty, \nu_\infty, c_\infty, h_{c,\infty}).\end{aligned}$$

由引理 F.1 可知，T_2 是上半连续的。此外，对于固定的 $(\mu', \nu', \mathrm{c}')$，映射 $\tilde{\mathrm{T}}_2$ 关于 h_c 连续，因为对于任意 h_c，均有

$$|T_2(\mu, \nu, c, h_c) - T_2(\mu, \nu, c, \tilde{h}_c)| \le \left\|h_c - \tilde{h}_c\right\|_\infty.$$

为证明 T_3 的上半连续性，取一个序列 $(\mathrm{h}_{\mu,\ n}, \mathrm{h}_{\nu,\ n})_{n\in\mathbb{N}}$，其极限为 $(\mathrm{h}_{\mu,\ \infty}, \mathrm{h}_{\nu,\ \infty})$。此外，根据 C 的定义，由引理 2.1 可知，任意 $\mathrm{c}_n \in \mathrm{C}$ 均满足 $\mathcal{H}_{c_n} \subseteq \mathcal{F}_{c_n c^n} \subseteq \mathcal{F}$。取一个序列 $\mathrm{f}_n \in \mathrm{S}_{c_n}(\mu_n, \nu_n) \subseteq \mathcal{F}$，使得

$$T_3(\mu_n, \nu_n, c_n, h_{\mu,n}, h_{\nu,n}) \le h_\mu(f_n) + h_\nu(f_n) + 1/n.$$

由 $\mathcal{F}$的紧致性可知，存在一个一致收敛的子序列，记为 $(f_{n_k})_{k\in\mathbb{N}}$，其极限为 $f_\infty \in \mathcal{F}$。接下来，我们证明 $f_\infty \in S_{c_\infty}(\mu_\infty, \nu_\infty)$。为此，我们注意到

$$\begin{aligned}
OT(\mu_\infty, \nu_\infty, c_\infty) &\geq \mu_\infty(f_\infty^{c_\infty c_\infty}) + \nu_\infty(f_\infty^{c_\infty}) \\
&= \lim_{k\to\infty} \mu_{n_k}(f_\infty^{c_\infty c_\infty}) + \nu_{n_k}(f_\infty^{c_\infty}) \\
&\geq \lim_{k\to\infty} \mu_{n_k}(f_{n_k}^{c_{n_k} c_{n_k}}) + \nu_{n_k}(f_{n_k}^{c_{n_k}}) - \left\| f_\infty^{c_\infty c_\infty} - f_{n_k}^{c_{n_k} c_{n_k}} \right\|_\infty - \left\| f_\infty^{c_\infty} - f_{n_k}^{c_{n_k}} \right\|_\infty \\
&= \lim_{k\to\infty} OT(\mu_{n_k}, \nu_{n_k}, c_{n_k}) - \left\| f_\infty^{c_\infty c_\infty} - f_{n_k}^{c_{n_k} c_{n_k}} \right\|_\infty - \left\| f_\infty^{c_\infty} - f_{n_k}^{c_{n_k}} \right\|_\infty \\
&= OT(\mu_\infty, \nu_\infty, c_\infty),
\end{aligned}$$

其中，最后一个等号由 T_1 的连续性得出。因此，我们得到 $f_\infty \in S_{c_\infty}(\mu_\infty, \nu_\infty)$。

B由 h_μ 和 h_ν 在 $\mathcal{F}$ 上的连续性，并记 $C_u(\mathcal{F})$ 上的范数为 $\|\cdot\|_{\mathcal{F}}$，我们推得 tha t

$$\begin{aligned}
\limsup_{k\to\infty} T_3(\mu_{n_k}, \nu_{n_k}, c_{n_k}, h_{\mu,n_k}, h_{\nu,n_k}) &= \limsup_{k\to\infty} \sup_{f\in S_{c_{n_k}}(\mu_{n_k},\nu_{n_k})} h_{\mu,n_k}(f) + h_{\nu,n_k}(f) \\
&\leq \limsup_{k\to\infty} h_{\mu,n_k}(f_{n_k}) + h_{\nu,n_k}(f_{n_k}) + 1/n_k \\
&\leq \limsup_{k\to\infty} h_{\mu,\infty}(f_{n_k}) + h_{\nu,\infty}(f_{n_k}) + \left\| h_{\mu,\infty} - h_{\mu,n_k} \right\|_{\mathcal{F}} + \left\| h_{\nu,\infty} - h_{\nu,n_k} \right\|_{\mathcal{F}} + 1/n_k \\
&= h_{\mu,\infty}(f_\infty) + h_{\nu,\infty}(f_\infty) \leq T_3(\mu_\infty, \nu_\infty, c_\infty, h_{\mu,\infty}, h_{\nu,\infty})
\end{aligned}$$

