MATH3012–Statistical Methods II

Sketch Solutions: December 2006

1. **Exponential family.**

   (a) Write down the general form of the probability (density) function, $f_Y(y;\theta,\phi)$, of the exponential family of distributions.

   (b) Which is the canonical parameter?

   (c) Find the mean and variance of a random variable following the exponential family of distributions assuming that

$$E\left[\frac{\partial}{\partial\theta}\log f_Y(y;\theta,\phi)\right] = 0,$$

and

$$\text{Var}\left[\frac{\partial}{\partial\theta}\log f_Y(y;\theta,\phi)\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_Y(y;\theta,\phi)\right].$$

Solution 1. The pdf of a random variable Y following the exponential family is given by:

$$f_Y(y;\theta,\phi) = \exp\left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right] \quad (1)$$

for some suitable functions $a(\cdot), b(\cdot)$ and $c(\cdot,\cdot)$.

(b) θ is called the canonical parameter.

(c) Here

$$\begin{aligned} & \tfrac{\partial}{\partial\theta}\log f_Y(y;\theta,\phi) &=& \tfrac{1}{a(\phi)}[y - b'(\theta)] \\ \Longrightarrow\ & E\left[\tfrac{\partial}{\partial\theta}\log f_Y(Y;\theta,\phi)\right] &=& \tfrac{1}{a(\phi)}[E(Y) - b'(\theta)] \end{aligned}$$

Hence we have $E(Y) = b'(\theta)$.

$$\begin{aligned} & \tfrac{\partial^2}{\partial\theta^2}\log f_Y(Y;\theta,\phi) &=& -\tfrac{1}{a(\phi)}b''(\theta) \\ \Longrightarrow\ & -E\left[\tfrac{\partial^2}{\partial\theta^2}\log f_Y(Y;\theta,\phi)\right] &=& \tfrac{1}{a(\phi)}b''(\theta). \end{aligned}$$

From the given result

$$\text{Var}\left[\frac{\partial}{\partial\theta}\log f_Y(Y;\theta,\phi)\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_Y(Y;\theta,\phi)\right],$$

we see that

$$\begin{aligned} & \text{Var}\left[\tfrac{1}{a(\phi)}\{Y - b'(\theta)\}\right] &=& \tfrac{1}{a(\phi)}b''(\theta). \\ \Longrightarrow\ & \tfrac{1}{\{a(\phi)\}^2}\text{Var}(Y) &=& \tfrac{1}{a(\phi)}b''(\theta) \\ \Longrightarrow\ & \text{Var}(Y) &=& a(\phi)b''(\theta). \end{aligned}$$

2. **Poisson distribution** Suppose that Y_i, $i = 1, \ldots, n$ follow the Poisson distribution (independently) with mean μ_i and the generalised linear model is:

$$\log \mu_i = \beta_1 + \beta_2 x_i.$$

(a) Show that the Poisson distribution is a member of the exponential family of distributions.

(b) Write down the log-likelihood function of β_1 and β_2 explicitly. Hence, derive a pair of simultaneous equations, the solution of which are the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$.

(c) Express the above model as $\boldsymbol{\eta} = X\boldsymbol{\beta}$ where X is the appropriate $n \times 2$ matrix.

(d) Simplify the equations $\sum_{i=1}^n x_{ik} w_i z_i = 0, \quad k = 1, 2.$ where

$$w_i = \left(\mathrm{Var}(Y_i)[g'(\mu_i)]^2\right)^{-1}, \; z_i = (y_i - \mu_i) g'(\mu_i),$$

and $g(\cdot)$ is the link function. Do the two sets of equations match?

(e) Simplify the equations by taking $x_i = 0$ for $i = 1, \ldots, m$ and $x_i = 1$ for $i = m+1, \ldots, n$. Can you get exact solutions?

(f) Calculate the information matrix. Using all these calculations describe the Fisher scoring method in algorithmic steps.

Solution 2.
(a) The probability function is given by

$$\begin{aligned} f_Y(y;\mu) &= \tfrac{\exp(-\mu)\mu^y}{y!} \qquad y \in \{0, 1, \ldots\}; \quad \mu \in \mathcal{R}_+ \\ &= \exp\left(y \log \mu - \mu - \log y!\right) \end{aligned}$$

This is in the form of the exponential family, with $\theta = \log \mu$, $b(\theta) = \exp \theta$, $a(\phi) = 1$ and $c(y, \phi) = -\log y!$. Therefore

$$\begin{aligned} &E(Y) = b'(\theta) = \exp \theta = \mu \\ &Var(Y) = a(\phi) b''(\theta) = \exp \theta = \mu \\ &V(\mu) = \mu. \end{aligned}$$

