Linear programming II

Review: LP problem

- The standard form of LP problem is (primal problem):

$$
\begin{aligned}
&max\ z = c\mathbf{x} \\
&s.t.\ A\mathbf{x} \le b,\ \ \mathbf{x} \ge 0
\end{aligned}
$$

- The corresponding dual problem is:

$$
\begin{aligned}
min \quad & b^T y \\
s.t. \quad & A^T y \ge c^T, y \ge 0
\end{aligned}
$$

- **Strong Duality Theorem**: If the primal problem has an optimal solution, then the dual also has an optimal solution and there is no **duality gap**.

Complementary Slackness Theorem

The result obtained from proving the strong duality theorem is a theorem itself called "**Complementary Slackness Theorem**", which states:

If x^* and y^* are feasible solutions of primal and dual problems, then x^* and y^* are both optimal if and only if

1. $y^{*T}(b - Ax^*) = 0$
2. $(y^{*T}A - c)x^* = 0$

This implies that if a primal constraint is not "bounded", its corresponding variables in the dual problem must be 0, and vice versa.

This theorem is useful for solving LP problems, and the foundation of another class of LP solver called "**interior problem**" method.

Consider our simple LP problem:

$$
\begin{aligned}
max\ z &= x_1 + x_2, \\
s.t.\ x_1 + 2x_2 &\leq 100 \\
2x_1 + x_2 &\leq 100 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

Its dual problem is:

$$
\begin{aligned}
min\ z &= 100y_1 + 100y_2, \\
s.t.\ y_1 + 2y_2 &\geq 1 \\
2y_1 + y_2 &\geq 1 \\
y_1, y_2 &\geq 0
\end{aligned}
$$

The the Complementary Slackness Theorem states:

$$\begin{aligned} y_1(100 - x_1 - 2x_2) &= 0 \\ y_2(100 - 2x_1 - x_2) &= 0 \\ x_1(y_1 + 2y_2 - 1) &= 0 \\ x_2(2y_1 + y_2 - 1) &= 0 \end{aligned}$$

At the optimal, we have $\mathbf{x} = [100/3, 100/3]$, $\mathbf{y} = [1/3, 1/3]$. The complementary slackness holds, and all the constrains are bounded in both primal and dual problems.

Complementary Slackness Theorem example (cont.)

Modify the simple LP problem a bit:

$$
\begin{aligned}
max\ z &= 3x_1 + x_2, \\
s.t.\ x_1 + 2x_2 &\leq 100 \\
2x_1 + x_2 &\leq 100 \\
x_1, x_2 &\geq 0
\end{aligned}
$$

and its dual problem is:

$$
\begin{aligned}
min\ z &= 100y_1 + 100y_2, \\
s.t.\ y_1 + 2y_2 &\geq 3 \\
2y_1 + y_2 &\geq 1 \\
y_1, y_2 &\geq 0
\end{aligned}
$$

Now the the Complementary Slackness Theorem states:

$$
\begin{aligned}
y_1(100 - x_1 - 2x_2) &= 0 \\
y_2(100 - 2x_1 - x_2) &= 0 \\
x_1(y_1 + 2y_2 - 3) &= 0 \\
x_2(2y_1 + y_2 - 1) &= 0
\end{aligned}
$$

For this LP problem, the optimal solutions are $\mathbf{x} = [50, 0]$, $\mathbf{y} = [0, 1.5]$. The complementary slackness still holds. We observe that:

- In the primal problem:
  - The first constrain is unbounded, so its corresponding variable in the dual problem (y_1) has to be 0.
  - The second constrain is bounded, so its corresponding variable in the dual problem (y_2) can be non-zero.
- In the dual problem:
  - The first constrain is bounded, so its corresponding variable in the primal problem (x_1) is free.
  - The second constrain is unbounded, so its corresponding variable in the primal problem (x_2) has to be 0.

Given the primal and dual problem with slack/surplus variables added:

Primal:

$$\begin{aligned} & max \quad cx \\ & s.t. \quad Ax + w = b, x, w \geq 0 \end{aligned}$$

Dual:

$$\begin{aligned} & min \quad b^T y \\ & s.t. \quad A^T y - z = c^T, y, z \geq 0 \end{aligned}$$

- The Complementary Slackness Theorem states that at optimal solution, we should have: $x_j z_j = 0, \forall j$, and $w_i y_i = 0, \forall i$.
- To put this in matrix notation, define $X = diag(x)$, which means X is a diagonal matrix with x_j as diagonal elements.
- Define e as a vector of 1's.
- Now the complementary conditions can be written as: $XZe = 0$, $WYe = 0$.

