'MATH+ECON+CODE' MASTERCLASS ON MATCHING MODELS, OPTIMAL TRANSPORT AND APPLICATIONS

Alfred Galichon (New York University)

Spring 2019
Wednesday: "Optimal transport II"
Block 7. Continuous multivariate matching

Learning objectives: block 7

- Existence of potentials in the quadratic case
- Knott-Smith criterion and Brenier's and McCann's theorems
- Entropic regularization
- The iterated proportional fitting procedure

- [OTME], Ch. 6
- [TOT] Villani (2003). *Topics in Optimal Transportation*. AMS. Ch. 1 and 2.

Section 1

Theory

- As a consequence of the previous lecture, we have seen that if P is a continuous distribution over $\mathbb{R}^d$ (distribution of the inhabitants' locations), and if $Q = \sum_{k=1}^{M} q_k \delta_{y_k}$ is a discrete distribution over $\mathbb{R}^d$ (distribution of the fountains' locations), then there exists a mapping T such that $T\#P = Q$, that is

$$Y = T(X)$$

  where:
  - $X \sim P$ and $Y \sim Q$, and $T(x)$ is the location of the fountain assigned to the inhabitant at x.
  - $T(x) = \nabla u(x)$, where u is a convex function which is given by $u(x) = \max_k \{x^\intercal y_k - v_k\}$.
- Note the connection with Becker's model: when the dimension $d = 1$, T is piecewise constant and nondecreasing (positive assortative matching).
- In this lecture, we shall generalize these results to the case when Q is a general distribution (not necessarily discrete). P will have a density, and the support of P and Q will be assumed to be convex.

- Assume that $\mathcal{X}$ and $\mathcal{Y}$ are convex subsets of $\mathbb{R}^d$, and that

$$\Phi(x, y) = x^\intercal y.$$

and P and Q are two probability distributions on $\mathcal{X}$ and $\mathcal{Y}$.

- The Monge-Kantorovich theorem provides assumptions under which the value of the primal problem

$$\mathcal{W} = \sup_{\pi \in \mathcal{M}(P,Q)} \mathbb{E}_\pi \left[X^\intercal Y \right] \tag{1}$$

coincides with the value of the dual

$$\mathcal{W} = \inf_{u(x)+v(y) \geq x^\intercal y} \mathbb{E}_P \left[u(X) \right] + \mathbb{E}_Q \left[v(Y) \right]. \tag{2}$$

- Note, however, that the M-K theorem requires Φ to be bounded by above, which is not the case of $\Phi(x, y) = x^\intercal y$ unless we assume P and Q have bounded support. We could alternatively work with $\Phi(x, y) = -|x-y|^2/2$, in which case we should assume that P and Q have finite second moment and replace $u(x)$ by $u(x) + |x|^2/2$, and v by a similar quantity. We shall assume away these concerns for now.

The following result ensures that u and v exist as soon as P and Q have finite second moments.

THEOREM

If P and Q have finite second moments, then there exists a pair (u, v) solution to the dual Monge-Kantorovich problem

$$\inf_{u(x)+v(y)\geq x^\intercal y} \mathbb{E}_P\left[u\left(X\right)\right] + \mathbb{E}_Q\left[v\left(Y\right)\right].$$

See theorem 2.9 in [TOT].

- Assume that a dual minimizer (u, v) exists; if needed, redefine u and v so that they take value $+\infty$ outside of the support of P and Q, assumed to be convex. As argued, u and v are then related by

$$v(y) = \max_{x \in \mathbb{R}^d} \left\{ x^\intercal y - u(x) \right\} \tag{3}$$

$$u(x) = \max_{y \in \mathbb{R}^d} \left\{ x^\intercal y - v(y) \right\} \tag{4}$$

hence we see immediately that if (u, v) is a solution to the dual problem, then u and v are convex functions. Further, the expression of v as a function of u is the same as the expression of u as a function of v.

We want to relate the solutions to the primal and dual problems. Recall that in the finite-dimensional case, where the primal and the dual problems are related by a complementary slackness condition. In the present case, let $(X, Y) \sim \pi$ be a solution to the primal problem, and (u, u^*) be a solution to the dual problem. Then almost surely X and Y are willing to match, which, by the previous discussion, implies that

$$u(X) + u^*(Y) = X^\intercal Y, \tag{5}$$

that is, the support of π is included in the set $\{(x, y) : u(x) + u^*(y) = x^\intercal y\}$. This condition appears as the correct generalization of the complementary slackness condition in the finite-dimensional case. Without surprise, taking the expectation with respect to π of equality (5) yields the equality between the value of the dual problem on the left-hand side, and the value of the primal problem on the right-hand side.

The following statement provides a generalization of the complementary slackness condition in finite dimension.

THEOREM (KNOTT-SMITH)

Let $\pi \in \mathcal{M}(P, Q)$ and u be a convex function. Then π and (u, u^*) are respective solutions to the primal and the dual Monge-Kantorovich problems if and only if

$$u(x) + u^*(y) = x^\intercal y \text{ holds for } \pi\text{-almost all } (x, y). \tag{6}$$

PROOF.