(b)
Here the log-likelihood is given by:

$$\begin{aligned} \log f_{\mathbf{Y}}(\mathbf{y}; \beta_1, \beta_2) &\propto \textstyle\sum_{i=1}^n \left[-\mu_i + y_i \log \mu_i\right] \\ &= \textstyle\sum_{i=1}^n \left[-\exp(\beta_1 + \beta_2 x_i) + y_i(\beta_1 + \beta_2 x_i)\right]. \end{aligned}$$

By differentiating the above with respective to β_1 and β_2 and setting the derivatives to zero we obtain the maximum likelihood equations. We thus have,

$$\begin{aligned} \tfrac{\partial \log L}{\partial \beta_1} &= \textstyle\sum_{i=1}^n y_i - \sum_{i=1}^n \exp(\beta_1 + \beta_2 x_i) = 0, \\ \tfrac{\partial \log L}{\partial \beta_2} &= \textstyle\sum_{i=1}^n y_i x_i - \sum_{i=1}^n \exp(\beta_1 + \beta_2 x_i) x_i = 0. \end{aligned}$$

(c) Here the X matrix is given by:

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}.$$

(d) We have $w_i = \frac{1}{Var(Y_i)g'(\mu_i)^2}$. We know

$$Var(Y_i) = \mu_i \text{ and } g(\mu_i) = \log \mu_i.$$

By differentiating $g(\mu_i)$ we obtain,

$$g'(\mu_i) = \frac{1}{\mu_i}.$$

Hence

$$w_i = \mu_i.$$

Now

$$z_i = (y_i - \mu_i)g'(\mu_i) = \frac{y_i - \mu_i}{\mu_i}.$$

When $x_{i1} = 1$, $\sum_{i=1}^n x_{i1} w_i z_i = 0$, simplifies to

$$\sum_{i=1}^n (y_i - \mu_i) = \sum_{i=1}^n (y_i - \exp(\beta_1 + \beta_2 x_i)) = 0,$$

and when $x_{i2} = x_i$, $\sum_{i=1}^n x_{i2} w_i z_i = 0$, simplifies to

$$\sum_{i=1}^n x_i(y_i - \mu_i) = \sum_{i=1}^n x_i (y_i - \exp(\beta_1 + \beta_2 x_i)) = 0.$$

Hence the two sets of equations match.

(e) Putting the values of x_i we have the pair of equations:

$$\begin{aligned} \sum_{i=1}^n [y_i - \sum_{i=1}^n \exp(\beta_1 + \beta_2 x_i)] &= \sum_{i=1}^n y_i - \sum_{i=1}^m \exp(\beta_1) - \sum_{i=m+1}^n \exp(\beta_1 + \beta 2) = 0 \\ \sum_{i=1}^n y_i x_i - \sum_{i=1}^n \exp(\beta_1 + \beta_2 x_i) x_i &= \sum_{i=m+1}^n y_i - \sum_{i=m+1}^n \exp(\beta_1 + \beta_2) = 0. \end{aligned}$$

These two equations can be solved exactly. Let $t_0 = \sum_{i=1}^m y_i$ and $t_1 = \sum_{i=m+1}^n y_i$. The solutions are:

$$\hat{\beta}_1 = \log\left(\frac{t_0}{m}\right), \quad \hat{\beta}_2 = \log\left(\frac{m}{n-m}\frac{t_1}{t_0}\right).$$

(d) Here we see that $x_{i1} = 1$ and $x_{i2} = x_i$. Now we have,

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{pmatrix}.$$

Thus

$$X^T W X = \left(\begin{array}{cc} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i x_i \\ \sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i^2 \end{array} \right).$$

where $w_i = \mu_i i = 1, \ldots, n$.

(f) Let $\boldsymbol{\eta} = X\boldsymbol{\beta}$. The Fisher scoring algorithm proceeds as follows.

1. Choose an initial estimate $\boldsymbol{\beta}^{(m)}$ for $\hat{\boldsymbol{\beta}}$ at $m = 0$.
2. Evaluate $\boldsymbol{\eta}^{(m)}$, $\mathbf{W}^{(m)}$ and $\mathbf{z}^{(m)}$ at $\boldsymbol{\beta}^{(m)}$.
3. Calculate $\boldsymbol{\beta}^{(m+1)} = [\mathbf{X}^T\mathbf{W}^{(m)}\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{W}^{(m)}[\boldsymbol{\eta}^{(m)} + \mathbf{z}^{(m)}]$.
4. If $||\boldsymbol{\beta}^{(m+1)} - \boldsymbol{\beta}^{(m)}|| >$ some pre-specified (small) tolerance then set $m \to m + 1$ and go to 2.
5. Use $\boldsymbol{\beta}^{(m+1)}$ as the solution for $\hat{\boldsymbol{\beta}}$.

