Closed form of hessian of negative binomial GLM with model for dispersion and mean parameter

David Fischer

October 19, 2018

Contents

1 Model description

1.1 Likelihood

The negative binomial likelihood of the GLM is:

$$L(y|\theta_i^m, \theta_i^r) = \frac{\Gamma(r(\theta^r) + y)}{y!\Gamma(r(\theta^r))} (\frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)})^y (\frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)})^{r(\theta^r)} \tag{1}$$

The negative binomial log-likelihood of the GLM is:

$$\begin{aligned} LL(y|\theta_i^m, \theta_i^r) &= \log L(y|\theta_i^m, \theta_i^r) \\ &= \log(\Gamma(r(\theta^r) + y)) - \log(y!\Gamma(r(\theta^r))) + y * (\log(m(\theta^m)) - \log(r(\theta^r) + m(\theta^m))) \\ &+ r(\theta^r) * (\log(r(\theta^r)) - \log(r(\theta^r) + m(\theta^m))) \end{aligned} \tag{2}$$

1.2 Parameter models

The mean (m) and dispersion (r) model of the GLM with design matrices X^m and X^r look as follows:

$$r(\theta^r) = \exp(\langle X^r, \theta_i^r \rangle) \tag{3}$$

$$m(\theta^m) = \exp(\langle X^m, \theta_i^m \rangle) \tag{4}$$

1.3 Derivatives of parameter models

Using the chain rule $\frac{f(m)}{d\theta_i^m} = \frac{f(m)}{m} \frac{m}{d\theta_i^m}$ we can decompose the first derivative. Firstly, the derivative of the mean model with respect to its parameters is:

$$\begin{aligned}
\frac{d}{d\theta_i^m} m(\theta^m) &= \frac{d}{d\theta_i^m} \exp(\langle X^m, \theta_i^m \rangle) \\
&= \exp(\langle X^m * \theta^m \rangle) * X_i^m \\
&= m(\theta^m) * X_i^m
\end{aligned} \tag{5}$$

Where X_i^m is the column of X^m that corresponds to θ_i^m. Equivalently for the dispersion model:

$$\begin{aligned}
\frac{d}{d\theta_i^r} r(\theta^r) &= \frac{d}{d\theta_i^r} \exp(\langle X^r, \theta_i^r \rangle) \\
&= \exp(\langle X^r * \theta^r \rangle) * X_i^r \\
&= r(\theta^r) * X_i^r
\end{aligned} \tag{6}$$

2 Miscancellous

2.1 Polygamma function

In the following, we also need the polygamma function ψ to compute the derivative of the gamma function:

$$\frac{d^{(n+1)}}{dy^{(n+1)}} \log(\Gamma(y)) = \psi_n(y) \tag{7}$$

Note that due to the chain rule, both homogenous and heterogenous secondary derivatives of the polygamma function can be computed via ψ_1:

$$\frac{d}{d\theta_i^r} \log \Gamma(r(\theta^r) + y) = r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) \tag{8}$$

$$\begin{aligned}
\frac{d}{d\theta_j^r} \frac{d}{d\theta_i^r} \log \Gamma(r(\theta^r) + y) &= \frac{d}{d\theta_j^r} \Big(r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) \Big) \\
&= X_i^r * \frac{d}{d\theta_j^r} \Big(r(\theta^r) \Big) * \psi_0(r(\theta^r) + y) + X_i^r * r(\theta^r) * \frac{d}{d\theta_j^r} \Big(\psi_0(r(\theta^r) + y) \Big) \\
&= X_i^r * r(\theta^r) * X_j^r * \psi_0(r(\theta^r) + y) + X_i^r * r(\theta^r) * r(\theta^r) * X_j^r * \psi_1(r(\theta^r) + y) \Big) \\
&= r(\theta^r) * X_i^r * X_j^r * \Big(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) \Big)
\end{aligned} \tag{9}$$

