Chapter 6
Quantitative Relationships Between Minimal Distances and Minimal Norms

The goals of this chapter are to:

- Explore the conditions under which there is equality between the Kantorovich and the Kantorovich–Rubinstein functionals;
- Provide inequalities between the Kantorovich and Kantorovich–Rubinstein functionals;
- Provide criteria for convergence, compactness, and completeness of probability measures in probability spaces involving the Kantorovich and Kantorovich–Rubinstein functionals;
- Analyze the problem of uniformity between the two functionals.

Notation introduced in this chapter:

Notation	Description
$\mathcal{P}_\lambda = \mathcal{P}_\lambda(U)$	Space of laws with a finite λ-moment
$\widehat{\Lambda}$	Generalized Kantorovich functional
$\overset{\circ}{\Lambda}$	Generalized Kantorovich–Rubinstein functional

6.1 Introduction

In Chap. 5, we discussed the Kantorovich and Kantorovich–Rubinstein functionals. They generate minimal distances, $\widehat{\mu}_c$, and minimal norms, $\overset{\circ}{\mu}_c$, respectively, and we considered the problem of evaluating these functionals. The similarities between the two functionals indicate there can be quantitative relationships between them.

In this chapter, we begin by exploring the conditions under which $\widehat{\mu}_c = \overset{\circ}{\mu}_c$. It turns out that equality holds if and only if the cost function $c(x, y)$ is a metric

S.T. Rachev et al., *The Methods of Distances in the Theory of Probability and Statistics*,
DOI 10.1007/978-1-4614-4869-3_6,

itself. Under more general conditions, certain inequalities hold involving $\widehat{\mu}_c$, $\mathring{\mu}_c$, and other probability metrics. These inequalities imply criteria for convergence, compactness, and uniformity in the spaces of probability measures $(\mathcal{P}(U), \widehat{\mu}_c)$ and $(\mathcal{P}(U), \mathring{\mu}_c)$. Finally, we conclude with a generalization of the Kantorovich and Kantorovich–Rubinstein functionals.

6.2 Equivalence Between Kantorovich Metric and Kantorovich–Rubinstein Norm

Levin (1975) proved that if U is a compact, $c(x,x) = 0$, $c(x,y) > 0$, and $c(x,y) + c(y,x) > 0$ for $x \neq y$, then $\widehat{\mu}_c = \mathring{\mu}_c$ if and only if $c(x,y) + c(y,x)$ is a metric in U. In the case of an s.m.s. U, we have the following version of Levin's result.

Theorem 6.2.1 (Neveu and Dudley 1980). *Suppose U is an s.m.s. and $c \in \mathfrak{C}^*$ (Corollary 5.3.1). Then*

$$\widehat{\mu}_c(P_1, P_2) = \mathring{\mu}_c(P_1, P_2) \tag{6.2.1}$$

for all P_1 and P_2 with

$$\int_U c(x,a)(P_1 + P_2)(\mathrm{d}x) < \infty \tag{6.2.2}$$

if and only if c is a metric.

Proof. Suppose (6.2.1) holds and set $P_1 = \delta_x$ and $P_2 = \delta_y$ for $x, y \in U$. Then the set $\mathcal{P}^{(P_1,P_2)}$ of all laws in $U \times U$ with marginals P_1 and P_2 contains only $P_1 \times P_2 = \delta_{(x,y)}$, and by Theorem 5.4.2,

$$\begin{aligned}
\widehat{\mu}_c(P_1, P_2) = c(x,y) = \mathring{\mu}_c(P_1, P_2) &= \sup\left\{\int f\,\mathrm{d}(P_1 - P_2) : \|f\|_c \le 1\right\} \\
&= \sup\{|f(x) - f(y)| : \|f\|_c \le 1\} \\
&\le \sup\{|f(x) - f(z)| + |f(z) - f(y)| : \|f\|_c \le 1\} \\
&\le c(x,z) + c(z,y).
\end{aligned}$$

By assumption, $c \in \mathfrak{C}^*$, and therefore the triangle inequality implies that c is a metric in U.

Now define $\mathcal{G}(U)$ as the set of all pairs (f, g) of continuous functions $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ such that $f(x) + g(y) < c(x,y)$ $\forall x, y \in U$. Let $\mathcal{G}_B(U)$ be the set of all pairs $(f, g) \in \mathcal{G}(U)$ with f and g bounded.

Now suppose that $c(x,y)$ is a metric and that $(f,g) \in \mathcal{G}_B(U)$. Define $h(x) = \inf\{c(x,y) - g(y) : y \in U\}$. As the infimum of a family of continuous functions, h is upper semicontinuous. For each $x \in U$ we have $f(x) \le h(x) \le -g(x)$. Then

$$\begin{aligned} h(x) - h(x') &= \inf_u (c(x,u) - g(u)) - \inf_v (c(x',v) - g(v)) \\ &< \sup_v (g(v) - c(x',v) + c(x,v) - g(v)) \\ &= \sup_v (c(x,v) - c(x',v)) \le c(x,x'), \end{aligned}$$

so that $\|h\|_c \le 1$. Then for P_1, P_2 satisfying (6.2.2) we have

$$\int f \mathrm{d}P_1 + \int g \mathrm{d}P_2 \le \int h \mathrm{d}(P_1 - P_2),$$

so that (according to Corollary 5.3.1 and Theorem 5.4.2 of Chap. 5) we have

$$\begin{aligned} \widehat{\mu}_c(P_1, P_2) &= \sup \left\{ \int f \mathrm{d}P_1 + \int g \mathrm{d}P_2 : (f,g) \in \mathcal{G}_B(U) \right\} \\ &\le \sup \left\{ \int h \mathrm{d}(P_1 - P_2) : \|h\|_c \le 1 \right\} = \mathring{\mu}_c(P_1, P_2). \end{aligned}$$

Thus $\widehat{\mu}_c(P_1, P_2) = \mathring{\mu}_c(P_1, P_2)$. ☐

Corollary 6.2.1. *Let (U,d) be an s.m.s. and $a \in U$. Then*

$$\widehat{\mu}_d(P_1, P_2) = \mathring{\mu}_d(P_1, P_2) = \sup \left\{ \int f \mathrm{d}(P_1 - P_2) : \|f\|_L \le 1 \right\} \qquad (6.2.3)$$

whenever

$$\int d(x,a) P_i(\mathrm{d}x) < \infty, \qquad i = 1, 2. \qquad (6.2.4)$$

The supremum is attained for some optimal f_0 with $\|f_0\|_L := \sup_{x \neq y}\{|f(x) - f(y)|/d(x,y)\}$.

If P_1 and P_2 are tight, there are some $b_0 \in \mathcal{P}^{(P_1,P_2)}$ and $f_0 : U \to \mathbb{R}$ with $\|f_0\|_L \le 1$ such that

$$\widehat{\mu}_d(P_1, P_2) = \int d(x,y) b_0(\mathrm{d}x, \mathrm{d}y) = \int f_0 \mathrm{d}(P_1 - P_2),$$

where $f_0(x) - f_0(y) = d(x,y)$ for b_0-a.e. (x,y) in $U \times U$.

Proof. Set $c(x,y) = d(x,y)$. Application of the theorem proves the first statement. The second (existence of f_0) follows from Theorem 5.4.3.

