Homework 6

Benjamin Noland

Yes, yes. . . I know. It's the bitter end of the semester.

1. Assumptions:
   (1) $(X_1 - X_2\gamma_1^*)^T(1/\hat{\tau}_1 - 1/\tau_1^*)(Y - X\beta^*)/n = o_p(n^{-1/2})$
   (2) $\|\hat{\kappa}_1\|_\infty \leq O_p(1)\sqrt{\log(p)/n}$
   (3) $\|\hat{\beta} - \beta^*\|_1 \leq O_p(1)S_{\beta^*}\sqrt{\log(p)/n}$
   (4) $S_{\beta^*}\log(p)/\sqrt{n} = o(1)$
   (5) $\|X_2^T(Y - X\beta^*)/n\|_\infty \leq O_p(1)\sqrt{\log(p)/n}$
   (6) $\|\hat{\gamma}_1 - \gamma_1^*\|_1 \leq O_p(1)S_{\gamma_1^*}\sqrt{\log(p)/n}$
   (7) $S_{\gamma_1^*}\log(p)/\sqrt{n} = o(1)$
   (8) $1/\hat{\tau}_1 = O_p(1)$.
   (9) The model $X_1 = X_2\gamma_1 + \delta$ is correctly specified, with true parameter γ_1^*, and the lasso estimator $\hat{\gamma}_1$ for γ_1 is consistent. Thus $\hat{\gamma}_1 \xrightarrow{p} \gamma_1^*$.

   From an argument given in the lectures, paired with assumption (1), we have:

$$\begin{aligned}
\hat{\hat{\beta}}_1 - \beta_1^* &= -\hat{\kappa}_1^T(\hat{\beta}_1 - \beta_2^*)/\hat{\tau}_1 + (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*)/(n\hat{\tau}_1) - (\hat{\gamma}_1 - \gamma_1^*)X_2^T(Y - X\beta^*)/(n\hat{\tau}_1) \\
&= -\hat{\kappa}_1^T(\hat{\beta}_1 - \beta_2^*)/\hat{\tau}_1 + (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*)/(n\tau_1^*) \\
&\quad + (X_1 - X_2\gamma_1^*)^T(1/\hat{\tau}_1 - 1/\tau_1^*)(Y - X\beta^*)/n \\
&\quad - (\hat{\gamma}_1 - \gamma_1^*)X_2^T(Y - X\beta^*)/(n\hat{\tau}_1) \\
&= -\hat{\kappa}_1^T(\hat{\beta}_1 - \beta_2^*)/\hat{\tau}_1 + (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*)/(n\tau_1^*) \\
&\quad - (\hat{\gamma}_1 - \gamma_1^*)X_2^T(Y - X\beta^*)/(n\hat{\tau}_1) + o_p(n^{-1/2}).
\end{aligned}$$

   Assumptions (2)-(4) paired with Hölder's inequality yield

$$|\hat{\kappa}_1^T(\hat{\beta}_2 - \beta_2^*)| \leq \|\hat{\kappa}_1\|_\infty\|\hat{\beta} - \beta^*\|_1 \leq O_p(1)\frac{\log(p)}{n}S_{\beta^*} = O_p(1)\frac{\log(p)}{\sqrt{n}}S_{\beta_1^*}\frac{1}{\sqrt{n}} = o_p(n^{-1/2}),$$

   so that

$$\hat{\kappa}_1^T(\hat{\beta}_2 - \beta_2^*) = o_p(n^{-1/2}).$$

   In addition, assumptions (5)-(6) paired with Hölder's inequality yield

$$\begin{aligned}
|(\hat{\gamma}_1 - \gamma_1^*)X_2^T(Y - X\beta^*)/n| &\leq \|\hat{\gamma}_1 - \gamma_1^*\|_1\|X_2^T(Y - X\beta^*)/2\|_\infty \\
&\leq O_p(1)\frac{\log(p)}{n}S_{\gamma_1^*} = O_p(1)\frac{\log(p)}{\sqrt{n}}S_{\gamma_1^*}\frac{1}{\sqrt{n}} = o_p(n^{-1/2}),
\end{aligned}$$