The second form of the gamma function that we encounter in the negative binomial log-likelihood is:

$$\begin{aligned}
\frac{d}{d\theta_i^r}\log\left(y!\Gamma(r(\theta^r))\right) &= r(\theta^r) * X_i^r * \frac{1}{y!\Gamma(r(\theta^r))}\frac{d}{d\theta_i^r}\left(y!\Gamma(r(\theta^r))\right) \\
&= r(\theta^r) * X_i^r * \frac{1}{\Gamma(y!r(\theta^r))} y!\Gamma(r(\theta^r)) * \psi_0(r(\theta^r)) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r))
\end{aligned} \tag{10}$$

$$\begin{aligned}
\frac{d}{d\theta_j^r}\frac{d}{d\theta_i^r}\log\left(y!\Gamma(r(\theta^r))\right) &= \frac{d}{d\theta_j^r}\left(r(\theta^r) * X_i^r * \psi_0(r(\theta^r))\right) \\
&= X_i^r * \frac{d}{d\theta_j^r}\left(r(\theta^r)\right) * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * \frac{d}{d\theta_j^r}\left(\psi_0(r(\theta^r))\right) \\
&= X_i^r * r(\theta^r) * X_j^r * \psi_0(r(\theta^r)) + X_i^r * r(\theta^r) * r(\theta^r) * X_j^r * \psi_1(r(\theta^r))\Big) \\
&= r(\theta^r) * X_i^r * X_j^r * \left(\psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r))\right)
\end{aligned} \tag{11}$$

3 Jacobians

3.1 Jacobian mean model

The Jacobians of the GLM wrt to the m model is:

$$\begin{aligned}
\frac{d}{d\theta_i^m}LL(y|\theta^m,\theta^r) &= \frac{d}{d\theta_i^m}\log(\Gamma(r(\theta^r)+y)) - \frac{d}{d\theta_i^m}\log(y!\Gamma(r(\theta^r))) \\
&+ \frac{d}{d\theta_i^m}y * \left(\log(m(\theta^m)) - \log(r(\theta^r)+m(\theta^m))\right) \\
&+ \frac{d}{d\theta_i^m}r(\theta^r)\left(\log(r(\theta^r)) - \log(r(\theta^r)+m(\theta^m))\right) \\
&= y * \left(\frac{d}{d\theta_i^m}\log(m(\theta^m)) - \frac{d}{d\theta_i^m}\log(r(\theta^r)+m(\theta^m))\right) \\
&+ r(\theta^r) * \left(\frac{d}{d\theta_i^m}\log(r(\theta^r)) - \frac{d}{d\theta_i^m}(\log(r(\theta^r)) + m(\theta^m))\right) \\
&= y * (\frac{1}{m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m} - \frac{1}{r(\theta^r)+m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m}) \\
&- \frac{r(\theta^r)}{r(\theta^r)+m(\theta^m)}\frac{dm(\theta^m)}{d\theta_i^m} \\
&= \frac{dm(\theta^m)}{d\theta_i^m}(\frac{y}{m(\theta^m)} - \frac{y+r(\theta^r)}{r(\theta^r)+m(\theta^m)}) \\
&= m(\theta^m) * X_i^m(\frac{y}{m(\theta^m)} - \frac{y+r(\theta^r)}{r(\theta^r)+m(\theta^m)}) \\
&= X_i^m * y - X_i^m * (y+r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r)+m(\theta^m)}
\end{aligned} \tag{12}$$

3.2 Jacobian dispersion model

The Jacobians of the GLM wrt to the r model is:

$$
\begin{aligned}
\frac{d}{d\theta_i^r}LL(y|\theta^m,\theta^r) &= \frac{d}{d\theta_i^r}\log(\Gamma(r(\theta^r)+y)) - \frac{d}{d\theta_i^r}\log(y!\Gamma(r(\theta^r))) \\
&+ y * \frac{d}{d\theta_i^r}\Bigg(\log(m(\theta^m)) - \log(r(\theta^r)+m(\theta^m))\Bigg) \\
&+ \frac{d}{d\theta_i^r}\Bigg(r(\theta^r) * \Bigg(\log(r(\theta^r)) - \log(r(\theta^r)+m(\theta^m))\Bigg)\Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r)+y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- y * \frac{d}{d\theta_i^r}\Bigg(\log(r(\theta^r)+m(\theta^m))\Bigg) \\
&+ \frac{d}{d\theta_i^r}\Bigg(r(\theta^r) * \log(r(\theta^r))\Bigg) - \frac{d}{d\theta_i^r}\Bigg(r(\theta^r)\log(r(\theta^r)+m(\theta^m))\Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r)+y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- y * \frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X_i^r \\
&+ \Bigg(r(\theta^r) * X_i^r * \log(r(\theta^r)) + r(\theta^r) * r(\theta^r) * X_i^r * \frac{1}{r(\theta^r)}\Bigg) \\
&- \Bigg(r(\theta^r) * X_i^r * \log(r(\theta^r)+m(\theta^m)) + r(\theta^r) * r(\theta^r) * X_i^r * \frac{1}{r(\theta^r)+m(\theta^m)}\Bigg) \\
&= r(\theta^r) * X_i^r * \psi_0(r(\theta^r)+y) + r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&- \frac{1}{r(\theta^r)+m(\theta^m)} r(\theta^r) * X_i^r * (r(\theta^r)+y) \\
&+ r(\theta^r) * X_i^r * \Bigg(\log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m))\Bigg) \\
&= r(\theta^r) * X_i^r * \Bigg(\psi_0(r(\theta^r)+y) + \psi_0(r(\theta^r)) \\
&- \frac{r(\theta^r)+y}{r(\theta^r)+m(\theta^m)} + \log(r(\theta^r)) + 1 - \log(r(\theta^r)+m(\theta^m))\Bigg)
\end{aligned} \tag{13}
$$

