

Group 8 - GLM Exercises

November 10, 2015

1. A density that belongs in the one-parameter exponential family has the following (canonical) representation:

$$p(z) = \dots$$

where θ is called the natural parameter, and c and h are functions whose exact form depends on the particular density. For many members of this family, $q > 0$, is a further unknown parameter, often called a precision parameter.

1. The following densities belong to the exponential family. Identify θ , q and an appropriate $c(\theta)$, for each of them (take into account that for some z might be a transformation of t):

Normal

$$\begin{aligned} t^{\sim} N(\mu, -q) &= \exp\left\{\frac{-(t-\mu)^2}{2q}\right\} \\ &= \exp\left\{\frac{-(t^2 - 2\mu t + \mu^2)}{2q}\right\} \\ &= \exp\left\{-\frac{1}{2q}\left[t^2 - 2\mu t + \mu^2\right]\right\} \\ &= \exp\left\{-\frac{1}{q}\left[\frac{t^2}{2} - \mu t + \frac{\mu^2}{2}\right]\right\} \\ &= \exp\left\{\frac{1}{q}\left[-\frac{t^2}{2} + \mu t - \frac{\mu^2}{2}\right]\right\} \\ &= \exp\left\{\frac{1}{q}\left[\mu t - \frac{\mu^2}{2}\right] - \frac{t^2}{2q}\right\} \end{aligned}$$

Result:

$$\theta = \mu$$

$q = q$ (since q^{-1} is the variance parameter passed, than q the parameter is the precision and the q above is the variance for the normal distribution)

$$c(\theta) = \frac{\mu^2}{2}$$

Bernoulli

$$\begin{aligned} t^{\sim} \text{Bern}(n, p) &= p^z (1-p)^{1-z} \\ &= \exp\left\{\log(p^z) + \log((1-p)^{1-z})\right\} \end{aligned}$$

$$\begin{aligned}
&= \exp \left\{ z \log(p) + (1-z) \log(1-p) \right\} \\
&= \exp \left\{ z \log(p) + \log(1-p) - z \log(1-p) \right\} \\
&= \exp \left\{ z \log\left(\frac{p}{1-p}\right) + \log(1-p) \right\}
\end{aligned}$$

Result:

$$\theta = \log\left(\frac{p}{1-p}\right)$$

$$q = 1$$

$$c(\theta) = -\log(1-p)$$

Binomial

$$\begin{aligned}
t \sim \text{Bin}(n, p) &= \binom{n}{k} p^k (1-p)^{n-k} \\
&= \exp \left\{ \log\left(\binom{n}{k}\right) + \log(p^k) + \log((1-p)^{n-k}) \right\} \\
&= \exp \left\{ \log\left(\binom{n}{k}\right) + k \log(p) + (n-k) \log(1-p) \right\} \\
&= \exp \left\{ \log\left(\binom{n}{k}\right) + k \log\left(\frac{p}{1-p}\right) + n \log(1-p) \right\} \\
&= \exp \left\{ n \left[\frac{k}{n} \log\left(\frac{p}{1-p}\right) + \log(1-p) \right] + \log\left(\binom{n}{k}\right) \right\}
\end{aligned}$$

Result:

$$\theta = \log\left(\frac{p}{1-p}\right)$$

$$q = n$$

$$c(\theta) = \log\left(\binom{n}{k}\right)$$

Poisson

$$\begin{aligned}
t \sim \text{Poisson}(\lambda) &= \frac{\lambda^k e^{-\lambda}}{k!} \\
&= \exp \left\{ \log(\lambda^k e^{-\lambda}) - \log(k!) \right\} \\
&= \exp \left\{ k * \log(\lambda) - \lambda \log(e) - \log(k!) \right\} \\
&= \exp \left\{ k * \log(\lambda) - \lambda - \log(k!) \right\}
\end{aligned}$$

Result:

$$\theta = \log(\lambda)$$

$$q = 1$$

$$c(\theta) = -\lambda$$

2. Identify the canonical link function for each of the models given above.

Normal

$$\mu$$

Bernoulli

$$\text{logit}(\mu)$$

Binomial

$$\text{logit}(\mu)$$

Poisson

$$\log(\mu)$$

3. Consider now the generalised linear model:

$$t_n | x_n \sim \text{Ndef}(\theta(x_n, w), q\gamma_n) \text{ with}$$

$$\theta(x_n, w) = (c')^{-1}(g^{-1}(\phi(x_n)^T w)) =: f(\phi(x_n)^T w)$$

We know the log-likelihood of the Ndef family of distributions takes the form:

$$2\log(p(t|x, \gamma, w, q)) = 2q \sum_n \gamma_n \left[t_n \theta(x_n, w) - c(\theta(x_n, w)) \right]$$