因此，根据引理 F.1，T_3 的上半连续性得以成立。此外，对于固定的 (μ', ν', c')˜ ˜，映射 T_3 关于 (h_μ, h_ν) 连续，因为对于另一个 (h_μ, h_ν)，有

$$|T_3(\mu, \nu, c, h_\mu, h_\nu) - T_3(\mu, \nu, c, \tilde{h}_\mu, \tilde{h}_\nu)| \leq \left\| \tilde{h}_\mu - \tilde{h}_\mu \right\|_{\mathcal{F}^{cc}} + \left\| \tilde{h}_\nu - \tilde{h}_\nu \right\|_{\mathcal{F}^c}.$$

最后，对于 T_4，取 $(h_{1,\mu}, \tilde{h}_{1,\mu}, h_{1,\nu}, \tilde{h}_{1,\nu}), (h_{2,\mu}, \tilde{h}_{2,\mu}, h_{2,\nu}, \tilde{h}_{2,\nu}) \in C_u(\mathcal{F})^4$，并注意到

$$\begin{aligned}
&|T_4(h_{1,\mu}, \tilde{h}_{1,\mu}, h_{1,\nu}, \tilde{h}_{1,\nu}) - T_4(h_{2,\mu}, \tilde{h}_{2,\mu}, h_{2,\nu}, \tilde{h}_{2,\nu})| \\
&\leq \left\| h_{1,\mu} - h_{2,\mu} \right\|_{\mathcal{F}} + \left\| \tilde{h}_{1,\mu} - \tilde{h}_{2,\mu} \right\|_{\mathcal{F}} + \left\| h_{1,\nu} - h_{2,\nu} \right\|_{\mathcal{F}} + \left\| \tilde{h}_{1,\nu} - \tilde{h}_{2,\nu} \right\|_{\mathcal{F}},
\end{aligned}$$

这表明了连续性。 □

E.4 引理 6.4 的证明

对于 (i)，取 $f, g \in \mathcal{G}$，则 $|\mu(f) - \mu(g)| \leq \|f - g\|_\infty$，因此 $\mu: \mathcal{G} \to \mathbb{R}$ 在一致范数下定义了一个利普希茨映射，这表明 $\mu \in C_u(\mathcal{G})$。断言 (ii) 由 Giné 和 Nickl（2016，第 17 页）得出。最后，(iii) 由 (ii) 得出，因为对于任意 $g \in \mathcal{G}$，求值 $\mu_n(g) = n^{-1} \sum_{i=1}^n g(X_i)$ 和 $\mu_{n,k}^b(g) = k^{-1} \sum_{i=1}^k g(X_i^b)$ 均为博雷尔可测。□

E.5 引理 6.5 的证明

W首先证明，对于 n, m ∈ 自然数集，若 m/(n + m) ∈ (0, 1)，则假设（JW）蕴含

$$\begin{pmatrix} \sqrt{n}\big((\mu_n - \mu)(f^{cc})\big)_{f\in\mathcal{F}} \\ \sqrt{m}\big((\nu_m - \nu)(f^c)\big)_{f\in\mathcal{F}} \\ \sqrt{\frac{nm}{n+m}}(c_{n,m} - c) \end{pmatrix} = \begin{pmatrix} \big(\mathbb{G}_n^\mu(f^{cc})\big)_{f\in\mathcal{F}} \\ \big(\mathbb{G}_m^\nu(f^c)\big)_{f\in\mathcal{F}} \\ \mathbb{G}_{n,m}^c \end{pmatrix} \rightsquigarrow \begin{pmatrix} \big(\mathbb{G}^\mu(f^{cc})\big)_{f\in\mathcal{F}} \\ \big(\mathbb{G}^\nu(f^c)\big)_{f\in\mathcal{F}} \\ \mathbb{G}^c \end{pmatrix} \tag{E.1}$$

在波兰空间 $C_u(\mathcal{F}) \times C_u(\mathcal{F}) \times C(\mathcal{X} \times \mathcal{Y})$ 中。为此，考虑映射

$$\Psi\colon C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^{c}) \times C(\mathcal{X} \times \mathcal{Y}) \to C_u(\mathcal{F}) \times C_u(\mathcal{F}) \times C(\mathcal{X} \times \mathcal{Y}),$$
$$(\alpha, \beta, \gamma) \mapsto \left((\alpha(f^{cc}))_{f \in \mathcal{F}}, (\beta(f^{c}))_{f \in \mathcal{F}}, \gamma \right).$$

