

1 Example: linear model

Theorem 1 (Multivariate normality of linear model score statistics). Suppose we observe i.i.d. samples $(X_1, Y_1, Z_1), \dots, (X_n, Y_n, Z_n)$, where $X_i \in \mathbb{R}$, $Y_i \in \mathbb{R}$, and $Z_i = [Z_i^1, \dots, Z_i^{p-1}] \in \mathbb{R}^{p-1}$. Let $X = [X_1, \dots, X_n]^T$ and $Y = [Y_1, \dots, Y_n]^T$ denote the X_i s and Y_i s, respectively, assembled into vectors. Next, let $\mathbf{Z} = [(1, Z_1), \dots, (1, Z_n)] \in \mathbb{R}^{n \times p}$ denote the Z_i s assembled into a matrix (with an additional column of 1s added for the intercept term). Let Q denote the marginal distribution of X_i (assumed to be known). For $j \in \{1, \dots, B\}$ and $i \in \{1, \dots, n\}$, let $\tilde{X}_i^j \sim Q$ be i.i.d. draws from Q . (We should be able to weaken the i.i.d. assumption significantly.) Denote $\tilde{X}^j = [\tilde{X}_1^j, \dots, \tilde{X}_n^j]^T$. Assume that $\mathbb{E}(Z_i) = \mathbb{E}(X_i) = 0$ and that $\mathbb{V}(Z_i) = \mathbb{V}(X_i) = 1$ (otherwise, standardize these variables). Furthermore, suppose that X_i and Z_i have finite fourth moment, i.e. $\mathbb{E}(X_i^4) < \infty$ and $\mathbb{E}(Z_i^4) < \infty$.

Next, assume that Y is a linear model of Z_i :

$$Y_i = \alpha_0 + \sum_{j=1}^{p-1} \alpha_j Z_i^j + \epsilon_i,$$

where $\epsilon_1, \dots, \epsilon_n \sim N(0, 1)$ and $\alpha = [\alpha_0, \alpha_1, \dots, \alpha_{p-1}] \in \mathbb{R}^p$ is a vector of unknown constants. Let $\hat{\alpha}$ be the MLE for α obtained from regressing Y onto \mathbf{Z} . Let $\hat{r} = Y - \mathbf{Z}\hat{\alpha}$ denote the vector of residuals. Let T_0, T_1, \dots, T_B be the score statistics obtained from testing $X, \tilde{X}^1, \dots, \tilde{X}^B$ for inclusion in the model, i.e.

$$\begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ \vdots \\ T_B \end{bmatrix} = \begin{bmatrix} (1/\sqrt{n})X^T\hat{r}/\sqrt{1 - (1/n)\|\mathbf{Z}^T X\|_P^2} \\ (1/\sqrt{n})(\tilde{X}^1)^T\hat{r} \\ (1/\sqrt{n})(\tilde{X}^2)^T\hat{r} \\ \vdots \\ (1/\sqrt{n})(\tilde{X}^B)^T\hat{r} \end{bmatrix},$$

where $\|\mathbf{Z}^T X\|_P^2 = X^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T X$. Then under the null hypothesis of $Y_i \perp\!\!\!\perp X_i | Z_i$, the vector T converges in distribution to the $B + 1$ -dimensional standard Gaussian, i.e.

$$T \xrightarrow{d} N_{B+1}(0, \mathbf{I}).$$

Remark. The score statistic that corresponds to the observed treatment vector T_0 is normalized by a factor of $(1 - (1/n)\|\mathbf{Z}^T X\|_P^2)^{1/2}$, while the other score statistics are not. The \tilde{X}^i 's are independent of \mathbf{Z} . Thus,