For each $n \ge 1$ choose $b_n \in \mathcal{P}^{(P_1,P_2)}$ with

$$\int d(x,y) b_n(\mathrm{d}x, \mathrm{d}y) < \widehat{\mu}_d(P_1, P_2) + \frac{1}{n}.$$

If P_1 and P_2 are tight, then by Corollary 5.3.1 there exists $b_0 \in \mathcal{P}^{(P_1,P_2)}$ such that

$$\widehat{\mu}_d(P_1, P_2) = \int d(x, y) b_0(\mathrm{d}x, \mathrm{d}y),$$

i.e., that b_0 is optimal. Integrating both sides of $f_0(x) - f_0(y) \le d(x, y)$ with respect to b_0 yields $\int f_0 \mathrm{d}(P_1 - P_2) \le \int d(x, y) b_0(\mathrm{d}x, \mathrm{d}y)$. However, we know that we have equality of these integrals. This implies that $f_0(x) - f_0(y) = d(x, y)$ b_0-a.e. $\square$

6.3 Inequalities Between $\widehat{\mu}_c$ and $\mathring{\mu}_c$

In the previous section we looked at conditions under which $\widehat{\mu}_c = \mathring{\mu}_c$. In general, $\widehat{\mu}_c \ne \mathring{\mu}_c$. For example, if $U = \mathbb{R}$, $d(x, y) = |x - y|$,

$$c(x, y) = d(x, y) \max(1, d^{p-1}(x, a), d^{p-1}(y, a)), \quad p \ge 1, \tag{6.3.1}$$

then for any laws P_i $(i = 1, 2)$ on $\mathcal{B}(R)$ with distribution functions (DFs) F_i we have the following explicit expressions:

$$\widehat{\mu}_c(P_1, P_2) = \int_0^1 c(F_1^{-1}(t), F_2^{-1}(t))\mathrm{d}t, \tag{6.3.2}$$

where F_i^{-1} is the function inverse to the DF F_i (see Theorem 7.4.2 in Chap. 7). On the other hand,

$$\mathring{\mu}_c(P_1, P_2) = \int_{-\infty}^{\infty} |F_1(x) - F_2(x)| \max(1, |x - a|^{p-1})\mathrm{d}x \tag{6.3.3}$$

(see Theorem 5.5.1 in Chap. 5). However, in the space $\mathcal{M}_p = \mathcal{M}_p(U)$ [$U = (U, d)$ is an s.m.s.] of all Borel probability measures P with finite $\int d^p(x, a)P(\mathrm{d}x)$, the functionals $\widehat{\mu}_c$ and $\mathring{\mu}_c$ [where c is given by (6.3.1)] metrize the same exact topology, that is, the following $\widehat{\mu}_c$- and $\mathring{\mu}_c$-convergence criteria will be proved.

Theorem 6.3.1. *Let (U, d) be an s.m.s., let c be given by (6.3.1), and let $P, P_n \in \mathcal{M}_p$ $(n = 1, 2, \dots)$. Then the following relations are equivalent:*

(I)

$$\widehat{\mu}_c(P_n, P) \to 0;$$

(II)

$$\mathring{\mu}_c(P_n, P) \to 0;$$

(III)

$$P_n \text{ converges weakly to } P \; (P_n \xrightarrow{\text{w}} P) \text{ and}$$

$$\lim_{N\to\infty} \sup_n \int d^p(x,a) I\{d(x,a) > N\} P_n(\mathrm{d}x) = 0;$$

(IV)

$$P_n \xrightarrow{\text{w}} P \text{ and } \int d^p(x,a) P_n(\mathrm{d}x) \to \int d^p(x,a) P(\mathrm{d}x).$$

(The assertion of the theorem is an immediate consequence of Theorems 6.3.2–6.3.5 below and the more general Theorem 6.4.1).

Theorem 6.3.1 is a qualitative $\widehat{\mu}_c$ ($\overset{\circ}{\mu}_c$)-convergence criterion. One can rewrite (III) as

$$\boldsymbol{\pi}(P_n, P) \to 0 \text{ and } \lim_{\varepsilon\to 0} \sup_n \omega(\varepsilon; P_n; \lambda) = 0,$$

where $\boldsymbol{\pi}$ is the Prokhorov metric[1]

$$\boldsymbol{\pi}(P, Q) := \inf\{\varepsilon > 0 : P(A) \le Q(A^\varepsilon) + \varepsilon \quad \forall A \in \mathcal{B}(U)\}$$
$$(A^\varepsilon := \{x : d(x, A) < \varepsilon\}) \tag{6.3.4}$$

and $\omega(\varepsilon; P; \lambda)$ is the following modulus of λ-integrability:

$$\omega(\varepsilon; P; \lambda) := \int \lambda(x) I\left\{d(x,a) > \frac{1}{\varepsilon}\right\} P(\mathrm{d}x), \tag{6.3.5}$$

where $\lambda(x) := \max(d(x,a), d^p(x,a))$. Analogously, (IV) is equivalent to

(IV^*)

$$\boldsymbol{\pi}(P_n, P) \to 0 \text{ and } D(P_n, P; \lambda) \to 0,$$

where

$$D(P, Q; \lambda) := \left| \int \lambda(x)(P - Q)(\mathrm{d}x) \right|. \tag{6.3.6}$$

In this section we investigate quantitative relationships between $\widehat{\mu}_c$, $\overset{\circ}{\mu}_c$, $\boldsymbol{\pi}$, ω, and D in terms of inequalities between these functionals. These relationships yield convergence and compactness criteria in the space of measures w.r.t. the Kantorovich-type functionals $\widehat{\mu}_c$ and $\overset{\circ}{\mu}_c$ (see Examples 3.3.2 and 3.3.6 in Chap. 3) as well as the $\overset{\circ}{\mu}_c$-completeness of the space of measures.

[1] See Examples 3.3.3 and 4.3.2 in Chaps. 3 and 4, respectively.

In what follows, we assume that the cost function c has the form considered in Example 5.2.1:

$$c(x,y) = d(x,y)k_0(d(x,a),d(y,a)) \quad x,y \in U, \tag{6.3.7}$$

where $k_0(t,s)$ is a symmetric continuous function nondecreasing on both arguments $t \geq 0$, $s \geq 0$, and satisfying the following conditions:

(C1)

$$\alpha := \sup_{s \neq t} \frac{|K(t) - K(s)|}{|t-s|k_0(t,s)} < \infty,$$

where $K(t) := tk_0(t,t)$, $t \neq 0$;

(C2)

$$\beta := k(0) > 0,$$

where $k(t) = k_0(t,t)$ $t \geq 0$; and

(C3)

$$\gamma := \sup_{t \geq 0, s \geq 0} \frac{k_0(2t,2s)}{k_0(t,s)} < \infty.$$

If c is given by (6.3.1), then c admits the form (6.3.7) with $k_0(t,s) = \max(1, t^{p-1}, s^{p-1})$, and in this case $\alpha = p$, $\beta = 1$, $\gamma = 2^{p-1}$. Further, let $\mathcal{P}_\lambda = \mathcal{P}_\lambda(U)$ be the space of all probability measures on the s.m.s. (U,d) with finite λ-moment

$$\mathcal{P}_\lambda(U) = \left\{ P \in \mathcal{P}(U) : \int_U \lambda(x)P(\mathrm{d}x) < \infty \right\}, \tag{6.3.8}$$

where $\lambda(x) = K(d(x,a))$ and a is a fixed point of U.

In Theorems 6.3.2–6.3.5 we assume that $P_1 \in \mathcal{P}_\lambda$, $P_2 \in \mathcal{P}_\lambda$, $\varepsilon > 0$, and denote $\widehat{\mu}_c := \widehat{\mu}_c(P_1,P_2)$ [see (5.2.16)], $\mathring{\mu}_c := \mathring{\mu}_c(P_1,P_2)$ [see (5.2.17)], $\boldsymbol{\pi} := \boldsymbol{\pi}(P_1,P_2)$,

$$\omega_i(\varepsilon) := \omega(\varepsilon; P_i; \lambda) := \int \lambda(x) I\{d(x,a) > 1/\varepsilon\} P_i(\mathrm{d}x), \quad P_i \in \mathcal{P}_\lambda$$

$$D := D(P_1,P_2;\lambda) := \left| \int \lambda \mathrm{d}(P_1 - P_2) \right|,$$

and the function c satisfies conditions (C1)–(C3). We begin with an estimate of $\widehat{\mu}_c$ from above in terms of $\boldsymbol{\pi}$ and $\omega_i(\varepsilon)$.