该映射定义良好（即其值域正确），因为对于任意 $(\alpha, \beta) \in C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^{c})$，存在˜连续模 w_α、w_β：$\mathbb{R}_+ \to \mathbb{R}_+$ 使得对于 f，$f \in \mathcal{F}$ 由引理 6.1 可得

$$|\alpha(f^{cc}) - \alpha(\tilde{f}^{cc})| \le w_\alpha(\left\| f^{cc} - \tilde{f}^{cc} \right\|_\infty) \le w_\alpha(\left\| f - \tilde{f} \right\|_\infty),$$
$$|\beta(f^{c}) - \beta(\tilde{f}^{c})| \le w_\beta(\left\| f^{c} - \tilde{f}^{c} \right\|_\infty) \le w_\beta(\left\| f - \tilde{f} \right\|_\infty),$$

这表明 $((\alpha(f^{cc}))_{f \in \mathcal{F}}, (\beta(f^{c}))_{f \in \mathcal{F}}, \gamma) \in C_u(\mathcal{F}) \times C_u(\mathcal{F}) \times C(\mathcal{X} \times \mathcal{Y})$。此外，对于任意 (α, β)、$(\tilde{\alpha}, \tilde{\beta}) \in C_u(\mathcal{F}^{cc}) \times C_u(\mathcal{F}^{c})$，有

$$\sup_{f \in \mathcal{F}} |\alpha(f^{cc}) - \tilde{\alpha}(f^{cc})| = \sup_{\tilde{f} \in \mathcal{F}^{cc}} |\alpha(\tilde{f}) - \tilde{\alpha}(\tilde{f})| \quad \text{and} \quad \sup_{f \in \mathcal{F}} |\beta(f^{c}) - \tilde{\beta}(f^{c})| = \sup_{\tilde{f} \in \mathcal{F}^{c}} |\beta(\tilde{f}) - \tilde{\beta}(\tilde{f})|,$$

因此该映射 Ψ 连续。由此，假设（JW）和连续映射定理（van der Vaart 和 Wellner，1996年，定理 1.11.1）推出弱收敛（E.1）。

此外，由 Varadarajan（1958），经验测度 (μ_n, ν_n) 在 $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 中几乎必然弱收敛至 (μ, ν)。注意，由于 $\mathcal{X}$ 和 $\mathcal{Y}$ 的紧致性，$\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$ 是一个可分完备度量空间（Bolley，2008）。结合（E.1）应用斯卢茨基引理（van der Vaart 和 Wellner，1996年，例 1.4.7），即可得到第一个结论。特别地，由于 $c_{n,m}$ 的可测性以及引理 6.4，所有涉及的量都是博雷尔可测的。

对于第二个断言，根据引理 6.1 可知，μ_n, ν_m 和 $c_{n,m}$ 的任意一个实现都会使过程 $\mathbb{G}_n^\mu(f_{c_{n,m}, c_{n,m}})$ 和 $\mathbb{G}_m^\nu(f_{c_{n,m}})$ 分别关于 f 满足 $2\sqrt{n}$-利普希茨和 $2\sqrt{m}$-利普希茨条件。因此，它们关于 f 一致连续。此外，对于固定的 $f \in \mathcal{F}$，我们可以证明函数

$$\tilde{\mathbb{G}}_n^\mu\colon \mathcal{P}(\mathcal{X}) \times C(\mathcal{X} \times \mathcal{Y}) \to \mathbb{R}, \quad (\tilde{\mu}, \tilde{c}) \mapsto \sqrt{n}(\tilde{\mu} - \mu)(f^{\tilde{c}\tilde{c}})$$

是上半连续的（即特别是可测的）。事实上，对于 $\tilde{\mu}_k \rightsquigarrow \tilde{\mu}$ in $\mathcal{P}(\mathcal{X})$ 和 $\tilde{c}_k \to \tilde{c}$ in $C(\mathcal{X} \times \mathcal{Y})$，由引理 6.1、$f^{cc}$ 的上半连续性以及 Portmanteau 定理（van der Vaart 和 Wellner，1996年，定理 1.3.4）可得

$$\limsup_{k \to \infty} \sqrt{n}(\tilde{\mu}_k - \mu)(f^{\tilde{c}_k \tilde{c}_k}) \le \limsup_{k \to \infty} \sqrt{n}(\tilde{\mu}_k - \mu)(f^{\tilde{c}\tilde{c}}) + 2\sqrt{n} \left\| f^{\tilde{c}_k \tilde{c}_k} - f^{\tilde{c}\tilde{c}} \right\|_\infty \le \sqrt{n}(\tilde{\mu} - \mu)(f^{\tilde{c}\tilde{c}}).$$

