Notation and Assumptions

Suppose $\mathcal{D} = \{\boldsymbol{z}_i\}_{1:n}$ are independent and identically distributed observations. Consider a model for $\mathcal{D}$ parameterized by $\boldsymbol{\theta}$. Let $\ell(\boldsymbol{\theta})$ denote the log likelihood:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell(\boldsymbol{\theta}; \boldsymbol{z}_i).$$

Denote the true value of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0$. The score for $\boldsymbol{\theta}$ is:

$$\dot{\ell}_\theta(\boldsymbol{\theta}_0) \equiv \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

The Hessian for $\boldsymbol{\theta}$ is:

$$\ddot{\ell}_{\theta\theta'}(\boldsymbol{\theta}_0) \equiv \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

Partition $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$, where $\boldsymbol{\beta}$ is the target parameter, and $\boldsymbol{\alpha}$ is a nuisance parameter. The null hypothesis will take the form $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$. Let $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}})$ denote the unrestricted MLE of $\boldsymbol{\theta}$, which satisfies:

$$\dot{\ell}_\beta(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \boldsymbol{0}, \qquad \dot{\ell}_\alpha(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\alpha}}) = \boldsymbol{0}.$$

Let $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}})$ denote the restricted MLE of $\boldsymbol{\theta}$, which satisfies:

$$\dot{\ell}_\alpha(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) = \boldsymbol{0}.$$

Assume sufficient regularity that, under H_0, the restricted and unrestricted MLEs are each $\sqrt{n}$-consistent:

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \mathcal{O}(n^{-1/2}), \qquad \hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 + \mathcal{O}(n^{-1/2}).$$

The score equation converges in distribution as:

$$n^{-1/2}\dot{\ell}_\theta(\boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \dot{\ell}_{\theta,i}(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N\big\{\boldsymbol{0}, \boldsymbol{B}(\boldsymbol{\theta}_0)\big\},$$

where $\boldsymbol{B}$ is the expected outer product of the score:

$$\boldsymbol{B}_0 \equiv \boldsymbol{B}(\boldsymbol{\theta}_0) \equiv E_Z\big\{\dot{\ell}_{\theta,i}(\boldsymbol{\theta}_0) \otimes \dot{\ell}_{\theta,i}(\boldsymbol{\theta}_0)\big\}.$$

The negative Hessian converges in probability as:

$$-n^{-1}\ddot{\ell}_{\theta\theta'}(\boldsymbol{\theta}_0) = -\frac{1}{n} \sum_{i=1}^{n} \ddot{\ell}_{\theta\theta',i}(\boldsymbol{\theta}_0) = \boldsymbol{A}(\boldsymbol{\theta}_0) + o_p(1),$$

where $\boldsymbol{A}$ is the expected information matrix:

$$\boldsymbol{A}_0 \equiv \boldsymbol{A}(\boldsymbol{\theta}_0) = E_Z\big\{ -\ddot{\ell}_{\theta\theta',i}(\boldsymbol{\theta}_0)\big\}.$$

Moreover, suppose sufficient regularity that:

$$-n^{-1}\ddot{\ell}_{\theta\theta'}(\tilde{\boldsymbol{\theta}}) = \boldsymbol{A}(\boldsymbol{\theta}_0) + o_p(1), \qquad -n^{-1}\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}}) = \boldsymbol{A}(\boldsymbol{\theta}_0) + o_p(1).$$

Let $\boldsymbol{\Omega}_0$ denote the covariance matrix:

$$\boldsymbol{\Omega}_0 \equiv \boldsymbol{\Omega}(\boldsymbol{\theta}_0) = \boldsymbol{A}(\boldsymbol{\theta}_0)^{-1}\boldsymbol{B}(\boldsymbol{\theta}_0)\boldsymbol{A}(\boldsymbol{\theta}_0)^{-T}.$$

Within the exponential family $\boldsymbol{A}_0 = \boldsymbol{B}_0$ such that:

$$\boldsymbol{\Omega}_0 = \boldsymbol{A}_0^{-1},$$

which is the inverse expected information.

