# Solutions to GLM questions

The subscript $i$ indicates the $i$th sample. We can add or remove this subscript in the following proof. If the subscript is removed, it means a generic sample.

1. $g(Y_i)$ can be approximated by the Taylor series at $\mu_i$ as follows:

$$g(Y_i) \approx g(\mu_i) + g'(\mu_i)\frac{Y_i - \mu_i}{1!}.$$

Since $\mu_i$ is constant, we have

$$\mathrm{Var}(g(Y_i)) \approx \mathrm{Var}(Y_i)g'(\mu_i)^2 = \phi V(\mu_i)g'(\mu_i)^2.$$

2. From the calculus, assuming $g(y) = x$, we have

$$\frac{dg(y)}{dx} = \frac{dg(y)}{dy}\frac{dy}{dx} = 1.$$

That is,

$$\frac{dg^{-1}(x)}{dx} = \frac{1}{dg(y)/dy}.$$

Approximate $\hat{\mu}_i$ at $\mu_i = g^{-1}(\boldsymbol{x}_i^T\beta)$ using the following Taylor series:

$$\hat{\mu}_i = \mu_i + \frac{dg^{-1}(\boldsymbol{x}_i^T\beta)}{d\boldsymbol{x}_i^T\beta}\frac{\boldsymbol{x}_i^T\hat{\beta} - \boldsymbol{x}_i^T\beta}{1!}$$

Since $\mu_i, \boldsymbol{x}_i^T\beta$ are constant, we have

$$\mathrm{Var}(\hat{\mu}_i) \approx \frac{\mathrm{Var}(\boldsymbol{x}_i^T\hat{\beta})}{g'(\mu_i)^2} \approx \frac{\boldsymbol{x}_i^T\mathrm{Var}(\hat{\beta})\boldsymbol{x}_i}{g'(\hat{\mu}_i)^2}.$$

3. Note that $Y$ and $\hat{Y}$ are independent.

$$\mathbb{E}\left[\left(Y - \hat{Y}\right)^2\right] = \mathrm{Var}(Y - \hat{Y}) + \left(\mathbb{E}[Y - \hat{Y}]\right)^2 = \mathrm{Var}(\hat{Y}) + \mathrm{Var}(Y) + \left(\mathbb{E}[Y] - \mathbb{E}[\hat{Y}]\right)^2.$$

4. For the sample $i$, we have

$$\phi l_i(\hat{\beta}_{full}) - \phi l_i(\hat{\beta}) = \omega_i\left(y_i\hat{\theta}_i^{full} - b(\hat{\theta}_i^{full}) - y_i\hat{\theta}_i + b(\hat{\theta}_i)\right),$$

which is independent with $\phi$. So the deviance is independent with $\phi$ and the scaled deviance is dependent with $\phi$.

5. We need to prove

$$l_i(\hat{\beta}_{full}) - l_i(\hat{\beta}) \geq 0, \qquad \text{for } i = 1, \dots, n.$$

$\beta$ is related to the nature parameter (or mean parameter) of EDF. So we consider the log-likelihood in terms of nature parameter, $l_i(\theta_i)$. We have the following first derivative with respect to $\theta_i$:

$$l_i'(\theta_i) = \frac{dl_i(\theta_i)}{d\theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}.$$

Since $b'(\theta_i) = \mu_i$, $l_i(\theta_i)$ increases with $\mu_i$ in the interval $\mu_i < y_i$ and decreases with $\mu_i$ in the interva $\mu_i > y_i$, indicating that $l_i(\theta_i)$ reaches the maximum when $\mu_i = y_i$. So $l_i(\hat{\beta}_{full}) \geq l_i(\hat{\beta})$.

6. We have $\ln(\lambda(\boldsymbol{x})) = \boldsymbol{x}^T\beta$. Using the results from Q2, we have

$$\text{Var}(\hat{\lambda}(\boldsymbol{x})) \approx \lambda(\boldsymbol{x})^2 \text{Var}(\hat{\eta}_i) \approx \hat{\lambda}(\boldsymbol{x})^2 \boldsymbol{x}^T \text{Var}(\hat{\beta}) \boldsymbol{x}$$

7. We assume that the bias can be ignored. So MSEP is the sum of estimation variance and the process variance. Note that $\hat{Y}_i = \hat{\lambda}_i$.

$$\begin{aligned}
\mathbb{E}\left[\left(Y_i - \hat{Y}_i\right)^2\right] &\approx \text{Var}(\hat{Y}_i) + \text{Var}(Y_i) \\
&\approx \hat{Y}_i^2 \boldsymbol{x}_i^T \text{Var}(\hat{\beta}) \boldsymbol{x}_i + \text{Var}\left(\frac{N_i}{v_i}\right) \\
&= \hat{Y}_i^2 \boldsymbol{x}_i^T \text{Var}(\hat{\beta}) \boldsymbol{x}_i + \frac{v_i \hat{Y}_i}{v_i^2} \\
&= \hat{Y}_i^2 \boldsymbol{x}_i^T \text{Var}(\hat{\beta}) \boldsymbol{x}_i + \frac{\hat{Y}_i}{v_i}
\end{aligned}$$