Coursework 3: STAT 570

Philip Pham

October 14, 2018

1. Consider the Poisson-gamma random effects model given by

$$\begin{aligned} Y_i \mid \mu_i, \theta_i &\sim \operatorname{Poisson}\left(\mu_i \theta_i\right) && (1) \\ \theta_i &\sim \operatorname{Gamma}\left(b, b\right), && (2) \end{aligned}$$

which leads to a negative binomial marginal model with the variance a quadratic function of the mean. Design a simulation study, along the lines of that which produced Table 2.3 in the book (overdispersed Poisson example) to investigate the efficiency and robustness under

- a Poisson model;
- quasi-likelihood with $\mathbb{E}\left[Y\right] = \mu$ and $\operatorname{Var}\left(Y\right) = \alpha\mu$; and
- sandwich estimation.

Use a log-linear model

$$\log \mu_i = \beta_0 + \beta_1 x_i, \tag{3}$$

with $x_i \sim_{\text{iid}} \mathcal{N}\left(0, 1\right)$ for $i = 1, 2, \ldots, n$, and $\beta_0 = -2$ and $\beta_1 = \log 2$.

Simulate for:

- $b \in \{0.2, 1, 10, 1000\}$.
- $n \in \{10, 20, 50, 100, 250\}$.

Summarize what your take away message is after carrying out these simulations.

Solution: Note that

$$\begin{aligned} \mathbb{P}\left(Y_i = y \mid \mu_i\right) &= \int_0^\infty \mathbb{P}\left(Y_i = y \mid \mu_i, \theta_i = \theta\right) \mathbb{P}\left(\theta_i = \theta \mid b\right)\,\mathrm{d}\theta \\ &= \int_0^\infty \left(\frac{(\mu_i\theta)^y}{y!}\exp\left(-\mu_i\theta\right)\right)\left(\frac{b^b}{\Gamma(b)}\theta^{b-1}\exp\left(-b\theta\right)\right)\mathrm{d}\theta \\ &= \frac{\mu_i^y b^b}{y!\Gamma(b)}\int_0^\infty \theta^{b+y-1}\exp\left(-\theta(b+\mu_i)\right)\,\mathrm{d}\theta \\ &= \frac{\Gamma(y+b)}{y!\Gamma(b)}\frac{\mu_i^y b^b}{(\mu_i+b)^{b+y}} = \frac{\Gamma(y+b)}{y!\Gamma(b)}\left(\frac{b}{\mu_i+b}\right)^b\left(\frac{\mu_i}{\mu_i+b}\right)^y \\ &\sim \operatorname{NegativeBinomial}\left(b, \frac{\mu_i}{\mu_i+b}\right). && (4) \end{aligned}$$

By properties of the negative binomial distribution, we have that

$$\begin{aligned}\mathbb{E}\left[Y_i \mid x_i\right] &= \mu_i = \exp\left(\beta_0 + \beta_1 x_i\right) \\ \operatorname{Var}\left(Y_i \mid x_i\right) &= \mu_i \left(1 + \frac{\mu_i}{b}\right).\end{aligned} \tag{5}$$

Thus, smaller values of b correspond to more dispersion.

Poisson Model

In the Poisson model, we assume that $\operatorname{Var}\left(Y_i \mid x_i\right) = \mu_i$, e.g. $b \to \infty$, so we neglect the overdispersion parameter.
In this case, the log-likelihood function is

$$l\left(\beta\right) = \sum_{i=1}^{n} \left[y_i \left(\beta_0 + \beta_1 x_i\right) - \exp\left(\beta_0 + \beta_1 x_i\right) - \sum_{k=1}^{y_i} \log k\right], \tag{6}$$

which gives us the score function

$$S\left(\beta\right) = \sum_{i=1}^{n} \begin{pmatrix} y_i - \exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i y_i - x_i \exp\left(\beta_0 + \beta_1 x_i\right) \end{pmatrix}. \tag{7}$$

We can estimate β by solving for $S\left(\hat{\beta}\right) = \mathbf{0}$, numerically.
We can estimate the variance of the estimates from the Fisher information,

$$\begin{aligned}\operatorname{Var}\left(\hat{\beta}\right) &\approx I_n\left(\hat{\beta}\right)^{-1} \\ &= \left(\sum_{i=1}^{n} \begin{pmatrix} \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) & x_i \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \\ x_i \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) & x_i^2 \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right) \end{pmatrix}\right)^{-1} \\ &= \frac{1}{\left(\sum_{i=1}^{n} \hat{\mu}_i\right)\left(\sum_{i=1}^{n} x_i^2 \hat{\mu}_i\right) - \left(\sum_{i=1}^{n} x_i \hat{\mu}_i\right)^2} \begin{pmatrix} \sum_{i=1}^{n} x_i^2 \hat{\mu}_i & -\sum_{i=1}^{n} x_i \hat{\mu}_i \\ -\sum_{i=1}^{n} x_i \hat{\mu}_i & \sum_{i=1}^{n} \hat{\mu}_i \end{pmatrix},\end{aligned} \tag{8}$$

where $\hat{\mu}_i = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right)$.