$$(1/n)\|\mathbf{Z}^T \tilde{X}^i\|_P^2 \xrightarrow{P} 0$$

for all $i \in \{1, \dots, B\}$. It follows that the normalized statistic

$$\begin{bmatrix} T'_0 \\ T'_1 \\ \vdots \\ T'_B \end{bmatrix} = \begin{bmatrix} (1/\sqrt{n})X^T \hat{r} / \sqrt{1 - (1/n)\|\mathbf{Z}^T X\|_P^2} \\ (1/\sqrt{n})(\tilde{X}^1)^T \hat{r} / \sqrt{1 - (1/n)\|\mathbf{Z}^T \tilde{X}^1\|_P^2} \\ \vdots \\ (1/\sqrt{n})(\tilde{X}^B)^T \hat{r} / \sqrt{1 - (1/n)\|\mathbf{Z}^T \tilde{X}^B\|_P^2} \end{bmatrix}$$

is asymptotically equivalent to the unnormalized statistic proposed above. However, the normalized statistic requires many more matrix multiplications and is therefore less efficient to compute than the unnormalized statistic.

The major assumption in the theorem is that the linear model holds; the other assumptions (e.g., the moment assumptions) are minor technical conditions that we would expect to hold in practice.

Proof. We proceed in three steps.

Step 1: Deriving an expression for the score and information matrix.

Consider the following expanded model for Y_i :

$$Y_i = \alpha_0 + \sum_{j=1}^{p-1} \alpha_j Z_i^j + \beta(X_i/\sqrt{n}) + \sum_{j=1}^B \gamma_j(\tilde{X}_i^j/\sqrt{n}),$$

where $\beta \in \mathbb{R}$ and $\gamma = [\gamma_1, \dots, \gamma_B] \in \mathbb{R}^B$ are constants. We have added the observed treatment X and synthetic treatments $\tilde{X}^1, \dots, \tilde{X}^B$ to the model. We also have divided the X_i 's and \tilde{X}_i^j 's by \sqrt{n} for normalization purposes. If the conditional independence hypothesis holds, then $\beta = \gamma = 0$. Let $\theta = [\alpha, \beta, \gamma]^T$ be the full vector of model parameters. Define the matrix

$$\tilde{\mathbf{X}}_{\sqrt{n}} = [(1/\sqrt{n})\tilde{X}^1, \dots, (1/\sqrt{n})\tilde{X}^B].$$

Similarly, denote the vector $(1/\sqrt{n})X$ by $X_{\sqrt{n}}$.

Let \mathbf{D} be the “augmented” design matrix formed by concatenating \mathbf{Z} , $X_{\sqrt{n}}$, and $\tilde{\mathbf{X}}_{\sqrt{n}}$

$$\mathbf{D} = [\mathbf{Z} \quad X_{\sqrt{n}} \quad \tilde{\mathbf{X}}_{\sqrt{n}}].$$

Let $\mathcal{L}(\theta)$ denote the model log likelihood evaluated at θ . The score $\nabla \mathcal{L}(\theta) = u(\theta)$ and Fisher information $-\mathbb{E}[\nabla^2 \mathcal{L}(\theta)] = \mathcal{I}(\theta) = \mathcal{I}$ of the model are as follows:

$$\begin{cases} u(\theta) = -\mathbf{D}^T(Y - \mathbf{D}\theta) \\ \mathcal{I} = \mathbf{D}^T \mathbf{D}. \end{cases}$$

We partition the score into subvectors:

$$\begin{aligned} u(\theta) &= \begin{bmatrix} \nabla_{\alpha} \mathcal{L}(\theta) \\ \frac{\partial \mathcal{L}(\theta)}{\partial \beta} \\ \nabla_{\gamma} \mathcal{L}(\theta) \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{Z}^T \\ X_{\sqrt{n}}^T \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T \end{bmatrix} \left(Y - [\mathbf{Z}\alpha + X_{\sqrt{n}}\beta + \tilde{\mathbf{X}}_{\sqrt{n}}\gamma] \right) \\ &= \begin{bmatrix} \mathbf{Z}^T[Y - (\mathbf{Z}\alpha + X_{\sqrt{n}}\beta + \tilde{\mathbf{X}}_{\sqrt{n}}\gamma)] \\ (X_{\sqrt{n}})^T[Y - \mathbf{Z}\alpha + X_{\sqrt{n}}\beta + \tilde{\mathbf{X}}_{\sqrt{n}}\gamma] \\ (\tilde{\mathbf{X}}_{\sqrt{n}})^T[Y - \mathbf{Z}\alpha + X_{\sqrt{n}}\beta + \tilde{\mathbf{X}}_{\sqrt{n}}\gamma] \end{bmatrix}. \end{aligned}$$

