

Closed form of the Jacobian and the hessian matrix of negative binomial GLM

David Fischer

October 22, 2018

Contents

1	Model description	1
1.1	Likelihood	1
1.2	Parameter models	2
1.3	Derivatives of parameter models	2
2	Miscellaneous	2
2.1	Polygamma function	2
3	Jacobians	3
3.1	Jacobian mean model	3
3.2	Jacobian dispersion model	4
4	Block-wise entries of hessian matrix	4
4.1	Mean-model block $H^{m,m}$	5
4.2	Dispersion-model block $H^{r,r}$	5
4.3	Mean-dispersion-model block $H^{r,m}$	7

1 Model description

1.1 Likelihood

The negative binomial likelihood of the GLM is:

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

The negative binomial log-likelihood of the GLM is:

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

1.2 Parameter models

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

$$m(\theta^m) = \exp(\langle X^m, \theta^m \rangle + s) \quad (3)$$

$$r(\theta^r) = \exp(\langle X^r, \theta^r \rangle) \quad (4)$$

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:

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

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

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

2 Miscancellous

2.1 Polygamma function

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

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

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

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

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

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

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

3 Jacobians

3.1 Jacobian mean model

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

3.2 Jacobian dispersion model

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

4 Block-wise entries of hessian matrix

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 $H^{m,m}$

$$\begin{aligned}
H_{i,j}^{m,m} &= \frac{d^2}{d\theta_i^m d\theta_j^m} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^m} \frac{d}{d\theta_i^m} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^m} \left(X_i^m * y - X_i^m * (y + r(\theta^r)) * \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \right) \\
&= -X_i^m * (y + r(\theta^r)) \frac{d}{d\theta_j^m} \frac{m(\theta^m)}{r(\theta^r) + m(\theta^m)} \\
&= -X_i^m * (y + r(\theta^r)) * \frac{\frac{d}{d\theta_j^m} \left(m(\theta^m) \right) * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * \frac{d}{d\theta_j^m} \left(r(\theta^r) + m(\theta^m) \right)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X_i^m * (y + r(\theta^r)) * \frac{m(\theta^m) * X_j^m * (r(\theta^r) + m(\theta^m)) - m(\theta^m) * m(\theta^m) * X_j^m}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X_i^m * (y + r(\theta^r)) * m(\theta^m) * X_j^m * \frac{r(\theta^r) + m(\theta^m) - m(\theta^m)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X_i^m * (y + r(\theta^r)) * m(\theta^m) * X_j^m * \frac{r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \\
&= -X_i^m * X_j^m * m(\theta^m) * \frac{y/r(\theta^r) + 1}{(1 + m(\theta^m)/r(\theta^r))^2}
\end{aligned} \tag{14}$$

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

$$\begin{aligned}
H_{i,j}^{r,r} &= \frac{d^2}{d\theta_i^r d\theta_j^r} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^r} \frac{d}{d\theta_i^r} LL(y|\theta^m, \theta^r) \\
&= \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) \right. \\
&\quad - r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \\
&\quad - r(\theta^r) * X_i^r * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} \\
&\quad + r(\theta^r) * X_i^r * \log(r(\theta^r)) \\
&\quad \left. + r(\theta^r) * X_i^r - r(\theta^r) * X_i^r * \log(r(\theta^r) + m(\theta^m)) \right) \\
&= \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) \right) \\
&\quad + \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) \right) \\
&\quad - \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} \right) \\
&\quad + \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \log(r(\theta^r)) \right) \\
&\quad + \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r \right) \\
&\quad - \frac{d}{d\theta_j^r} \left(r(\theta^r) * X_i^r * \log(r(\theta^r) + m(\theta^m)) \right)
\end{aligned} \tag{15}$$

$$\begin{aligned}
H_{i,j}^{r,r} = & \left(X_j^r * r(\theta^r) * X_i^r * \psi_0(r(\theta^r) + y) + r(\theta^r) * X_i^r * X_j^r * r(\theta^r) * \psi_1(r(\theta^r) + y) \right) \\
& - \left(X_j^r * r(\theta^r) * X_i^r * \psi_0(r(\theta^r)) + r(\theta^r) * X_i^r * X_j^r * r(\theta^r) * \psi_1(r(\theta^r)) \right) \\
& - \left(X_j^r * r(\theta^r) * X_i^r * \frac{r(\theta^r) + y}{r(\theta^r) + m(\theta^m)} + r(\theta^r) * X_i^r \right. \\
& \quad \left. * \frac{X_j^r * r(\theta^r) * (r(\theta^r) + m(\theta^m)) - (r(\theta^r) + y) * X_j^r * r(\theta^r)}{(r(\theta^r) + m(\theta^m))^2} \right) \\
& + \left(X_j^r * r(\theta^r) * X_i^r * \log(r(\theta^r)) + r(\theta^r) * X_i^r * X_j^r * r(\theta^r) * \frac{1}{r(\theta^r)} \right) \\
& + \left(X_j^r * r(\theta^r) * X_i^r \right) \\
& - \left(X_j^r * r(\theta^r) * X_i^r * \log(r(\theta^r) + m(\theta^m)) + r(\theta^r) * X_i^r * X_j^r * r(\theta^r) * \frac{1}{r(\theta^r) + m(\theta^m)} \right) \\
= & X_i^r * X_j^r * r(\theta^r) \left(\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)) \right. \\
& - \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} \\
& \left. + \log(r(\theta^r)) + 1 + 1 - \log(r(\theta^r) + m(\theta^m)) - \frac{r(\theta^r)}{r(\theta^r) + m(\theta^m)} \right) \\
= & X_i^r * X_j^r * r(\theta^r) \left(\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)) \right. \\
& - \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} \\
& \left. + \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m)) \right) \\
= & X_i^r * X_j^r * r(\theta^r) \left(\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)) \right. \\
& - \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} \\
& \left. + \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m)) \right) \\
= & X_i^r * X_j^r * r(\theta^r) \left(\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)) \right. \\
& - \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} \\
& \left. + \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m)) \right) \\
= & X_i^r * X_j^r * r(\theta^r) \left(\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)) \right. \\
& - \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} \\
& \left. + \log(r(\theta^r)) + 2 - \log(r(\theta^r) + m(\theta^m)) \right)
\end{aligned} \tag{16}$$

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

Mean-dispersion-model block derived from the mean-model jacobian:

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