Wasserstein-metric

The ‘Wasserstein-metric’ has a colourful history with several quite different fields of applications. It also has various historical sources.

The term ‘Vasershtein distance’ appeared for the first time in Dobrushin’s paper (1970) on the existence and uniqueness of random fields. For probability measures P, Q on a metric space (U, d) Vasershtein (1969) had introduced the metric $\ell_1(P,Q) = \inf\{E\,\mathrm{d}(X,Y)\}$ where the inf is w.r.t. all random variables X, Y with distributions P, Q. His work was very influential in ergodic theory in connection with generalizations of Ornsteins (1973) isomorphism theorem (see Gray, Neuhoff and Shields (1975)). In the english literature the russian name was pronounced typically as Wasserstein and the notation $W(P,Q)$ is common for $\ell_1(P,Q)$.

The minimal L_1-metric ℓ_1 had been introduced and investigated already in 1940 by Kantorovich for compact metric spaces. This work was motivated by the classical Monge-Transportation Problem. Kantorovich later on generalized the transportation distance by general cost functionals; the special case $c(x,y) = \mathrm{d}^p(x,y)$ leads to the minimal L_p-metric $\ell_p(P,Q) = \inf\{\|\mathrm{d}(X,Y)\|_p\}$, $\|\ \|_p$ the usual L_p-norm. The famous Kantorovich-Rubinstein 1958 theorem gives a dual representation of ℓ_1 in terms of a Lipschitz-metric:

$$\ell_1(P,Q) = \sup\{\int f\,\mathrm{d}(P-Q); \mathrm{Lip} f \leq 1\}.$$

From this point of view the notion Kantorovich-metric or minimal L_1- (minimal L_p-) metric seems to be also appropriate historically.

In fact in 1914 Gini introduced in his introduction of a ‘simple index of dissimilarity’ the ℓ_1 metric in a discrete setting on the real line and Salvemini 1943 (in the discrete case) and Dall’ Aglio 1956 (in the general case) proved the basic representation

$$\ell_p^p(P,Q) = \int_0^1 |F^{-1}(u) - G^{-1}(u)|^p\,\mathrm{d}u, \quad p \geq 1$$

where F, G are the distribution functions of P, Q. Gini had given this formula for empirical distributions and $p = 1, 2$. This influential work initiated a lot of work on measures with given marginals in the Italian School of probability while Fréchet (1957) explicitly dealt with metric properties of these distances.

Mallows 1972 introduced the ℓ_2-metric in a statistical context. He used its properties for proving a central limit theorem and proved the representation above. Based on Mallows work, Bickel and Freedman (1981) described topological properties and investigated applications to statistical problems as the bootstrap. They introduced the notion Mallows metric for ℓ_2. This notion is used mainly in the statistics literature and in some literature on algorithms. So the minimal L_p-metric ℓ_p was invented historically several times from different perspectives. Maybe historically the notion Gini – Dall’ Aglio – Kantorovich – Vasersthein – Mallows metric would be correct for this class of metrics. For simplicity reasons and in order to not omit justified credits the author of this review prefers the notion minimal L_p-metric termed by $\ell_p(P,Q)$.

Literatur

[1] Bickel, P.J.; Freedman, D.A. (1981), Some asymptotic theory for the bootstrap, *Ann. Stat.*, 9:1196–1217.

[2] Dall' Aglio, G. (1956) Sugli estremi dei momenti delle funzioni di ripartizione doppia, *Ann. Scuola Normale Superiore Di Pisa*, Cl. Sci 3(1):33–74.

[3] Dobrushin, R.L. (1970), Prescribing a system of random variables by conditional distributions, *Theor. Prob. Appl.*, 15:458-486.

[4] Fréchet M. (1957), Les tableaux de corréllation dont les marges sont données, *Ann. Univ. Lyon S. A*, 20:13–31.

[5] Gray, R.M.; Neuhoff, D.L.; Shields, P.C. (1975), A generalization to Ornstein's d-distance with applications to information theory, *Ann. Prob.*, 3:315-328.

[6] Kantorovich, L.V. (1940), On one effective method of solving certain classes of extremal problems, *Dokl. Akad. Nauk, USSR*, 28:212–215.

[7] Kantorovich, L.V.; Rubinstein, G.Sh. (1958), On the space of completely additive functions, *Vestnic Leningrad Univ., Ser. Mat. Mekh. i Astron.*, 13(7):52–59. (In Russian).

[8] Mallows, C.L. (1972), A note on asymptotic joint normality, *Ann. Math.l Stat.*, 43:508–515.

[9] Salvemini, T. (1943), *Sul calcolo degli indici di concordanza tra due caratteri quantitativi*, Atti della VI Riunione della Soc. Ital. di Statistica, Roma.

[10] Wasserstein, L.N. (1969), Markov processes over denumerable products of spaces describing large systems of automata, *Prob. Inf. Transmission*, 5:47–52.

AMS 1980 Subject Classification : 60 E 99

Prof. Dr. Ludger Rüschendorf
Institut für Mathematische Stochastik
Eckerstr. 1
79104 FREIBURG
Germany