

## Statistical Modeling and Inference – Problem Set #5

GROUP 3: NITI MISHRA · MIQUEL TORRENS · BÁLINT VÁN

November 23<sup>rd</sup>, 2015

Solution to proposed exercises.

### Exercise 1

Part (a)

We define:

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

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

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

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

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

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

For the **Bernoulli** distribution:

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

Mimetizing the analysis in (3), we have that  $q = 1$ ,  $\theta = \log\left(\frac{p}{1-p}\right)$  and that  $c(\theta) = -\log(1-p)$ . Also note that:

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

For the **binomial** distribution:

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

We adapted the objective variable to proportion of successes  $z/n$ , obtaining for the binomial:  $q = n$ ,  $\theta = \log \left( \frac{p}{1-p} \right)$  and  $c(\theta) = -\log (1-p)$ .

Finally, for the **Poisson** distribution:

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

And using (1) that means that for Poisson  $q = 1$ ,  $\theta = \log \lambda$  and  $c(\theta) = \lambda$ .

Part (b)

The procedure to obtain the canonical links is the following:

- Compute the variance function:  $V(\mathbb{E}[z]) = c''(\theta)$ .
- Obtain the canonical link function  $g(\mathbb{E}[z]) = \int V^{-1}(\mathbb{E}[z]) d\mathbb{E}[z]$ .

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

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

This means that  $c'(\theta) = \theta$ ,  $c''(\theta) = 1$  and  $V^{-1}(\mu) = 1$ . Then:

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

For the **Bernoulli** 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:

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

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:

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

This implies:

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

Part (c)

The likelihood function is the following:

$$\begin{aligned} \mathcal{L}(\text{Ndef}(t_n | \theta(\mathbf{x}_n, \mathbf{w}), q\gamma_n)) &= \prod_{n=1}^N \exp\{\gamma_n q[t_n \theta_n - c(\theta_n)] + h(t_n, q\gamma_n)\} \\ &= \exp\left\{\sum_{n=1}^N \gamma_n q[t_n \theta_n - c(\theta_n)] + h(t_n, q\gamma_n)\right\}. \end{aligned}$$

Then the log-likelihood:

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

Under the canonical link:  $\theta(\mathbf{x}_n, \mathbf{w}) = \phi(\mathbf{x}_n)^T \mathbf{w}$ , therefore:

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

The left part of this result, the positive sum, is linear. If the right-handed sum is convex, then the negative of this sum is concave. We use the second derivative:

$$\begin{aligned} \frac{\partial c(\theta)}{\partial \theta_n} &= \mathbb{E}[t_n] \\ \frac{\partial^2 c(\theta)}{\partial \theta_n^2} &= \mathbb{E}[t_n^2] - \mathbb{E}[t_n]^2 = \mathbb{V}[t_n] \\ \frac{\partial^2 c(\theta)}{\partial \theta_n \partial \theta_m} &= \mathbb{E}[t_n t_m] - \mathbb{E}[t_n] \mathbb{E}[t_m] = \text{cov}[t_n, t_m] \end{aligned}$$

This final derivative evaluated for all pairs  $n, m \in [1, N]$  produces the Hessian covariance matrix. The covariance matrix is positive semi-definite, and thus log-convex. Its negative is log-concave. Therefore, the log-likelihood function is log-concave.

## Exercise 2

### Part (a)

From the slides on MLE regression<sup>1</sup>, we know that:

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

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:

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

Thus proved.

### Part (b)

The problem in Part (a) also yielded:

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

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

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

This implies:

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

Also:

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

Finally:

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

Hence proved.

---

<sup>1</sup>Slide 8

Part (c)

Using the results in Exercise 2.1:

$$\begin{aligned} D_0 &= -2 \log p(\mathbf{t} | \mathbf{X}_{m_0} \mathbf{w}_{m_0, MLE}, q_{m_0, MLE}) \\ &= N \log(\mathbf{t} - \bar{t})^T (\mathbf{t} - \bar{t}) + c \\ &= N \log q^{-1} + c \\ &= -N \log q + c \end{aligned}$$

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

Also note that:

$$R^2 = 1 - \frac{\text{var}(\mathbf{t} | \mathbf{x})}{\text{var}(\mathbf{t})} = 1 - \frac{(\mathbf{t} - \hat{\mathbf{t}})^T (\mathbf{t} - \hat{\mathbf{t}})}{(\mathbf{t} - \bar{t})^T (\mathbf{t} - \bar{t})} = 1 - \frac{\mathbf{e}^T \mathbf{e}}{q^{-1}} = 1 - q \mathbf{e}^T \mathbf{e}$$

Now we calculate:

$$\begin{aligned} D_0 - D_1 &= (-N \log q + c) - (N \log \mathbf{e}^T \mathbf{e} + c) \\ &= -N \log q - N \log \mathbf{e}^T \mathbf{e} \\ &= -N \log q \mathbf{e}^T \mathbf{e} \\ &= -N \log (1 - (1 - q \mathbf{e}^T \mathbf{e})) \\ &= -N \log (1 - R^2). \end{aligned}$$

This finishes the proof.