# Homework 5 - Statistical modelling and inference
# Group 7

November 23, 2015



| **Students** | **Student IDs** |
|---|---|
| Max van Esso | 73539 |
| Marco Fayet | 125593 |
| Felix Gutmann | 125604 |



# 1 Exercise 1

## 1.1

**To show:** $p(z) = \exp\{(q[z\theta - c(\theta)] + h(z,q)\}$ for the following distributions:

---

**(a) Normal Distribution:** $t \sim \mathcal{N}(\mu, q^{-1})$

We start from the pdf of the Normal Distribution:

$$\begin{aligned}
\mathcal{N}(t|\mu,q) &= \frac{1}{\sqrt{2\pi q^{-1}}} \exp\{-\frac{1}{2q^{-1}}(t-\mu)^2\} \\
&= \frac{1}{\sqrt{2\pi q^{-1}}} \exp\{-\frac{1}{2q^{-1}}(t^2 - 2\mu t + \mu^2)\} \\
&= \frac{1}{\sqrt{2\pi q^{-1}}} \exp\{-\frac{1}{2}qt^2 + q(-\frac{1}{2}\mu^2 + \mu t)\}
\end{aligned}$$

Note that the normalizing constant can be expressed as $c = e^{\log c}$. Using that this terms adds to the exponent.

$$= \exp\{q(-\frac{1}{2}\mu^2 + \mu t) - \frac{1}{2}qt^2 + \log(\frac{1}{\sqrt{2\pi q^{-1}}})\}$$

**Indentification for Normal Distribution:**

| | | |
|---|---|---|
| $\theta$ | $\Longrightarrow$ | $\mu$ |
| $q$ | $\Longrightarrow$ | $q$ |
| $\mathrm{c}(\theta)$ | $\Longrightarrow$ | $\frac{1}{2}\mu^2 = \frac{1}{2}\theta^2$, (see also next exercise) |

---

### (b) Bernoulli Distribution: $t \sim Bern(p)$

We start from the pdf of Bernoulli Distribution, bring it into the exponential form and simplyfie it:

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

**Indentification for Bernoulli Distribution:**

| | | |
|---|---|---|
| $\theta$ | $\Longrightarrow$ | $\log(\frac{p}{1-p})$ |
| $q$ | $\Longrightarrow$ | 1 |
| $\mathrm{c}(\theta)$ | $\Longrightarrow$ | $\log(1-p) = \log(1+e^\theta)$, see next exercise for more details |

---

### (c) Binomial Distribution: $t \sim Bin(n, p)$

We start with the pdf of the Binomial Distribution, bring it into the exponential form and simplyfie it:

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

**Indentification for Binomial distribution:**

| | | |
|---|---|---|
| $\theta$ | $\Longrightarrow$ | $\log(\frac{p}{1-p})$ |
| $q$ | $\Longrightarrow$ | $n$ |
| c($\theta$) | $\Longrightarrow$ | $\log(1-p) = \log(1+e^{\theta})$, see next exercise for more details |

---

**(d) Poisson Distribution:** $t \sim Pois(\lambda)$

Again we start with the pdf of the Poisson distribution and bring it into the exponential form:

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

**Indentification for Poisson distribution:**

| | | |
|---|---|---|
| $\theta$ | $\Longrightarrow$ | $\log\lambda$ |
| $q$ | $\Longrightarrow$ | 1 |
| c($\theta$) | $\Longrightarrow$ | $\lambda = e^{\theta}$, see next exercise for more details |

## 1.2

**To show:** Canonical link functions $g(\cdot)$ for the last four distributions

**(a) Normal Distribution:**

From the last exercise we know $\theta = \mu$ and hence $c(\theta) = \frac{1}{2}\mu^2 = \frac{1}{2}\theta^2$. We take the derivative and invert it to obtain the result

$$\begin{aligned} &c'(\theta) = \theta \\ &g(y) = (c')^{-1}(\theta) = \theta = \mu \end{aligned}$$

**(b) Bernoulli and Binomial Distribution:**

We are using the assumed $\theta$ from the last exercise to compute p in order to obtain $c(\theta)$ in terms of $\theta$

$$\begin{aligned}
\theta &= \log \frac{p}{1-p} \\
e^{\theta} &= \frac{p}{1-p} \\
p &= \frac{e^{\theta}}{1+e^{\theta}}
\end{aligned}$$

We plug that p into the equation for $c(\theta)$ and take the derivative:

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

We set this equal to p to compute the inverse function to obtain $g(\cdot)$:

$$\begin{aligned}
p &= \frac{e^{\theta}}{1+e^{\theta}} \\
\theta &= \log \frac{p}{1-p}
\end{aligned}$$

