High-dimension Simultaneous Heterogeneity Inference and Divide-and-conquer Algorithms

Xiang Lyu

Department of Statistics
Purdue University

Nov 10, 2015
(Group Meeting)

A joint work with Prof. Guang Cheng

Outline

High-dimension Linear Model

- Consider a massive data set, (X, Y), which consists of $\{(X^{(1)}, Y^{(1)}), \ldots, (X^{(s)}, Y^{(s)})\}$.
- There may exist *heterogeneity* among them.
- In this paper, we focus on

$$Y^{(j)} = X^{(j)}\beta^{(j)} + \varepsilon^{(j)}.$$

  $X^{(j)}$ is independent of $\varepsilon^{(j)}$ and $\beta^{(j)}$. ε_i has zero mean and known variance σ^2.
- The null hypothesis is

$$\beta^{(1)} = \cdots = \beta^{(j)} = \cdots = \beta^{(G)}, \forall j \in [G], G \subseteq \{1, \ldots, s\}.$$

Model Assumption

Assumption A.1. (Regularity Condition).

(i) X has i.i.d sub-Gaussian rows. ε is i.i.d. subexponential;

(ii) X is uniformly bounded, $\|X\|_\infty = O(1)$. ε is i.i.d. with $\sigma^2 = O(1)$.

Assumption A.2. (Boundess condition).

(i) The smallest eigenvalue of Σ is $\Lambda_{\min}$, $c_1 \leq |\Lambda_{\min}| \leq C_1$ for two positive constants c_1 and C_1.

(ii) the elements of X's covariance Σ is bounded, $\|\Sigma\|_\infty = O(1)$.

Desparsifed Lasso

Van de Geer et al. (2014)

- Recall the lasso estimator by Tibshirani (1996)

$$\hat{\beta} = \arg\min_{\beta \in \mathbb{R}^p} (\|Y - X\beta\|_2^2/n + 2\lambda\|\beta\|_1)$$

- The Karush-Kuhn-Tucker condition of lasso,

$$-X^T(Y - X\hat{\beta})/n + \lambda\hat{\delta}, \tag{1}$$

$$\|\hat{\delta}\|_\infty \leq 1 \text{ and } \hat{\delta}_k = \text{sign}(\hat{\beta}_k) \text{ if } \hat{\beta}_k \neq 0, \quad \forall k \in [p] \tag{2}$$

- We get the despasified lasso by inverting the KKT condition,

$$\check{\beta} = \hat{\beta} + \hat{\Theta}(Y - X\hat{\beta})/n, \tag{3}$$

where $\hat{\Theta}$ is a approximation of the inverse of $\hat{\Sigma} = X^T X/n$ and $\Theta = \Sigma^{-1}$.

Constructing $\hat{\Theta}$ via Nodewise Lasso

Meinshausen and Bühlmann (2006)

- For each $k \in [p]$, consider

$$\hat{\gamma}_k := \arg\min_{\gamma \in \mathbb{R}^{p-1}} (\|X_k - X_{-k}\gamma\|_2^2 + 2\lambda_k \|\gamma\|_1), \tag{4}$$

  where X_{-k} denotes X without the k-th column. Compute p times on each X_k versus X_{-k} regression.
- $\hat{C} = (\hat{c}_{i,j})$ with $\hat{c}_{i,i} = 1$ and $\hat{c}_{i,j} = -\hat{\gamma}_{i,j}$ for $i \neq j$.
- $\hat{T}^2 = \text{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2)$ is diagonal matrix with $\hat{\tau}_k^2 := \|X_k - X_{-k}\hat{\gamma}_k\|_2^2/n + \lambda_k \|\hat{\gamma}_k\|_1$.
- We get the $\hat{\Theta} = \hat{T}^{-2}\hat{C}$.

