SUPPLEMENTARY MATERIAL

Supplement B: Proof of Lemma 3.1

(doi: [10.1214/17-AOAS1028SUPPB](https://doi.org/10.1214/17-AOAS1028SUPPB); .pdf).

PROOF. Consider the estimating equation of ridge regression for the Poisson family with a log link

$$(1) \quad \frac{1}{n} \mathbf{X}^T (\mathbf{Y} - e^{\log(t) + \mathbf{X}\beta}) + \mathbf{\Lambda}\beta \triangleq \mathbf{U}_n(\beta) + \mathbf{\Lambda}\beta = \mathbf{0},$$

where we define

$$\mathbf{U}_n(\beta) = \frac{1}{n} \mathbf{X}^T (\mathbf{Y} - e^{\log(t) + \mathbf{X}\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i (Y_i - e^{\log(t_i) + \mathbf{X}_i^T \beta})$$

is the estimating function of the generalized linear model

$$(2) \quad \log[E(\mathbf{Y})] = \log(t) + \mathbf{X}\beta.$$

We first state some classical results for the generalized linear model. For the random variables $(\mathbf{X}, Y, t)$, we define the function

$$\mathbf{U}(\beta) = \mathbf{X}(Y - e^{\log(t) + \mathbf{X}^T \beta}).$$

Since β^* is the population coefficient satisfying (2), we have $E[\mathbf{U}(\beta^*)] = 0$. For notational convenience, we let $\mathbf{\Omega}(\beta^*) = \text{Var}[\mathbf{U}(\beta^*)]$. By the Central Limit Theorem, $\sqrt{n}\mathbf{U}_n(\beta^*) \xrightarrow{d} N(0, \mathbf{\Omega}(\beta^*))$. We will use these results in the following proof.

We now prove Lemma 3.1. By second order Taylor expansion on (1), we obtain

$$(3) \quad 0 = [\mathbf{H}_n(\hat{\beta}_\lambda) + \mathbf{\Lambda}](\hat{\beta}_\lambda - \beta^*) + \mathbf{U}_n(\beta^*) + \mathbf{\Lambda}\beta^* + o_p(1),$$

where $\mathbf{H}_n(\hat{\beta}_\lambda) = \mathbf{X}^T \text{diag}(e^{\log(t) + \mathbf{X}\hat{\beta}_\lambda}) \mathbf{X} / n$ is obtained by taking derivative of $\mathbf{U}_n(\hat{\beta}_\lambda)$ with respect to $\hat{\beta}_\lambda$. By rearranging the terms in (3) and multiplying both sides of the equation by $\sqrt{n}$, we obtain

$$\sqrt{n}\{[\mathbf{H}_n(\hat{\beta}_\lambda) + \mathbf{\Lambda}]\hat{\beta}_\lambda - \mathbf{H}_n(\hat{\beta}_\lambda)\beta^*\} = -\sqrt{n}\mathbf{U}_n(\beta^*) + o_p(1).$$

Equivalently, we have

$$(4) \quad \sqrt{n}[\mathbf{H}_n(\hat{\beta}_\lambda) + \mathbf{\Lambda}]\{\hat{\beta}_\lambda - [\mathbf{H}_n(\hat{\beta}_\lambda) + \mathbf{\Lambda}]^{-1}\mathbf{H}_n(\hat{\beta}_\lambda)\beta^*\} \xrightarrow{d} N(0, \mathbf{\Omega}(\beta^*)),$$

where we use the fact that $\sqrt{n}\mathbf{U}_n(\boldsymbol{\beta}^*) \xrightarrow{d} N(0, \boldsymbol{\Omega}(\boldsymbol{\beta}^*))$.

For a given $\boldsymbol{\Lambda}$, let $\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*$ be the population ridge coefficient satisfying $E[\mathbf{U}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*) + \boldsymbol{\Lambda}\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*] = 0$. By continuous mapping theorem, we have $\mathbf{H}_n(\hat{\boldsymbol{\beta}}_{\lambda}) \xrightarrow{p} \mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*)$. Thus, by Slutsky's theorem and (4), we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\lambda} - [\mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*) + \boldsymbol{\Lambda}]^{-1} \mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*) \boldsymbol{\beta}^*) \xrightarrow{d} N(0, [\mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*) + \boldsymbol{\Lambda}]^{-1} \boldsymbol{\Omega}(\boldsymbol{\beta}^*) [\mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*) + \boldsymbol{\Lambda}]^{-1}).$$

Note that $\hat{\boldsymbol{\beta}}_{\lambda}$ is not an unbiased estimator of $\boldsymbol{\beta}^*$. We propose the following de-bias estimator for estimating $\boldsymbol{\beta}^*$:

$$(5) \quad \hat{\boldsymbol{\beta}}_{\text{debias}} = \{[\mathbf{H}_n(\hat{\boldsymbol{\beta}}_{\lambda}) + \boldsymbol{\Lambda}]^{-1} \mathbf{H}_n(\hat{\boldsymbol{\beta}}_{\lambda})\}^{-1} \hat{\boldsymbol{\beta}}_{\lambda}.$$

By (4) and (5),

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{debias}} - \boldsymbol{\beta}^*) \xrightarrow{d} N(0, [\mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*)]^{-1} \boldsymbol{\Omega}(\boldsymbol{\beta}^*) [\mathbf{H}(\boldsymbol{\beta}_{\boldsymbol{\Lambda}}^*)]^{-1}).$$

□