**Poisson Distribution:**

Using the relation between $\theta$ and $\lambda$ we can indentify the derivative of $c(\theta)$ and hence the canonical link:

$$\begin{aligned}
\theta &= \log \lambda \implies e^{\theta} = \lambda \\
c(\theta) &= \lambda = e^{\theta} \\
c'(\theta) &= e^{\theta} = \lambda \implies g(\cdot) = \theta = \log \lambda
\end{aligned}$$

## 1.3

**To show:** The log-likelihood under the conical link is concave

### Prerequisite information:

- $t_n | \boldsymbol{x}_n \sim NdEF(\theta(\boldsymbol{x}_n, \boldsymbol{w}), q\gamma_n))$
- $\theta(\boldsymbol{x}_n, \boldsymbol{w}) =: f(\phi(\boldsymbol{x}_n)\boldsymbol{w})$

**Solution:**

Under the canonical link $f(x) = x$ and hence we get:

$$\begin{aligned}\theta(\boldsymbol{x}_n, \boldsymbol{w}) &= f(\phi(\boldsymbol{x}_n)^T \boldsymbol{w}) \\ &= \phi(\boldsymbol{x}_n)^T \boldsymbol{w} \end{aligned} \tag{1}$$

We start from the genral formulation and take the log of it:

$$\begin{aligned} p(t_n|\boldsymbol{x}_n, \boldsymbol{w}, \gamma_n, q) &= \exp\{q\gamma_n[t_n\theta(\boldsymbol{x}_n, \boldsymbol{w}) - c(\theta(\boldsymbol{x}_n, \boldsymbol{w}))] + h(t_n, q)\} \\ \log p(t_n|\boldsymbol{x}_n, \boldsymbol{w}, \gamma_n, q) &= q\gamma_n[t_n\theta(\boldsymbol{x}_n, \boldsymbol{w}) - c(\theta(\boldsymbol{x}_n, \boldsymbol{w}))] + h(t_n, q) \end{aligned}$$

Pluggin in the result from (1) leads to:

$$= q\gamma_n[t_n\phi(\boldsymbol{x}_n)^T \boldsymbol{w} - c(\phi(\boldsymbol{x}_n)\boldsymbol{w})] + h(t_n, q)$$

In the next steps we take the derivatives with respect to $\boldsymbol{w}$

$$\nabla p(t_n|\boldsymbol{x}_n, \boldsymbol{w}, \gamma_n, q) = q\gamma_n[t_n\phi(\boldsymbol{x}_n) - \nabla c(\phi(\boldsymbol{x}_n)^T \boldsymbol{w})]$$

Second derivative:

$$\nabla\nabla p(t_n|\boldsymbol{x}_n, \boldsymbol{w}, \gamma_n, q) = -(\gamma_n q \nabla\nabla c(\phi(\boldsymbol{x}_n)^T \boldsymbol{w})) \tag{2}$$

From (2) we notice the last factor is a variance term implying that it is positve or zero. Furthermore q and $\gamma_n$ are also positve or zero. Multiplying all three factors by minus one makes the whole term negative, which implies concavity

# 2

## 2.1

**To show:** $-2 \log p(\boldsymbol{t}|\boldsymbol{X}, \boldsymbol{w}_{MLE}, q_{MLE}) = N \log \boldsymbol{e}^T \boldsymbol{e} + const$

**Prerequisite information:**

- $q_{MLE} = (\frac{1}{N} \boldsymbol{e}^T \boldsymbol{e})^{-1}$
- $\boldsymbol{e} = (\boldsymbol{t} - \boldsymbol{\Phi} \boldsymbol{w}_{MLE})$

**Solution:**

We start from the expanded log likelihood function with the MLE parameters:

$$= -N \log q_{MLE} + q_{MLE} (\boldsymbol{t} - \boldsymbol{\Phi} \boldsymbol{w}_{MLE})^T (\boldsymbol{t} - \boldsymbol{\Phi} \boldsymbol{w}_{MLE}) + const$$

We replace $q_{MLE}$ by the mentioned definition:

$$= -N \log(\frac{1}{N} \boldsymbol{e}^T \boldsymbol{e})^{-1} + (\frac{1}{N} \boldsymbol{e}^T \boldsymbol{e})^{-1} (\boldsymbol{t} - \boldsymbol{\Phi} \boldsymbol{w}_{MLE})^T (\boldsymbol{t} - \boldsymbol{\Phi} \boldsymbol{w}_{MLE}) + const$$

