Wasserstein 距离的渐近性

沈梓梁　2021213218

2022/6/15

1　引言

在极限理论课程的学习之中，我们讨论当样本量 n 趋向于无穷时，统计量所表现的渐近性质. 概率论中的各种收敛形式: 依概率收敛，几乎必然收敛，依分布收敛，都是帮助刻画这种渐近性质的重要工具. 在课程中我们花费了非常多的篇幅去讨论它们，其中依分布收敛是我们非常关心的一种收敛形式. 无论是借助中心极限定理 (一般的形式或鞅形式的)，还是利用其它手段，最终的目的是研究渐近分布的存在性以及导出具体形式. 渐近分布这对我们应用上的假设检验 (设计检验统计量，计算 P 值)，或者评价估计量的渐近方差都是至关重要的.

对于传统的中心极限定理，此时极限分布是正态分布. 但很多情况下，极限分布并不是正态的分布，例如极值分布，极限分布可能是指数分布或者其他. 在点过程 (The Point Process) 研究中，极限点过程可能是泊松点过程.

回顾一般的依分布收敛的定义 (在 $\mathbb{R}^1$ 上)

$$F_n(x) \Rightarrow F_\infty(x)$$

假设分布函数存在，分布的极限如果拓展到高维的情况 ($\mathbb{R}^k$ 上)，我们会发现分布函数不一定能用简单的形式表示出. 例如对于 k 维正态分布，如果协方差矩阵不可逆，此时密度函数不存在，分布函数很难用简单形式表示出来 (甚至不能够用)，但我们可以借助特征函数来帮助我们描述这种依分布收敛.

连续性定理 (Continuity Theorem)([1] Resnick[1,Theorem 9.5.2]):

(a) $\lim_{n\to\infty} \phi_n(t)$ 存在对任意的 t. 我们将这个极限记作 $\phi_\infty(t)$.

(b) $\phi_\infty(t)$ 在 0 处连续. 于是存在分布函数 F_∞ 我们有:

$$F_n \Rightarrow F_\infty$$

当我们更加进一步，讨论更加一般的随机元 (例如随机函数等)，我们需要更加广泛的弱收敛的定义，此时需要如下的 Portmanteau 定理.

Portmanteau Theorem([2]Billingsley [2, Theorem 2.1]): $\mathscr{X}$ 是一个完备可分空间，测度 $\mu, \mu_n \in P(\mathscr{X})$. 下列各项等价:

- $\mu_n \xrightarrow{\text{w}} \mu$ 弱收敛;
- $F_n(x) \to F(x)$ 对所有 F 的连续点 x. 如果 $\mathscr{X}=\mathbb{R}^d, F_n$ 是测度 μ_n 的分布函数，F 是 μ 的分布函数;
- 对任意的开集 $G \subseteq \mathscr{X}, \liminf \mu_n(G) \geq \mu(G)$;
- 对任意的闭集 $F \subseteq \mathscr{X}$, $\limsup \mu_n(F) \leq \mu(F)$;
- $\int h \, \mathrm{d}\mu_n \to \int h \, \mathrm{d}\mu \quad \forall h \in C_b(\mathscr{X})$.

需要注意的是, 我们在讨论距离的弱与强之时, 实际上是讨论它们所形成的拓扑的强弱, 如果在一种距离 d_1 下的收敛蕴含着另一种距离 d_2 下的收敛，我们称距离 d_1 形成的拓扑比距离 d_2 形成的拓扑要强，也就是说 d_1 比 d_2 强，d_2 比 d_1 要弱.

