Statistical Modeling and Inference – Problem Set #5

Group 3: Niti Mishra · Miquel Torrens · Bálint Ván

November 23rd, 2015

Solution to proposed exercises.

Exercise 1

Part (a)

We define:

$$\text{NdEF}(z|\theta,q) \equiv p(z|\theta,q) = \exp\{q[z\theta - c(\theta)] + h(z,q)\} \tag{1}$$

Let's start by the **normal** distribution:

$$\begin{aligned}
\mathcal{N}(t|\mu,q) &= \frac{q^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{q}{2}(t-\mu)^2\right\} \\
&= \frac{q^{1/2}}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}q(t^2+\mu^2-2z\mu)\right\} \\
&= \frac{q^{1/2}}{\sqrt{2\pi}} \exp\left\{q\left[t\mu - \frac{1}{2}\left(z^2+\mu^2\right)\right]\right\} \\
&= \exp\left\{q\left[t\mu - \frac{1}{2}\mu^2\right] + \log\left(\frac{q^{1/2}}{\sqrt{2\pi}}\right) - \frac{1}{2}qz^2\right\}
\end{aligned} \tag{2}$$

Now using (1) we can see the following in (2):

$$\mathcal{N}(t|\mu,q) = \exp\left\{\underbrace{q}_{=q}\left[\underbrace{t}_{=z}\underbrace{\mu}_{=\theta} - \underbrace{\frac{1}{2}\mu^2}_{=c(\theta)}\right] + \underbrace{\log\left(\frac{q^{1/2}}{\sqrt{2\pi}}\right) - \frac{1}{2}qz^2}_{=h(z,q)}\right\} \tag{3}$$

Thus with the normal distribution: $z=t$, $q=q$, $\theta=\mu$ and $c(\theta)=\frac{1}{2}\mu^2=\frac{1}{2}\theta^2$.

For the **Bernouilli** distribution:

$$\begin{aligned}
\text{Bern}(t|p) &= p^t(1-p)^{1-t} \\
&= \exp\{t\log p + (1-t)\log(1-p)\} \\
&= \exp\left\{t\log\left(\frac{p}{1-p}\right) + \log(1-p)\right\}
\end{aligned}$$

Mimetizing the analysis in (3), we have that $z=t$, $q=1$, $\theta=\log\left(\frac{p}{1-p}\right)$ and that $c(\theta)=-\log(1-p)=\log(1+e^\theta)$, because:

$$\theta = \log\left(\frac{p}{1-p}\right) \Leftrightarrow p = \frac{e^\theta}{1+e^\theta}. \tag{4}$$

For the **binomial** distribution:

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

Then for the binomial: $z = t/n$, $q = n$, $\theta = \log\left(\frac{p}{1-p}\right)$ and $c(\theta) = -\log(1-p) = \log(1+e^{\theta})$.

Finally, for the **Poisson** distribution:

$$\begin{aligned}
\text{Pois}(t|\lambda) &= \frac{\lambda^t e^{-\lambda}}{t!} \\
&= \exp\{t\log\lambda - \lambda - \log t!\} \\
&= \exp\{[t\log\lambda - \lambda] - \log t!\}
\end{aligned}$$

And using (1) that means that for Poisson $z = t$, $q = 1$, $\theta = \log\lambda$ and $c(\theta) = \lambda = e^{\theta}$.

Part (b)

To obtain the canonical links we apply the following formula[1]:

$$g(\mathbb{E}[z]) = \int \frac{1}{c''(\theta)} d\mathbb{E}[z].$$

For the **normal** distribution, where $\theta = \mu$:

$$c(\theta) = \frac{1}{2}\theta^2.$$

This means that $c'(\theta) = \theta$, $c''(\theta) = 1$ and $1/c''(\theta) = 1$. Then:

$$g(\mu) = \int d\mu = \mu.$$

For the **Bernouilli** and **binomial** distributions, which share $c(\theta)$, recall (4):

$$c(\theta) = -\log(1-p) = -\log\left(1 - \frac{e^{\theta}}{1+e^{\theta}}\right) = -\log\left(\frac{1}{1+e^{\theta}}\right) = \log(1+e^{\theta})$$