Theorem 6.3.2.

$$\widehat{\mu}_c \leq \boldsymbol{\pi}[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2k(1)] + 5\omega_1(\varepsilon) + 5\omega_2(\varepsilon). \tag{6.3.9}$$

Proof. Recall that $\mathcal{P}^{(P_1,P_2)}$ is the space of all laws P on $U \times U$ with prescribed marginals P_1 and P_2. Let $\mathbf{K} = \mathbf{K}_1$ be the Ky Fan metric with parameter 1 (see Example 3.4.2 in Chap. 3)

$$\mathbf{K}(P) := \inf\{\delta > 0 : P(d(x,y) > \delta) < \delta\} \quad P \in \mathcal{P}_\lambda(U). \tag{6.3.10}$$

Claim 1. For any $N > 0$ and for any measure P on U^2 with marginals P_1 and P_2, i.e., $P \in \mathcal{P}^{(P_1,P_2)}$, we have

$$\int_{U\times U} c(x,y)P(\mathrm{d}x,\mathrm{d}y) \leq \mathbf{K}(P)\left[4K(N) + \int_U k(d(x,a))(P_1+P_2)(\mathrm{d}x)\right] + 5\omega_1(1/N) + 5\omega_2(1/N). \tag{6.3.11}$$

Proof of Claim 1. Suppose $\mathbf{K}(P) < \sigma \leq 1$, $P \in \mathcal{P}^{(P_1,P_2)}$. Then by (6.3.7) and (C3),

$$\int c(x,y)P(\mathrm{d}x,\mathrm{d}y) \leq \int d(x,y)k(\max\{d(x,a),d(y,a)\})P(\mathrm{d}x,\mathrm{d}y) \leq I_1 + I_2,$$

where

$$I_1 := \int_{U\times U} d(x,y)k(d(x,a))P(\mathrm{d}x,\mathrm{d}y)$$

and

$$I_2 := \int_{U\times U} d(x,y)k(d(y,a))P(\mathrm{d}x,\mathrm{d}y).$$

Let us estimate I_1:

$$\begin{aligned} I_1 &:= \int d(x,y)k(d(x,a))[I\{d(x,y) < \delta\} + I\{d(x,y) \geq \delta\}]P(\mathrm{d}x,\mathrm{d}y) \\ &\leq \delta \int k(d(x,a))P(\mathrm{d}x,\mathrm{d}y) \\ &\quad + \int d(x,y)k(d(x,a))I\{d(x,y) \geq \delta\}P(\mathrm{d}x,\mathrm{d}y) \\ &\leq I_{11} + I_{12} + I_{13}, \end{aligned} \tag{6.3.12}$$

where

$$I_{11} := \delta \int_U k(d(x,a))[I\{d(x,a) \geq 1\} + I\{d(x,a) \leq 1\}]P_1(\mathrm{d}x),$$

$$I_{12} := \int_{U\times U} d(x,a)k(d(x,a))I\{d(x,y) \geq \delta\}P(\mathrm{d}x,\mathrm{d}y), \quad \text{and}$$

$$I_{13} := \int_{U\times U} d(y,a)k(d(x,a))I\{d(x,y) \geq \delta\}P(\mathrm{d}x,\mathrm{d}y).$$

Obviously, by $\lambda(x) := K(d(x,a))$, $I_{11} \le \delta \int k(d(x,a))I\{d(x,a) \ge 1\}P_1(\mathrm{d}x) + \delta k(1) \le \delta\omega_1(1) + \delta k(1)$. Further,

$$\begin{aligned}
I_{12} &= \int K(d(x,a))I\{d(x,y)\ge\delta\}[I\{d(x,a) > N\}+I\{d(x,a)\le N\}]P(\mathrm{d}x,\mathrm{d}y)\\
&\le \int_U \lambda(x)I\{d(x,a) > N\}P_1(\mathrm{d}x) + K(N)\int_{U\times U} I\{d(x,y) \ge \delta\}P(\mathrm{d}x,\mathrm{d}y)\\
&\le \omega_1(1/N) + K(N)\delta.
\end{aligned}$$

Now let us estimate the last term in estimate (6.3.12):

$$\begin{aligned}
I_{13} = &\int_{U\times U} d(y,a)k(d(x,a))I\{d(x,y) \ge \delta\}[I\{d(x,a) \ge d(y,a) > N\}\\
&+ I\{d(y,a) > d(x,a) > N\} + I\{d(x,a) > N, d(y,a) \le N\}\\
&+ I\{d(x,a)\le N, d(y,a) > N\}+I\{d(x,a) \le N, d(y,a) \le N\}]P(\mathrm{d}x,\mathrm{d}y)\\
\le &\int_{U\times U} \lambda(x)I\{d(x,a) > d(y,a) > N\}P(\mathrm{d}x,\mathrm{d}y)\\
&+ \int_{U\times U} \lambda(y)I\{d(y,a) \ge d(x,a) \ge N\}P(\mathrm{d}x,\mathrm{d}y)\\
&+ \int_U \lambda(x)I\{d(x,a) > N\}P_1(\mathrm{d}x) + \int_U \lambda(y)I\{d(y,a) > N\}P_2(\mathrm{d}y)\\
&+ K(N)\int_{U\times U} I\{d(x,y) \ge \delta\}P(\mathrm{d}x,\mathrm{d}y)\\
\le\ & 2\omega_1(1/N) + 2\omega_2(1/N) + K(N)\delta.
\end{aligned}$$

Summarizing the preceding estimates we obtain $I_1 \le \delta\omega_1(1)+\delta k(1)+3\omega_1(1/N)+2\omega_2(1/N)+2K(N)\delta$. By symmetry we have $I_2 \le \delta\omega_2(1)+\delta k(1)+3\omega_2(1/N)+2\omega_1(1/N)+2K(N)\delta$. Therefore, the last two estimates imply

$$\begin{aligned}
\int c(x,y)P(\mathrm{d}x,\mathrm{d}y) &\le I_1 + I_2\\
&\le \delta(\omega_1(1) + \omega_2(1) + 2k(1) + 4K(N))\\
&\quad + 5\omega_1(1/N) + 5\omega_2(1/N).
\end{aligned}$$

Letting $\delta \to \mathbf{K}(P)$ we obtain (6.3.11), which proves the claim.

Claim 2 (Strassen–Dudley Theorem).

$$\inf\{\mathbf{K}(P) : P \in \mathcal{P}^{(P_1,P_2)}\} = \boldsymbol{\pi}(P_1, P_2). \tag{6.3.13}$$

Proof of Claim 2. See Dudley (2002) (see also further Corollary 7.5.2 in Chap. 7).
Claims 1 and 2 complete the proof of the theorem. □

The next theorem shows that $\widehat{\mu}_c$-convergence and $\overset{\circ}{\mu}_c$-convergence imply the weak convergence of measures.

Theorem 6.3.3.

$$\beta\boldsymbol{\pi}^2 \le \overset{\circ}{\mu}_c \le \widehat{\mu}_c. \tag{6.3.14}$$

Proof. Obviously, for any continuous nonnegative function c,

$$\overset{\circ}{\mu}_c \le \widehat{\mu}_c \tag{6.3.15}$$

and

$$\overset{\circ}{\mu}_c \ge \boldsymbol{\zeta}_c, \tag{6.3.16}$$

where $\boldsymbol{\zeta}_c$ is the Zolatarev simple metric with a ζ-structure (Definition 4.4.1)

$$\begin{aligned}\boldsymbol{\zeta}_c &:= \boldsymbol{\zeta}_c(P_1, P_2)\\ &:= \sup\left\{\left|\int_U f\,\mathrm{d}(P_1 - P_2)\right| : f : U \to \mathbb{R}, |f(x) - f(y)| \le c(x,y) \forall x, y \in U\right\}.\end{aligned} \tag{6.3.17}$$

Now, using assumption (C2) we have that $c(x,y) \ge \beta d(x,y)$ and, hence, $\boldsymbol{\zeta}_c \ge \beta\boldsymbol{\zeta}_d$. Thus, by (6.3.16),

$$\overset{\circ}{\mu}_c \ge \beta\boldsymbol{\zeta}_d. \tag{6.3.18}$$

Claim 3.

$$\boldsymbol{\zeta}_d \ge \boldsymbol{\pi}^2. \tag{6.3.19}$$

Proof of Claim 3. Using the dual representation of $\widehat{\mu}_d$ [see (6.2.3)] we are led to

$$\widehat{\mu}_d = \boldsymbol{\zeta}_d, \tag{6.3.20}$$

which in view of the inequality

$$\int d(x,y)P(\mathrm{d}x, \mathrm{d}y) \ge \mathbf{K}^2(P) \ \text{ for any } \ P \in \mathcal{P}^{(P_1,P_2)} \tag{6.3.21}$$

establishes (6.3.19). The proof of the claim is now completed.