我们关注弱收敛，一个重要原因在于弱拓扑要求更少的收敛条件，因此具有更广泛的适用范围, 使得许多随机对象能够建立稳定的极限理论。与此同时, 弱收敛往往需要通过适当的距离进行刻画, 而 Wasserstein 距离由于兼具最优传输结构与概率测度几何性质, 在统计推断与机器学习中得到了广泛应用. 实际上，Portmanteau 定理的最后一条命题就可以是视作为对弱收敛的度量，我们通过一个具体的实数列的收敛来刻画随机元的弱收敛. 如下这些距离 ([3]Villani) 都是在对弱收敛进行度量:

1.Lévy-Prokhorov 距离 (或称为 Prokhorov 距离):

$$d_P(\mu,\nu)=\inf\{\varepsilon>0; \exists X,Y; \inf \mathbb{P}\lfloor d(X,Y)>\varepsilon\rfloor \leq \varepsilon\};$$

2.The Bounded Lipschitz 距离 (或称为 Fortet-Mourier 距离):

$$d_{bL}(\mu,\nu)=\sup\left\{\int \varphi d\mu-\int \varphi d\nu; \quad \|\varphi\|_{\infty}+\|\varphi\|_{\mathrm{Lip}} \leq 1\right\}$$

3.weak-* 距离 (在局部紧的度量空间):

$$d_{w*}(\mu,\nu)=\sum_{k\in\mathbb{N}} 2^{-k}\left|\int \varphi_k d\mu-\int \varphi_k d\nu\right|$$

$(\varphi_k)_{k\in\mathbb{N}}$ 是 $C_0(\mathscr{X})$ 中的稠密序列，值得提及的是，weak-* 距离是最弱的一种距离 (在我们讨论的这几种距离之中);

4.Toscani 距离 (在 $P_2\left(\mathbb{R}^n\right)$) :

$$d_T(\mu,\nu)=\sup_{\xi\in\mathbb{R}^n\setminus\{0\}}\left(\frac{\left|\int e^{-ix\cdot\xi} d\mu(x)-\int e^{-ix\cdot\xi} d\nu(x)\right|}{|\xi|^2}\right).$$

(此处我们含蓄的假设了 μ,ν 具有相同的均值, 否则 $d_T(\mu,\nu)$ 可能会无穷; 我们也可以引入 d_T 通过改变分母的指数 2.)

5.Wasserstein 距离:

Polish 度量空间 $(\mathscr{X},d)$ 上，满足 $\mu,\nu \in \mathscr{P}_p(\mathscr{X})$ 的两个概率测度之间的 p 阶 Wasserstein 距离 (WD) 定义如下

$$W_p(\mu,\nu) = \left(\inf_{\pi} \int_{\mathscr{X}\times\mathscr{X}} d(x,y)^p \mathrm{d}\pi(x,y)\right)^{1/p}$$

其中 $p \in [1,\infty)$, 下界 inf 取尽所有在乘积空间 $\mathscr{X} \times \mathscr{X}$ 边际为 μ 和 ν 的联合测度 π .

上述的距离定义都在某些条件下能够和弱收敛等价，本质上它们都是对弱收敛性的度量，其中 Wasserstein 距离使用得非常广泛，它方便地运用最优传输理论 (optimal transport)，而且有非常多的性质.

2 W_p 的渐近分布

现在我们已经是在一个 Polish 空间上讨论问题，依分布收敛已经拓展为一般的弱收敛，上面以及提到，我们可以用一些距离去度量弱收敛性 (在课堂中也介绍了 L^p 空间和 Skorohod 空间上的度量).

虽然 Wasserstein 距离具有非常好的性质，但需要注意的是：Wasserstein 拓扑比弱拓扑更强。因为 $W_p(\mu_n,\mu) \to 0$ 蕴含 $\mu_n \Rightarrow \mu$ 外加矩收敛。所以它不是“弱收敛的另一种刻画”，而是在有限 p 阶矩空间上诱导比弱收敛更强的拓扑。另外，Wasserstein 距离的计算是困难的，除了一些如 $\mathscr{X} = \mathbb{R}^1$ 特殊的情况，Wasserstein 距离是没有显式的表达的，这让我们在研究它的渐近性时会遇见很大的困难.