Quasi-likelihood

In a quasi-likelihood model, we specify the mean and variance as

$$\begin{aligned}\mathbb{E}\left[Y_i \mid x_i\right] &= \mu_i = \exp\left(\beta_0 + \beta_1 x_i\right) \\ \operatorname{Var}\left(Y_i \mid x_i\right) &= \alpha \mu_i\end{aligned} \tag{9}$$

From Equation 5, we see that this is not quite correct, still, but it is closer to the real model than the Poisson model.
Then, by Equation 2.30 of Wakefield's *Bayesian and Frequentist Regression Methods* our estimating function is

$$\begin{aligned}U\left(\beta\right) &= D^{\intercal} V^{-1} \left(y - \mu\right)/\alpha \\ &= \sum_{i=1}^{n} \begin{pmatrix} \exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i \exp\left(\beta_0 + \beta_1 x_i\right) \end{pmatrix} \frac{y_i - \exp\left(\beta_0 + \beta_1 x_i\right)}{\alpha \exp\left(\beta_0 + \beta_1 x_i\right)} \\ &= \frac{1}{\alpha} \sum_{i=1}^{n} \begin{pmatrix} y_i - \exp\left(\beta_0 + \beta_1 x_i\right) \\ x_i y_i - x_i \exp\left(\beta_0 + \beta_1 x_i\right) \end{pmatrix} = \frac{1}{\alpha} S\left(\beta\right)\end{aligned} \tag{10}$$

from Equation 7. Thus, the maximum quasi-likelihood estimate will be the same as the maximum likelihood estimate from the Poisson model.
Having solved for $\hat{\beta}$, we have

$$\hat{\mu} = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_i\right). \tag{11}$$

by Equation 2.31 of Wakefield's *Bayesian and Frequentist Regression Methods*, we can then compute

$$\hat{\alpha}_n = \frac{1}{n-2}\sum_{i=1}^{n}\frac{\left(y_i - \hat{\mu}_i\right)^2}{\hat{\mu}_i} \tag{12}$$

Then, the variance of our estimates is

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}\right) &\approx \hat{\alpha}_n\left(\hat{D}^\intercal \hat{V}^{-1}\hat{D}\right)^{-1} \\
&= \hat{\alpha}_n\left(\sum_{i=1}^{n}\begin{pmatrix}\hat{\mu}_i & x_i\hat{\mu}_i \\ x_i\hat{\mu}_i & x_i^2\hat{\mu}_i.\end{pmatrix}\right)^{-1} \\
&= \hat{\alpha}_n I_n\left(\hat{\beta}\right)^{-1} \qquad (13)
\end{aligned}$$

from Equation 8.

Sandwich Estimation

In sandwich estimation, we only need to specify an estimating function $G(\beta)$. Then, we can apply Equation 2.43 of Wakefield's *Bayesian and Frequentist Regression Methods* to compute the variance of our estimates:

$$\begin{aligned}
\operatorname{Var}\left(\hat{\beta}\right) &= \hat{A}^{-1}\hat{B}\left(\hat{A}^{-1}\right)^\intercal \\
\hat{A} &= -\frac{\partial}{\partial\beta}G\left(\hat{\beta}\right) \\
\hat{B} &= G\left(\hat{\beta}\right)G\left(\hat{\beta}\right)^\intercal.
\end{aligned}$$

We can reuse the score function from the quasi-likelihood estimate in Equation 10. Thus, our estimate for $\hat{\beta}$ will remain the same.
From Equation 9, we have that

$$\hat{A} = \hat{D}\hat{V}^{-1}\hat{D} = \frac{I_n\left(\hat{\beta}\right)}{\hat{\alpha}_n}. \tag{14}$$

From Equation 10, we have that

$$\begin{aligned}
\hat{B} &= \hat{D}\begin{pmatrix}\frac{\left(y_1-\hat{\mu}_1\right)^2}{\hat{\alpha}_n^2\hat{\mu}_1^2} \\ \vdots \\ \frac{\left(y_n-\hat{\mu}_n\right)^2}{\hat{\alpha}_n^2\hat{\mu}_n^2}\end{pmatrix}\hat{D}^\intercal = \sum_{i=1}^{n}\frac{\left(y_i-\hat{\mu}_i\right)^2}{\hat{\alpha}_n^2\hat{\mu}_i^2}\begin{pmatrix}\hat{\mu}_i^2 & x_i\hat{\mu}_i^2 \\ x_i\hat{\mu}_i^2 & x_i^2\hat{\mu}_i^2\end{pmatrix} \\
&= \frac{1}{\hat{\alpha}_n^2}\sum_{i=1}^{n}\left(y_i-\hat{\mu}_i\right)^2\begin{pmatrix}1 & x_i \\ x_i & x_i^2\end{pmatrix}. \qquad (15)
\end{aligned}$$