因此，由引理 6.4(ii) 可知 $(\mathbb{G}_n^\mu(f_{c_{n,m}, c_{n,m}}))_{f \in \mathcal{F}}$ 是博雷尔可测的。同样，我们可以得出 $(\mathbb{G}_m^\nu(f_{c_{n,m}}))_{f \in \mathcal{F}}$ 是博雷尔可测的。

因此，根据 (Sup)，我们推得，对于 $n, m \to \infty$，

$$\left(\mathbb{G}_n^\mu(f^{cc}) - \mathbb{G}_n^\mu(f^{c_{n,m} c_{n,m}}), \mathbb{G}_m^\nu(f^{c}) - \mathbb{G}_m^\nu(f^{c_n}) \right)_{f \in \mathcal{F}} \xrightarrow{P} (0, 0) \quad \text{in } C_u(\mathcal{F})^2.$$

结合斯卢茨基引理和连续映射定理，便可得出该结论（van der Vaart 和 Wellner，1996年，例 1.4.7，定理 1.11.1）。□

E.6 引理 6.8 的证明

第一项结论源自 Römisch（2006）中的一个观察，因为概率测度$\mathcal{P}(\mathcal{X})$的集合是凸集。更多见解请参阅 Aubin 和 Frankowska（1990，命题 4.2.1）

对于第二项结论，考虑一个序列 $\Delta_n = (\tilde{\mu}_n - \mu)/t_n$，其中 $t_n > 0$，且 $\tilde{\mu}_n \in \mathcal{P}(\mathcal{X})$，满足$\|\Delta_n - \Delta\|_{\tilde{\mathcal{F}}} = \sup_{f\in\tilde{\mathcal{F}}} |\Delta_n(f) - \Delta(f)| \to 0$。那么，根据三角不等式可得

$$\begin{aligned}\left|\Delta(f) - \Delta(f')\right| &= \left|\Delta_n(f) - \Delta_n(f') + (\Delta - \Delta_n)(f) + (\Delta - \Delta_n)(f')\right| \\ &\le \left|\Delta_n(f - f')\right| + 2\|\Delta - \Delta_n\|_{\tilde{\mathcal{F}}}\end{aligned}$$

在此，第一项趋于零，因为 $\Delta_n(f - f') = (\tilde{\mu}_n - \mu)(\kappa)/t_n = 0$，而第二项在 $n \to \infty$ 趋于零时收敛。因此，$\Delta(f) = \Delta(f')$。

第三个结论依赖于 Portemanteau定理（van der Vaart 和 Wellner，1996年，引理 1.3.4），该定理借助外概率 P^* 的概念断言：

$$\mathbb{P}\left(\mathbb{G}^{\mu} \in T_{\mu}\mathcal{P}(\mathcal{X})\right) \ge \limsup_{n\to\infty} P^*\left(\sqrt{n}(\mu_n - \mu) \in T_{\mu}\mathcal{P}(\mathcal{X})\right) = 1. \qquad \square$$

E.7 引理B.1的证明

我们先证明 (i)。注意，对于 $\kappa \in \mathbb{R}$，有

$$(f + \kappa)^c(y) = \inf_{x\in\mathcal{X}} c(x, y) - f(x) - \kappa = f^c(y) - \kappa,$$

由此可得所述结论。为证明断言 (ii)，根据引理 6.1，注意到

$$\left\|g^{(c+\Delta^c)(c+\Delta^c)}\right\|_{\infty} \le \|g\|_{\infty} + 2\left\|c + \Delta^c\right\|_{\infty} \le B. \tag{E.2}$$

此外，我们发现

$$-\|c\|_{\infty} - \sup_{x\in\mathcal{X}} g^{(c+\Delta^c)(c+\Delta^c)}(x) \le g^{(c+\Delta^c)(c+\Delta^c)c}(y) \le \|c\|_{\infty} - \sup_{x\in\mathcal{X}} g^{(c+\Delta^c)(c+\Delta^c)}(x). \tag{E.3}$$

利用该引理的 (i) 部分，我们得到

$$g^{(c+\Delta^c)(c+\Delta^c)cc} = \left(\left(g^{(c+\Delta^c)(c+\Delta^c)}\right)^c + \sup_{x\in\mathcal{X}} g^{(c+\Delta^c)(c+\Delta^c)}(x)\right)^c + \sup_{x\in\mathcal{X}} g^{(c+\Delta^c)(c+\Delta^c)}(x).$$