From this we obtain:

$$\begin{aligned}
c'(\theta) &= \frac{e^{\theta}}{1+e^{\theta}} \\
c''(\theta) &= \frac{e^{\theta}}{(1+e^{\theta})^2} \\
\frac{1}{c''(\theta)} &= \frac{(1+e^{\theta})^2}{e^{\theta}} = \frac{\left(1+\frac{p}{1-p}\right)^2}{\frac{p}{1-p}}
\end{aligned}$$

[1] GLMs lecture, slide number 11.

And finally recover:

$$g(p) = \int \frac{(1 + \frac{p}{1-p})^2}{\frac{p}{1-p}} dp = \log p - \log(1-p) = \log\left(\frac{p}{1-p}\right).$$

In the case of the **Poisson** distribution, we have that $\theta = \log\lambda$ and $c(\theta) = \lambda = e^\theta$, thus:

$$\begin{aligned} c'(\theta) &= e^\theta \\ c''(\theta) &= e^\theta = \lambda \\ \frac{1}{c''(\theta)} &= \lambda^{-1} \end{aligned}$$

This implies:

$$g(\lambda) = \int \lambda^{-1} d\lambda = \log\lambda.$$

Part (c)

From the log-likelihood function:

$$\log\mathcal{L}\left(\text{NdEF}(t_n|\theta(\mathbf{x}_n, \mathbf{w}), q\gamma_n)\right) = \sum_{n=1}^{N} \gamma_n q[t_n\theta(\mathbf{x}_n, \mathbf{w}) - c(\theta(\phi(\mathbf{x}_n)^T\mathbf{w}))] + \sum_{n=1}^{N} h(t_n, q\gamma_n)$$

Under the canonical link:

$$\theta(\mathbf{x}_n, \mathbf{w}) = \phi(\mathbf{x}_n)^T\mathbf{w}$$

This implies:

$$\log\mathcal{L} = \sum_{n=1}^{N} \gamma_n q[t_n\phi(\mathbf{x}_n)^T\mathbf{w} - c(\phi(\mathbf{x}_n)^T\mathbf{w})] + \sum_{n=1}^{N} h(t_n, q\gamma_n)$$

We can differentiate with respect to $\mathbf{w}$ to find out about log-concavity:

$$\begin{aligned} \nabla_{\mathbf{w}} \log\mathcal{L} &= \sum_{n=1}^{N} \gamma_n q[t_n\phi(\mathbf{x}_n) - \nabla c(\phi(\mathbf{x}_n)^T\mathbf{w})] \\ \nabla\nabla_{\mathbf{w}} \log\mathcal{L} &= -\sum_{n=1}^{N} \gamma_n q \nabla\nabla c(\phi(\mathbf{x}_n)^T\mathbf{w}) \end{aligned}$$

Observe that the matrix of second derivatives $\nabla\nabla c(\phi(\mathbf{x}_n)^T\mathbf{w})$ is positive-semi-definite, thereby log-convex, as $c''(\theta)$ is the variance function. Also, q and η_n are non-negative by construction, and so the multiplication remains log-convex. The minus sign in front makes it negative and thus the log-likelihood is log-concave.

Exercise 2

Part (a)

From the slides on MLE regression[2], we know that:

$$-2\log p(\mathbf{t}|\mathbf{X},\mathbf{w},q) = -N\log q + q(\mathbf{t}-\mathbf{\Phi}\mathbf{w})^T(\mathbf{t}-\mathbf{\Phi}\mathbf{w}) + c,$$

where c gathers all constant terms not dependent on $\mathbf{X}$. This yielded:

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

Then:

$$\begin{aligned}
-2\log p(\mathbf{t}|\mathbf{X},\mathbf{w}_{MLE},q_{MLE}) &= -N\log\left(\frac{1}{N}\mathbf{e}^T\mathbf{e}\right)^{-1} + \left(\frac{1}{N}\mathbf{e}^T\mathbf{e}\right)^{-1}(\mathbf{t}-\mathbf{\Phi}\mathbf{w}_{MLE})^T(\mathbf{t}-\mathbf{\Phi}\mathbf{w}_{MLE}) + c \\
&= N\log\mathbf{e}^T\mathbf{e} + N\left(\mathbf{e}^T\mathbf{e}\right)^{-1}\left(\mathbf{e}^T\mathbf{e}\right) + c \\
&= N\log\mathbf{e}^T\mathbf{e} + c.
\end{aligned}$$

Thus proved.