Notations

- $|\{k : \beta_k^{(j)} \neq 0, \forall k \in [p]\}| = s_0^j$ and $s_0^* = \max_{j \in [G] \cup \{0\}} s_0^j$.
- $|\{l : \Theta_{k,l} \neq 0, \forall l \in [p]\}| = s_k, \forall k \in [p]$ and $s^* = \max_{k \in [p]} s_k$.
- $\hat{\Upsilon} = (\hat{\omega}_{kl})_{k,l \in [p]} = 2\sigma^2 (\hat{\Theta}\hat{\Sigma}\hat{\Theta})_{kl}$.

Main Results

Consistency of T_G

- We utilize $T_G = \max_{j \in [G-1], k \in [p]} \sqrt{n}(\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)})$ as test statistic
- The critical value is obtained by multiplier bootstrap method by Chernozhukov et al. (2013). Denote

$$W_G := \max_{j \in [G-1], k \in [p]} \frac{1}{\sqrt{n}} \sum_{i_1 \in L_j} (\hat{\Theta}^{(j)})_k X_{i_1}^{(j)} \epsilon_{i_1} - \frac{1}{\sqrt{n}} \sum_{i_2 \in L_{j+1}} (\hat{\Theta}^{(j+1)})_k X_{i_2}^{(j+1)} \epsilon_{i_2},$$

where ϵ_i is i.i.d. normal distribution with zero mean and variance σ^2. The bootstrap critical value is obtained by $c_G(\alpha) = \inf\{t \in \mathbb{R} : P(W_G \leq t | X) \geq 1 - \alpha\}$.

Main Results

Consistency of T_G

Theorem 1

Suppose Assumption A.1. - A.2. hold. together with $s_0^* \log(p)/\sqrt{n} = o(1)$, $s^* \log(p)/n = o(1)$, $\lambda_k^j \asymp \sqrt{\log(p)/n}$ *and* $\lambda_j \asymp \sqrt{\log(p)/n}$. *In addition, we assume that* $p\sqrt{s^* \log(pd) \log(p)/n} = o(1)$, $\log(p(d-1))/n = o(1)$ *and* $(\log(p(d-1)n))^7/n \leq Cn^{-c}$ *for some constants* c *and* C. *Then under* H_0, *we have*

$$\sup_{\alpha \in (0,1)} |P(T_G > c_G(\alpha)) - \alpha| = o(1).$$

Main Results

Consistency of T_G

Remak I The consistency of $T_{G,I} = \max_{j\in[G-1],k\in I} \sqrt{n}(\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)})$ holds for the testing $H_0 : \beta_I^{(1)} = \cdots = \beta_I^{(j)} = \cdots = \beta_I^{(G)}, \forall j \in [G], G \subseteq \{1, \ldots, s\}, I \subseteq \{1, \ldots, p\}$.

Remak II Denote $t_G = \max_{j\in[G-1],k\in[p]} \frac{1}{\sqrt{n}} \sum_{i_1\in L_j} \hat{\Theta}^{(1)} X_{i_1}^{(j)} \varepsilon_{i_1}^{(j)} - \frac{1}{\sqrt{n}} \sum_{i_2\in L_{j+1}} \hat{\Theta}^{(1)} X_{i_2}^{(j+1)} \varepsilon_{i_2}^{(j+1)}$.

Then

$$P(|t_G - T_G| \geq \varsigma_1) < \varsigma_2$$

Main Results

Consistency of T_G^*

Remak III Define $T_G^* = \max_{j \in [G-1], k \in [p]} \sqrt{n} |\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)}| = \sqrt{n} \max_{j \in [G-1], k \in [p]} \max\{\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)}, \check{\beta}_k^{(j+1)} - \check{\beta}_k^{(j)}\}$.

$$\sup_{\alpha \in (0,1)} \left| P\left(\max_{j \in [G-1], k \in [p]} \sqrt{n} |\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)}| > c_G^*(\alpha)\right) - \alpha \right| = o(1),$$

where $c_G^*(\alpha) = \inf\{t \in \mathbb{R} : P(W_G^* \leq t | X) \geq 1 - \alpha\}$ with $W_G^* = \max_{j \in [G-1], k \in [p]} \left| \sum_{i_1 \in L_j} (\hat{\Theta}^{(j)})_k X_{i_1}^{(j)} \epsilon_{i_1} / \sqrt{n} - \sum_{i_2 \in L_{j+1}} (\hat{\Theta}^{(j+1)})_k X_{i_2}^{(j+1)} \epsilon_{i_2} / \sqrt{n} \right|$.

