A Monge-Kantorovich approach to multivariate quantile regression

Guillaume Carlier [a] .

Joint work with Victor Chernozhukov (MIT) and Alfred Galichon (Sciences Po, Paris),

Conference on Optimization, Transportation and Equilibrium in Economics,

Fields Institute, Toronto, september 2014.

[a]CEREMADE, Université Paris Dauphine

Econometricians are typically interested in modeling the dependence between a certain variable Y and explanatory variables X. Standard linear regression estimates the conditional expectation $\mathbf{E}(Y|X = x)$ assuming a linear in x form by least squares. There are many reasons to be rather interested in modelling conditional median (or other quantiles) rather than conditional means, for instance quantiles are more robust to outliers than means and the whole conditional quantile function gives the whole conditional distribution not only its mean...

Many applications in economics: wage structure, program evaluation, demand analysis, income inequality, finance, and other areas (ecology, biometrics).

Quantile regression as pioneered by Koenker and Bassett (1978) provides a very convenient and powerful tool to estimate conditional quantiles, assuming a linear form in the explanatory variables. Quantile regression relies very much on convex optimization (with an L^1-criterion instead of quadratic programming used for linear regression).

However, one strong limitation of the method is that Y should be univariate (what is the median of a multivariate variable?).

Aim of this talk:

- recall the standard univariate quantile regression approach, relate it to problems of optimal transport (OT) type, clarify the case where the conditional quantile is not linear in the explanatory variables,
- extend the analysis to the multivariate case by means of optimal transport arguments.

Outline

① Classical quantile regression: old and new

- Quantiles, conditional quantiles
- Quantiles and polar factorizations,
- Specified and quasi specified quantile regression
- General case

② Multivariate quantile regression

- Multivariate quantiles
- Specified case
- General case and duality
- Quantile regression as optimality conditions

Quantiles, conditional quantiles

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be some nonatomic probability space and Y be some (univariate) random variable defined on this space. Denoting by F_Y the distribution function of Y:

$$F_Y(\alpha) := \mathbf{P}(Y \leq \alpha), \ \forall \alpha \in \mathbf{R}$$

the *quantile* function of Y, $Q_Y = F_Y^{-1}$ is the generalized inverse of F_Y given by:

$$Q_Y(t) := \inf\{\alpha \in \mathbf{R} \ : \ F_Y(\alpha) > t\} \text{ for all } t \in (0, 1). \tag{1}$$

Two well-known facts about quantiles:

- $\alpha = Q_Y(t)$ is a solution of the convex minimization problem

$$\min_{\alpha}\{\mathbf{E}((Y-\alpha)_+) + \alpha(1-t)\} \tag{2}$$

- there exists a uniformly distributed random variable U such that $Y = Q_Y(U)$ (polar factorization). Moreover, among uniformly distributed random variables, U is maximally correlated to Y in the sense that it solves

$$\max\{\mathbf{E}(VY),\ \mathrm{Law}(V) = \mu\} \tag{3}$$

where $\mu := \mathrm{uniform}([0,1])$ is the uniform measure on $[0,1]$.

gives two different approaches to study or estimate quantiles:

- the *local* or "t by t" approach which consists, for a fixed probability level t, in using directly formula (1) or the minimization problem (2), this can be done very efficiently in practice but has the disadvantage of forgetting the fundamental global property of the quantile function: it should be monotone in t,

- the global approach (or polar factorization approach), where quantiles of Y are defined as all nondecreasing functions Q for which on one can write $Y = Q(U)$ with U uniformly distributed; in this approach, one rather tries to recover directly the whole monotone function Q (or the uniform variable U that is maximally correlated to Y), in this global approach, one should rather use the OT problem (3).

Conditional quantiles Assume now that, in addition to the random variable Y, we are also given a random vector $X \in \mathbf{R}^N$ which we may think of as being a list of explanatory variables for Y. We are therefore interested in the dependence between Y and X and in particular the conditional quantiles. In the sequel we shall denote by ν the joint law of (X, Y), $\nu := \mathrm{Law}(X, Y)$ and assume that ν is compactly supported on $\mathbf{R}^{N+1}$ (i.e. X and Y are bounded). We shall also denote by m the first marginal of ν i.e. $m := \Pi_{X\#}\nu = \mathrm{Law}(X)$. We shall denote by $F(x, y)$ the conditional cdf:

$$F(x, y) := \mathbf{P}(Y \leq y | X = x)$$

and $Q(x, t)$ the conditional quantile

$$Q(x, t) := \inf\{\alpha \in \mathbf{R} \ : \ F(x, \alpha) > t\}.$$

For the sake of simplicity we shall also assume that:

- for m-a.e. x, $t \mapsto Q(x,t)$ is continuous and increasing (so that for m-a.e. x, identities $Q(x, F(x,y)) = y$ and $F(x, Q(x,t)) = t$ hold for every y and every t)
- the law of (X,Y) does not charge nonvertical hyperplanes i.e. for every $(\alpha,\beta) \in \mathbf{R}^{1+N}$, $\mathbf{P}(Y = \alpha + \beta \cdot X) = 0$.