We similarly partition the information matrix into submatrices:

$$\begin{aligned} \mathcal{I} &= [\mathbf{Z} \quad X_{\sqrt{n}} \quad \tilde{\mathbf{X}}_{\sqrt{n}}]^T [\mathbf{Z} \quad X_{\sqrt{n}} \quad \tilde{\mathbf{X}}_{\sqrt{n}}] \\ &= \begin{bmatrix} \mathbf{Z}^T \mathbf{Z} & \mathbf{Z}^T X_{\sqrt{n}} & \mathbf{Z}^T \tilde{\mathbf{X}}_{\sqrt{n}} \\ X_{\sqrt{n}}^T \mathbf{Z} & X_{\sqrt{n}}^T X_{\sqrt{n}} & X_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{Z} & \tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} & \tilde{\mathbf{X}}_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \end{bmatrix}. \end{aligned}$$

Define

$$\begin{aligned} \mathcal{I}_{11} &= \mathbf{Z}^T \mathbf{Z}, \quad \mathcal{I}_{21} = \begin{bmatrix} X_{\sqrt{n}}^T \mathbf{Z} \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{Z} \end{bmatrix}, \quad \mathcal{I}_{12} = \mathcal{I}_{21}^T \\ \mathcal{I}_{22} &= \begin{bmatrix} X_{\sqrt{n}}^T X_{\sqrt{n}} & X_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} & \tilde{\mathbf{X}}_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \end{bmatrix}. \end{aligned}$$

Let $\mathcal{I}_X = \mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12}$. Also, let $\hat{\alpha}$ denote the MLE for α (obtained by an OLS regression of Y onto \mathbf{Z}). Let $u_X(\theta)$ denote the pieces of the score corresponding to β and γ , i.e.

$$u_X(\theta) = \begin{bmatrix} \frac{\partial \mathcal{L}(\theta)}{\partial \beta} \\ \nabla_{\gamma} \mathcal{L}(\theta) \end{bmatrix}.$$

Our goal is to find the distribution of $u_X(\hat{\alpha}, 0, 0)$, i.e. the distribution of the score subvector under the null hypothesis after estimating the nuisance parameters. First, we see that

$$u_X(\hat{\alpha}, 0, 0) = \begin{bmatrix} X_{Y\sqrt{n}}^T [Y - \mathbf{Z}\hat{\alpha}] \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T [Y - \mathbf{Z}\hat{\alpha}] \end{bmatrix} = \begin{bmatrix} (X_{Y\sqrt{n}}^T)\hat{r} \\ (\tilde{\mathbf{X}}_{\sqrt{n}}^T)\hat{r} \end{bmatrix}.$$

Next, standard asymptotic results indicate that

$$\mathcal{I}_X^{-1/2} u_X(\hat{\alpha}, 0, 0) \xrightarrow{d} N_{B+1}(0, \mathbf{I}).$$

Step 2: Simplifying the expression for \mathcal{I}_X

Next, we simplify the expression for $\mathcal{I}_X = \mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12}$. We begin by studying the matrix \mathcal{I}_{11}^{-1} . We have that

$$(1/n)\mathcal{I}_{11} = (1/n)\mathbf{Z}^T \mathbf{Z} \xrightarrow{P} \Sigma,$$

where $\Sigma \in \mathbb{R}^{p \times p}$ is the correlation matrix of Z . By the continuous mapping theorem,

$$n[\mathcal{I}_{11}^{-1}] = n[\mathbf{Z}^T \mathbf{Z}]^{-1} \xrightarrow{P} \Sigma^{-1},$$