Main Results

Consistency of $\bar{T}_G$

Consider the studentized statistic

$$\bar{T}_G = \max_{j \in [G-1], k \in [p]} \sqrt{n}(\check{\beta}_k^{(j)} - \check{\beta}_k^{(j+1)})/\sqrt{\hat{\omega}_{kk}}.$$

The bootstrap critical value is

$$\bar{c}_G(\alpha) = \inf\{t \in \mathbb{R} : P(\bar{W}_G \leq t | (Y, X)) \geq 1 - \alpha\},$$

where $\bar{W}_G = \max_{j \in [G-1], k \in [p]} \sum_{i_1 \in L_j} (\hat{\Theta}^{(j)})_k X_{i_1}^{(j)} \epsilon_{i_1}/\sqrt{n\hat{\omega}_{kk}} - \sum_{i_2 \in L_{j+1}} (\hat{\Theta}^{(j+1)})_k X_{i_2}^{(j+1)} \epsilon_{i_2}/\sqrt{n\hat{\omega}_{kk}}$.

Main Results

Consistency of $\bar{T}_G$

Theorem 2

Under the same conditions of Theorem 1. We have

$$\sup_{\alpha \in (0,1)} \left| P(\bar{T}_G > \bar{c}_G(\alpha)) - \alpha \right| = o(1)$$

under H_0*.*

Main Results

Consistency of $\bar{T}_G^*$

Remak. $\bar{T}_{G,I} = \max_{j\in[G-1],k\in I} \sqrt{n}(\breve{\beta}_k^{(j)} - \breve{\beta}_k^{(j+1)})$.

Remak. Denote $\bar{t}_G =$

$$\max_{j\in[G-1],k\in[p]} \sum_{i_1\in L_j} (\hat{\Theta}^{(1)})_k^T X_{i_1}^{(j)} \varepsilon_{i_1}^{(j)} / \sqrt{n\hat{\omega}_{kk}} - \sum_{i_2\in L_{j+1}} (\hat{\Theta}^{(1)})_k^T X_{i_2}^{(j+1)} \varepsilon_{i_2}^{(j+1)} / \sqrt{n\hat{\omega}_{kk}}.$$

Then we have $P(|\bar{t}_G - \bar{T}_G| \geq \varsigma_1) < \varsigma_2$.

Remark. Define $\bar{T}_G^* = \max_{j\in[G-1],k\in[p]} \sqrt{n}|\breve{\beta}_k^{(j)} - \breve{\beta}_k^{(j+1)}|/\sqrt{\hat{\omega}_{kk}}$ and $\bar{c}_G^*(\alpha) = \inf\{t \in \mathbb{R} : P(\bar{W}_G^* \leq t|X) \geq 1-\alpha\}$ where $\bar{W}_G^* =$

$$\max_{j\in[G-1],k\in[p]} |\sum_{i_1\in L_j} (\hat{\Theta}^{(j)})_k X_{i_1}^{(j)} \epsilon_{i_1} / \sqrt{n\hat{\omega}_{kk}} - \sum_{i_2\in L_{j+1}} (\hat{\Theta}^{(j+1)})_k X_{i_2}^{(j+1)} \epsilon_{i_2} / \sqrt{n\hat{\omega}_{kk}}|.$$

Main Results

Minimax-optimality of $\bar{T}_G^*$

- Denote α as Type I error and δ as Type II error.
- The null hypothesis is $H_0 : \theta_0 = 0_p$ and the alternative is $H_1 : \theta_0 \in \Theta_0/\{0_p\}$for the high dimensional linear model $Y = X\theta_0 + \epsilon$.
- The variance σ^2 of ϵ is known. And X is Gaussian design or fixed design.