Finally we denote by ν^x the conditional probability of Y given $X = x$ so that $\nu = m \otimes \nu^x$.

Quantiles and polar factorizations

Let us define the random variable $U := F(X, Y)$, then by construction:

$$\begin{aligned}\mathbf{P}(U < t|X = x) &= \mathbf{P}(F(x, Y) < t|X = x) = \mathbf{P}(Y < Q(x, t)|X = x) \\ &= F(x, Q(x, t)) = t.\end{aligned}$$

From this elementary observation we deduce that

- U is independent from X (since its conditional cdf does not depend on x),
- U is uniformly distributed,
- $Y = Q(X, U)$ where $Q(x, .)$ is increasing.

This easy remark leads to a conditional polar factorization of Y with an independence condition between U and X. There is a variational principle behind this conditional decomposition. Recall that we have denoted by μ the uniform measure on $[0,1]$. Let us consider the variant of the optimal transport problem (3) where one further requires U to be independent from the vector of regressors X:

$$\max\{\mathbf{E}(VY),\ \mathrm{Law}(V) = \mu,\ V \perp\!\!\!\perp X\}. \tag{4}$$

which in terms of joint law $\theta = \mathrm{Law}(X, Y, U)$ can be written as

$$\max_{\theta \in I(\nu,\mu)} \int u \cdot y \ \theta(dx, dy, du) \tag{5}$$

where $I(\mu, \nu)$ consists of probability measures θ on $\mathbf{R}^{N+1} \times [0,1]$ such that the (X,Y) marginal of θ is ν and the (X,U) marginal of θ is $m \otimes \mu$.

In the previous conditional polar factorization, it is very demanding to ask that U is independent from the regressors X, but the function $Q(X,.)$ is just monotone nondecreasing, its dependence in x is arbitrary. In practice, the econometrician rather looks for a specific form of Q (linear in X for instance), which by duality will amount to relaxing the independence constraint. We shall develop this idea in details next and relate it to classical quantile regression.

Specified and quasi specified quantile regression

From now, on we normalize X to be centered i.e. assume (and this is without loss of generality) that

$$\mathbf{E}(X) = 0.$$

We also assume that $m := \text{Law}(X)$ is nondegenerate in the sense that its support contains some ball centered at $\mathbf{E}(X) = 0$.

Since the seminal work of Koenker and Bassett, it has been widely been accepted that a convenient way to estimate conditional quantiles is to stipulate an affine form with respect to x for the conditional quantile.

Since a quantile function should be monotone in its second argument, this leads to the following definition

Definition 1 *Quantile regression is specified if there exist* $(\alpha, \beta) \in C([0,1], \mathbf{R}) \times C([0,1], \mathbf{R}^N)$ *such that for m-a.e. x*

$$t \mapsto \alpha(t) + \beta(t) \cdot x \text{ is increasing on } [0,1] \tag{6}$$

and

$$Q(x,t) = \alpha(t) + x \cdot \beta(t). \tag{7}$$

for m-a.e. x and every $t \in [0,1]$. *If (6)-(7) hold, quantile regression is specified with regression coefficients* (α, β).

Specification of quantile regression can be characterized by

Proposition 1 *Let* (α, β) *be continuous and satisfy (6). Quantile regression is specified with regression coefficients* (α, β) *if and only if there exists* U *such that*

$$Y = \alpha(U) + X \cdot \beta(U) \ \textit{a.s.}, \ \mathrm{Law}(U) = \mu, \ U \perp\!\!\!\perp X. \qquad (8)$$

Interpretation: linear model with a random factor independent from the explanatory variables.