where $\mathbf{P} := \Sigma^{-1}$ is the precision matrix of Z . (This matrix exists if Σ is positive definite, which we assume to be the case.) Let $\hat{\mathbf{P}} = n(\mathbf{Z}^T \mathbf{Z})^{-1}$ denote the empirical precision matrix of Z . We then can write \mathcal{I}_X as

$$\mathcal{I}_X = \begin{bmatrix} X_{Y\sqrt{n}}^T X_{\sqrt{n}} & X_{Y\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \\ \tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} & \tilde{\mathbf{X}}_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \end{bmatrix} - \begin{bmatrix} (1/n)X_{Y\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} & (1/n)X_{Y\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \\ (1/n)\tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} & (1/n)\tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \end{bmatrix},$$

where $\mathbf{A} = \mathbf{Z}\hat{\mathbf{P}}\mathbf{Z}^T$. We can compute the $n \times n$ matrix \mathbf{A} elementwise:

$$\begin{aligned} \mathbf{Z}\hat{\mathbf{P}}\mathbf{Z}^T &= \begin{bmatrix} z_1^T \hat{\mathbf{P}} z_1 & z_1^T \hat{\mathbf{P}} z_2 & z_1^T \hat{\mathbf{P}} z_3 & \dots & z_1^T \hat{\mathbf{P}} z_n \\ z_2^T \hat{\mathbf{P}} z_1 & z_2^T \hat{\mathbf{P}} z_2 & z_2^T \hat{\mathbf{P}} z_3 & \dots & z_2^T \hat{\mathbf{P}} z_n \\ \vdots & \vdots & \vdots & \dots & \vdots \\ z_n^T \hat{\mathbf{P}} z_1 & z_n^T \hat{\mathbf{P}} z_2 & z_n^T \hat{\mathbf{P}} z_3 & \vdots & z_n^T \hat{\mathbf{P}} z_n \end{bmatrix} \\ &= \begin{bmatrix} \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_1^k z_1^l & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_1^k z_2^l & \dots & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_1^k z_n^l \\ \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_2^k z_1^l & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_2^k z_2^l & \dots & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_2^k z_n^l \\ \vdots & \vdots & \dots & \vdots \\ \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_n^k z_1^l & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_n^k z_2^l & \dots & \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_n^k z_n^l \end{bmatrix}. \end{aligned}$$

Note that \mathbf{A} is symmetric. We frequently will consider bilinear forms and quadratic forms of \mathbf{A} . Specifically, for vectors $L = [L_1, \dots, L_n]^T \in \mathbb{R}^n$ and $M = [M_1, \dots, M_n]^T \in \mathbb{R}^n$, the bilinear form $(1/n)(L/\sqrt{n})^T \mathbf{A} (M/\sqrt{n})$ is given by

$$\begin{aligned} \frac{1}{n} (L/\sqrt{n})^T \mathbf{A} (M/\sqrt{n}) &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} L_i M_j \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} z_i^k z_j^l L_i M_j \\ &= \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} \left(\frac{1}{n} \sum_{i=1}^n z_i^k L_i \right) \left(\frac{1}{n} \sum_{i=1}^n z_i^l M_i \right). \quad (1) \end{aligned}$$

We see that the bilinear form of \mathbf{A} decomposes into a sum of p^2 terms. Each term is the product of three pieces: the (k, l) th entry of the empirical precision matrix, the average over the $Z_i^k L_i$ s, and the average over the $Z_i^l M_i$ s. Meanwhile, the linear form $(1/n)(L/\sqrt{n})^T \mathbf{A} (L/\sqrt{n})$ is obtained by substituting L_i for M_i above:

$$\frac{1}{n} (L/\sqrt{n})^T \mathbf{A} (L/\sqrt{n}) = \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} \left(\frac{1}{n} \sum_{i=1}^n z_i^k L_i \right) \left(\frac{1}{n} \sum_{i=1}^n z_i^l L_i \right).$$

Finally, note that for given $k, l \in \{1, \dots, p\}$,

$$\hat{P}_{kl} \xrightarrow{P} P_{kl},$$

i.e. the empirical precision matrix converges to the theoretical precision matrix elementwise.