Lemma 3 (Verzelen et al. (2012))

Assume that $\alpha + \delta \leq 33\%$, $p > n > C(\alpha, \delta)$, and that $n \geq 8\log(2/\delta)$. The rate of minimax separation distance for Gaussian, τ_G^2, and fixed design, τ_F^2, is

$$\frac{k}{n}\log(p) \wedge \frac{1}{\sqrt{n}}.$$

Main Results

Minimax-optimality of $\bar{T}_G^*$

- Recall that our null hypothesis is

$$H_0 : \beta^{(1)} = \cdots = \beta^{(j)} = \cdots = \beta^{(G)}, \forall j \in [G], G \subseteq \{1, \ldots, s\}.$$

- Rewrite it in another expression,

$$H_0 : \delta^{(j)} = 0, \forall j \in [G-1], G \subseteq \{1, \ldots, s\},$$

  where $\delta^{(j)} = \beta_0^{(j)} - \beta_0^{(j+1)}$.
- Denote $\bar{\beta} = ((\beta^{(1)})^T, \ldots, (\beta^{(d-1)})^T)^T$. The equivalent form of null hypothesis is

$$H_0 : \bar{\beta} = ((\beta_0^{(2)})^T, \ldots, (\beta_0^{(d)})^T)^T.$$

  Define $\bar{Y} = Y^{(1)} + \ldots + Y^{(d-1)}$, $\bar{X} = (X^{(1)}, \ldots, X^{(d-1)})$, and $\bar{\varepsilon} = \varepsilon^{(1)} + \ldots + \varepsilon^{(d-1)}$. Thus the linear model becomes $\bar{Y} = \bar{X}\bar{\beta} + \bar{\varepsilon}$. The minimax rate is $\sqrt{\log(p(d-1))/n}$.

Assumption A.3 Denote $\rho = (\rho_{kl})_{k,l\in[p]} = \Theta_{kl}/\sqrt{\omega_{kk}\omega_{ll}}$.

(i) $\max\limits_{1\le i\ne j\le p} |\rho_{ij}| \le c_4 < 1$ for some positive constant c_4.

(ii) $\max\limits_{j\in[p]} \sum_{i=1}^{p} \rho_{ij}^2 \le C_4$ for some positive constant C_4.

Main Results

Minimax-optimality of $\bar{T}_G^*$

Define the separation set $U_G(c_0) =$

$$\{\beta_0^{(j)} : \max_{j\in[G-1],k\in[p]} |\beta_{0,k}^{(j)} - \beta_{0,k}^{(j+1)}|/\sqrt{\omega_{kk}} > c_0\sqrt{\log(p(d-1))/n}\}.$$

Theorem 4

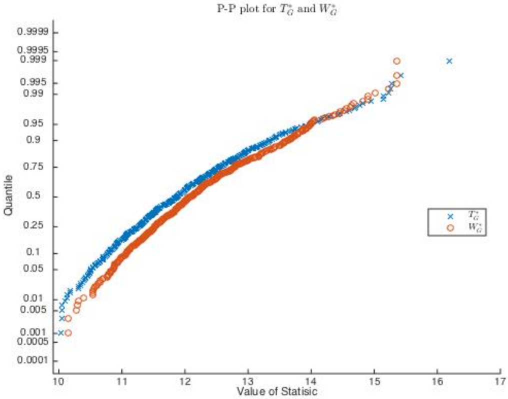
Under the same conditions of Theorem 2 together with Assumption A.3. Then for any $\epsilon_0 > 0$, we have

$$\inf_{\beta_0^{(j)}\in U_G(\sqrt{2}+\epsilon_0),\forall j\in[G]} P\Big(\max_{j\in[G-1],k\in[p]} \sqrt{n}\left|\check{\beta}_k^{(j+1)} - \check{\beta}_k^{(j)}\right|/\sqrt{\hat{\omega}_{kk}} > \bar{c}_G^*(\alpha)\Big) \to 1,$$

as $p(d-1) \to \infty$.