Part (b)

The problem in Part (a) also yielded:

$$\mathbf{w}_{MLE} = \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{t}.$$

In the null model, $\mathbf{\Phi}$ is a column vector with N components all of which are equal to 1. Consequently $\mathbf{\Phi}^T$ is a row vector also of length N. Then:

$$\mathbf{\Phi}^T\mathbf{\Phi} = \sum_{i=1}^{N} 1\times 1 = N.$$

This implies:

$$\left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1} = \frac{1}{N}.$$

Also:

$$\mathbf{\Phi}^T\mathbf{t} = \sum_{i=1}^{N} 1\times t_i = \sum_{i=1}^{N} t_i.$$

Finally:

$$w_{0,MLE} = \left(\mathbf{\Phi}^T\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^T\mathbf{t} = \frac{1}{N}\sum_{i=1}^{N} t_i = \bar{t}.$$

Hence proved.

[2] Slide 8

Part (c)

Using the results in Exercise 2.1:

$$\begin{aligned} D_0 &= -2\log p(\mathbf{t}|\mathbf{X}_{m_0}\mathbf{w}_{m_0,MLE}, q_{m_0,MLE}) \\ &= N\log(\mathbf{e}_0^T\mathbf{e}_0) + c \end{aligned}$$

$$\begin{aligned} D_1 &= -2\log p(\mathbf{t}|\mathbf{X}_{m_1}\mathbf{w}_{m_1,MLE}, q_{m_1,MLE}) \\ &= N\log(\mathbf{e}_1^T\mathbf{e}_1) + c \end{aligned}$$

Also note that given that D_0 is for the intercept model it has $\hat{t} = \bar{t}$ and we can write:

$$R^2 = 1 - \frac{\mathrm{var}(\mathbf{t}|\mathbf{x})}{\mathrm{var}(\mathbf{t})} = 1 - \frac{\frac{1}{N}\sum_{n=1}^{N}(t_n - \hat{t})^2}{\frac{1}{N}\sum_{n=1}^{N}(t_n - \bar{t})^2} = 1 - \frac{\sum_{n=1}^{N}(t_n - \hat{t})^2}{\sum_{n=1}^{N}(t_n - \bar{t})^2} = 1 - \frac{(\mathbf{t} - \hat{\mathbf{t}})^T(\mathbf{t} - \hat{\mathbf{t}})}{(\mathbf{t} - \bar{\mathbf{t}})^T(\mathbf{t} - \bar{\mathbf{t}})} = 1 - \frac{\mathbf{e}_1^T\mathbf{e}_1}{\mathbf{e}_0^T\mathbf{e}_0}$$

Now we calculate:

$$\begin{aligned} D_0 - D_1 &= (N\log\mathbf{e}_0^T\mathbf{e}_0 + c) - (N\log\mathbf{e}_1^T\mathbf{e}_1 + c) \\ &= -N\log(\mathbf{e}_0^T\mathbf{e}_0)^{-1} - N\log\mathbf{e}_1^T\mathbf{e}_1 \\ &= -N\log\left[\mathbf{e}_1^T\mathbf{e}_1(\mathbf{e}_0^T\mathbf{e}_0)^{-1}\right] \\ &= -N\log\left[1 - \left(1 - \frac{\mathbf{e}_1^T\mathbf{e}_1}{\mathbf{e}_0^T\mathbf{e}_0}\right)\right] \\ &= -N\log\left(1 - R^2\right). \end{aligned}$$

This finishes the proof.