We wil have the **optimality conditions** for the primal/dual problems as:

$$
\begin{aligned}
Ax + w - b &= 0 \\
A^T y - z - c^T &= 0 \\
XZe &= 0 \\
WYe &= 0 \\
x, y, w, z &\geq 0
\end{aligned}
$$

- The first two conditions are simply the constraints for primal/dual problems.
- The next two are complementary slackness.
- The last one is the non-negativity constraint.

Ignoring the non-negativity constraints, this is a set of $2n + 2m$ equations with $2n + 2m$ unknowns (n and m are the number of unknowns and constraints in the primal problem), which can be solved using Newton's method.

Such approach is called "**primal-dual interior point method**".

- The primal-dual interior point method finds the primal-dual optimal solution (x^*, y^*, w^*, z^*) by applying Newton's method to the primal-dual optimality conditions.
- The direction and length of the steps are modified in each step so that the non-negativity condition is strictly satisfied in each iteration.

To be specific, define the following function $\mathbf{F} : \mathbb{R}^{2n+2m} \rightarrow \mathbb{R}^{2n+2m}$:

$$\mathbf{F}(\mathbf{x}, \mathbf{y}, \mathbf{w}, \mathbf{z}) = \begin{bmatrix} A\mathbf{x} + \mathbf{w} - \mathbf{b} \\ A^T\mathbf{y} - \mathbf{z} - \mathbf{c}^T \\ XZe \\ WYe \end{bmatrix}$$

The goal is to find solution for $\mathbf{F} = \mathbf{0}$.

Applying Newton's method, if at iteration k the variables are $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{w}^k, \mathbf{z}^k)$, we obtain a search direction $(\delta\mathbf{x}, \delta\mathbf{y}, \delta\mathbf{w}, \delta\mathbf{z})$ by solving the linear equations:

$$\mathbf{F}'(\mathbf{x}^k, \mathbf{y}^k, \mathbf{w}^k, \mathbf{z}^k)\begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{y} \\ \delta\mathbf{w} \\ \delta\mathbf{z} \end{bmatrix} = -\mathbf{F}(\mathbf{x}^k, \mathbf{y}^k, \mathbf{w}^k, \mathbf{z}^k).$$

Here $\mathbf{F}'$ is the Jacobian. At iteration k, the equations are:

$$\begin{bmatrix} A & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & A^T & \mathbf{0} & -\mathbf{I} \\ Z & \mathbf{0} & \mathbf{0} & X \\ \mathbf{0} & W & Y & \mathbf{0} \end{bmatrix}\begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{y} \\ \delta\mathbf{w} \\ \delta\mathbf{z} \end{bmatrix} = \begin{bmatrix} -A\mathbf{x}^k - \mathbf{w}^k + \mathbf{b} \\ -A^T\mathbf{y}^k + \mathbf{z}^k + \mathbf{c}^T \\ -X^kZ^ke \\ -W^kY^ke \end{bmatrix}$$

Then the update will be obtained as: $(\mathbf{x}^k, \mathbf{y}^k, \mathbf{w}^k, \mathbf{z}^k) + \alpha(\delta\mathbf{x}, \delta\mathbf{y}, \delta\mathbf{w}, \delta\mathbf{z})$ with $\alpha \in (0, 1]$. α is chosen so that the result from the next iteration is feasible.

Given that at current iteration, both primal and dual are strictly feasible, the first two terms on the right hand side are 0.

An improved algorithm

The algorithm in its current setup is not ideal because often only a small step can be taken before the positivity constraints are violated. A more flexible version is proposed as follow.

- The value of $XZe + WYe$ represents the duality gap.
- Instead of trying to eliminate the duality gap, reducing the duality gap by some factor in each step.

In order word, we replace the complementary slackness by:

$$XZe = \mu_x e$$
$$WYe = \mu_y e$$

When $\mu_x, \mu_y \to 0$ as $k \to \inf$, the solution from this system will converge to the optimal solution of the original LP problem. Easy selections of μ's are $\mu_x^k = (\mathbf{x}^k)^T z/n$ and $\mu_y^k = (\mathbf{w}^k)^T y/m$. Here n and m are dimensions of $\mathbf{x}$ and $\mathbf{y}$ respectively.