Numerical Results

Bootstrap Approximation



Numerical Results

Consistency and Power

		$\lvert G_1 \rvert = 5$		$\lvert G_2 \rvert = 10$	
		Type I	Power	Type I	Power
$p_1 = 550$	$n_1 = 400$	0.022	1	0.024	1
	$n_2 = 500$	0.042	1	0.040	1
$p_2 = 600$	$n_1 = 400$	0.022	1	0.024	1
	$n_2 = 500$	0.050	1	0.034	1

A divide-and-conquer framework for the next step of Heterogeneity Inference

Distributed Decorrelated Score Test

- We consider the step after heterogeneity inference.
- Assume that a collection of $|G| = d$ subpopulations accept the null hypothesis.
- Merge these subpopulation into a pool and get a extreme large data set, (X, Y).
- Conditional on X_i, we assume that Y_i are i.i.d generated from the distribution $P = \{P_\beta : \beta \in \Omega\}$.

Divide-and-conquer Setup

- Denote machine number m as $N^{1-\gamma}$.
- Each machine receives $n = N/m = N^{\gamma}$ local data set.
- In high dimensional problem, one general approach to estimate β in machine k, $k \in [m]$, is given by the following penalized M-estimator,

$$\hat{\beta}_k = \arg\min_{\beta \in \Omega} \ell_k(\beta) + P_\lambda(\beta),\ k \in [m]$$

where $\ell_k(\beta)$ in many cases corresponds to the empirical negative log-likelihood of Y conditioned on X and $P_\lambda(\beta)$ is a penalty function with a tuning parameter λ.

Divide-and-conquer Setup

- We partition the parameter β as $\beta = (\theta, \delta)$, where θ is a univariate parameter of interest and δ is a $p-1$ dimensional nuisance parameter.
- Note that we get $\hat{\theta}$ and $\hat{\delta}$ from $\hat{\beta}$.
- We consider $\ell_k(\theta, \gamma)$ as the empirical log-likelihood of the local data set in machine k,

$$\ell_k(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^{n} \ell_{k,i}(\theta, \gamma),$$

  where $\ell_{k,i}(\theta, \gamma) = -\log f(U_{k,i}; \theta, \gamma)$ denotes the negative log-likelihood for the i-th observation in machine k's local data set and f is the density function corresponding to the model P_β.

Decorrelated Score Function

- We use the decorrelated score function proposed by Ning and Liu (2014) as the score function

$$S(\theta, \gamma) = \nabla_\theta \ell(\theta, \gamma) - w^T \nabla_\gamma \ell(\theta, \gamma),$$

  with $w^T = I_{\theta,\gamma} I_{\gamma,\gamma}^{-1}$. I is the information matrix for β, defined as $I = E_\beta \nabla^2(\ell(\beta))$. The partial information matrix is defined as $I_{\theta|\gamma} = I_{\theta\theta} - I_{\theta\gamma} I_{\gamma\gamma}^{-1} I_{\gamma\theta}$, where $I_{\theta\theta}, I_{\theta\gamma}, I_{\gamma\gamma}$ and $I_{\gamma\theta}$ are the corresponding partition of I.
- The point estimator of θ is one-step estimator $\tilde{\theta}$, defined as

$$\tilde{\theta} = \hat{\theta} - \hat{S}(\hat{\beta}) / \hat{I}_{\theta|\gamma},$$

  where $\hat{I}_{\theta,\gamma} = \nabla^2_{\theta\theta} \ell(\hat{\beta}) - \hat{w}^T \nabla^2_{\gamma\theta} \ell(\hat{\beta})$.

Distributed Score Test

- In each machine, we compute the test statistic $\hat{S}_k$ and point estimator $\tilde{\theta}_k$, $k \in [m]$. The algorithm for $\hat{S}_k$ and $\tilde{\theta}_k$ is given in Ning and Liu (2014) Algorithm 1.
- Define the distributed score test statistic $\bar{U} =$

$$\bar{U} = \frac{1}{m} \sum_{k=1}^{m} \hat{U}_k,$$

where $\hat{U}_k = n^{1/2} \hat{S}_k(0, \hat{\gamma}_k) \hat{I}_{\theta|\gamma;k}^{-1/2}$.