Koenker and Bassett showed that, for a fixed probability level t, the regression coefficients (α, β) can be estimated by quantile regression i.e. the minimization problem

$$\inf_{(\alpha,\beta)\in\mathbf{R}^{1+N}} \mathbf{E}(\rho_t(Y - \alpha - \beta \cdot X)) \tag{9}$$

where the penalty[a] ρ_t is given by $\rho_t(z) := tz_- + (1-t)z_+$ with z_- and z_+ denoting as usual the negative and positive parts of z. For further use, note that (9) can be conveniently be rewritten as

$$\inf_{(\alpha,\beta)\in\mathbf{R}^{1+N}} \{\mathbf{E}((Y - \alpha - \beta \cdot X)_+) + (1-t)\alpha\}. \tag{10}$$

[a]It is worth noting here the difference with ordinary least squares (quadratic penalty) for the estimation of conditional expectations.

As already noticed by Koenker and Bassett, this convex program admits as dual formulation

$$\sup\{\mathbf{E}(U_t Y)) \ : \ U_t \in [0,1], \ \mathbf{E}(U_t) = (1-t), \ \mathbf{E}(U_t X) = 0\} \quad (11)$$

An optimal (α, β) for (10) and an optimal U_t in (11) are related by the complementary slackness condition:

$$Y > \alpha + \beta \cdot X \Rightarrow U_t = 1, \text{ and } \ Y < \alpha + \beta \cdot X \Rightarrow U_t = 0. \quad (12)$$

Note that α appears naturally as a Lagrange multiplier associated to the constraint $\mathbf{E}(U_t) = (1-t)$ and β as a Lagrange multiplier associated to $\mathbf{E}(U_t X) = 0$. Since $\nu = \mathrm{Law}(X, Y)$ gives zero mass to nonvertical hyperplanes, we may simply write

$$U_t = \mathbf{1}_{\{Y > \alpha + \beta \cdot X\}}. \quad (13)$$

The constraints $\mathbf{E}(U_t) = (1-t)$, $\mathbf{E}(XU_t) = 0$ then read

$$\mathbf{E}(\mathbf{1}_{\{Y>\alpha+\beta\cdot X\}}) = \mathbf{P}(Y > \alpha+\beta\cdot X) = (1-t),\ \mathbf{E}(X\mathbf{1}_{\{Y>\alpha+\beta\cdot X\}}) = 0 \tag{14}$$

which simply are the first-order conditions for (10).

Any pair (α, β) which solves[a] the optimality conditions (14) for the Koenker and Bassett approach will be denoted

$$\alpha = \alpha^{QR}(t), \beta = \beta^{QR}(t)$$

and the variable U_t solving (11) given by (13) will similarly be denoted U_t^{QR}

$$U_t^{QR} := \mathbf{1}_{\{Y>\alpha^{QR}(t)+\beta^{QR}(t)\cdot X\}}. \tag{15}$$

[a]Uniqueness will be discussed later on

Note that in the previous considerations the probability level t is fixed, this is what we called the "t by t" approach. For this approach to be consistent with conditional quantile estimation, if we allow t to vary we should add an additional monotonicity requirement:

Definition 2 *Quantile regression is quasi-specified if there exists for each* t*, a solution* $(\alpha^{QR}(t), \beta^{QR}(t))$ *of (14) (equivalently the minimization problem (9)) such that* $t \in [0,1] \mapsto (\alpha^{QR}(t), \beta^{QR}(t))$ *is continuous and, for* m*-a.e.* x

$$t \mapsto \alpha^{QR}(t) + \beta^{QR}(t) \cdot x \text{ is increasing on } [0,1]. \tag{16}$$

A first consequence of quasi-specification is given by

Proposition 2 *If quantile regression is quasi-specified and if we define* $U^{QR} := \int_0^1 U_t^{QR} dt$ *(recall that* U_t^{QR} *is given by (15)) then:*

- U^{QR} *is uniformly distributed,*
- X *is mean-independent from* U^{QR} *i.e.* $\mathbf{E}(X|U^{QR}) = \mathbf{E}(X) = 0$,
- $Y = \alpha^{QR}(U^{QR}) + \beta^{QR}(U^{QR}) \cdot X$ *a.s.*

Moreover U^{QR} *solves the correlation maximization problem with a mean-independence constraint:*

$$\max\{\mathbf{E}(VY),\ \mathrm{Law}(V) = \mu,\ \mathbf{E}(X|V) = 0\}. \tag{17}$$

One has uniqueness for the mean-independent conditional polar factorization in proposition 2:

Proposition 3 *Let us assume that*

$$Y = \alpha(U) + \beta(U) \cdot X = \overline{\alpha}(\overline{U}) + \overline{\beta}(\overline{U}) \cdot X$$

with:

- *both* U *and* $\overline{U}$ *uniformly distributed,*
- X *is mean-independent from* U *and* $\overline{U}$*:* $\mathbf{E}(X|U) = \mathbf{E}(X|\overline{U}) = 0,$
- $\alpha, \beta, \overline{\alpha}, \overline{\beta}$ *are continuous on* $[0, 1]$*,*
- (α, β) *and* $(\overline{\alpha}, \overline{\beta})$ *satisfy the monotonicity condition (6),*

then

$$\alpha = \overline{\alpha},\ \beta = \overline{\beta},\ U = \overline{U}.$$

To sum up, we have shown that quasi-specification is equivalent to the validity of the linear factor model:

$$Y = \alpha(U) + \beta(U) \cdot X$$

for (α, β) continuous and satisfying the monotonicity condition (6) and U, uniformly distributed and mean-independent from X. In the specified case, U is independent from X. In the general case, the conditional polar factorization gives $Y = Q(X, U)$ where U is required to be independent from X but the dependence of Y with respect to U, given X, is given by any nondecreasing function of U.

General case

Now we wish to address quantile regression in the case where neither specification nor quasi-specification can be taken for granted. From what we saw, we can think of two natural approaches.

The first one consists in studying directly the correlation maximization with a mean-independence constraint (17):

$$\max\{\mathbf{E}(VY),\ \mathrm{Law}(V) = \mu,\ \mathbf{E}(X|V) = 0\}. \tag{18}$$

The second one consists in getting back to the Koenker and Bassett t by t problem (11) but adding as an additional global consistency constraint that U_t should be nonincreasing with respect to t:

$$\sup \mathbf{E}(\int_0^1 U_t Y \, dt)$$

subject to:

$$U_t \text{ nonincreasing}, U_t \in [0,1], \ \mathbf{E}(U_t) = (1-t), \ \mathbf{E}(U_t X) = 0. \tag{19}$$

In fact, these two approaches are equivalent (they have the same dual in fact). Let us remark that (17) can directly be considered in the multivariate case whereas the monotonicity constrained problem (19) makes sense only in the univariate case.

Multivariate quantiles

We now consider the case where the endogenous Y variable belongs to $\mathbf{R}^d$. The idea then is to define the multivariate quantile of Y as Brenier's map. Set $\mu := \text{uniform}([0,1]^d)$ and consider the correlation maximization problem

$$\max\{\mathbf{E}(V \cdot Y),\ \text{Law}(V) = \mu\} \tag{20}$$

i.e. the quadratic optimal transport problem

$$\inf \int_{\mathbf{R}^d \times \mathbf{R}^d} |u - y|^2 \gamma(du, dy)\ \gamma \in \Pi(\mu, \text{Law}(Y)).$$

Brenier's theorem says that if Y is a squared-integrable d-dimensional random variable, there is a unique map of the form $T = \nabla\varphi$ with φ convex on $[0,1]^d$ such that $\nabla\varphi_{\#}\mu = \text{Law}(Y)$. This map is the optimal transport from the uniform law to Law(Y).

By definition, we call this map the quantile function of Y.

Polar factorization

$$Y = \nabla\varphi(U),\ \varphi \text{ convex },U \text{ uniform.}$$

Conditional quantile

Now, take a N-dimensional random vector X of regressors, $\nu := \text{Law}(X, Y)$, $m := \text{Law}(X)$, $\nu = m \otimes \nu^x$ where ν^x is the law of Y given $X = x$. One can consider $Q(x, u) = \nabla\varphi(x, u)$ as the optimal transport between μ and ν^x. $Q(x, .)$ is then the conditional multivariate quantile of Y given $X = x$.

Under some regularity assumptions on ν^x, one can invert $Q(x,.)$: $Q(x,.)^{-1} = \nabla_y \varphi(x,.)^*$ (where the Legendre transform is taken for fixed x) and one can define U through

$$U = \nabla_y \varphi^*(X,Y),\ Y = Q(X,U) = \nabla_u \varphi(X,U).$$

As in the unidimensional case, this U is uniformly distributed, independent from X and solves:

$$\max\{\mathbf{E}(V \cdot Y),\ \mathrm{Law}(V) = \mu, V \perp\!\!\!\perp X\}. \tag{21}$$

Note that the additional mean-independence constraint looks a little bit like the martingale constraint of Henry-Labordère-Galichon-Touzi.