Under the new algorithm, at the kth iteration, the Newton equations become:

$$\begin{bmatrix} A & \mathbf{0} & \mathbf{I} & -\mathbf{0} \\ \mathbf{0} & A^T & \mathbf{0} & -\mathbf{I} \\ Z & \mathbf{0} & \mathbf{0} & X \\ \mathbf{0} & W & Y & \mathbf{0} \end{bmatrix} \begin{bmatrix} \delta\mathbf{x} \\ \delta\mathbf{y} \\ \delta\mathbf{w} \\ \delta\mathbf{z} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ -X^k Z^k e + \mu_x^k e \\ -W^k Y^k e + \mu_y^k e \end{bmatrix} \tag{1}$$

This provides the **general primal-dual interior point method** as follow:

1. Choose strictly feasible initial solution $(\mathbf{x}^0, \mathbf{y}^0, \mathbf{w}^0, \mathbf{z}^0)$, and set $k = 0$. Then Repeat following two steps until convergence.

2. Solve system (1) to obtain the updates $(\delta\mathbf{x}, \delta\mathbf{y}, \delta\mathbf{w}, \delta\mathbf{z})$.

3. Update the solution: $(\mathbf{x}^{k+1}, \mathbf{y}^{k+1}, \mathbf{w}^{k+1}, \mathbf{z}^{k+1}) = (\mathbf{x}^k, \mathbf{y}^k, \mathbf{w}^k, \mathbf{z}^k) + \alpha^k(\delta\mathbf{x}, \delta\mathbf{y}, \delta\mathbf{w}, \delta\mathbf{z})$. α^k is chosen so that all variables are greater than or equal to 0.

The interior point algorithm is closely related to the **Barrier Problem**. Go back to the primal problem:

$$\begin{aligned} &max\ z = cx \\ &s.t.\ \ Ax + w = b, x, w \geq 0. \end{aligned}$$

The non-negativity constraints can be replaced by adding two **barrier terms** in the objective function. The barrier term is defined as $B(\mathbf{x}) = \sum_j \log x_j$, which is finite as long as x_j is positive. Then the primal problem becomes:

$$\begin{aligned} &max\ z = cx + \mu_x B(\mathbf{x}) + \mu_y B(\mathbf{w}) \\ &s.t.\ Ax + w = b. \end{aligned}$$

The barrier terms make sure x and w won't become negative.

Before trying to solve this problem, we need some knowledge about **Lagrange multiplier**.

The method Lagrange multiplier is a general algorithm for optimization problems with equality constraints. For example, consider a problem:

$$
\begin{aligned}
&max\ f(x, y) \\
&s.t.\ g(x, y) = c
\end{aligned}
$$

We introduce a new variable λ called **Lagrange multiplier** and form the following new objective function :

$$L(x, y, \lambda) = f(x, y) + \lambda[g(x, y) - c]$$

We will then optimize L with respect to x, y and λ using typical method. Note that the condition $\partial L/\partial\lambda = 0$ at optimal solution guarantees that the constraints will be satisfied.

The Barrier Problem (cont.)

Go back to the barrier problem, the Lagrangian for this problem is (using $\mathbf{y}$ as the multiplier):

$$L(\mathbf{x}, \mathbf{y}, \mathbf{w}) = c\mathbf{x} + \mu_x B(\mathbf{x}) + \mu_y B(\mathbf{w}) + \mathbf{y}^T(b - w - Ax).$$

The optimal solution for the problem satisfies (check this!):

$$\begin{aligned} c + \mu_x X^{-1} e - A^T \mathbf{y} &= 0 \\ \mu_y W^{-1} e - \mathbf{y} &= 0 \\ b - w - Ax &= 0 \end{aligned}$$

Define new variables $z = \mu_x X^{-1} e$ and rewrite these conditions, we obtain exactly the same set of equations as the relaxed optimality conditions for primal-dual problem.

Introduction to quadratic programming

We have discussed linear programming, where both the objective function and constraints are linear functions of the unknowns.

The **quadratic programming (QP)** problem has quadratic objective function and linear constraints:

$$\begin{aligned} \max \quad & f(x) = \frac{1}{2}x^T Bx + cx \\ s.t. \quad & Ax \le b, x \ge 0 \end{aligned}$$

The algorithm for solving QP problem is very similar to that for LP. But first we need to introduce the KKT condition.