Sommerfeld 与 Munk[4] 在有限元素的测度空间导出了 Wasserstein 距离的渐近性质, Tameling, Sommerfeld 与 Munk[5] 进一步地将其推广到可数元素的测度空间. 本文主要目标包括：

1. 介绍经验 Wasserstein 距离的渐近分布；
2. 说明无限维线性规划在 Wasserstein 问题中的作用；
3. 阐释 Hadamard 方向微分与 Delta 方法如何用于渐近理论证明。

我将介绍最基本的结果, 其他结果以及实例可以阅读 ([4]Sommerfeld and Munk) 与 ([5]Tameling, Sommerfeld and Munk). 首先我们需要介绍 EWD——经验 Wasserstein 距离 (Empirical Wasserstein distance).

μ 与 ν 之间的 Wasserstein 距离 (WD) 一般使用经验测度进行估计:

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta(X_i), \quad \hat{\nu}_m = \frac{1}{m}\sum_{j=1}^{m} \delta(Y_j)$$

其中 $X_1,\ldots,X_n \overset{i.i.d.}{\sim} \mu$, $Y_1,\ldots,Y_m \overset{i.i.d.}{\sim} \nu$, $\delta(x)$ 表示在 x 处的 Dirac 测度 (支撑仅在 x 点上). 而 $W_p(\hat{\mu}_n,\hat{\nu}_m)$ 就称之为 μ 与 ν 之间的经验 Wasserstein 距离 (EWD).

我们假设 $\mathscr{X}=\{x_1,x_2,\dots\}$ 是一个可数空间，并且赋予一个度量 $d:\mathscr{X}\times\mathscr{X}\to\mathbb{R}_+$. 于是 $\mathscr{X}$ 上的概率测度就是一个无穷维向量 r ，全体的 $\mathscr{X}$ 上的概率测度构成了如下集合:

$$\mathscr{P}(\mathscr{X})=\left\{r=(r_x)_{x\in\mathscr{X}}:r_x\geq 0,\forall x\in\mathscr{X},\sum_{x\in\mathscr{X}}r_x=1\right\}$$

为了刻画极限分布，定义了如下的协方差结构:

$$\Sigma(r)=\begin{cases}r_x(1-r_x) & \text{if } x=x^{'}\\ -r_xr_{x'} & \text{if } x\neq x^{'}.\end{cases} \tag{1}$$

定理 1 ([5]Tameling, Sommerfeld and Munk 定理 2.4)**.** *假设 $(\mathscr{X},d)$ 是一个可数元素的空间，$r,s\in\mathscr{P}_p(\mathscr{X})$, $p\geq 1$, 且 $\hat{r}_n$ 由 $X_1,\dots,X_n\overset{i.i.d.}{\sim} r$ 生成，$\hat{s}_m$ 由 $Y_1,\dots,Y_m\overset{i.i.d.}{\sim} s$ 生成 (与 $X_1,\dots,X_n$ 独立). 令 $G\sim N(0,\Sigma(r))$ 为一个高斯过程，其协方差结构 $\Sigma(r)$(定义如 (1)).*

$H\sim N(0,\Sigma(s))$, 与 G 独立. 满足 $\sum_{x\in\mathscr{X}}d^p(x,x_0)\sqrt{r_x}<\infty$, $\sum_{x\in\mathscr{X}}d^p(x,x_0^{'})\sqrt{s_x}<\infty$ 对存在的 $x_0,x_0^{'}\in\mathscr{X}$ 成立. 我们有:

(a) 令 $\rho_{n,m}=(nm/(n+m))^{1/2}$. 若 $r=s$ 且 $min(n,m)\to\infty$ ，$m/(n+m)\to\alpha\in[0,1]$ 我们有:

$$\rho_{n,m}W_p^p(\hat{r}_n,\hat{s}_m)\xrightarrow{d}\max_{\lambda\in S^*(r)}\langle G,\lambda\rangle \tag{2}$$