4 Block-wise entries of hessian matriy

The hessian can be decomposed into a block-wise hessian of the form $H = [[H^{m,m}, H^{m,r}], [H^{r,m}, H^{r,r}]]$, where $H^{r,m} = H^{m,r}$ as H is symmetric.

4.1 Mean-model block $H^{m,m}$

$$
\begin{aligned}
H^{m,m}_{i,j} &= \frac{d^2}{d\theta^m_i d\theta^m_j} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j}\frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^m_j}\left(X^m_i * y - X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X^m_i * (y + r(\theta^r)) \frac{d}{d\theta^m_j} \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{\frac{d}{d\theta^m_j}\Big(m(\theta^m) \Big) * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * \frac{d}{d\theta^m_j}\Big(r(\theta^r) + m(\theta^m) \Big)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * \frac{m(\theta^m) * X^m_j * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * m(\theta^m) * X^m_j}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_j * \frac{r(\theta^r) + m(\theta^m) - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * (y + r(\theta^r)) * m(\theta^m) * X^m_j * \frac{r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X^m_i * X^m_j * m(\theta^m) * \frac{y/r(\theta^r) + 1}{(1 + m(\theta^m)/r(\theta^r))^2}
\end{aligned} \tag{14}
$$

4.2 Dispersion-model block $H^{r,r}$

$$
\begin{aligned}
H^{r,r}_{i,j} &= \frac{d^2}{d\theta^r_i d\theta^r_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j}\frac{d}{d\theta^m_i} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta^r_j}\Bigg(r(\theta^r) * X^r_i * \psi_0(r(\theta^r) + y) \\
&\quad - r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) \\
&\quad - r(\theta^r) * X^r_i * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} \\
&\quad + r(\theta^r) * X^r_i * \log(r(\theta^r)) \\
&\quad + r(\theta^r) * X^r_i - r(\theta^r) * X^r_i * \log(r(\theta^r) + m(\theta^m)) \Bigg) \\
&= \frac{d}{d\theta^r_j}\Big(r(\theta^r) * X^r_i * \psi_0(r(\theta^r) + y) \Big) \\
&\quad + \frac{d}{d\theta^r_j}\Big(r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) \Big) \\
&\quad - \frac{d}{d\theta^r_j}\left(r(\theta^r) * X^r_i * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} \right) \\
&\quad + \frac{d}{d\theta^r_j}\Big(r(\theta^r) * X^r_i * \log(r(\theta^r)) \Big) \\
&\quad + \frac{d}{d\theta^r_j}\Big(r(\theta^r) * X^r_i \Big) \\
&\quad - \frac{d}{d\theta^r_j}\Big(r(\theta^r) * X^r_i * \log(r(\theta^r) + m(\theta^m)) \Big)
\end{aligned} \tag{15}
$$