Distributed Score Test

Uniform Convergency under Null Hypothesis

- Denote

$$\Omega_0 = \{(0, \gamma_0) : \|\gamma\|_0 \leq s^*, \text{ for some } s^* \ll n\}$$

as the parameter space of β_0.

Theorem 5

Assume that the Assumption Set C II. holds. It also holds that $\sup_{\beta_0 \in \Omega_0} \|w_0\|_1 \eta_5(n) = o(1)$, $\sup_{\beta_0 \in \Omega_0} \eta_2(n) \|I_{\theta\gamma,0}\|_\infty = o(1)$, *and* $n^{1/2}(\eta_2(n)\eta_3(n) + \eta_1(n)\eta_4(n)) = o(1)$. *Then, we have*

$$\lim_{n \to \infty} \sup_{\beta_0 \in \Omega_0} \sup_{t \in \mathbb{R}} \left| P_{\beta_0}(\bar{U} \leq t) - \Phi(\sqrt{m}t) \right| = 0.$$

Distributed Score Test

Uniform Convergency under Alternatives

- Define the sequence of alternative hypothesis $H_{1n} : \theta_0 \tilde{C} n^{-\phi}$, where $\tilde{C}$ is a constant and ϕ is a positive constant.
- Assume that the parameter space for local alternatives is

$$\Omega_1(\tilde{C}, \phi) = \{(\theta_0, \gamma_0) : \|\gamma\|_0 \leq s^*, \text{ for some } s^* \ll n\}.$$

Distributed Score Test

Uniform Convergency under Alternatives

Theorem 6

Under the Assumptions C III., we also assume, $\sup_{\beta_0 \in \Omega_1(\tilde{C},\phi)} \|w_0\|_1 \eta_5(n) = o(1)$, $\eta_2(n) \sup_{\beta_0 \in \Omega_1(\tilde{C},\phi)} \|I_{\theta\gamma,0}\|_\infty = o(1)$, *and* $n^{1/2}(\eta_2(n)\eta_3(n) + \eta_1(n)\eta_4(n)) = o(1)$. *Then, we have*

$$\lim_{n\to\infty} \sup_{\beta_0 \in \Omega_1(\tilde{C},\phi)} \sup_{t\in\mathbb{R}} |P_{\beta_0}(\bar{U} \le t) - \Phi(\sqrt{m}t)| = 0, \text{ if } \phi > 1/2,$$

$$\lim_{n\to\infty} \sup_{\beta_0 \in \Omega_1(\tilde{C},\phi)} \sup_{t\in\mathbb{R}} |P_{\beta_0}(\bar{U} \le t) - \Phi(\sqrt{m}t + \sqrt{m}\tilde{C} I^{1/2}_{\theta|\gamma,0})| = 0 \quad , \text{ if } \phi = 1/2,$$

$$\lim_{n\to\infty} \sup_{\beta_0 \in \Omega_1(\tilde{C},\phi)} |P_{\beta_0}(\bar{U} \le t)| = 0, \text{ if } \phi < 1/2. \tag{5}$$

where Eq. (5) *holds for any fixed* $t \in \mathbb{R}$ *and* $\tilde{C} \neq 0$.

- To construct confidence interval in high-dimension models, we consider the distributed one-step estimator

$$\bar{\theta} = \frac{1}{m} \sum_{k=1}^{m} \tilde{\theta}_k,$$

where $\tilde{\theta}_k$ is the one-step estimator in machine k based on its local data set, $\tilde{\theta}_k = \hat{\theta}_k - \hat{S}_k(\hat{\beta})/\hat{I}_{\theta|\gamma;k}$ and $\hat{I}_{\theta,\gamma;k} = \nabla^2_{\theta\theta}\ell_k(\hat{\beta}) - \hat{w}_k^T \nabla^2_{\gamma\theta}\ell_k(\hat{\beta}_k)$.