Computing the limit of \mathcal{I}_X

We now compute the limit in probability of the matrix \mathcal{I}_X . We proceed block-by-block.

Lower right block. The lower right block $\text{LR}(\mathcal{I}_X)$ of \mathcal{I}_X is

$$\text{LR}(\mathcal{I}_X) = (\tilde{\mathbf{X}}_{\sqrt{n}}^T) \tilde{\mathbf{X}}_{\sqrt{n}} - (1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \in \mathbb{R}^{B \times B}.$$

Let

$$\tilde{X}_{\sqrt{n}}^r = \frac{1}{\sqrt{n}} [\tilde{X}_1^r, \dots, \tilde{X}_n^r]^T$$

denote the r th column of $\tilde{\mathbf{X}}_{\sqrt{n}}$; meanwhile, let

$$\tilde{X}_{\sqrt{n}}^s = \frac{1}{\sqrt{n}} [\tilde{X}_1^s, \dots, \tilde{X}_n^s]$$

denote the s th column of $\tilde{\mathbf{X}}_{\sqrt{n}}$. By the decomposition (1), we have

$$\left[(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \right]_{r,s} = \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} \left(\frac{1}{n} \sum_{i=1}^n z_i^k \tilde{X}_i^r \right) \left(\frac{1}{n} \sum_{i=1}^n z_i^l \tilde{X}_i^s \right).$$

Because \tilde{X}_i^r is independent of z_i^k , we see

$$\mathbb{E} [z_i^k \tilde{X}_i^r] = \mathbb{E} [z_i^k] \mathbb{E} [\tilde{X}_i^r] = 0.$$

We additionally can upper-bound the variance of $z_i^k \tilde{X}_i^r$:

$$\mathbb{V} [z_i^k \tilde{X}_i^r] = \mathbb{E} [(z_i^k)^2 (\tilde{X}_i^r)^2] = \mathbb{E} [(z_i^k)^2] \mathbb{E} [(\tilde{X}_i^r)^2] = 1.$$

Finally, $z_{i_1}^k \tilde{X}_{i_1}^r$ and $z_{i_2}^k \tilde{X}_{i_2}^r$ are independent and thus uncorrelated for $i_1 \neq i_2$. By WLLN,

$$\frac{1}{n} \sum_{i=1}^n z_i^k \tilde{X}_i^r \xrightarrow{P} 0.$$

A similar argument shows that

$$\frac{1}{n} \sum_{i=1}^n z_i^l \tilde{X}_i^s \xrightarrow{P} 0.$$

Moreover, $\hat{P}_{kl} \xrightarrow{P} P_{kl}$. Hence,

$$\left[(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \right]_{r,s} \xrightarrow{P} \sum_{k=1}^p \sum_{l=1}^p P_{kl} \cdot 0 \cdot 0 = 0.$$

Because r and s were chosen arbitrarily,

$$(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \xrightarrow{P} 0.$$

Finally, it is easy to check that $\tilde{\mathbf{X}}_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} \xrightarrow{P} \mathbf{I}$. We conclude that the lower right block of I_X , $\text{LR}(\mathbf{I}_X)$, converges in probability to \mathbf{I} :

$$\text{LR}(\mathbf{I}_X) = \tilde{\mathbf{X}}_{\sqrt{n}}^T \tilde{\mathbf{X}}_{\sqrt{n}} - (1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} \tilde{\mathbf{X}}_{\sqrt{n}} \xrightarrow{P} \mathbf{I}.$$

Lower left and upper right blocks. The lower left block $\text{LL}(\mathcal{I}_X)$ of \mathcal{I}_X is

$$\text{LL}(\mathcal{I}_X) = \tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} - (1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \in \mathbb{R}^B.$$