The desired inequalities (6.3.14) are the consequence of (6.3.15), (6.3.16), (6.3.18), and Claim 3. □

The next theorem establishes the uniform λ-integrability

$$\lim_{\varepsilon\to 0} \sup_n \omega(\varepsilon, P_n, \lambda) = 0$$

of the sequence of measures $P_n \in \mathcal{P}_\lambda$ $\overset{\circ}{\mu}_c$-converging to a measure $P \in \mathcal{P}_\lambda$.

Theorem 6.3.4.

$$\omega_1(\varepsilon/2) \le \alpha(2\gamma+1)\mathring{\mu}_c + 2(\gamma+1)\omega_2(\varepsilon). \tag{6.3.22}$$

Proof. For any $N > 0$, by the triangle inequality, we have

$$\omega_1(1/2N) := \int \lambda(x) I\{d(x,a) > 2N\} P_1(\mathrm{d}x) \le \mathcal{T}_1 + \mathcal{T}_2, \tag{6.3.23}$$

where

$$\mathcal{T}_1 := \left| \int \lambda(x) I\{d(x,a) > 2N\}(P_1 - P_2)(\mathrm{d}x) \right|$$

and

$$\mathcal{T}_2 := \int \lambda(x) I\{d(x,a) > N\} P_2(\mathrm{d}x) = \omega_2(1/N).$$

Claim 4.

$$\mathcal{T}_1 \le \alpha\mathring{\mu}_c + K(2N)\int I\{d(x,a) > 2N\}(P_1 + P_2)(\mathrm{d}x). \tag{6.3.24}$$

Proof of Claim 4. Denote $f_N(x) := (1/\alpha)\max(\lambda(x), K(2N))$. Since $\lambda(x) = K(d(x,a)) = d(x,a)k_0(d(x,a), d(x,a))$, then by (C1),

$$\begin{aligned} |f_N(x) - f_N(y)| &\le (1/\alpha)|\lambda(x) - \lambda(y)| \\ &\le |d(x,a) - d(y,a)| k_0(d(x,a), d(y,a)) \le c(x,y) \end{aligned}$$

for any $x, y \in U$. Thus the inequalities

$$\left| \int_U f_N(x)(P_1 - P_2)(\mathrm{d}x) \right| \le \boldsymbol{\zeta}_c(P_1, P_2) \le \mathring{\mu}_c(P_1, P_2) \tag{6.3.25}$$

follow from (6.3.16) and (6.3.17). Since $\alpha f_N(x) = \max(K(d(x,a)), K(2N))$ and (6.3.25) holds, then

$$\begin{aligned} \mathcal{T}_1 < &\left| \int_U K(d(x,a)) I\{d(x,a) > 2N\}(P_1 - P_2)(\mathrm{d}x) \right. \\ &\left. - \int_U K(2N) I\{d(x,a) \le 2N\}(P_1 - P_2)(\mathrm{d}x) \right| \\ &+ K(2N) \left| \int_U I\{d(x,a) \le 2N\}(P_1 - P_2)(\mathrm{d}x) \right| \end{aligned}$$

$$= \left|\int_U \alpha f_N(x)(P_1 - P_2)(\mathrm{d}x)\right| + K(2N)\left|\int_U I\{d(x,a) > 2N\}(P_1 - P_2)(\mathrm{d}x)\right|$$

$$\leq \alpha\overset{\circ}{\mu}_c + K(2N)\int I\{d(x,a) > 2N\}(P_1 + P_2)(\mathrm{d}x),$$

which proves the claim.

Claim 5.

$$A(P_1) := K(2N)\int_U I\{d(x,a) > 2N\}P_1(\mathrm{d}x) \leq 2\alpha\gamma\overset{\circ}{\mu}_c + 2\gamma\omega_2(1/N). \quad (6.3.26)$$

Proof of Claim 5. As in the proof of Claim 4, we choose an appropriate Lipschitz function. That is, write

$$g_N(x) = (1/(2\alpha\gamma))\min\{K(2N), K(2d(x, O(a, N)))\},$$

where $O(a, N) := \{x : d(x,a) < N\}$. Using (C1) and (C3),

$$|g_N(x) - g_N(y)| \leq (1/(2\alpha\gamma))|K(2d(x, O(a, N))) - K\{2d(y, O(a, N)))|$$

by (C1)

$$\leq (1/\gamma)|d(x, O(a, N))$$
$$-d(y, O(a, N))|k_0(2d(x, O(a, N)), 2d(y, O(a, N)))$$

by (C3)

$$\leq d(x, y)k_0(d(x, O(a, N)), d(y, O(a, N))) \leq c(x, y).$$

Hence

$$\left|\int g_N(P_1 - P_2)(\mathrm{d}x)\right| \leq \zeta_c \leq \overset{\circ}{\mu}_c. \quad (6.3.27)$$

Using (6.3.27) and the implications

$$d(x,a) > 2N \Rightarrow d(x, O(a, N)) > N \Rightarrow K(2d(x, O(a, N))) \geq K(2N)$$

we obtain the following chain of inequalities:

$$A(P_1) \leq 2\alpha\gamma\int g_N(x)P_1(\mathrm{d}x)$$

$$\leq 2\alpha\gamma\left|\int g_N(x)(P_1 - P_2)(\mathrm{d}x)\right| + 2\alpha\gamma\int_U g_N(x)P_2(\mathrm{d}x)$$

$$\leq 2\alpha\overset{\circ}{\mu}_c + \int K(2d(x, O(a, N)))I\{d(x,a) \geq N\}P_2(\mathrm{d}x),$$

$$\left(\text{by C3, } \frac{K(2t)}{K(t)} = \frac{2tk_0(2t,2t)}{tk_0(t,t)} \le 2\gamma\right)$$

$$\le 2\alpha\gamma\mathring{\mu}_c + 2\gamma \int K(d(x, O(a,N)))I\{d(x,a) \ge N\}P_2(\mathrm{d}x)$$

$$\le 2\alpha\gamma\mathring{\mu}_c + 2\gamma\omega_2(1/N), \tag{6.3.28}$$

which proves the claim.

For $A(P_2)$ [see (6.3.26)] we have the following estimate:

$$A(P_2) \le \int_U K(d(x,a))I\{d(x,a) > 2N\}P_2(\mathrm{d}x) \le \omega_2(1/N). \tag{6.3.29}$$

Summarizing (6.3.23), (6.3.24), (6.3.26), and (6.3.29) we obtain

$$\omega_1(1/2N) \le \alpha\mathring{\mu}_c + A(P_1) + A(P_2) + \omega_2(1/N)$$
$$\le (\alpha + 2\alpha\gamma)\mathring{\mu}_c + (2\gamma + 2)\omega_2(1/N)$$

for any $N > 0$, as desired. □

The next theorem shows that $\mathring{\mu}_c$-convergence implies convergence of the λ-moments.

Theorem 6.3.5.

$$D \le \alpha\mathring{\mu}_c. \tag{6.3.30}$$

Proof. By (C1), for any finite nonnegative measure Q with marginals P_1 and P_2 we have

$$D := \left|\int_U \lambda(x)(P_1 - P_2)(\mathrm{d}x)\right| = \left|\int_{U\times U} \lambda(x) - \lambda(y)Q(\mathrm{d}x,\mathrm{d}y)\right|$$
$$\le \int_{U\times U} \alpha|d(x,a) - d(y,a)|k_0(d(x,a),d(y,a))Q(\mathrm{d}x,\mathrm{d}y)$$
$$\le \alpha\int_{U\times U} c(x,y)Q(\mathrm{d}x,\mathrm{d}y)$$

which completes the proof of (6.3.30). □

Inequalities (6.3.9), (6.3.14), (6.3.22), and (6.3.30), described in Theorems 6.3.2–6.3.5, imply criteria for convergence, compactness, and uniformity in the spaces of probability measures $(\mathcal{P}(U), \widehat{\mu}_c)$ and $(\mathcal{P}(U), \mathring{\mu}_c)$ (see also the next section). Moreover, the estimates obtained for $\widehat{\mu}_c$ and $\mathring{\mu}_c$ may be viewed as quantitative results demanding conditions that are necessary and sufficient for

$\widehat{\mu}_c$-convergence and $\mathring{\mu}_c$-convergence. Note that, in general, quantitative results require assumptions additional to the set of necessary and sufficient conditions implying the qualitative results. The classic example is the central limit theorem, where the uniform convergence of the normalized sum of i.i.d. RVs can be at any low rate assuming only the existence of the second moment.