The Karush-Kuhn-Tucker (KKT) conditions are a set of necessary conditions for a solution to be optimal in a general non-linear programming problem.
Consider the following problem :

$$\begin{aligned} \max \quad & f(x) \\ s.t. \quad & g_i(x) \leq 0, i = 1, \ldots, I \\ & h_j(x) = 0, j = 1, \ldots, J \end{aligned}$$

The Lagrangian is: $L(x, \mu, \lambda) = f(x) - \sum_i y_i g_i(x) - \sum_j z_j h_j(x)$. Then at the optimal solution, following **KKT** conditions must be satisfied:

- Primal feasibility: $g_i(x^*) \leq 0, \ h_j(x^*) = 0$.
- Dual feasibility: $y_i \geq 0$. (what about z_j?)
- Complementary slackness: $y_i g_i(x^*) = 0$.
- Stationary: $\nabla f(x^*) - \sum_i y_i \nabla g_i(x) - \sum_j z_j \nabla h_j(x) = 0$.

Optimal solution for QP

Following the same procedure, the Lagrangian for the QP problem can be expressed as : $L(x, \mu, \lambda) = \frac{1}{2}x^T Bx + cx - y^T(Ax - b) + z^T x$.

Then the KKT conditions for the QP problem is:

- Primal feasibility: $Ax \leq b, x \geq 0$.
- Dual feasibility: $y \geq 0, z \geq 0$ (pay attention to the sign of z).
- Complementary slackness: $Y(Ax - b) = \mathbf{0}$, $Zx = 0$.
- Stationary: $Bx + c - A^T y + z = 0$.

Y and Z are diagonal matrices with y and z at diagonal.

This can be solved using the interior-point method.

To be specific, add slack variable w ($= b - Ax$), the optimality conditions become:

$$
\begin{aligned}
Ax + w - b &= 0 \\
Bx + c - A^T y + z &= 0 \\
Zx &= 0 \\
Yw &= 0 \\
x, y, z, w &\geq 0
\end{aligned}
$$

The unknowns are x, y, z, w. We can then obtain the Jacobians, form the Newton equation and solve for the optimal solution iteratively.

The `quadprog` package provide functions (`solve.QP.compact`) to solve quadratic programming problem.
Pay attention to the definition of function parameters. They are slightly different from what I have used in the standard form!

```
solve.QP               package:quadprog               R Documentation

Solve a Quadratic Programming Problem

Description:

     This routine implements the dual method of Goldfarb and Idnani
     (1982, 1983) for solving quadratic programming problems of the
     form min(-d^T b + 1/2 b^T D b) with the constraints A^T b >= b_0.

Usage:
     solve.QP(Dmat, dvec, Amat, bvec, meq=0, factorized=FALSE)

Arguments:

    Dmat: matrix appearing in the quadratic function to be minimized.
```

```
    dvec: vector appearing in the quadratic function to be minimized.

    Amat: matrix defining the constraints under which we want to
          minimize the quadratic function.

    bvec: vector holding the values of b_0 (defaults to zero).

     meq: the first meq constraints are treated as equality
          constraints, all further as inequality constraints (defaults
          to 0).

factorized: logical flag: if TRUE, then we are passing R^(-1) (where
          D = R^T R) instead of the matrix D in the argument Dmat.
```

QP example

To solve $min\ \ \frac{1}{2}(x_1^2 + x_2^2), \quad s.t.\ \ 2x_1 + x_2 \geq 1$.

```
> Dmat = diag(rep(1,2))
> dvec = rep(0,2)
> Amat = matrix(c(2,1))
> b = 1
> solve.QP(Dmat=Dmat,dvec=rep(0,2),Amat=Amat, bvec=b)
$solution
[1] 0.4 0.2

$value
[1] 0.1

$unconstrained.solution
[1] 0 0

$iterations
[1] 2 0

$Lagrangian
[1] 0.2

$iact
[1] 1
```

Review

We have covered in previous two classes:

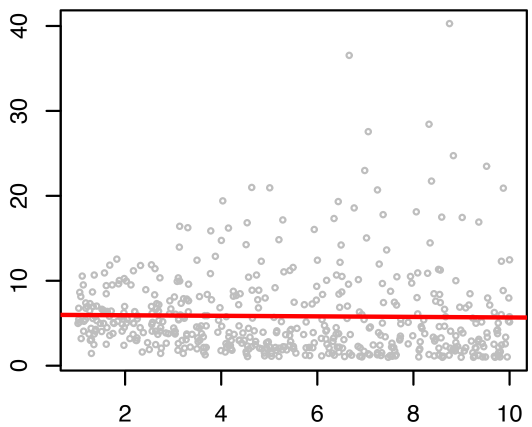
- LP problem set up.
- Simplex method.
- Duality.
- Interior point algorithm.
- Quadratic programming.