Again, let $\tilde{X}_{\sqrt{n}}^r$ denote the r th column of $\tilde{\mathbf{X}}_{\sqrt{n}}$. By the decomposition (1),

$$\left[(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \right]_r = \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} \left(\frac{1}{n} \sum_{i=1}^n \tilde{X}_i^r Z_i^k \right) \left(\frac{1}{n} \sum_{j=1}^n X_j Z_j^l \right).$$

As we showed above, the first sum converges to 0:

$$\frac{1}{n} \sum_{i=1}^n \tilde{X}_i^r Z_i^k \xrightarrow{P} 0.$$

Consider now the second sum. The mean $\mathbb{E}(X_j Z_j^l)$ exists, as $|\mathbb{E}(X_j Z_j^l)| \leq \sqrt{\mathbb{E}[X_j^2] \mathbb{E}[(Z_j^l)^2]} = 1$. Furthermore,

$$\mathbb{V}[X_j(Z_j^l)] \leq \mathbb{E}[X_j^2(Z_j^l)^2] \leq \sqrt{\mathbb{E}[X_j^4] \mathbb{E}[(Z_j^l)^4]}.$$

Additionally, the terms in the sequence are independent and thus uncorrelated. Thus, by WLLN,

$$\frac{1}{n} \sum_{i=1}^n X_j Z_j^l \xrightarrow{P} \mathbb{E} [X_j^2 (Z_j^l)^2].$$

Finally, $\hat{P}_{kl} \xrightarrow{P} P_{kl}$. Piecing these results together,

$$\left[(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \right]_r \xrightarrow{P} \sum_{k=1}^p \sum_{l=1}^p P_{kl} \cdot \mathbb{E} [X_j^2 (Z_j^l)^2] \cdot 0 = 0.$$

Because the elements of $(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}}$ are equal in distribution,

$$(1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \xrightarrow{P} 0.$$

Next, by the independence of $\tilde{\mathbf{X}}$ and X , it is clear that

$$\tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} \xrightarrow{P} 0.$$

Hence, the lower left block of \mathcal{I}_X converges to the zero vector, i.e.

$$\text{LL}(\mathcal{I}_X) = \tilde{\mathbf{X}}_{\sqrt{n}}^T X_{\sqrt{n}} - (1/n) \tilde{\mathbf{X}}_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \xrightarrow{P} 0.$$

The upper right block of \mathcal{I}_X , $\text{UR}(\mathcal{I}_X)$, is the transpose of the lower-left block. Thus, it too converges to the (transposed) zero vector:

$$\text{UR}(\mathcal{I}_X) = \text{LL}(\mathcal{I}_X)^T \xrightarrow{P} 0.$$

Upper left block

The upper left block $\text{UL}(\mathcal{I}_X)$ of \mathcal{I}_X is a scalar.

$$\text{UL}(\mathcal{I}_X) = X_{\sqrt{n}}^T X_{\sqrt{n}} - \frac{1}{n} X_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}}.$$

By the decomposition (1),

$$\frac{1}{n} X_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} = \sum_{k=1}^p \sum_{l=1}^p \hat{P}_{kl} \left(\frac{1}{n} \sum_{i=1}^n z_i^k X_i \right) \left(\frac{1}{n} \sum_{j=1}^n z_j^l X_j \right).$$

Arguing as above, by WLNN,

$$\left(\frac{1}{n} \sum_{i=1}^n z_i^k X_i\right) \left(\frac{1}{n} \sum_{j=1}^n z_j^l X_j\right) \xrightarrow{P} (\mathbb{E}(Z_i^k X_i)) \mathbb{E}(Z_i^l X_i).$$

Next, because $\hat{P}_{kl} \xrightarrow{P} P_{kl}$, we see that

$$\frac{1}{n} X_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}} \xrightarrow{P} \sum_{k=1}^p \sum_{l=1}^p P_{kl} \mathbb{E}(Z_i^k X_i) \mathbb{E}(Z_i^l X_i).$$

Let $\rho_x \in \mathbb{R}^p$ denote the vector

$$\begin{bmatrix} \text{Cov}(X_i Z_i^1) \\ \text{Cov}(X_i Z_i^2) \\ \vdots \\ \text{Cov}(X_i Z_i^p) \end{bmatrix}.$$

We can express the limit (in probability) of $(1/n) X_{\sqrt{n}}^T \mathbf{A} X_{\sqrt{n}}$ as the following quadratic form:

$$\sum_{k=1}^p \sum_{l=1}^p P_{kl} \mathbb{E}(Z_i^k X_i) \mathbb{E}(Z_i^l X_i) = \rho_x^T P \rho_x = \|\rho_x\|_P^2,$$

where $\|a\|_P^2 = a^T P a$ denotes the weighted norm associated with matrix P .