We will prove it is log-concave if it is the negative of a log-convex function, so will prove:

$$f(\theta(x_n, w), q\gamma_n) = -2q \sum_n \gamma_n \left[t_n \theta(x_n, w) - c(\theta(x_n, w)) \right]$$

is log-convex, taking into account the link function constructed as:

$$g(y) = c'^{-1}(\theta) \Rightarrow \theta(x_n, w) = \phi(x_n)^T w$$

We will prove it is log-convex by showing the second derivative is positive semi-definite for all $\theta(x_n, w)$

The first derivative of the negative log likelihood is:

$$f'(\theta(x_n, w), q\gamma_n) = -2q \left[\sum_n \gamma_n t_n \phi(x_n)^T - \sum_n \gamma_n c'(\phi(x_n)^T w) \right]$$

$$f''(\theta(x_n, w), q\gamma_n) = 2q \sum_n \gamma_n c''(\phi(x_n)^T w)$$

We know that c'' is the variance function and γ_n and q are positive by definition. So the second derivative is always positive and therefore:

$$x^T f'' x \geq 0 \text{ for all } x.$$

2. R^2 and deviance

1. Show that for any linear regression model:

$$-2\log p(t|X, w_{MLE}, q_{MLE}) = N \log e^T e + \text{const}$$

where “const” does not depend on M or X .

We know that:

$$-2\log p(t|X, w, q) = -N\log q + q(t - \phi w)^T(t - \phi w) + \text{const}$$

Maximizing with respect to w gives us:

$$w_{mle} = (\phi^T \phi)^{-1} \phi^T t$$

We plug this w_{mle} back into the log likelihood to get the likelihood function given w_{mle} :

$$-2\log p(t|X, w_{mle}, q) = -N\log(q) + qe^T e$$

Now, we can take the derivative wrt to q to solve for q_{mle} which gives us:

$$q_{mle} = \left(\frac{1}{N} e^T e \right)^{-1}$$

We use this to solve for the deviance of the maximum likelihood function:

$$\begin{aligned} -2\log p(t|X, w_{mle}, q_{mle}) &= -N\log(q_{mle}) + q_{mle}e^T e \\ -N\log(q_{mle}) + q_{mle}e^T e &= -N\log\left(\left(\frac{1}{N}e^T e\right)^{-1}\right) + \left(\frac{1}{N}e^T e\right)^{-1} e^T e \\ &= N\log\left(\left(\frac{1}{N}e^T e\right)\right) + N \\ &= N[\log(e^T e) - \log(N)] + N \\ &= N\log(e^T e) - N(\log(N) + 1) \\ &= N\log(e^T e) + \text{const} \end{aligned}$$

2. Show that in the null model,

$$w_{0,MLE} = \bar{t}$$

We know from 1:

$$w_{mle} = (\phi^T \phi)^{-1} \phi^T t$$

In the case of the null model, ϕ is a vector of ones so this becomes:

$$\frac{1}{N} \sum_{n=1}^N t_n = \bar{t}$$

3. The null model is nested within the saturated model, and it corresponds to the special case where $w_1 = \dots w_M = 0$. Let D_0 be the deviance of the null model and D_1 be that of the saturated model. Show that:

$$D_0 - D_1 = -N\log(1 - R^2)$$

where R^2 is the coefficient R^2 for the saturated model.

We know from part 1 that:

$$-2\log p(t|X, w_{mle}, q_{mle}) = N\log(e^T e) + \text{const}$$

Which also what we equate as D_M for model M , so:

$$\begin{aligned} D_0 - D_1 &= N\log(e_0^T e_0) - N\log(e_1^T e_1) \\ &= N\log\left(\frac{e_0^T e_0}{e_1^T e_1}\right) \end{aligned}$$

We use the following properties:

$$e = t - \hat{t}, \text{ and,}$$

$$R^2 = 1 - \frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2}$$

$$1 - R^2 = \frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2}$$

To show that

$$= N \log \left(\frac{e_0^T e_0}{e_1^T e_1} \right)$$

$$= N \log \left(\frac{\sum_i (t_i - \bar{t})^2}{\sum_i (t_i - \hat{t}_i)^2} \right)$$

$$= N \log(\sum_i (t_i - \bar{t})^2) - N \log(\sum_i (t_i - \hat{t}_i)^2)$$

$$= -N \log \left(\frac{\sum_i (t_i - \hat{t}_i)^2}{\sum_i (t_i - \bar{t})^2} \right)$$

$$= -N \log(1 - R^2)$$