(b) 若 $r\neq s$ 且 $min(n,m)\to\infty$ ，$m/(n+m)\to\alpha\in[0,1]$ 我们有:

$$\rho_{n,m}(W_p^p(\hat{r}_n,\hat{s}_m)-W_p^p(r,s))\xrightarrow{d}\max_{(\lambda,\mu)\in S^*(r,s)}\{\sqrt{\alpha}\langle G,\lambda\rangle+\sqrt{1-\alpha}\langle H,\mu\rangle\} \tag{3}$$

其中 $\langle r,\lambda\rangle=\sum_{x\in\mathscr{X}}r_x\lambda_x$，而 $\max_{\lambda\in S^*(r)}\langle G,\lambda\rangle$ 也可以写成高斯宽度（Gaussian width）的形式. 我们不难发现除非 $S^*(r)$ 是单点集，极限分布不是正态的. 我们暂时没有详细介绍 $S^*(r)$ 的定义，但在后面我们会详细说明.

备注 1. *1. 定理 1 给出了 $r=s$ 与 $r\neq s$ 两种极限分布的形式，我们可以根据极限分布设计一个检验，去判别两个测度是否相等. 由于极限分布实际是某个集合上的高斯宽度，导出具体极限分布是困难的. 所以，在实际工程上我们会用极限分布的上界来代替，详细可参考 ([5]Tameling, Sommerfeld and Munk 第 3 部分).*

2. 我们将 $\sum_{x\in\mathscr{X}}d^p(x,x_0)\sqrt{r_x}<\infty$ 称之为可和性 (summability) 条件，我们可以使用一些替代条件去验证可和性条件，例如在指数分布族中我们可以参考 ([5]Tameling, Sommerfeld and Munk 定理 2.10).

3 定理 1 的证明

在这一节我们介绍定理 1（即原定理 2.4）的证明过程，特别是运用的方法.

3.1 弱收敛

首先，我们的底空间是可数空间，$\mathscr{X}=\{x_1,x_2,...\}$ 是一个可数空间，那么它上面的任意概率测度可以视作为一个无穷维向量 (参照我们前面定义的 $\mathscr{P}(\mathscr{X})$)，依据泛函分析的知识，概率测度的弱收敛可以转换为它的对偶空间上泛函作用上下的数列收敛 (对应 Portmanteau 定理的最后一条). 下面介绍 $l^1(\mathscr{X})$ 与 $l^\infty(\mathscr{X})$ 空间，而 $l^1(\mathscr{X})$ 的对偶空间即为 $l^\infty(\mathscr{X})$.

$$l^1(\mathscr{X})=\left\{(a_x)_{x\in\mathscr{X}}\in\mathbb{R}^{\mathscr{X}}:\sum_{x\in\mathscr{X}}|a_x|<\infty\right\}$$

$$l^\infty(\mathscr{X})=\left\{(a_x)_{x\in\mathscr{X}}\in\mathbb{R}^{\mathscr{X}}:\sup_{x\in\mathscr{X}}|a_x|<\infty\right\}$$

$\forall r_n,r\in\mathscr{P}(\mathscr{X})\subseteq l^1(\mathscr{X}),r_n\xrightarrow{\mathrm{w}}r\iff\langle r_n,\lambda\rangle\to\langle r,\lambda\rangle,\forall\lambda\in l^\infty(\mathscr{X})(n\to+\infty)$

实际上，并不是所有 $\mathscr{P}(\mathscr{X})$ 的元素都是我们关注的，考虑底空间上定义的距离函数 d，以及渐近性定理所要求的可和性条件，我们引入如下加权 l^1 范数，从而将 $l^1(\mathscr{X})$ 拓展为加权 l^1 空间 $l^1_{d^p_{x_0}}(\mathscr{X})$，以及对偶空间加权 l^∞ 空间 $l^\infty_{d^{-p}_{x_0}}(\mathscr{X})$.