Specified case

If the conditional quantile function is affine in X (specified case), then $Y = Q(X, U) = \alpha(U) + \beta(U)X$ where U is uniform and independent from X, the function $u \mapsto \alpha(u) + \beta(u)x$ should be the gradient of some function of u which requires

$$\alpha = \nabla\varphi,\ \beta = Db^T$$

for some potential φ and some vector-valued function b in which case, $Q(x, .)$ is the gradient of $u \mapsto \varphi(u) + b(u) \cdot x$. Moreover, since quantiles are gradients of convex potentials one should also have

$$u \in [0,1]^d \mapsto \varphi(u) + b(u) \cdot x \text{ is convex .}$$

As in the unidimensional case, one can weaken the specification assumption: quasi-specification holds when $Y = \nabla\varphi(U) + Db^T(U)X$, with U mean independent from X and

$$u \in [0,1]^d \mapsto \Phi_x(u) := \varphi(u) + b(u) \cdot x \text{ is convex .}$$

In such as case, U solves:

$$\max\{\mathbf{E}(V \cdot Y),\ \mathrm{Law}(V) = \mu,\ \mathbf{E}(X|V) = 0\}. \tag{22}$$

Indeed, $Y = \nabla\Phi_X(U)$ hence $UY = \varphi(U) + b(U) \cdot X + \Phi_X^*(Y)$, integrating and using the fact that U is mean independent from X then gives

$$\mathbf{E}(UY) = \mathbf{E}(\varphi(U)) + \mathbf{E}(\Phi_X^*(Y))$$

and similarly for V uniform and such that $\mathbf{E}(X|V) = 0$ one has

$$\mathbf{E}(VY) \leq \mathbf{E}(\varphi(V)) + \mathbf{E}(\Phi_X^*(Y)).$$

General case and duality

We now consider the general case where quasi-specification does not necessarily hold. What does the optimal problem with a mean-independence condition

$$\max\{\mathbf{E}(V \cdot Y),\ \mathrm{Law}(V) = \mu,\ \mathbf{E}(X|V) = 0\}.$$

say about the depedence between X and Y? Regression interpretation?

As usual, a good starting point is duality.

Formal derivation of the dual. Recall notations $\mu = \text{uniform}([0,1]^d)$, $\nu := \text{Law}(X, Y)$ (on $\mathbf{R}^N \times \mathbf{R}^d$), (with $m := \text{Law}(X)$, centered). Rewrite then the mean-independent correlation maximization problem in terms of joint law:

$$\sup_{\theta \in MI(\mu,\nu)} \int_{\mathbf{R}^N \times \mathbf{R}^d \times [0,1]^d} u \cdot y \, \theta(dx, dy, du) \tag{23}$$

where $MI(\mu, \nu)$ consists of the probability measures θ on $\mathbf{R}^N \times \mathbf{R}^d \times \mathbf{R}^d$ such that that $\Pi_{X,Y\,\#}\theta = \nu$, $\Pi_{U\,\#}\theta = \mu$ and according to θ, x is mean independent of u i.e.

$$\int_{\mathbf{R}^N \times \mathbf{R}^d \times [0,1]^d} (b(u) \cdot x) \theta(dx, dy, du) = 0, \ \forall b \in C([0,1]^d, \mathbf{R}^d). \tag{24}$$

The constraints on the marginals can be rewritten as usual as

$$\int \varphi(u)\theta(dx, dy, du) = \int_{[0,1]^d} \varphi(u)du, \ \forall\varphi,$$

$$\int \psi(x, y)\theta(dx, dy, du) = \int_{\mathbf{R}^N \times \mathbf{R}^d} \psi(x, y)du, \ \forall\psi.$$

One can then rewrite (23) as

$$\sup_{\theta\geq 0} \inf_{\varphi,\psi,b} \int_{\mathbf{R}^N \times \mathbf{R}^d \times [0,1]^d} \Big(u\cdot y - \psi(x, y) - \varphi(u) - b(u)\cdot x \Big) \theta(dx, dy, du).$$

Switching the sup and the inf gives the (formal) dual:

$$\inf_{\psi,\varphi,b} \int_{\mathbf{R}^N \times \mathbf{R}^d} \psi(x,y)\nu(dx,dy) + \int_{[0,1]^d} \varphi(u)du$$

subject to the pointwise constraint:

$$\psi(x,y) + \varphi(u) \geq u \cdot y + b(u) \cdot x.$$

The existence of optimal functions ψ, φ and b is not totally obvious. Assume

- the support of ν, is of the form $\mathrm{spt}(\nu) := \overline{\Omega}$ where Ω is an open bounded convex subset of $\mathbf{R}^N \times \mathbf{R}^d$,
- $\nu \in L^\infty(\Omega)$,
- ν is bounded away from zero on compact subsets of Ω that is for every K compact, included in Ω there exists $\alpha_K > 0$ such that $\nu \geq \alpha_K$ a.e. on K.