6.4 Convergence, Compactness, and Completeness in $(\mathcal{P}(U), \widehat{\mu}_c)$ and $(\mathcal{P}(U), \mathring{\mu}_c)$

In this section, we assume that the cost function c satisfies conditions (C1)–(C3) in the previous section and $\lambda(x) = K(d(x,a))$. We begin with the criterion for $\widehat{\mu}_c$- and $\mathring{\mu}_c$-convergence.

Theorem 6.4.1. *If P_n, and $P \in \mathcal{P}_\lambda(U)$, then the following statements are equivalent*

(A)

$$\widehat{\mu}_c(P_n, P) \to 0;$$

(B)

$$\mathring{\mu}_c(P_n, P) \to 0;$$

(C)

$$P_n \xrightarrow{\text{w}} P \ (P_n \textit{ converges weakly to } P) \textit{ and } \int \lambda \mathrm{d}(P_n - P) \to 0 \textit{ as } n \to \infty;$$

(D)

$$P_n \xrightarrow{\text{w}} P \ \textit{ and } \ \lim_{\varepsilon \to 0} \sup_n \omega_n(\varepsilon) = 0,$$

where $\omega_n(\varepsilon) := \omega(\varepsilon; P_n; \lambda) = \int \lambda(x)\{d(x,a) > 1/\varepsilon\} P_n(\mathrm{d}x)$.

Proof. From inequality (6.3.14) it is apparent that $A \Rightarrow B$ and $B \Rightarrow P_n \xrightarrow{\text{w}} P$. Using (6.3.30) we obtain that B implies $\int \lambda \mathrm{d}(P_n - P) \to 0$, and thus $B \Rightarrow C$. Now, let C hold.

Claim 6. $C \Rightarrow D$.

Proof of Claim 6. Choose a sequence $\varepsilon_1 > \varepsilon_2 > \cdots \to 0$ such that $P(d(x,a) = 1/\varepsilon_n) = 0$ for any $n = 1, 2, \ldots$. Then for fixed n

$$\int \lambda(x) I\{d(x,a) \le 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \to 0 \ \text{ as } \ k \to \infty$$

by Billingsley (1999, Theorem 5.1). Since $P \in \mathcal{P}_\lambda$, $\omega(\varepsilon_n) := \omega(\varepsilon_n; P; c) \to 0$ as $n \to \infty$, and hence

$$\begin{aligned}
\limsup_{k\to\infty} \omega_k(\varepsilon_n) &\leq \limsup_{k\to\infty} \left| \int \lambda(x)\{d(x,a) > 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \right| + \omega(\varepsilon_n) \\
&\leq \limsup_{k\to\infty} \left| \int \lambda(x)(P_k - P)(\mathrm{d}x) \right| \\
&\quad + \limsup_{k\to\infty} \left| \int \lambda(x) I\{d(x,a) \leq 1/\varepsilon_n\}(P_k - P)(\mathrm{d}x) \right| \\
&\quad + \omega(\varepsilon_n) \to 0 \text{ as } n \to \infty.
\end{aligned}$$

The last inequality and $P_k \in \mathcal{P}_\lambda$ imply $\lim_{\varepsilon\to 0} \sup_n \omega_n(\varepsilon) = 0$, and hence D holds.

The claim is proved.

Claim 7. $D \Rightarrow A$.

Proof of Claim 7. By Theorem 6.3.2,

$$\widehat{\mu}_c(P_n, P) \leq \boldsymbol{\pi}(P_n, P)[4K(1/\varepsilon_n) + \omega_n(1) + \omega(1) + 2k(1)] + 5\omega_n(\varepsilon_n) + 5\omega(\varepsilon_n),$$

where ω_n and ω are defined as in Claim 6 and, moreover, $\varepsilon_n > 0$ is such that

$$4K(1/\varepsilon_n) + \sup_{n\geq 1} \omega_n(1) + \omega(1) + 2k(1) \leq (\boldsymbol{\pi}(P_n, P))^{-1/2}.$$

Hence, using the last two inequalities we obtain

$$\widehat{\mu}_c(P_n, P) \leq \sqrt{\boldsymbol{\pi}(P_n, P)} + 5 \sup_{n\geq 1} \omega_n(\varepsilon_n) + 5\omega(\varepsilon_n),$$

and hence $D \Rightarrow A$, as we claimed. □

The Kantorovich–Rubinstein functional $\mathring{\mu}_c$ is a metric in $\mathcal{P}_\lambda(U)$, while $\widehat{\mu}_c$ is not a metric except for the case $c = d$ (see the discussion in the previous section). The next theorem establishes a criterion for $\mathring{\mu}_c$-relative compactness of sets of measures. Recall that a set $\mathcal{A} \subset \mathcal{P}_\lambda$ is said to be $\mathring{\mu}_c$*-relatively compact* if any sequence of measures in $\mathcal{A}$ has a $\mathring{\mu}_c$-convergent subsequence and the limit belongs to $\mathcal{P}_\lambda$. Recall that the set $\mathcal{A} \subset \mathcal{P}(U)$ is *weakly compact* if $\mathcal{A}$ is $\boldsymbol{\pi}$-relatively compact, i.e., any sequence of measures in $\mathcal{A}$ has a weakly ($\boldsymbol{\pi}$-) convergent subsequence.

Theorem 6.4.2. *The set $\mathcal{A} \subset \mathcal{P}_\lambda$ is $\mathring{\mu}_c$-relatively compact if and only if $\mathcal{A}$ is weakly compact and*

$$\lim_{\varepsilon\to 0} \sup_{P\in\mathcal{A}} \omega(\varepsilon; P; \lambda) = 0. \tag{6.4.1}$$

Proof. "If" part: If $\mathcal{A}$ is weakly compact, (6.4.1) holds and $\{P_n\}_{n\geq 1} \subset \mathcal{A}$, then we can choose a subsequence $\{P_{n'}\} \subset \{P_n\}$ that converges weakly to a probability measure P.

Claim 8. $P \in \mathcal{P}_\lambda$.

Proof of Claim 8. Let $0 < \alpha_1 < \alpha_2 < \cdots$, $\lim \alpha_n = \infty$ be such a sequence that $P(d(x,a) = \alpha_n) = 0$ for any $n \geq 1$. Then, by Billingsley (1999, Theorem 5.1) and (6.4.1),

$$\int \lambda(x) I\{d(x,a) \leq \alpha_{n'}\} P(\mathrm{d}x) = \lim_{n\to\infty} \int \lambda(x) I\{d(x,a) \leq \alpha_{n'}\} P_{n'}(\mathrm{d}x)$$
$$\leq \liminf_{n\to\infty} \int \lambda(x) P_{n'}(\mathrm{d}x) < \infty,$$

which proves the claim.

Claim 9.

$$\overset{\circ}{\mu}_c(P_{n'}, P) \to 0.$$

Proof of Claim 9. Using Theorem 6.3.2, Claim 8, and (6.4.1) we have, for any $\delta > 0$,

$$\overset{\circ}{\mu}_c(P_{n'}, P) \leq \widehat{\mu}_c(P_{n'}, P) \leq \boldsymbol{\pi}(P_{n'}, P)[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2K(1)]$$
$$+5 \sup_{n'} \omega(P_{n'}; \varepsilon; \lambda) + \omega(P; \varepsilon; \lambda)$$
$$\leq \boldsymbol{\pi}(P_{n'}, P)[4K(1/\varepsilon) + \omega_1(1) + \omega_2(1) + 2K(1)] + \delta$$

if $\varepsilon = \varepsilon(\delta)$ is small enough. Hence, by $\boldsymbol{\pi}(P_{n'}, P) \to 0$, we can choose $N = N(\delta)$ such that $\overset{\circ}{\mu}_c(P_{n'}, P) < 2\delta$ for any $n' \geq N$, as desired.

Claims 8 and 9 establish the "if" part of the theorem.

"Only if" part: If $\mathcal{A}$ is $\overset{\circ}{\mu}_c$-relatively compact and $\{P_n\} \subset \mathcal{A}$, then there exists a subsequence $\{P_{n'}\} \subset \{P_n\}$ that is convergent w.r.t. $\overset{\circ}{\mu}_c$, and let P be the limit. Hence, by Theorem 6.3.3, $\overset{\circ}{\mu}_c(P_n, P) \geq \beta \boldsymbol{\pi}^2(P_n, P) \to 0$, which demonstrates that the set $\mathcal{A}$ is weakly compact.