$$\|a\|_{l^1_{d^p_{x_0}}}=\sum_{x\in\mathscr{X}}d^p(x,x_0)|r_x|+|a_{x_0}|$$

$$\|a\|_{l^\infty_{d^{-p}_{x_0}}}=\max\left(|a_{x_0}|,\sup_{x\neq x_0\in\mathscr{X}}|d^{-p}(x,x_0)a_x|\right)$$

$$l^1_{d^p_{x_0}}(\mathscr{X})=\left\{(a_x)_{x\in\mathscr{X}}\in\mathbb{R}^{\mathscr{X}}:\|a\|_{l^1_{d^p_{x_0}}}<\infty\right\}$$

$$l^\infty_{d^{-p}_{x_0}}(\mathscr{X})=\left\{(a_x)_{x\in\mathscr{X}}\in\mathbb{R}^{\mathscr{X}}:\|a\|_{l^\infty_{d^{-p}_{x_0}}}<\infty\right\}$$

加权 l^1 范数可以拓展到乘积空间 $\mathscr{X}\times\mathscr{X}$ 上，对 $\omega\in\mathscr{P}_p(\mathscr{X}\times\mathscr{X})$:

$$\begin{aligned}\|\omega\|_{l^1_{d^p_{x_0}}}=&\sum_{x,x'\in\mathscr{X}}d^p(x,x_0)|\omega_{x,x'}|+|\omega_{x_0,x'}|\\&+\sum_{x,x'\in\mathscr{X}}d^p(x',x_0)|\omega_{x,x'}|+|\omega_{x_0,x}|\end{aligned}$$

注意加权 $l^1_{d^p_{x_0}}(\mathscr{X})$ 空间以及其对偶空间都依赖于 $x_0\in\mathscr{X}$.

3.2 无穷维线性规划

因为我们假定了底空间只有可数个元素，Wasserstein 距离可以改写为如下的形式:

$$W_p^p(r,s)=\min_{\omega\in\Pi(r,s)}\sum_{x,x'\in\mathscr{X}}d^p(x,x')\omega_{x,x'}$$

其中 $\Pi(r,s)=\left\{\omega\in P(\mathscr{X}\times\mathscr{X}):\sum_{x\in\mathscr{X}}\omega_{x,x'}=s(x^{'}),\sum_{x'\in\mathscr{X}}\omega_{x,x'}=r(x),\forall x,x^{'}\in\mathscr{X}\right\}$

也就是如下无穷维线性规划问题:

$$
\begin{aligned}
&\min_{\omega\in l^1_{d^p_{x_0}}(\mathscr{X}\times\mathscr{X})}\sum_{x,x'\in\mathscr{X}}d^p(x,x^{'})\omega_{x,x'}\\
&s.t.\quad \sum_{x\in\mathscr{X}}\omega_{x,x'}=s(x^{'}),\forall x^{'}\in\mathscr{X}\\
&\qquad\quad \sum_{x^{'}\in\mathscr{X}}\omega_{x,x'}=r(x),\forall x\in\mathscr{X}\\
&\qquad\quad \omega_{x,x'}>0,\forall x,x^{'}\in\mathscr{X}
\end{aligned}
\tag{4}
$$

写成矩阵形式就是:

$$
\begin{aligned}
&\min_{\omega\in l^1_{d^p_{x_0}}(\mathscr{X}\times\mathscr{X})}f(\omega)\\
&s.t.\quad C(\omega,(r,s))\in K\\
&\qquad\quad C(\omega,(r,s))=\begin{pmatrix}\omega\\ \Sigma_1\omega-r\\ \Sigma_2\omega-s\end{pmatrix}
\end{aligned}
\tag{5}
$$

其中 $\Sigma_1\omega=\sum_{x'\in\mathscr{X}}\omega_{x,x'}$, $\Sigma_2\omega=\sum_{x\in\mathscr{X}}\omega_{x,x'}$，闭凸集 $K=l^1_{d^p_{x_0}}(\mathscr{X}\times\mathscr{X})_+\times\{0\}\times\{0\}$.