Assume that (6) holds. Then, note that (u, u^*) satisfies the constraints of the dual; further, taking expectation with respect to π yields $\mathbb{E}_P[u(X)] + \mathbb{E}_Q[v(Y)] = \mathbb{E}_\pi[X^\intercal Y]$, which implies that π is an optimal primal solution and (u, u^*) is an optimal dual solution. Conversely, assume that π is an optimal primal solution and (u, u^*) is an optimal dual solution. Then $\mathbb{E}_\pi[u(X) + u^*(Y) - X^\intercal Y] = 0$; but $(x, y) \to u(x) + u^*(y) - x^\intercal y$ is nonnegative, thus (6) holds. □

THEOREM (BRENIER)

Assume that P and Q have finite second moments, and P has a density. Then the solution $(X, Y) \sim \pi \in \mathcal{M}(P, Q)$ to the primal problem is represented by

$$Y = \nabla u(X)$$

where (u, u^*) is a solution to the dual problem. Such u is unique up to a constant.

Intuition of the proof: if u is differentiable, then y is matched with x that maximizes $\{x^\intercal y - u(x)\}$ over $x \in \mathbb{R}^d$. By first order conditions, such x satisfy $\nabla u(x) = y$. It turns out, however, that differentiability is not a serious concern (at least, almost never).

While we evoked the case when the Kantorovich potentials u and v are differentiable, there is no a-priori guarantee that they are so. However, an important result in Analysis called Rademacher's theorem implies that the set of non-differentiable points of a convex function is of zero Lebesgue measure, and hence can be ignored for practical purposes as soon as P is continuous. Thus the Monge map solution, $T(x)$, can be defined as $T(x) = \nabla u(x)$ wherever the latter quantity exists, and $T(x)$ can be defined arbitrarily elsewhere, without affecting the distributional properties of $T(X)$.

The previous result allows to provide a representation of a large class of probability distributions Q over $\mathbb{R}^d$ as the probability distribution of $\nabla u(X)$, for X with a fixed distribution P. There is however a limitation, in the sense that it requires that Q has finite second moments, which is needed to interpret u as entering the solution to the dual problem. Fortunately, McCann's theorem addresses this issue:

Theorem (McCann)

Assume that P and Q are probability distributions such that P has a density. Then there is a unique (up to a constant) function u such that

$$Y = \nabla u(X)$$

holds almost surely with $X \sim P$ and $Y \sim Q$.

Section 2

Applications and examples

- Brenier-McCann's theorem allows us to describe a model of the heterosexual marriage market where P and Q are contiuous distributions that stand for the distributions of the men and the women's characteristics, and the surplus function is

$$\Phi\left(x,y\right) = x^{\intercal} A y$$

i.e. $\Phi\left(x,y\right) = \sum_{1\leq k,l\leq d} A_{kl} x_k y_l$, that is A_{kl} stand for the "affinity" between characteristics x_k of the man and y_l of the woman. Recall this model is equivalent to $\Phi\left(x,y\right) = \sum_{1\leq k,l\leq d} A_{kl} \left|x_k - y_l\right|^2/2$.

EXERCISE

Assume A is invertible. Show that the optimal matching can be given by $y = T\left(x\right)$ where $T = A^{-1}\nabla u\left(x\right)$, where u is a convex function. Characterize u as the solution of a minimization problem.

- Consider the a particular case of the previous model when $d = 2$ and A is diagonal, i.e. $A = diag\left(\lambda_1, \lambda_2\right)$. Then $\Phi^{\lambda}\left(x, y\right) = \lambda_1 x_1 y_1 + \lambda_2 x_2 y_2$. Assume x_1 and y_1 are interpreted as the man and woman's income, and x_2 and y_2 are interpreted as the man and woman's education.

EXERCISE

Consider $\mathcal{C} = \left\{\left(Cov\left(X_1, Y_1\right), Cov\left(X_2, Y_2\right)\right), \pi \in \mathcal{M}\left(P, Q\right)\right\}$.
(a) Show that $\mathcal{C}$ *is a convex set.*
(b) Show that $(0, 0) \in \mathcal{C}$ *and interpret this point.*
(c) Show that the boundary points of $\mathcal{C}$ *are the solution to an optimal transport problem with some surplus* Φ^{λ}.
(d) What can be said of $\max\left\{C_1 : \left(C_1, C_2\right) \in \mathcal{C}\right\}$ *and* $\max\left\{C_2 : \left(C_1, C_2\right) \in \mathcal{C}\right\}$*?*
(e) Characterize the solution to the optimal transport problem with surplus Φ^{λ} *when* $\lambda = (1, \varepsilon)$*, for* $\varepsilon \to 0$.

- When $P = \mathcal{N}(0, \Sigma_X)$ and $Q = \mathcal{N}(0, \Sigma_Y)$ and $\Phi(x, y) = x^\intercal A y$, and Σ_X, Σ_Y and A are invertible, one can get a solution in closed form.

EXERCISE

(a) Consider first the case when $\Sigma_X = I_d$ and $A = I_d$. Then show that the optimal transport map is given by

$$T(x) = \Sigma_Y^{1/2} x.$$

(b) Using the result in (a), show that when $A = I_d$, but with general Σ_X and Σ_Y, the solution is obtained by

$$T(x) = \Sigma_X^{-1/2} \left(\Sigma_X^{1/2} \Sigma_Y \Sigma_X^{1/2} \right)^{1/2} \Sigma_X^{-1/2} x.$$

(c) Using the result in (b), show that when A, Σ_X and Σ_Y are general invertible matrices, the solution is obtained by

$$T(x) = A^{-1} \Sigma_X^{-1/2} \left(\Sigma_X^{1/2} A^\intercal \Sigma_Y A \Sigma_X^{1/2} \right)^{1/2} \Sigma_X^{-1/2} x.$$