Further, if (6.4.1) is not valid, then there exists $\delta > 0$ and a sequence $\{P_n\}$ such that

$$\omega(1/n; P_n; \lambda) > \delta \quad \forall n \geq 1. \tag{6.4.2}$$

Let $\{P_{n'}\}$ be a $\overset{\circ}{\mu}_c$-convergent subsequence of $\{P_n\}$, and let $P \in \mathcal{P}_\lambda$ be the corresponding limit. By Theorem 6.3.4, $\omega(1/n'; P_{n'}; \lambda) \geq (2\gamma + 2)(\alpha \overset{\circ}{\mu}_c(P_{n'}, P) + \omega(1/n'; P; \lambda)) \to 0$ as $n' \to \infty$, which is in contradiction with (6.4.2). □

In the light of Theorem 6.4.1, we can now interpret Theorem 6.4.2 as a criterion for $\mathring{\mu}_c$-relative compactness of sets of measures in $\mathcal{P}$ by simply changing $\mathring{\mu}_c$ with $\widehat{\mu}_c$ in the formation of the last theorem.

The well-known Prokhorov theorem says that (U, d) *is a complete s.m.s; then the set of all laws on* U *is complete w.r.t. the Prokhorov metric* $\boldsymbol{\pi}$.[2] The next theorem is an analog of the Prokhorov theorem for the metric space $\mathcal{P}_\lambda, \mathring{\mu}_c$).

Theorem 6.4.3. *If* (U, d) *is a complete s.m.s., then* $(\mathcal{P}_\lambda(U), \mathring{\mu}_c)$ *is also complete.*

Proof. If $\{P_n\}$ is a $\mathring{\mu}_c$-fundamental sequence, then by Theorem 6.3.3, $\{P_n\}$ is also $\boldsymbol{\pi}$-fundamental, and hence there exists the weak limit $P \in \mathcal{P}(U)$.

Claim 10. $P \in \mathcal{P}_\lambda$.

Proof of Claim 10. Let $\varepsilon > 0$ and $\mathring{\mu}_c(P_n, P_m) \leq \varepsilon$ for any $n, m \geq n_\varepsilon$. Then, by Theorem 6.3.5, $\left|\int \lambda(x)(P_n - P_{n_\varepsilon})(\mathrm{d}x)\right| < \alpha\varepsilon$ for any $n > n_\varepsilon$; hence,

$$\sup_{n \geq n_\varepsilon} \int \lambda(x) P_n(\mathrm{d}x) < \alpha\varepsilon + \int \lambda(x) P_{n_\varepsilon}(\mathrm{d}x) < \infty.$$

Choose the sequence $0 < \alpha_1 < \alpha_2 < \cdots$, $\lim_{k\to\infty} \alpha_k = \infty$, such that $P(d(x,a) = \alpha_k) = 0$ for any $k > 1$. Then

$$\begin{aligned}
\int \lambda(x) I\{d(x,a) \leq \alpha_k\} P(\mathrm{d}x) &= \lim_{n\to\infty} \int \lambda(x) I\{d(x,a) \leq \alpha_k\} P_n(\mathrm{d}x) \\
&\leq \liminf_{n\to\infty} \int \lambda(x) P_n(\mathrm{d}x) \\
&\leq \sup_{n \geq n_\varepsilon} \int_U \lambda(x) P_n(\mathrm{d}x) < \infty.
\end{aligned}$$

Letting $k \to \infty$ the assertion follows.

Claim 11.

$$\mathring{\mu}_c(P_n, P) \to 0.$$

Proof of Claim 11. Since $\mathring{\mu}_c(P_n, P_{n_\varepsilon}) \leq \varepsilon$ for any $n \geq n_\varepsilon$, then, by Theorem 6.3.4,

$$\sup_{n \geq n_\varepsilon} \omega(\delta; P_n; \lambda) \leq 2(\gamma + 1)(\alpha\varepsilon + \omega(2\delta; P_{n_\varepsilon}; \lambda))$$

for any $\delta > 0$. The last inequality and Theorem 6.3.2 yield

$$\mathring{\mu}_c(P_n, P) \leq \widehat{\mu}_c(P_n, P) \leq \boldsymbol{\pi}(P_n, P)[4K(1/\delta)$$

[2] See, for example, Hennequin and Tortrat (1965) and Dudley (2002, Theorem 11.5.5).

$$+\sup_{n\geq n_\varepsilon}\omega(1;P_n;\lambda)+\omega(1;P;\lambda)+2K(1)]$$
$$+10(\gamma+1)(\alpha\varepsilon+\omega(2\delta;P_{n_\varepsilon};\lambda)+5\omega(\delta;P_{n_\varepsilon};\lambda)) \quad (6.4.3)$$

for any $n \geq n_\varepsilon$ and $\delta > 0$. Next, choose $\delta_n = \delta_{n,\varepsilon} > 0$ such that $\delta_n \to 0$ as $n \to \infty$ and

$$4K(1/\delta_n)+\sup_{n\geq n_\varepsilon}\omega(1;P_n;\lambda)+\omega(1;P;\lambda)+2k(1)\leq\frac{1}{(\boldsymbol{\pi}(P_n,P))^{1/2}}. \quad (6.4.4)$$

Combining (6.4.3) and (6.4.4) we have that $\overset{\circ}{\mu}_c(P_n,P) \leq$ const. ε for n large enough, which proves the claim. □

6.5 $\overset{\circ}{\mu}_c$- and $\widehat{\mu}_c$-Uniformity

In the previous section, we saw that $\overset{\circ}{\mu}_c$ and $\widehat{\mu}_c$ induce the same exact convergence in $\mathcal{P}_\lambda$. Here we would like to analyze the uniformity of $\overset{\circ}{\mu}_c$ and $\widehat{\mu}_c$-convergence. Namely, if for any $P_n, Q_n \in \mathcal{P}_\lambda$, the equivalence

$$\overset{\circ}{\mu}_c(P_n,Q_n) \iff \widehat{\mu}_c(P_n,Q_n)\to 0 \quad n\to\infty \quad (6.5.1)$$

holds. Obviously, $\Leftarrow$, by $\overset{\circ}{\mu}_c(P_n,Q_n) \leq \widehat{\mu}_c$. So, if

$$\widehat{\mu}_c(P,Q)\leq\phi(\overset{\circ}{\mu}_c(P,Q)) \qquad P,Q\in\mathcal{P}_\lambda \quad (6.5.2)$$

for a continuous nondecreasing function, $\phi(0) = 0$, then (6.5.1) holds.

Remark 6.5.1. Given two metrics, say μ and ν, in the space of measures, the equivalence of μ- and ν-convergence does not imply the existence of a continuous nondecreasing function ϕ vanishing at 0 and such that $\mu \leq \phi(\nu)$. For example, both the Lévy metric $\mathbf{L}$ [see (4.2.3)] and the Prokhorov metric $\boldsymbol{\pi}$ [see (3.3.18)] metrize the weak convergence in the space $\mathcal{P}(\mathbb{R})$. Suppose there exists ϕ such that

$$\boldsymbol{\pi}(X,Y)\leq\phi(\mathbf{L}(X,Y)) \quad (6.5.3)$$

for any real-valued r.v.s X and Y. (Recall our notation $\mu(X,Y) := \mu(\mathrm{Pr}_X,\mathrm{Pr}_Y)$ for any metric μ in the space of measures.) Then, by (4.2.4) and (3.3.23),

$$\mathbf{L}(X/\lambda,Y/\lambda)=\mathbf{L}_\lambda(X,Y)\to\boldsymbol{\rho}(X,Y) \quad \text{as} \quad \lambda\to 0 \quad (6.5.4)$$

and

$$\boldsymbol{\pi}(X/\lambda, Y/\lambda) = \boldsymbol{\pi}_\lambda(X,Y) \to \boldsymbol{\sigma}(X,Y) \quad \text{as} \quad \lambda \to 0, \tag{6.5.5}$$

where $\boldsymbol{\rho}$ is the Kolmogorov metric [see (4.2.6)] and $\boldsymbol{\sigma}$ is the total variation metric [see (3.3.13)]. Thus, (6.5.3)–(6.5.5) imply that $\boldsymbol{\sigma}(X,Y) \le \phi(\boldsymbol{\rho}(X,Y))$. The last inequality simply is, however, not true because in general $\boldsymbol{\rho}$-convergence does not yield $\boldsymbol{\sigma}$-convergence. [For example, if X_n is a random variable taking values k/n, $k = 1,\dots,n$ with probability $1/n$, then $\boldsymbol{\rho}(X_n, Y) \to 0$ where Y is a $(0,1)$-uniformly distributed random variable. On the other hand, $\boldsymbol{\sigma}(X_n, Y) = 1$.]