$$
\begin{aligned}
H^{r,r}_{i,j} = &\Bigg(X^r_j * r(\theta^r) * X^r_i * \psi_0(r(\theta^r) + y) + r(\theta^r) * X^r_i * X^r_j * r(\theta^r) * \psi_1(r(\theta^r) + y)\Bigg) \\
&- \Bigg(X^r_j * r(\theta^r) * X^r_i * \psi_0(r(\theta^r)) + r(\theta^r) * X^r_i * X^r_j * r(\theta^r) * \psi_1(r(\theta^r))\Bigg) \\
&- \Bigg(X^r_j * r(\theta^r) * X^r_i * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} + r(\theta^r) * X^r_i * \frac{X^r_j * r(\theta^r) * (r(\theta^r) + m(\theta^m)) - (r(\theta^r) + y) * X^r_j * r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2}\Bigg) \\
&+ \Bigg(X^r_j * r(\theta^r) * X^r_i * \log(r(\theta^r)) + r(\theta^r) * X^r_i * X^r_j * r(\theta^r) * \frac{1}{r(\theta^r)}\Bigg) \\
&+ \Bigg(X^r_j * r(\theta^r) * X^r_i\Bigg) \\
&- \Bigg(X^r_j * r(\theta^r) * X^r_i * \log(r(\theta^r) + m(\theta^m)) + r(\theta^r) * X^r_i * X^r_j * r(\theta^r) * \frac{1}{r(\theta^r) + m(\theta^m)}\Bigg) \\
&= X^r_i * X^r_j * r(\theta^r)\Bigg(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) - \psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r)) \\
&- \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} - \frac{r(\theta^r) * (m(\theta^m)) - y)}{(r(\theta^r) + m(\theta^m))^2} \\
&+ \log(r(\theta^r)) + 1 + 1 - \log(r(\theta^r) + m(\theta^m)) - \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)}\Bigg) \\
&= X^r_i * X^r_j * r(\theta^r)\Bigg(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) - \psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r)) \\
&- \frac{(r(\theta^r) + y) * (r(\theta^r) + m(\theta^m))}{(r(\theta^r) + m(\theta^m))^2} - \frac{r(\theta^r) * (m(\theta^m)) - y)}{(r(\theta^r) + m(\theta^m))^2} - \frac{r(\theta^r) * (r(\theta^r) + m(\theta^m))}{(r(\theta^r) + m(\theta^m))^2} \\
&+ \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m))\Bigg) \\
&= X^r_i * X^r_j * r(\theta^r)\Bigg(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) - \psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r)) \\
&- \frac{2 * r(\theta^r)^2 + y * r(\theta^r) + 2 * r(\theta^r) * m(\theta^m) + y * m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} - \frac{r(\theta^r) * m(\theta^m)) - r(\theta^r) * y}{(r(\theta^r) + m(\theta^m))^2} \\
&+ \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m))\Bigg) \\
&= X^r_i * X^r_j * r(\theta^r)\Bigg(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) - \psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r)) \\
&- \frac{2 * r(\theta^r)^2 + y * r(\theta^r) + 2 * r(\theta^r) * m(\theta^m) + y * m(\theta^m) + r(\theta^r) * m(\theta^m))}{(r(\theta^r) + m(\theta^m))^2} \\
&+ \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m))\Bigg) \\
&= X^r_i * X^r_j * r(\theta^r)\Bigg(\psi_0(r(\theta^r) + y) + r(\theta^r) * \psi_1(r(\theta^r) + y) - \psi_0(r(\theta^r)) + r(\theta^r) * \psi_1(r(\theta^r)) \\
&- \frac{2 * r(\theta^r) * (r(\theta^r) + m(\theta^m)) + m(\theta^m) * (y + r(\theta^r))}{(r(\theta^r) + m(\theta^m))^2} \\
&+ \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m))\Bigg)
\end{aligned} \tag{16}
$$

4.3 Offdiagonal-model block $H^{r,m}$

Offdiagonal-model block derived from mean-model jacobian:

$$
\begin{aligned}
H_{i,j}^{r,m} &= \frac{d^2}{d\theta_i^m d\theta_j^r} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^r} \frac{d}{d\theta_i^m} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^r} \left(X_i^m * y - X_i^m * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X_i^m * \frac{d}{d\theta_j^r} \left((y + r(\theta^r)) \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X_i^m * \left(r(\theta^r) * X_j^r * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} - (y + r(\theta^r)) * r(\theta^r) * X_j^r * \frac{m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \right) \\
&= -m(\theta^m) * X_i^m * r(\theta^r) * X_j^r * \left(\frac{1}{r(\theta^r) + m(\theta^m)} - \frac{y + r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \right) \\
&= -m(\theta^m) * X_i^m * r(\theta^r) * X_j^r * \frac{r(\theta^r) + m(\theta^m) - y - r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -m(\theta^m) * X_i^m * r(\theta^r) * X_j^r * \frac{m(\theta^m) - y}{(r(\theta^r) + m(\theta^m))^2} \\
&= X_i^m * X_j^r * r(\theta^r) * m(\theta^m) * \frac{y - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2}
\end{aligned}
\tag{17}
$$