LASSO/Poisson DML implementation

Thomas R. Covert, Yixin Sun

July 20, 2020

1 Setup from Chernozhukov et al

Let θ be the thing we care about and β be the nuisance parameters (location, time etc). The data is $W = (Y, D, X)$ where Y is an outcome, D is the vector of stuff we care about and X is the stuff we don't care about. The true values of θ and β, denoted as θ_0 and β_0, fit the data best, in the sense that

$$(\theta_0, \beta_0) = \arg\max_{\theta,\beta} \mathbb{E}_W \left[l(W, \theta, \beta) \right]$$

where $l(W, \theta, \beta)$ is some criterion (squared deviation, log likelihood etc).

The *Neyman Orthogonal Score* ψ is defined by:

$$\psi(W, \theta, \beta, \mu) = \frac{\partial}{\partial \theta} l(W, \theta, \beta) - \mu \frac{\partial}{\partial \beta} l(W, \theta, \beta)$$

The vector μ above is defined by the hessian of this criterion function. Let J be:

$$J = \begin{pmatrix} J_{\theta,\theta} & J_{\theta,\beta} \\ J_{\beta,\theta} & J_{\beta,\beta} \end{pmatrix} = \frac{\partial}{\partial \theta \partial \beta} \mathbb{E}_W \left[\frac{\partial}{\partial \theta \partial \beta} l(W, \theta, \beta) \right]$$

Then we define μ as $\mu = J_{\theta,\beta} J_{\beta,\beta}^{-1}$.

2 The Poisson Setting

In Poisson regression, the function l is

$$l(Y, D, X, \theta, \beta) = Y(D\theta + X\beta) - \exp(D\theta + X\beta)$$

and its associated gradients needed for the definition of ψ are

$$\frac{\partial}{\partial\theta} l(W,\theta,\beta) = (Y - \exp(D\theta + X\beta))D$$
$$\frac{\partial}{\partial\beta} l(W,\theta,\beta) = (Y - \exp(D\theta + X\beta))X$$

The entries in the Hessian matrix that we need to compute μ are:

$$J_{\theta,\theta} = -\mathbb{E}\left[D'D\exp(D\theta + X\beta)\right]$$
$$J_{\theta,\beta} = -\mathbb{E}\left[D'X\exp(D\theta + X\beta)\right]$$
$$J_{\beta,\beta} = -\mathbb{E}\left[X'X\exp(D\theta + X\beta)\right]$$

yielding this expression for μ:

$$\mu = \mathbb{E}\left[D'X\exp(D\theta + X\beta)\right]\left(\mathbb{E}\left[X'X\exp(D\theta + X\beta)\right]\right)^{-1}$$

I *think* this constructing is revealing, since it looks like weighted least squares, with D as the outcome, X as the covariates, and weights equal to $\exp(D\theta + X\beta)$.

The Neyman Orthogonal moment for Poisson regression is then:

$$\psi = (Y - \exp(D\theta + X\beta))(D - X\mu)$$

How would we implement this?

1. In the fitting sample, do a full poisson LASSO of Y on D and X, but only LASSO over the X terms, to pick which ones count. Then, in the fitting sample, do a regular poisson regression to get initial estimates for β and θ using the subset of X picked out in LASSO. For the estimation sample, use β to construct $s = X\beta$.

2. Compute weights $w = \exp(D\theta + X\beta)$ in the fitting sample, and then. Compute a linear LASSO of D on X using those weights, and then do weighted OLS, again with those weights, on the selected covariates. The coefficients of this are μ.

3. Finally, the moment is $(Y - \exp(D\theta + s))(D - X\mu)$, and we only evaluate/fit this in the **hold out sample**.

4. If D is univariate, we can just do root-finding. If D is multivariate, we won't be able to match this exactly, so lets minimize squared deviations from zero. In the optimization routine, the

2.1 How would the linear version work?

Exactly the same way. Get rid of the exp's, and wherever the above says "Poisson" replace with "OLS". Moment is now $(Y - D\theta - s)(D - X\mu)$.