考虑对偶问题，拉格朗日函数可以写成如下 ((ν,λ,μ) 为拉格朗日乘子)

$$
L(\omega,\lambda,\mu)=\sum_{x,x'\in\mathscr{X}}d^p(x,x^{'})\omega_{x,x'}+\langle\nu,\omega\rangle+\langle\lambda,\omega^T\cdot\vec{1}-r\rangle+\langle\mu,\omega\cdot\vec{1}-s\rangle \tag{6}
$$

问题 (4) 是一个无穷维的线性规划问题，根据 [6]Bonnans 与 Shapiro(定义 2.163)，这是一个凸问题. 而且根据 [3]Villani(定理 4.1)，问题 (4) 的最优解一定存在 (但不一定唯一). 根据 [3]Villani(定理 5.10)，我们还可得悉问题 (4) 的强对偶性.

对 $s,r\in\mathscr{P}_p(\mathscr{X})$, 我们定义如下两个凸集:

$$
\begin{aligned}
S^*(r,s)&=\left\{(\lambda,\mu)\in l^\infty_{d^{-p}_{x_0}}(\mathscr{X})\times l^\infty_{d^{-p}_{x_0}}(\mathscr{X}):\langle r,\lambda\rangle+\langle s,\mu\rangle=W^p_p(r,s),\lambda_x+\mu_{x'}\le d^p(x,x^{'})\right\}\\
S^*(r)&=\left\{\lambda\in l^\infty_{d^{-p}_{x_0}}(\mathscr{X}):\lambda_x-\lambda_{x'}\le d^p(x,x^{'}),\forall x,x^{'}\in supp(r)\right\}
\end{aligned}
\tag{7}
$$

对偶问题的最优解就在这两个凸集之上.

3.3 Δ 方法 (基于 Hadamard 方向微分)

下面将介绍 Hadamard 方向微分的基本定义, 以及基于 Hadamard 方向微分的 Δ 方法. 特别地我们将探讨 Wasserstein 距离的 Hadamard 方向可微性. 这是证明渐近性质的重要一步.

定义 1 (Hadamard 方向可微性)**.** *([6]Bonnans 与 Shapiro 定义 2.45). $\mathscr{U}$ 与 $\mathscr{Y}$ 为赋范空间. 一个映射 $f: D_f \subseteq \mathscr{U} \to \mathscr{Y}$ 称之为在 $u \in D_f$ 处 Hadamard 方向可微，当对于任意收敛到 h 的序列 h_n 和任意的数列 $t_n \searrow 0$，满足 $u + t_n h_n \in D_f$ 有极限:*

$$f_u^{'}(h) = \lim_{n\to\infty} \frac{f(u + t_n h_n) - f(u)}{t_n} \tag{8}$$

定义 2 (对集合切向的 Hadamard 方向可微性)**.** *([6]). K 是 D_f 的子集, f 对 K 切向在 u 处 Hadamard 切向可微的当且仅当对所有收敛到 h 的序列 h_n，极限 (8) 存在，h_n 具有如下的形式 $h_n = t_n^{-1}(k_n - u)$，其中 $k_n \in K$ 且 $t_n \to 0$. 这个导数是定义在 u 处对 K 的 contingent (Bouligand) 锥.*

$$Tan_K(u) = \left\{ h \in \mathscr{U} : h = \lim_{n\to\infty} t_n^{-1}(k_n - u), k_n \in K, t_n \searrow 0 \right\}$$

定理 2 (Δ 方法)**.** *([7]Romisch 定理 1). K 是 $\mathscr{U}$ 的子集, 映射 $f: K \to \mathscr{Y}$ 满足如下两个条件:*

(a) 映射 f 在 $u \in K$ 处切向 KHadamard 方向可微，且导数为 $f_u^{'}(\cdot): T_K(u) \to \mathscr{Y}$.

(b) 对每个 n, $X_n : \Omega_n \to K$ 是一组映射，且存在序列 $a_n \to +\infty$ 满足 $a_n(X_n - u) \xrightarrow{d} X$ 且随机元 X 取值在 $T_K(u)$ 上.