Wald Test

Proposition 2.1. Under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the Wald statistic converges as:

$$n\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)\boldsymbol{\Omega}_0^{-1}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \xrightarrow{\mathcal{L}} \chi_r^2,$$

where $r = \dim(\boldsymbol{\theta}_0)$. ♦

Proof. Taylor expand of the score for $\boldsymbol{\theta}$ about the truth:

$$\mathbf{0} = \dot{\ell}_\theta(\hat{\boldsymbol{\theta}}) = \dot{\ell}(\boldsymbol{\theta}_0) + \ddot{\ell}_{\theta\theta'}(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||^2)$$

Solving for $(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$:

$$(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \big\{ - \ddot{\ell}_{\theta\theta'}(\boldsymbol{\theta}_0)\big\}^{-1}\dot{\ell}_\theta(\boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1}).$$

Scaling by $\sqrt{n}$:

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \big\{ - n^{-1}\ddot{\ell}_{\theta\theta'}(\boldsymbol{\theta}_0)\big\}^{-1} n^{-1/2}\dot{\ell}_\theta(\boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-1/2}) \\ &= \boldsymbol{A}_0^{-1} n^{-1/2}\dot{\ell}_\theta(\boldsymbol{\theta}_0) + o_p(1) \\ &\xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-T}\big).\end{aligned}$$

■

Corollary 2.1. The Wald test of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is:

$$n\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big)\boldsymbol{\Omega}_{0,\beta\beta'}^{-1}\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big) \xrightarrow{\mathcal{L}} \chi_p^2,$$

where $p = \dim(\boldsymbol{\beta}_0)$ and $\boldsymbol{\Omega}_{0,\beta\beta'}$ is the sub-matrix of $\boldsymbol{\Omega}_0$ corresponding to $\boldsymbol{\beta}$. ♣

Remark 2.1. Within the exponential family $\boldsymbol{\Omega}_{0,\beta\beta'}^{-1}$ is the efficient information for $\boldsymbol{\beta}$:

$$\boldsymbol{\Omega}_{0,\beta\beta'}^{-1} = \boldsymbol{A}_{\beta\beta'} - \boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{A}_{\alpha\beta'}.$$

♦

Score Test

Proposition 3.1. Under $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the score statistic converges as:

$$\frac{1}{n}\dot{\ell}_\theta(\boldsymbol{\theta}_0)'\boldsymbol{B}_0^{-1}\dot{\ell}_\theta(\boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} \chi^2_r,$$

where $r = \dim(\boldsymbol{\theta}_0)$. ♦

Proposition 3.2. The score test of $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is:

$$\frac{1}{n}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}})'\big\{\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0'\big\}^{-1}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}})' \xrightarrow{\mathcal{L}} \chi^2_p,$$

where $p = \dim(\boldsymbol{\beta}_0)$ and

$$\boldsymbol{C}_0 = \big(\boldsymbol{I}, -\boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\big).$$

♦

Proof. Taylor expand the scores for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ about the constrained MLE:

$$\begin{aligned}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) &= \dot{\ell}_\beta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) + \ddot{\ell}_{\beta\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \mathcal{O}_p(n^{-1}),\\ 0 = \dot{\ell}_\alpha(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) &= \dot{\ell}_\alpha(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) + \ddot{\ell}_{\alpha\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)(\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + \mathcal{O}_p(n^{-1}).\end{aligned}$$

Substituting the second expansion into the first:

$$\begin{aligned}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) &= \dot{\ell}_\beta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) - \ddot{\ell}_{\beta\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\big\{\ddot{\ell}_{\alpha\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\big\}^{-1}\dot{\ell}_\alpha(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\\ &= \Big(\boldsymbol{I}, -\ddot{\ell}_{\beta\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\big\{\ddot{\ell}_{\alpha\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\big\}^{-1}\Big)\begin{pmatrix}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\\ \dot{\ell}_\alpha(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\end{pmatrix}\\ &= \Big(\boldsymbol{I}, -n^{-1}\ddot{\ell}_{\beta\alpha'}\big\{n^{-1}\ddot{\ell}_{\alpha\alpha'}(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\big\}^{-1}\Big)\dot{\ell}_\theta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0)\\ &= \big(\boldsymbol{I}, -\boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\big)\dot{\ell}_\theta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) + o_p(1).\end{aligned}$$

Let $\boldsymbol{C}_0 = \big(\boldsymbol{I}, -\boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\big)$, then:

$$\frac{1}{\sqrt{n}}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) = \boldsymbol{C}_0 \cdot \frac{1}{\sqrt{n}}\dot{\ell}_\theta(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0) + o_p(1) \xrightarrow{\mathcal{L}} N\big(\boldsymbol{0}, \boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0'\big).$$

The asymptotic covariance is:

$$\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0' = \boldsymbol{B}_{\beta\beta'} - \boldsymbol{B}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{A}_{\beta\alpha'}^T - \boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{B}_{\alpha\beta'} + \boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{B}_{\alpha\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{A}_{\beta\alpha'}^T.$$