Putting the pieces together

Piecing together $\text{LR}(\mathcal{I}_X)$, $\text{LL}(\mathcal{I}_X)$, $\text{UR}(\mathcal{I}_X)$, and $\text{UL}(\mathcal{I}_X)$, we conclude that

$$\mathcal{I}_X \xrightarrow{P} \text{diag}\{\|\rho_x\|_P^2, 1, \dots, 1\} := \mathcal{I}_{\text{lim}}.$$

By the continuous mapping theorem,

$$\mathcal{I}_X^{-1/2} \xrightarrow{P} \text{diag}\left\{\frac{1}{\|\rho_x\|_P}, 1, \dots, 1\right\} = \mathcal{I}_{\text{lim}}^{-1/2}.$$

Furthermore, $\mathcal{I}_{\text{lim}}^{-1/2} \mathcal{I}_X^{1/2} \xrightarrow{P} \mathbf{I}$. Hence, by Slutsky's theorem,

$$\begin{aligned} & \mathcal{I}_{\text{lim}}^{-1/2} \mathcal{I}_X^{1/2} \mathcal{I}_X^{-1/2} u_X(\hat{\alpha}, 0, 0) \\ &= \mathcal{I}_{\text{lim}}^{-1/2} u_X(\hat{\alpha}, 0, 0) \xrightarrow{d} \mathbf{I} N_{B+1}(0, \mathbf{I}) = N_{B+1}(0, \mathbf{I}). \quad (2) \end{aligned}$$

But

$$\text{diag}\{1 - \|\mathbf{Z}X_{\sqrt{n}}\|_P^2, 1, \dots, 1\} \xrightarrow{P} \mathbf{I}_{\text{lim}},$$

implying

$$\text{diag}\left\{\frac{1}{\sqrt{1 - \|\mathbf{Z}^T X_{\sqrt{n}}\|_P^2}}, 1, \dots, 1\right\} \mathbf{I}_{\text{lim}}^{1/2} \xrightarrow{P} \mathbf{I}.$$

Hence, applying Slutsky's theorem to (2),

$$\begin{aligned} \text{diag}\left\{\frac{1}{\sqrt{1 - \|\mathbf{Z}^T X_{\sqrt{n}}\|_P^2}}, 1, \dots, 1\right\} \mathbf{I}_{\text{lim}}^{1/2} \mathbf{I}_{\text{lim}}^{-1/2} u_X(\hat{\alpha}, 0, 0) \\ \xrightarrow{P} \mathbf{I} N_{B+1}(0, \mathbf{I}) = N_{B+1}(0, \mathbf{I}). \end{aligned}$$

We conclude that

$$\begin{bmatrix} T_0 \\ T_1 \\ T_2 \\ \vdots \\ T_B \end{bmatrix} = \begin{bmatrix} X_{\sqrt{n}}^T \hat{r} / \sqrt{1 - \|\mathbf{Z}^T X_{\sqrt{n}}\|_P^2} \\ \left(\tilde{X}_{\sqrt{n}}^1\right)^T \hat{r} \\ \left(\tilde{X}_{\sqrt{n}}^2\right)^T \hat{r} \\ \vdots \\ \left(\tilde{X}_{\sqrt{n}}^B\right)^T \hat{r} \end{bmatrix} \xrightarrow{d} N_{B+1}(0, \mathbf{I}).$$

2 Connection between GLM score statistic and distilled score statistic

In progress