Now you should be able to formulate a LP/QP problem and solve it. But how are these useful in statistics?

- Remember LP is essentially an optimization algorithm.
- There are plenty of optimization problems in statistics, e.g., MLE.
- It's just a matter of formulating the objective function and constraints.

Motivation:

- Goal of regression: to tease out the relationship between outcome and covariates. Traditional regression: mean of the outcome depends on covariates.
- Problem: data are not always well-behaved. Are mean regression methods sufficient in all circumstances?



Quantile regression:

- provides a much more exhaustive description of the data.
- The collection of regressions at all quantiles would give a complete picture of outcome-covariate relationships.

Regress conditional quantiles of response on the covariates. Assume the outcome Y is continuous and that X is the vector of covariates.

- Classical model: $Q_\tau(Y|X) = X\beta_\tau$
- $Q_\tau(Y|X)$ is the τ^{th} conditional quantile of Y given X.
- β_τ is the parameter of interest.

The above model is equivalent to specifying

$$Y = X\beta_\tau + \epsilon,\ Q_\tau(\epsilon|X) = 0\ .$$

In comparison, the mean regression is:

$$Y = X\beta + \epsilon,\ E[\epsilon|X] = 0\ .$$

Pros and cons

Advantages:

- Regression at a sequence of quantiles provides a more complete view of data.
- Inference is robust to outliers.
- Estimation is more efficient when residual normality is highly violated.
- Allows interpretation in the outcome's original scale of measurement.

Disadvantages:

- To be useful, needs to regress on a set of quantiles: computational burden.
- Solution has no closed form.
- Adaptation to non-continuous outcomes is difficult.

The loss function

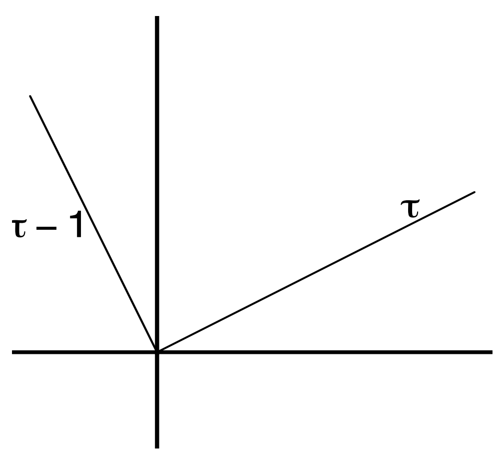
Link between estimands and loss functions.

- To obtain sample mean of $\{y_1, y_2, ..., y_n\}$, minimize $\sum_i (y_i - b)^2$.
- To obtain sample median of $\{y_1, y_2, ..., y_n\}$, minimize $\sum_i |y_i - b|$.

It can be shown that to obtain the sample τ^{th} quantile, one needs to minimize asymmetric absolute loss, that is, compute

$$\hat{Q}_\tau(\mathbf{Y}) = \underset{b}{\operatorname{argmin}} \left\{ \sum_{i:y_i \geq b} \tau |y_i - b| + \sum_{i:y_i < b} (1-\tau)|y_i - b| \right\} .$$

For convenience, defined $\rho_\tau(x) = x[\tau - \mathbb{1}(x < 0)]$.



Estimator

The classical linear quantile regression model is fitted by determining

$$\hat{\beta}_\tau = \operatorname*{argmin}_b \sum_{i=1}^{n} \rho_\tau(y_i - x_i b).$$

The estimator have all "expected" properties:

- Scale equivariance:

$$\hat{\beta}_\tau(ay, X) = a\hat{\beta}_\tau(y, X), \quad \hat{\beta}_\tau(-ay, X) = -a\hat{\beta}_{1-\tau}(y, X)$$

- Shift (or regression) equivariance:

$$\hat{\beta}_\tau(y + X\gamma, X) = \hat{\beta}_\tau(y, X) + \gamma$$

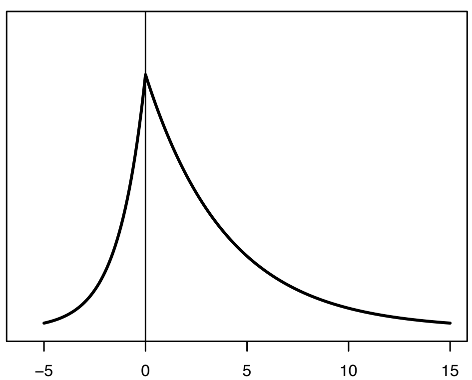
- Equivariance to reparametrization of design:

$$\hat{\beta}_\tau(y, XA) = A^{-1}\hat{\beta}_\tau(y, X)$$

Asymmetric Double Exponential (ADE) distribution

- Least-squares estimator $\Leftrightarrow$ MLE if residuals are normal.
- QR estimator $\Leftrightarrow$ MLE if residuals are ADE.
- Density function for ADE: $f(y;\mu,\sigma,\tau) = \frac{\tau(1-\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\}$.