We replace also the expression in the parenthesis with the mentioned definition and simplyfie it

$$\begin{aligned} &= -N \log(\frac{1}{N} \boldsymbol{e}^T \boldsymbol{e}) + N (\boldsymbol{e}^T \boldsymbol{e})^{-1} (\boldsymbol{e}^T \boldsymbol{e}) + const \\ &= N \log(\boldsymbol{e}^T \boldsymbol{e}) - \log N + N + const \end{aligned}$$

Since the last two terms neither depend on w nor q we assign them the constant to obtain the final result:

$$= N \log(\boldsymbol{e}^T \boldsymbol{e}) + const$$

## 2.2

**To show:** For the null model: $w_{0,MLE} = \bar{t}$

**Solution:**

We start with the equation for the MLE estimator $\boldsymbol{w}$:

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

Under the null model $\mathbf{\Phi}$ is a column vector containing ones ($\mathbf{\Phi}_0 = \mathbf{1}$). We have a look on the two parts of the equation separatly:

$$(\mathbf{\Phi}_0^T \mathbf{\Phi}_0)^{-1} = \frac{1}{1_1^2 + 1_2^2 + \cdots + 1_n^2} = \frac{1}{N}$$
$$(\mathbf{\Phi}_0^T \boldsymbol{t}) = t_1 + t_2 + \cdots + t_n = \sum_{i=1}^{N} t_i$$

Combining results we see that the estimator for the null model is equal to the sample mean:

$$w_0 = \frac{1}{N} \sum_{i=1}^{N} t_i$$

## 2.3

**To show:** $D_0 - D_1 = -N \log(1 - R^2)$

### Prerequisite information

- In general: $q_{MLE}^{-1} = var(t|x) = \frac{1}{N} \boldsymbol{e}^T \boldsymbol{e}$
- For the null model we have the variance of t: $q_0^{-1} = var(t) = \frac{1}{N} \boldsymbol{e}_0^T \boldsymbol{e}_0$
- Definition of R-Squared: $R^2 = 1 - \frac{var(t|x)}{var(t)} = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(t_i - \hat{t}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(t_i - \bar{t})^2} = 1 - \frac{\sum_{i=1}^{N}(t_i - \hat{t}_i)^2}{\sum_{i=1}^{N}(t_i - \bar{t})^2}$

### Solution

From exercise 2.1 we know $D_1$:

$$D_1 = N \log(\boldsymbol{e}^T \boldsymbol{e}) + const$$

For $D_0$ we define new particular residuals $\boldsymbol{e}_0$:

$$\begin{aligned} D_0 &= N \log(\boldsymbol{t} - \mathbf{\Phi}_0 w_0)^T (\boldsymbol{t} - \mathbf{\Phi}_0 w_0) + const \\ &= N \log(\boldsymbol{t} - \bar{\boldsymbol{t}})^T (\boldsymbol{t} - \bar{\boldsymbol{t}}) + const \\ &= N \log(\boldsymbol{e}_0^T \boldsymbol{e}_0) + const \end{aligned}$$

We notice $\boldsymbol{e}_0^T \boldsymbol{e}_0$ is the unormalized variance term of t:

$$\boldsymbol{e}_0^T \boldsymbol{e}_0 = (\boldsymbol{t} - \bar{\boldsymbol{t}})^T (\boldsymbol{t} - \bar{\boldsymbol{t}}) = \sum_{i=1}^{N} (t_i - \bar{t})^2$$

Combining results we get:

$$\begin{aligned} D_0 - D_1 &= N \log \boldsymbol{e}_0^T \boldsymbol{e}_0 + const - (N \log \boldsymbol{e}^T \boldsymbol{e} + const) \\ &= N \log \boldsymbol{e}_0^T \boldsymbol{e}_0 - N \log \boldsymbol{e}^T \boldsymbol{e} \end{aligned}$$

By factorizing -N and merging the log terms we get:

$$\begin{aligned} &= -N(\log \boldsymbol{e}^T \boldsymbol{e} - \log \boldsymbol{e}_0^T \boldsymbol{e}_0) \\ &= -N(\log((\boldsymbol{e}^T \boldsymbol{e})(\boldsymbol{e}_0^T \boldsymbol{e}_0)^{-1}) \end{aligned}$$

Note that this describes the variation of the residuals divided by the variation of t. We introduce one last manipulation to obtain the final result stated at the begining:

$$\begin{aligned} &= -N \log(1 - (1 - (\boldsymbol{e}^T \boldsymbol{e})(\boldsymbol{e}_0^T \boldsymbol{e}_0)^{-1}) \\ &= -N \log(1 - (1 - \frac{\sum_{i=1}^N (t_i - \hat{t}_i)^2}{\sum_{i=1}^N (t_i - \bar{t})^2}) \\ &= -N \log(1 - R^2) \end{aligned}$$