Theorem 1 *Under the assumptions above, the dual problem admits at least a solution (and its value coincides with that of the mean-independent correlation maximization problem (23)) .*

In the standard OT problem, the dual potentials are convex conjugates, one therefore have control on their regularity. Here, we have no control on the additional Lagrange multiplier b. Proof uses Komlos' theorem and gives a b and a φ which are no better than L^1.

Quantile regression as optimality conditions

In the dual problem, one can impose

$$\psi(x,y) = \sup_{t\in[0,1]^d} \{t\cdot y - b(t)\cdot x - \varphi(t)\} \tag{25}$$

so that ψ is convex.

Let U solve the mean-independent OT problem and (ψ, φ, b) solve the dual.

The primal-dual relations give

$$\psi(x,y) + \varphi(t) + b(t)\cdot x \geq t\cdot y \text{ pointwise } (x,y,t) \in \Omega \times [0,1]$$

and almost-surely

$$\psi(X,Y) + \varphi(U) + b(U)\cdot X = U\cdot Y.$$

Since ψ is convex given by (25), this gives

$$(-b(U), U) \in \partial\psi(X, Y),$$

or, equivalently

$$(X, Y) \in \partial\psi^*(-b(U), U)$$

almost surely.

If ψ was smooth and b continuous, we would then have

$$U = \nabla_y \psi(X, Y), \; -b(U) = \nabla_x \psi(x, y).$$

In this case, ψ solves the vectorial Hamilton-Jacobi equation:

$$\nabla_x \psi(x, y) + b(\nabla_y \psi(x, y)) = 0 \tag{26}$$

Furthermore, if φ and b were smooth then

$$Y = \nabla \varphi(U) + Db(U)^T X = \nabla \Phi_X(U)$$

(where $\Phi_x(t) := \varphi(t) + b(t) \cdot x$). We then see that φ and b are consistent with multivariate quantile regression estimation.

But such regularity cannot be taken for granted.

Still, problems (22) and its dual thus enabled us to find:

- U uniformly distributed and mean-independent from X,
- a map b from $[0,1]^d$ to $\mathbf{R}^d$ and a convex function ψ,

such that $(X, Y) \in \partial \psi^*(-b(U), U)$.

Specification of multivariate quantile regression rather asks whether one can write $Y = \nabla\varphi(U) + Db(U)^T X := \nabla\Phi_X(U)$ with $u \mapsto \Phi_x(u) := \varphi(u) + b(u)x$ is convex in u for fixed x.

In general, one gets from our optimization problems, a relaxation of the affine in X specifiation of the conditional quantile.

Indeed, we have

$$Y \in \partial\psi_X^*(U), Y \in \partial_u \psi^*(-b(U), U).$$

Setting $\psi_x := \psi(x, .)$, $\Phi_x := \varphi(.) + b(.) \cdot x$, the constraint in the dual can be rewritten as

$$\psi_x \geq \Phi_x^* \text{ hence } \psi_x^* \leq (\Phi_x)^{**} \leq \Phi_x$$

(where Φ_x^{**} denotes the convex envelope of Φ_x). Although Φ_X is not convex in general, the duality relations also give the following

Proposition 4

$$\Phi_X(U) = \Phi_X^{**}(U) \text{ and } U \in \partial\Phi_X^*(Y) \text{ i.e. } Y \in \partial\Phi_X^{**}(U)$$

almost surely.

Which is the natural relaxation of the relation $Y = \nabla \Phi_X(U)$ which holds in the specified case in the general case where Φ_X is neither smooth nor convex.

We have seen that quantile regression is tightly related to an OT-like problem with a mean-independence constraint. This enabled us to introduce a multivariate extension of the classical Koenker and Basset framework. In this talk, we did not address computational issues and applications to real data. Still, in the discrete setting, the mean-independent OT problem can be attacked by linear-programming techniques and there are efficient methods to solve it. Whether one can obtain some regularity of the solution of the dual remains to be investigated.