We are going to prove (6.5.2) for the special but important case where $\mathring{\mu}_c$ is the Fortet–Mourier metric on $\mathcal{P}_\lambda(\mathbb{R})$, i.e., $\mathring{\mu}_c(P,Q) = \boldsymbol{\zeta}(P,Q;\mathcal{G}^p)$ [see (4.4.34)]; in other words, for any $P, Q \in \mathcal{P}_\lambda$,

$$\mathring{\mu}_c(P,Q) = \sup\left\{\int f\,\mathrm{d}(P-Q) : f : \mathbb{R} \to \mathbb{R}, |f(x) - f(y)| \le c(x,y) \forall x, y \in \mathbb{R}\right\},$$

where

$$c(x,y) = |x-y| \max(1, |x|^{p-1}, |y|^{p-1}) \quad p \ge 1. \tag{6.5.6}$$

Since $\lambda(x) := 2\max(|x|, |x|^p)$, then $\mathcal{P}_\lambda(\mathbb{R})$ is the space of all laws on $\mathbb{R}$, with finite pth absolute moment.

Theorem 6.5.1. *If c is given by (6.5.6), then*

$$\widehat{\mu}_c(P,Q) \le p\mathring{\mu}_c(P,Q) \quad \forall P, Q \in \mathcal{P}_\lambda(\mathbb{R}). \tag{6.5.7}$$

Proof. Denote $h(t) = \max(1, |t|^{p-1})$, $t \in \mathbb{R}$, and $H(x) = \int_0^x h(t)\mathrm{d}t$, $x \in \mathbb{R}$. Let X and Y be real-valued RVs on a nonatomic probability space $(\Omega, \mathcal{A}, \Pr)$ with distributions P and Q, respectively. Theorem 5.5.1 gives us explicit representation of $\mathring{\mu}_c$, namely,

$$\mathring{\mu}_c(P,Q) = \int_{-\infty}^{\infty} h(t)|F_X(t) - F_Y(t)|\mathrm{d}t, \tag{6.5.8}$$

and thus

$$\mathring{\mu}_c(P,Q) = \int_{-\infty}^{\infty} |F_{H(X)}(x) - F_{H(Y)}(x)|\mathrm{d}x. \tag{6.5.9}$$

Claim 12. Let X and Y be real-valued RVs with distributions P and Q, respectively. Then

$$\mathring{\mu}_c(P,Q) = \inf\{E|H(\widetilde{X}) - H(\widetilde{Y})| : F_{\widetilde{X}} = F_X, F_{\widetilde{Y}} = F_Y\}. \tag{6.5.10}$$

Proof of Claim 12. Using the equality $\widehat{\mu}_d = \mathring{\mu}_d$ [see (6.2.3) and (5.5.5)] with $H(t) = t$ we have that

$$\mathring{\mu}_d(F,G) = \widehat{\mu}_d(F,G) = \inf\{E|X'-Y'| : F_{X'} = F, F_{Y'} = G\}$$
$$= \int_{-\infty}^{\infty} |F(x) - G(x)|\mathrm{d}x \tag{6.5.11}$$

for any DFs F and G. Hence, by (6.5.9)

$$\mathring{\mu}_c(P,Q) = \inf\{E|X'-Y'| : F_{X'} = F_{H(X)}, F_{Y'} = F_{H(Y)}\}$$
$$= \inf\{E|H(\widetilde{X}) - H(\widetilde{Y})| : F_{\widetilde{X}} = F_X, F_{\widetilde{Y}} = F_Y\}$$

which proves the claim.

Next we use Theorem 2.7.2, which claims that on a nonatomic probability space, the class of all joint distributions $\Pr_{X,Y}$ coincides with the class of all probability Borel measures on $\mathbb{R}^2$. This implies

$$\widehat{\mu}_c(P,Q) = \inf\{Ec(\widetilde{X},\widetilde{Y}) : F_{\widetilde{X}} = F_X, F_{\widetilde{Y}} = F_Y\}. \tag{6.5.12}$$

Claim 13. For any $x, y \in \mathbb{R}$, $c(x,y) \le p|H(x) - H(y)|$.

Proof of Claim 13.

(a) Let $y > x > 0$. Then

$$c(x,y) = (y-x)h(y) = yh(y) - xh(y) \le yh(y) - xh(x)$$
$$\le (H(y) - H(x)) \sup_{y>x>0} \frac{yh(y) - xh(x)}{H(y) - H(x)}.$$

Since $H(t)$ is a strictly increasing continuous function,

$$B := \sup_{y>x>0} \frac{yh(y) - xh(x)}{H(y) - H(x)} = \sup_{t>s>0} \frac{f(t) - f(s)}{t-s},$$

where $f(t) := H^{-1}(t)h(H^{-1}(t))$ and H^{-1} is a function inverse to H; hence, $B = \operatorname{ess\,sup}_t |f'(t)| \le p$.

(b) Let $y > 0 > x > -y$. Then $c(x,y) = |x-y|h(y) = (y + (-x))h(y) = yh(y) + (-x)h(|x|) + ((-x)h(y) - (-x)h(|x|)) \le yh(y) + (-x)h(|x|)$. Since

$$th(t) = \begin{cases} t & \text{if } t \le 1, \\ t^p & \text{if } t \ge 1, \end{cases} \qquad H(t) = \begin{cases} t & \text{if } 0 < t \le 1, \\ \dfrac{p-1}{p} + \dfrac{1}{p}t^p & \text{if } t \ge 1, \end{cases}$$

then $yh(y) + (-x)h(|x|) \le p(H(y) + H(-x)) = p(H(y) - H(x))$. By symmetry, the other cases are reduced to (a) or (b). The claim is shown. Now, (6.5.7) is a consequence of Claims 12, 13, and (6.5.12). □

6.6 Generalized Kantorovich and Kantorovich–Rubinstein Functionals

In this section, we consider a generalization of the Kantorovich-type functionals $\widehat{\mu}_c$ and $\overset{\circ}{\mu}_c$ [see (5.2.16) and (5.2.17)].

Let $U = (U, d)$ be an s.m.s. and $\mathcal{M}(U \times U)$ the space of all nonnegative Borel measures on the Cartesian product $U \times U$. For any probability measures P_1 and P_2 define the sets $\mathcal{P}^{(P_1,P_2)}$ and $\mathcal{Q}^{(P_1,P_2)}$ as in Sect. 5.2 [see (5.2.2) and (5.2.13)].

Let $\Lambda : \mathcal{M}(U \times U) \to [0, \infty]$ satisfy the conditions

1. $\Lambda(\alpha P) = \alpha\Lambda(P) \quad \forall \alpha \geq 0$,
2. $\Lambda(P + Q) \leq \Lambda(P) + \Lambda(Q) \quad \forall P$ and Q in $\mathcal{M}(U \times U)$.