■

Corollary 3.1. Within the exponential family:

$$\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0' = \boldsymbol{A}_{\beta\beta'} - \boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\boldsymbol{A}_{\alpha\beta'} = \big(\boldsymbol{A}_0^{-1}\big)^{-1}_{\beta\beta'},$$

which is the efficient information for $\boldsymbol{\beta}$. ♣

Likelihood Ratio Test

Remark 4.1. This section assumes $\boldsymbol{A}_0$ is symmetric and positive definite. ♦

Proposition 4.1. Under the $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, the likelihood ratio statistic converges as:

$$2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)\big\} \xrightarrow{\mathcal{L}} \sum_{j=1}^{r} \lambda_j \chi_1^2,$$

where $(\lambda_1, \cdots, \lambda_r)$ are the eigenvalues of $\boldsymbol{A}_0^{-1/2}\boldsymbol{B}_0\boldsymbol{A}_0^{-1/2}$ and $r = \dim(\boldsymbol{\theta}_0)$. ♦

Proof. Taylor expand the log likelihood at the truth about the unconstrained MLE:

$$\ell(\boldsymbol{\theta}_0) = \ell(\hat{\boldsymbol{\theta}}) + \dot{\ell}_\theta(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) + \mathcal{O}_p(n^{-3/2}).$$

Since $\dot{\ell}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}$, upon rearranging:

$$\begin{aligned}
2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)\big\} &= \big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)'\big\{-\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-3/2}) \\
&= n\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)'\big\{-n^{-1}\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big\}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + \mathcal{O}_p(n^{-3/2}) \\
&= n\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)'\boldsymbol{A}_0\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p(1).
\end{aligned}$$

Recall from the Wald statistic that:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) \xrightarrow{\mathcal{L}} N\big(\boldsymbol{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\big).$$

Consequently, the quadratic form

$$\begin{aligned}
2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)\big\} &= n\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big)'\boldsymbol{A}_0\big(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\big) + o_p(1) \\
&= \big\{\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big\}\boldsymbol{A}_0\big\{\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big\} \xrightarrow{\mathcal{L}} \boldsymbol{\omega}'\boldsymbol{A}_0\boldsymbol{\omega}
\end{aligned}$$

where $\boldsymbol{\omega} \sim N\big(\boldsymbol{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\big)$. Let $\boldsymbol{L}\boldsymbol{L}'$ denote the Cholesky decomposition of $\boldsymbol{A}_0$, then:

$$\boldsymbol{\omega}'\boldsymbol{A}_0\boldsymbol{\omega} = \boldsymbol{\omega}'\boldsymbol{L}\boldsymbol{L}'\boldsymbol{\omega} \stackrel{d}{=} \sum_{j=1}^{r} \lambda_j \chi_1^2,$$

where $(\lambda_1, \cdots, \lambda_r)$ are the eigenvalues of $\boldsymbol{L}'\boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}\boldsymbol{L}$. ■

Corollary 4.1. Within the exponential family:

$$\boldsymbol{A}_0^{-1/2}\boldsymbol{B}_0\boldsymbol{A}_0^{-1/2} = \boldsymbol{A}_0^{-1/2}\boldsymbol{A}_0\boldsymbol{A}_0^{-1/2} = \boldsymbol{I}.$$

Consequently, $\lambda_j = 1$ for each j, and :

$$2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\boldsymbol{\theta}_0)\big\} \xrightarrow{\mathcal{L}} \chi_r^2.$$

♣

Proposition 4.2. The likelihood ratio test $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ is:

$$2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\big\} \xrightarrow{\mathcal{L}} \sum_{j=1}^{p} \lambda_j \chi_1^2,$$

where $(\lambda_1, \cdots, \lambda_p)$ are the eigenvalues of:

$$\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}^{1/2} \boldsymbol{C}_0 \boldsymbol{B}_0 \boldsymbol{C}_0' \big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'},$$

$\boldsymbol{C}_0 = \big(\boldsymbol{I}, -\boldsymbol{A}_{\beta\alpha'}\boldsymbol{A}_{\alpha\alpha'}^{-1}\big)$, and $p = \dim(\boldsymbol{\beta}_0)$. ♦

Proof. Taylor expand the log likelihood at the constrained MLE about the unconstrained MLE:

$$\ell(\tilde{\boldsymbol{\theta}}) = \ell(\hat{\boldsymbol{\theta}}) + \dot{\ell}_\theta(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \frac{1}{2}\big(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\big)' \ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}\big) + \mathcal{O}_p(n^{-3/2}).$$

Since $\dot{\ell}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, upon rearranging:

$$\begin{aligned}
2\big\{\ell(\hat{\boldsymbol{\theta}}) - \ell(\tilde{\boldsymbol{\theta}})\big\} &= \big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big)'\big\{-\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big\}\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big) + \mathcal{O}_p(n^{-3/2}) \\
&= n\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big)'\big\{-n^{-1}\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big\}\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big) + \mathcal{O}_p(n^{-3/2}) \\
&= n\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big)'\boldsymbol{A}_0\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big) + o_p(1).
\end{aligned}$$