Distributed Score Test

Optimal Confidence Region

Theorem 7

Assume that the Assumption C.1 and Assumption Set C IV. hold. It also holds that $\|w_0\|_1 \eta_5(n) = o(1)$, $\eta_2(n)\|I_{\theta\gamma,0}\|_\infty = o(1)$ *and* $n^{1/2}(\eta_2(n)\eta_3(n) + \eta_1(n)\eta_4(n)) = o(1)$. *In addition, assume that* $n^{1/2}(\hat{\theta}_k - \theta_0)\|w_0\|_1 \eta_5(n) = o_p(1)$. *Then, we have*

$$n^{1/2}(\bar{\theta} - \theta_0)\bar{\hat{I}}^{1/2}_{\theta|\gamma} = n^{1/2}(\bar{\theta} - \theta_0) I^{1/2}_{\theta|\gamma,0} + o_p(1) = \frac{1}{\sqrt{m}} N + o_p(1),$$

where $N \sim N(0,1)$ *and*
$\bar{\hat{I}}^{1/2}_{\theta|\gamma} = \frac{1}{m}\sum_{k=1}^{m} \hat{I}^{1/2}_{\theta|\gamma;k} = \frac{1}{m}\sum_{k=1}^{m}\left[\nabla^2_{\theta\theta}\ell_k(\hat{\beta}_k) - \hat{w}_k^T \nabla^2_{\gamma\theta}\ell_k(\hat{\beta}_k)\right]$.

Distributed Score Test

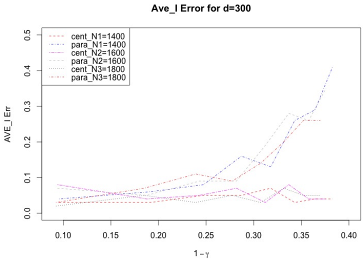
Optimal Confidence Region

Remark I $\bar{\theta}$ is semiparametric efficient by Chapter 25 Van der Vaart (2000). The $(1-\alpha)\times 100\%$ confidence interval for θ_0 is given by $\left[\bar{\theta} - N^{-1/2}\bar{\hat{I}}_{\theta|\gamma}^{-1/2}\Phi^{-1}(1-\frac{\alpha}{2}), \bar{\theta} + N^{-1/2}\bar{\hat{I}}_{\theta|\gamma}^{-1/2}\Phi^{-1}(1-\frac{\alpha}{2})\right]$. This confidence interval is optimal in the criterion of semiparametric efficiency, inherited from $\bar{\theta}$.

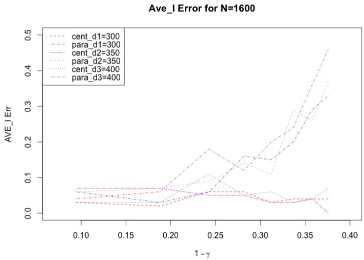
Remark II Note that the asymptotic interval is not depends on machine numbers. To be specific, the width relies on total data size N and significant level α.

Numerical Results

Type I error



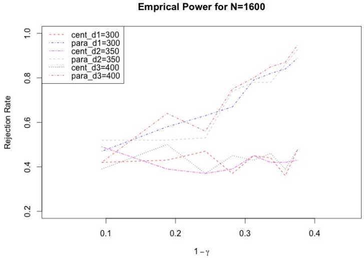
(a) Averaged Type I Error



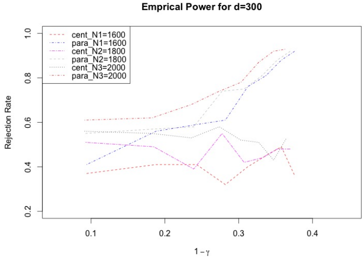
(b) Averaged Type I Error

Numerical Results

Power



(c) Empirical Power for $N = 1600$



(d) Empirical Power for $p = 300$

Reference I

Chernozhukov, V., Chetverikov, D., Kato, K., et al. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786–2819.

Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, pages 1436–1462.

Ning, Y. and Liu, H. (2014). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *arXiv preprint arXiv:1412.8765*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Reference II

Van de Geer, S., Bühlmann, P., Ritov, Y., Dezeure, R., et al. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Verzelen, N. et al. (2012). Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electronic Journal of Statistics*, 6:38–90.