结合 (E.2)、(E.3) 与上述等式可得 $g^{(c+\Delta^c)(c+\Delta^c)cc} \in \mathcal{H}_c + [-B, B]$，从而得证。□

F 初等分析结果

引理 F.1. 设 $(a_n)_{n\in\mathbb{N}}$ 是一个实值序列，且 $K \in \mathbb{R}$。

(i) 若对于任意子序列 $(a_{n_k})_{k\in\mathbb{N}}$，都存在一个子序列 $(a_{n_{k_l}})_{l\in\mathbb{N}}$，使得 $\limsup_{l\to\infty} a_{n_{k_l}} \le K$，则可得 $\limsup_{n\to\infty} a_n \le K$。(ii) 若对于任意子序列 $(a_{n_k})_{k\in\mathbb{N}}$，都存在一个子序列 $(a_{n_{k_l}})_{l\in\mathbb{N}}$，使得 $\liminf_{l\to\infty} a_{n_{k_l}} \ge K$，则可得 $\liminf_{n\to\infty} a_n \ge K$。(iii) 若对于任意子序列 $(a_{n_k})_{k\in\mathbb{N}}$，都存在一个子序列 $(a_{n_{k_l}})_{l\in\mathbb{N}}$，使得 $\lim_{l\to\infty} a_{n_{k_l}} = K$，则可得 $\lim_{n\to\infty} a_n = K$。

证明。我们只证明 (i)，并指出 (ii) 和 (iii) 可用类似方法证明。假设对于某个 $\varepsilon > 0$，有 $\limsup_{n\to\infty} a_n = \inf_{n\in\mathbb{N}}(\sup_{m\ge n} a_m) \ge K + \varepsilon$。由于 $(\sup_{m\ge n} a_m)_{n\in\mathbb{N}}$ 随 n 单调递减，因此这意味着对所有 $n \in \mathbb{N}$，都有 $\sup_{m\ge n} a_m \ge K + \varepsilon$。于是，存在 $(a_n)_{n\in\mathbb{N}}$ 的一个子序列，记作 $(a_{n_l})_{l\in\mathbb{N}}$，使得对所有 $l \in \mathbb{N}$，都有 $a_{n_l} \ge K + \varepsilon/2$。然而，这将意味着 $\liminf_{l\to\infty} a_{n_l} \ge K + \varepsilon/2 > K$，与假设矛盾。因此，$\limsup_{n\to\infty} a_n \le K$。□

引理 F.2. 设 $(\mathcal{X}, \mathrm{d}_{\mathcal{X}})$ 是一个紧致度量空间，并考虑 $\mathcal{X}$ 上的连续（伪）度量$\tilde{}$ $\tilde{}\mathrm{d}_{\mathcal{X}}$。那么，$(\mathcal{X}, \mathrm{d}_{\mathcal{X}})$ 是一个紧致（伪）度量空间。此外，给定一个波兰空间 $\mathcal{Y}$ it$\tilde{}$，可得 $\mathrm{C}((\mathcal{X}, \mathrm{d}_{\mathcal{X}}) \times \mathcal{Y}) \subseteq \mathrm{C}((\mathcal{X}, \mathrm{d}_{\mathcal{X}}) \times \mathcal{Y})$。

$\tilde{}$ $\tilde{}$证明。对于 $(\mathcal{X}, \mathrm{d}_{\mathcal{X}})$，(伪)度量的性质显然成立。由 $\mathrm{d}_{\mathcal{X}}$ 关于 $\mathrm{d}_{\mathcal{X}}\tilde{}$ 的连续性，典范包含映射 $\iota\colon (\mathcal{X}, \mathrm{d}_{\mathcal{X}}) \to (\mathcal{X}, \mathrm{d}_{\mathcal{X}}), \mathrm{x} \mapsto \mathrm{x}$ 是连续的。紧致空间在连续映射下的像仍然紧致，因此第一个结论成立。对于第二个结论，取 $\mathrm{h} \in \mathrm{C}((\mathcal{X}, \tilde{\mathrm{d}}_{\mathcal{X}}) \times \mathcal{Y})$。于是，复合映射 $\mathcal{X} \times \mathcal{Y} \to \mathbb{R}, (\mathrm{x}, \mathrm{y}) \mapsto \mathrm{h}(\iota(\mathrm{x}), \mathrm{y})$ 是连续的，因此 h 的典范嵌入 $\mathrm{h} \circ (\iota, \mathrm{Id}_{\mathcal{Y}})$ 属于 $\mathrm{C}((\mathcal{X}, \mathrm{d}_{\mathcal{X}}) \times \mathcal{Y})$。□