Taylor expand the score at the constrained MLE to obtain:

$$\dot{\ell}_\theta(\tilde{\boldsymbol{\theta}}) = \dot{\ell}_\theta(\hat{\boldsymbol{\theta}}) + \ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \mathcal{O}_p(n^{-1}).$$

Since $\dot{\ell}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$, upon rearranging:

$$\begin{aligned}
\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big) &= \big\{-n^{-1}\ddot{\ell}_{\theta\theta'}(\hat{\boldsymbol{\theta}})\big\}^{-1}\frac{1}{\sqrt{n}}\dot{\ell}_\theta(\tilde{\boldsymbol{\theta}}) + \mathcal{O}_p(n^{-1}) \\
&= \boldsymbol{A}_0^{-1}\frac{1}{\sqrt{n}}\dot{\ell}_\theta(\tilde{\boldsymbol{\theta}}) + o_p(1) \\
&= \boldsymbol{A}_0^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) \\ \dot{\ell}_\alpha(\boldsymbol{\beta}, \tilde{\boldsymbol{\alpha}})\end{pmatrix} + o_p(1) \\
&= \boldsymbol{A}_0^{-1}\frac{1}{\sqrt{n}}\begin{pmatrix}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\alpha}}) \\ \mathbf{0}\end{pmatrix} + o_p(1).
\end{aligned}$$

Recall from the score statistic that:

$$\frac{1}{\sqrt{n}}\dot{\ell}_\beta(\boldsymbol{\beta}_0, \tilde{\alpha}) \xrightarrow{\mathcal{L}} N\big(\mathbf{0}, \boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0'\big).$$

Thus:

$$\sqrt{n}\big(\hat{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}\big) \xrightarrow{\mathcal{L}} \boldsymbol{A}_0^{-1}\begin{pmatrix}N(\mathbf{0}, \boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0') \\ \mathbf{0}\end{pmatrix}.$$

The limiting distribution of the quadratic form is:

$$\begin{aligned}2\big\{\ell(\hat{\boldsymbol{\theta}})-\ell(\tilde{\boldsymbol{\theta}})\big\} &= n\big(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\big)'\boldsymbol{L}\boldsymbol{L}'\big(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}}\big)+o_p(1),\\ &= \big\{\sqrt{n}(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}})\big\}'\boldsymbol{A}_0\big\{\sqrt{n}(\hat{\boldsymbol{\theta}}-\tilde{\boldsymbol{\theta}})\big\}+o_p(1)\\ &\xrightarrow{\mathcal{L}} \big(\boldsymbol{\omega},\boldsymbol{0}\big)'\boldsymbol{A}_0^{-1}\boldsymbol{A}_0\boldsymbol{A}_0^{-1}\begin{pmatrix}\boldsymbol{\omega}\\ \boldsymbol{0}\end{pmatrix} = \boldsymbol{\omega}'\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}\boldsymbol{\omega},\end{aligned}$$

where $\boldsymbol{\omega}\sim N(\boldsymbol{0},\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0')$ and $\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}$ is the $p\times p$ sub-matrix of $\boldsymbol{A}_0^{-1}$ corresponding to $\boldsymbol{\beta}$. Let $\boldsymbol{L}\boldsymbol{L}'$ denote the Cholesky decomposition of $\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}$. Then:

$$\boldsymbol{\omega}'\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}\boldsymbol{\omega} = \boldsymbol{\omega}'\boldsymbol{L}\boldsymbol{L}'\boldsymbol{\omega} \overset{d}{=} \sum_{j=1}^{p}\lambda_j\chi_1^2,$$

where $(\lambda_1,\cdots,\lambda_p)$ are the eigenvalues of $\boldsymbol{L}'\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0'\boldsymbol{L}$.

■

Corollary 4.2. Within the exponential family $\boldsymbol{C}_0\boldsymbol{B}_0\boldsymbol{C}_0' = (\boldsymbol{A}_0^{-1})_{\beta\beta'}^{-1}$, such that:

$$\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}^{1/2}(\boldsymbol{A}_0^{-1})_{\beta\beta'}^{-1}\big(\boldsymbol{A}_0^{-1}\big)_{\beta\beta'}^{1/2} = \boldsymbol{I}.$$

Consequently, $\lambda_j = 1$ for each j, and:

$$2\big\{\ell(\hat{\boldsymbol{\theta}})-\ell(\tilde{\boldsymbol{\theta}})\big\} \xrightarrow{\mathcal{L}} \chi_p^2.$$

♣