so that

$$(\hat{\gamma}_1 - \gamma_1^*) X_2^T (Y - X\beta^*)/n = o_p(n^{-1/2}).$$

Since $1/\hat{\tau}_1 = O_p(1)$ by assumption (8), we therefore have

$$\begin{aligned}
\hat{\hat{\beta}}_1 - \beta_1^* &= -\hat{\kappa}_1^T(\hat{\beta}_1 - \beta_2^*)/\hat{\tau}_1 + (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*)/(n\tau_1^*) \\
&\quad - (\hat{\gamma}_1 - \gamma_1^*)X_2^T(Y - X\beta^*)/(n\hat{\tau}_1) + o_p(n^{-1/2}) \\
&= o_p(n^{-1/2})O_p(1) + (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*)/(n\tau_1^*) + o_p(n^{-1/2})O_p(1) + o_p(n^{1/2}) \\
&= (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*) + o_p(n^{-1/2}).
\end{aligned}$$

Thus we can write

$$\begin{aligned}
\hat{\hat{\beta}}_1 - \beta_1^* &= (X_1 - X_2\gamma_1^*)^T(Y - X\beta^*) + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n (X_{i1} - X_{i2}^T\gamma_1^*)(Y_i - X_i^T\beta^*)/\tau_1^* + o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \psi(Y_i, X_i; \beta^*, \gamma_1^*) + o_p(n^{-1/2}),
\end{aligned}$$

where $\psi(Y_i, X_i; \beta^*, \gamma_1^*) = (X_{i1} - X_{i2}^T\gamma_1^*)(Y_i - X_i^T\beta^*)/\tau_1^*$.

2. Assumptions:
   (1) The model $Y = X\beta + \epsilon$ is correctly specified, with true parameter β^*, and the lasso estimator $\hat{\beta}$ for β is consistent. Thus $\hat{\beta} \xrightarrow{p} \beta^*$.
   (2) The model $X_1 = X_2\gamma_1 + \delta$ is correctly specified, with true parameter γ_1^*, and the lasso estimator $\hat{\gamma}_1$ for γ_1 is consistent. Thus $\hat{\gamma}_1 \xrightarrow{p} \gamma_1^*$.

   We have

$$\begin{aligned}
V &= \mathrm{Var}[\psi(Y_0, X_0; \beta^*, \gamma_1^*)] = \mathrm{Var}[(X_{01} - X_{02}^T\gamma_1^*)(Y_0 - X_0^T\beta^*)/\tau_1^*] \\
&= \mathrm{E}\{[(X_{01} - X_{02}^T\gamma_1^*)(Y_0 - X_0^T\beta^*)/\tau_1^*]^2\}.
\end{aligned}$$

   Thus, since $\hat{\beta} \xrightarrow{p} \beta^*$ and $\hat{\gamma}_1 \xrightarrow{p} \gamma_1^*$,

$$\begin{aligned}
\hat{V} = \frac{1}{n}\sum_{i=1}^n [\psi(Y_i, X_i; \hat{\beta}, \hat{\gamma}_1)]^2 &= \frac{1}{n}\sum_{i=1}^n [(X_{i1} - X_{i2}^T\hat{\gamma}_1)(Y_i - X_i^T\hat{\beta})/\hat{\tau}_1]^2 \\
&\xrightarrow{p} \mathrm{E}\{[(X_{01} - X_{02}^T\gamma_1^*)(Y_0 - X_0^T\beta^*)/\tau_1^*]^2\} = V.
\end{aligned}$$

   Thus $\hat{V} = V + o_p(1)$.

3. Theorem 2.2 in van der Geer et al. provides conditions for asymptotic normality of the desparsified lasso estimator under random design, and provides an explicit expression for the asymptotic variance. Problems 1 and 2 together do the same, though with stronger assumptions (there is almost certainly quite a lot of redundancy) than those given in van der Geer et al.