ADE(0, 1, 0.25)



- If residuals are iid $ADE(0,1,\tau)$, then the log-likelihood for β_τ is

$$\ell(\beta_\tau;\mathbf{Y},\mathbf{X},\tau) = -\sum_{i=1}^{n}\rho_\tau\left(y_i - x_i\beta_\tau\right) + c_0$$

The QR question $\hat{\beta}_\tau = \mathrm{argmin}_b \sum_{i=1}^n \rho_\tau(y_i - x_i b)$ can be framed into an LP problem.

First define a set of new variables:

$$
\begin{aligned}
u_i &\equiv [y_i - x_i b]_+ \\
v_i &\equiv [y_i - x_i b]_- \\
b_+ &\equiv [b]_+ \\
b_- &\equiv [b]_-
\end{aligned}
$$

The the problem can be formulated as:

$$
\begin{aligned}
\max \quad & -\sum_{i=1}^n [\tau u_i + (1-\tau) v_i] \\
s.t. \quad & y_i = x_i b_+ - x_i b_- + u_i - v_i \\
& u_i, v_i \geq 0, \quad i = 1, \ldots, n \\
& b_+, b_- \geq 0
\end{aligned}
$$

This is a standard LP problem can be solved by Simplex/Interior point method.

Written in matrix notation, and make u_i, v_i, b_+, b_- as unknowns, get

$$\max \quad -[\mathbf{0}, \mathbf{0}, \tau, \mathbf{1}-\tau] \begin{bmatrix} \mathbf{b}_+ \\ \mathbf{b}_- \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix}$$

$$s.t. \quad [\mathbf{X}, -\mathbf{X}, \mathbf{I}, -\mathbf{I}] \begin{bmatrix} \mathbf{b}_+ \\ \mathbf{b}_- \\ \mathbf{u} \\ \mathbf{v} \end{bmatrix} = \mathbf{y}$$

$$\mathbf{b}_+, \mathbf{b}_-, \mathbf{u}, \mathbf{v} \geq 0$$

The dual problem is:

$$\min \quad \mathbf{y}^T \mathbf{d}$$

$$s.t. \quad \begin{bmatrix} \mathbf{X}^T \\ -\mathbf{X}^T \\ \mathbf{I} \\ -\mathbf{I} \end{bmatrix} \mathbf{d} \geq - \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \tau \\ \mathbf{1}-\tau \end{bmatrix}$$

$\mathbf{d}$ is unrestricted

Manipulating the constraints, get

$$\begin{aligned} \mathbf{X}^T\mathbf{d} &= \mathbf{0} \\ -\tau \le \mathbf{d} &\le 1-\tau \end{aligned}$$

Define a new variable $\mathbf{a} = 1 - \tau - \mathbf{d}$, the original LP problem can be formulated as:

$$\begin{aligned} \max \quad & \mathbf{y}^T\mathbf{a} \\ s.t. \quad & \mathbf{X}^T\mathbf{a} = \mathbf{0} \\ & 0 \le \mathbf{a} \le 1 \end{aligned}$$

Adding slack variables $\mathbf{s}$ for the $\le$ constraints, the problem can be formulated in the standard form, and can be solved by using either Simplex or Interior Point methods.

$$\begin{aligned} \max \quad & \mathbf{y}^T\mathbf{a} \\ s.t. \quad & \mathbf{X}^T\mathbf{a} = \mathbf{0} \\ & \mathbf{a} + \mathbf{s} = 1 \\ & \mathbf{a}, \mathbf{s} \ge 0 \end{aligned}$$

In this form, $\mathbf{y}$ (n-vector) are outcomes, $\mathbf{X}$ ($n \times p$ matrix) are predictors, $\mathbf{a}$ and $\mathbf{s}$ (n-vectors) are unknowns. There are $2n$ unknowns and $p + n$ constraints.

Once we have optimal $\mathbf{a}$, $\mathbf{d}$ can be obtained given τ (the quantile). Then depending on which constraints are hit in the dual problem, one can determine the set of basic variables in primal problem, and then solve for β.