We introduce the *generalized Kantorovich functional*

$$\widehat{\Lambda}(P_1, P_2) := \inf\{\Lambda(P) : P \in \mathcal{P}^{(P_1,P_2)}\} \tag{6.6.1}$$

and the *generalized Kantorovich–Rubinstein functional*

$$\overset{\circ}{\Lambda}(P_1, P_2) := \inf\{\Lambda(P) : P \in \mathcal{Q}^{(P_1,P_2)}\}. \tag{6.6.2}$$

Example 6.6.1. The Kantorovich metric[3]

$$\ell_1(P_1, P_2) := \sup\left\{\left|\int f \mathrm{d}(P_1 - P_2)\right| : f : U \to \mathbb{R}, |f(x) - f(y)| \leq d(x, y), x, y \in U\right\}$$

in the space of measures P with finite "first moment," $\int d(x, a)P(\mathrm{d}x) < \infty$, has the dual representations $\ell_1(P_1, P_2) = \overset{\circ}{\Lambda}(P_1, P_2) = \widehat{\Lambda}(P_1, P_2)$, where

$$\Lambda(P) = \Lambda_1(P) := \int_{U\times U} d(x, y)P(\mathrm{d}x, \mathrm{d}y). \tag{6.6.3}$$

Example 6.6.2. Let $U = \mathbb{R}$, $d(x, y) = |x - y|$. Then

$$\ell_1(P_1, P_2) = \int_{\mathbb{R}} |F_1(t) - F_2(t)|\mathrm{d}t,$$

[3] See Example 3.3.2 in Chap. 3.

where F_i is the DF of P_i and

$$\begin{aligned}\Lambda_1(P) &= \int_{\mathbb{R}} (\Pr(X \le t < Y) + \Pr(Y \le t < X))\mathrm{d}t \\ &= \int_{\mathbb{R}} \Pr(X \le t) + \Pr(Y \le t) - 2\Pr(\max(X,Y) \le t)\mathrm{d}t \\ &= E(2\max(X,Y) - X - Y) = E|X-Y|\end{aligned}$$

for RVs X and Y with $\Pr_{X,Y} = P$. We generalize (6.6.3) as follows: for any $1 \le p \le \infty$, define

$$\Lambda(P) := \Lambda_p(P) := \begin{cases} \left\{\int_{\mathbb{R}} \left[\int_{\mathbb{R}} c_t(x,y)P(\mathrm{d}x,\mathrm{d}y)\right]^p \lambda(\mathrm{d}t)\right\}^{1/p} & 1 \le p < \infty \\ \operatorname{ess\,sup}_\lambda \int_{\mathbb{R}^2} c_t(x,y)P(\mathrm{d}x,\mathrm{d}y) & \\ := \inf\left\{\varepsilon > 0 : \lambda\left\{t : \int_{\mathbb{R}^2} c_t \mathrm{d}P > \varepsilon\right\} = 0\right\} & p = \infty, \end{cases} \tag{6.6.4}$$

where c_t $(t \in \mathbb{R})$ is the following semimetric in $\mathbb{R}$

$$c_t(x,y) := I\{x \le t \le y\} + I\{y \le t \le x\} \forall x, y \in \mathbb{R}, \tag{6.6.5}$$

and $\lambda(\cdot)$ is a nonnegative measure on $\mathbb{R}$. In the space $\mathfrak{X} = \mathfrak{X}(\mathbb{R})$ of all real-valued RVs on a nonatomic probability space $(\Omega, \mathcal{A}, \Pr)$, the minimal metric w.r.t. Λ is given by

$$\widehat{\Lambda}_p(P_1,P_2) = \begin{cases} \inf\left\{\left[\int_{\mathbb{R}} \phi_t^p(X,Y)\lambda(\mathrm{d}t)\right]^{1/p} : X, Y \in \mathfrak{X}, \Pr_X = P_1, \Pr_Y = P_2\right\} & \\ & 1 \le p < \infty \\ \inf\left\{\sup_{t\in\mathbb{R}} \phi_t(X,Y) : X,Y \in \mathfrak{X}, \Pr_X = P_1, \Pr_Y = P_2\right\} & \\ & p = \infty. \end{cases} \tag{6.6.6}$$

Similarly, the minimal norm with respect to Λ is

$$\overset{\circ}{\Lambda}_p(P_1,P_2) = \begin{cases} \inf\left\{\alpha\left[\int_{\mathbb{R}} \phi_t^p(X,Y)\lambda(\mathrm{d}t)\right]^{1/p} : \alpha > 0, \quad X,Y \in \mathfrak{X}, \right. \\ \qquad\qquad \left. \alpha(\Pr_X - \Pr_Y) = P_1 - P_2 \right\} \quad \text{if } p < \infty \\ \inf\left\{\alpha \sup_\lambda \phi_t(X,Y) : \alpha > 0,\ X,Y \in \mathfrak{X}, \right. \\ \qquad\qquad \left. \alpha(\Pr_X - \Pr_Y) = P_1 - P_2 \right\} \quad \text{if } p = \infty, \end{cases} \tag{6.6.7}$$

where in (6.6.6) and (6.6.7)

$$\phi_t(X,Y) := \Pr(X \le t < Y) + \Pr(Y \le t < X). \tag{6.6.8}$$

The next theorem gives the explicit form of $\widehat{\Lambda}_p$ and $\overset{\circ}{\Lambda}_p$.

Theorem 6.6.1. *Let F_i be the DF of P_i ($i = 1,2$). Then*

$$\widehat{\Lambda}_p(P_1,P_2) = \overset{\circ}{\Lambda}_p(P_1,P_2) = \boldsymbol{\lambda}_p(F_1,F_2), \tag{6.6.9}$$

where

$$\boldsymbol{\lambda}_p(F_1,F_2) = \begin{cases} \left(\int_{\mathbb{R}} |F_1(t) - F_2(t)|^p \lambda(\mathrm{d}t)\right)^{1/p} & 1 \le p < \infty \\ \operatorname{ess\,sup}_\lambda |F_1 - F_2| = \inf\{\varepsilon > 0 : \lambda(t : |F_1(t) - F_2(t)| > \varepsilon) = 0\} & p = \infty. \end{cases} \tag{6.6.10}$$

Claim 14. $\boldsymbol{\lambda}_p(F_1,F_2) \le \overset{\circ}{\Lambda}_p(P_1,P_2)$.

Proof of Claim 14. Let $P \in \mathcal{Q}^{(P_1,P_2)}$. Then in view of Remark 2.7.2 in Chap. 2, there exist $\alpha > 0$, $X \in \mathfrak{X}$, $Y \in \mathfrak{X}$, such that $\alpha \Pr_{X,Y} = P$ and $\alpha(F_X - F_Y) = F_1 - F_2$; thus

$$\begin{aligned} |F_1(x) - F_2(x)| &= \alpha|F_X(t) - F_Y(t)| \\ &= \alpha[\max(F_X(t) - F_Y(t), 0) + \max(F_Y(t) - F_X(t), 0)] \\ &\le \alpha\phi_t(X,Y). \end{aligned} \tag{6.6.11}$$

By (6.6.7) and (6.6.11), it follows that $\boldsymbol{\lambda}_p(F_1,F_2) \le \overset{\circ}{\Lambda}_p(P_1,P_2)$, as desired.

Further

$$\overset{\circ}{\Lambda}_p(P_1, P_2) \le \widehat{\Lambda}_p(P_1, P_2) \tag{6.6.12}$$

by the representations (6.6.6) and (6.6.7).

Claim 15.

$$\widehat{\Lambda}_p(P_1, P_2) \le \boldsymbol{\lambda}_p(F_1, F_2).$$

Proof of claim 15. Let $\widetilde{X} := F_1^{-1}(V), \widetilde{Y} := F_2^{-1}(V)$, where F_i^{-1} is the generalized inverse to the DF F_i [see (3.3.16) in Chap. 3] and V is a $(0, 1)$-uniformly distributed RV. Then $F_{\widetilde{X},\widetilde{Y}}(t, s) = \min(F_1(t), F_2(s))$ for all $t, s \in \mathbb{R}$. Hence, $\phi_t(\widetilde{X}, \widetilde{Y}) = |F_1(t) - F_2(t)|$, which proves the claim by using (6.6.6) and (6.6.7).

Combining Claims 14, 15, and (6.6.12) we obtain (6.6.9). □

Problem 6.6.1. In general, dual and explicit solutions of $\widehat{\Lambda}_p$ and $\overset{\circ}{\Lambda}_p$ in (6.6.1) and (6.6.2) are not known.

References

Billingsley P (1999) Convergence of probability measures, 2nd edn. Wiley, New York
Dudley RM (2002) Real analysis and probability, 2nd edn. Cambridge University Press, New York
Hennequin PL, Tortrat A (1965) Théorie des probabilités et quelques applications. Masson, Paris
Levin VL (1975) On the problem of mass transfer. Sov Math Dokl 16:1349–1353
Neveu J, Dudley RM (1980) On Kantorovich–Rubinstein theorems (transcript)