于是我们有 $a_n(f(X_n) - f(u)) \xrightarrow{d} f_u^{'}(X)$.

考虑如下的参数规划问题 P_u，$u \in U$，U 为 Banach 空间，X 为 Hausdorff 拓扑空间，$\Phi(u) \in X$ 非空闭集，$f: X \times U \mapsto \mathbb{R}$ 连续:

$$\min_{x\in X} f(x,u) \quad s.t.\ x \in \Phi \tag{9}$$

其中可行域为 Φ 与 u 无关，我们用 $v(u)$ 表示最后的最优值.

inf-紧性条件：存在 $\alpha \in \mathbb{R}$，紧集 $C \subseteq X$, 对于 u_0 附近的每一个 u，水平集

$$lev_\alpha f(\cdot, u) := \{x \in \Phi : f(x,u) \le \alpha\}$$

是非空的，并包含在 C 中.

方向正则性 ([6]Bonnans 与 Shapiro 命题 4.12).

(i) 函数 $f(x,u)$ 在 $X \times U$ 连续，

(ii) inf-紧性条件成立，

(iii) 对任意的 $x \in \Phi$，函数 $f_x(\cdot) := f(x,\cdot)$ 在 u_0 方向可微,

(iv) 若 $d \in U, t_n \downarrow 0$ 且 $\{x_n\}$ 是 C 中序列，于是 $\{x_n\}$ 有极限点 $\bar{x}$ 满足:

$$\limsup_{n\to\infty} \frac{f(x_n, u_0 + t_n d) - f(x_n, u_0)}{t_n} \ge f_{\bar{x}}'(u_0, d)$$

最优值函数 $v(u)$ 在 u_0 处 Hadamard 可微，且导数为:

$$v'(u_0, d) = \inf_{x\in\mathcal{S}(u_0)} f_x'(u_0, d)$$

定理 3. *([6]Bonnans 与 Shapiro 定理 4.24)*

(i) 凸性与最优解的存在性：规划问题是凸问题，且最优解存在，

(ii) 方向正则性：对于所有 $x_0 \in \mathcal{S}(u_0)$ 在方向 d 满足方向正则性，

(iii) 原问题最优解稳定性：若 $u_n := u_0 + t_n d + o(t_n)$，对序列 $t_n \downarrow 0$，则 (P_{u_n}) 有 $o(t_n)$ -最优解 x_n ：序列 $\{x_n\}$ 有极限 *(*强拓扑意义下*)* $x_0 \in \mathcal{S}(u_0)$*(*仍为最优解*)*.

最优值函数 $v(\cdot)$ 在 u_0 处是沿着 d *Hadamard* 方向可微的，

$$v'(u_0, d) = \inf_{x \in \mathcal{S}(u_0)} \sup_{\lambda \in \Lambda(u_0)} D_u L(x, \lambda, u_0) d$$

备注 2. 定理 *3* 将约束的线性规划最优值的 *Hadamard* 导数的求解，转化为了其拉格朗日函数的导数求解，再进行 sup 与 inf 运算.

我们定义：

$$\mathscr{W}_p(\mathscr{X}) = \left\{ r \in \mathscr{P} : \sum_{x \in \mathscr{X}} d^p(x, x_0) r_x < \infty \right\}.$$

其中随意取定点 $x_0 \in \mathscr{X}$，且集合与 x_0 的取定无关，该集合表示所有通过距离函数 d 定义的 p 阶矩有限的概率测度 (实际上这里就是与 x_0 处的 Dirac 测度之间的 Wasserstein 距离有限的所有概率测度).

为了应用定理 3，需要验证：(1) Wasserstein 问题对应凸规划；(2) 最优解存在且稳定；(3) 强对偶成立；(4) 参数扰动满足方向正则性。由于这些条件已经由 Villani 与 Bonnans-Shapiro 理论保证，因此可以推出最优值函数 W_p^p 具有 Hadamard 方向可微性。借助定理 3，我们可以得出 W_p^p 是 $(\mathscr{W}_p(\mathscr{X}) \times \mathscr{W}_p(\mathscr{X}), \|\cdot\|_{l^1_{d^p_{x_0}}})$ 到 $\mathbb{R}^+$ 的映射，$(r, s) \mapsto W_p^p(r, s)$ 是切向 $(\mathscr{W}_p(\mathscr{X}) \times \mathscr{P}_p(\mathscr{X}))$ Hadamard 方向可微的.

导数定义的 contingent 锥:

$$\mathscr{D}(r, s) = \mathscr{D}(r) \times \mathscr{D}(s)$$

$$\mathscr{D}(r) := \left\{ d \in l^1_{d^p_{x_0}}(\mathscr{X}) \setminus \{0\} : \sum_{x \in \mathscr{X}} d_x = 0, d_x \in [-r_x, 1 - r_x] \right\}$$

由 3.2 节公式 (6)，我们知拉格朗日函数为：

$$L(\omega, (\nu, \lambda, \mu), (r, s)) = \sum_{x, x' \in \mathscr{X}} d^p(x, x') \omega_{x,x'} + \langle \nu, \omega \rangle + \langle \lambda, \omega^T \cdot \vec{1} - r \rangle + \langle \mu, \omega \cdot \vec{1} - s \rangle$$

在 Fréchet 意义上对 (r, s) 微分，我们用 (d_1, d_2) 代入，结果是

$$D_{(r,s)} L(\omega, (\nu, \lambda, \mu), (r, s))(d_1, d_2) = -(\langle \lambda, d_1 \rangle + \langle \mu, d_2 \rangle)$$

所以 Wasserstein 距离 $W_p^p(r, s)$ 在 (r, s) 处的 Hadamard 方向导数为:

$$(d_1, d_2) \mapsto \sup_{(\lambda, \mu) \in S^*(r,s)} -(\langle \lambda, d_1 \rangle + \langle \mu, d_2 \rangle) \tag{10}$$

其中 $S^*(r,s)$(定义见上) 是对偶问题的最优解.

由 [8](推论 1)，当可和性条件满足时，

$$\begin{aligned}\rho_{n,m}((\hat{r}_n,\hat{s}_m)-(r,s)) &= \left(\sqrt{\frac{m}{n+m}}\sqrt{n}(\hat{r}_n-r),\sqrt{\frac{n}{n+m}}\sqrt{m}(\hat{s}_m-s)\right)\\ &\xrightarrow{d}(\sqrt{\alpha}G,\sqrt{1-\alpha}H)\end{aligned} \tag{11}$$

G、H 为高斯过程 (定义见定理 1).

最后根据 3.3.3 中的 Delta 方法，定理 1 得证，

$$\rho_{n,m}(W_p^p(\hat{r}_n,\hat{s}_m)-W_p^p(r,s))\xrightarrow{d}\max_{(\lambda,\mu)\in S^*(r,s)}\{\sqrt{\alpha}\langle G,\lambda\rangle+\sqrt{1-\alpha}\langle H,\mu\rangle\}$$

参考文献

[1] S. Resnick, *A probability path.* Springer, 2019.

[2] P. Billingsley, *Convergence of probability measures.* John Wiley & Sons, 2013.

[3] C. Villani, *Optimal transport: old and new.* Springer, 2009, vol. 338.

[4] M. Sommerfeld and A. Munk, "Inference for empirical wasserstein distances on finite spaces," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 80, no. 1, pp. 219–238, 2018.

[5] C. Tameling, M. Sommerfeld, and A. Munk, "Empirical optimal transport on countable metric spaces: Distributional limits and statistical applications," *The Annals of Applied Probability*, vol. 29, no. 5, pp. 2744–2781, 2019.

[6] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems.* Springer Science & Business Media, 2013.

[7] W. Römisch, "Delta method, infinite dimensional," 2004.

[8] N. C. Jain, "Central limit theorem in a banach space," in *Probability in Banach spaces.* Springer, 1976, pp. 113–130.