3. **Bernoulli** Suppose that Y_i, $i = 1, \ldots, n$ follow the Bernoulli distribution independently with probability p_i and the generalised linear model is:

$$\log \frac{p_i}{1 - p_i} = \beta_1 + \beta_2 x_i.$$

(a) Write down the log-likelihood function of β_1 and β_2 explicitly. Hence, derive a pair of simultaneous equations, the solution of which are the maximum likelihood estimates for $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$.

(b) Express the above model as $\boldsymbol{\eta} = X\boldsymbol{\beta}$ where X is the appropriate $n \times 2$ matrix.

(c) Simplify the equations $\sum_{i=1}^n x_{ik} w_i z_i = 0, \quad k = 1, 2.$ where

$$w_i = \left(\mathrm{Var}(Y_i)[g'(\mu_i)]^2\right)^{-1}, \; z_i = (y_i - \mu_i) g'(\mu_i),$$

and $g(\cdot)$ is the link function. Do the two sets of equations match?

(d) Calculate the information matrix. Using all these calculations describe the Fisher scoring method in algorithmic steps.

Solution 3. We know that $E(Y_i) = p_i = \mu_i$. Here the canonical link function is given by:

$$\theta_i = \log \frac{\mu_i}{1 - \mu_i} = \beta_1 + \beta_2 x_i.$$

Now,

$$\begin{array}{rcl} \theta_i & = & \log \frac{\mu_i}{1-\mu_i} \\ \implies \frac{\mu_i}{1-\mu_i} & = & e^{\theta_i} \\ \implies \mu_i & = & \frac{e^{\theta_i}}{1+e^{\theta_i}} \end{array}$$

We have

$$\log \frac{\mu_i}{1-\mu_i} = \beta_1 + \beta_2 x_i.$$

By solving we get

$$\mu_i = \frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}},\ 1-\mu_i = \frac{1}{1+e^{\beta_1+\beta_2 x_i}}.$$

The likelihood function is:

$$\begin{aligned} L(\beta_1, \beta_2; \mathbf{y}, \mathbf{x}) &= \prod_{i=1}^n (\mu_i)^{y_i}(1-\mu_i)^{1-y_i} \\ &= \prod_{i=1}^n \left(\frac{\mu_i}{1-\mu_i}\right)^{y_i} (1-\mu_i). \end{aligned}$$

Hence, the log-likelihood function is:

$$\begin{aligned} \log L(\beta_1, \beta_2; \mathbf{y}, \mathbf{x}) &= \sum_{i=1}^n \left\{ y_i \log \frac{\mu_i}{1-\mu_i} + \log(1-\mu_i) \right\} \\ &= \sum_{i=1}^n \left\{ y_i(\beta_1 + \beta_2 x_i) - \log(1+e^{\beta_1+\beta_2 x_i}) \right\}. \end{aligned}$$

By differentiating the above with respective to β_1 and β_2 and setting the derivatives to zero we obtain the maximum likelihood equations. We thus have,

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_1} &= \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}} = 0, \\ \frac{\partial \log L}{\partial \beta_2} &= \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \frac{e^{\beta_1+\beta_2 x_i}}{1+e^{\beta_1+\beta_2 x_i}} = 0. \end{aligned}$$

We have $w_i = \frac{1}{Var(Y_i) g'(\mu_i)^2}$. We know

$$Var(Y_i) = \mu_i(1-\mu_i) \text{ and } g(\mu_i) = \log \frac{\mu_i}{1-\mu_i}.$$

By differentiating $g(\mu_i)$ we obtain,

$$g'(\mu_i) = \frac{1}{\mu_i(1-\mu_i)}.$$

Hence

$$w_i = \mu_i(1-\mu_i).$$

Now

$$z_i = (y_i - \mu_i) g'(\mu_i) = \frac{y_i - \mu_i}{\mu_i(1-\mu_i)}.$$

Here we see that $x_{i1} = 1$ and $x_{i2} = x_i$. Now we have,

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad W = \begin{pmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{pmatrix}.$$

Thus

$$X^T W X = \begin{pmatrix} \sum_{i=1}^n w_i & \sum_{i=1}^n w_i x_i \\ \sum_{i=1}^n w_i x_i & \sum_{i=1}^n w_i x_i^2 \end{pmatrix}.$$

(d) See before in Question 2.

4. **Exponential distribution** The time to failure (Y) of a certain type of electrical component is thought to follow a negative exponential distribution, with probability density of the form

$$f_Y(y;\theta) \quad = \quad \theta \exp(-\theta y), \qquad y > 0; \quad \theta > 0.$$

It is believed that the distribution of failure time for a component is related to its electrical resistance (x) by the relationship

$$\theta \quad = \quad \beta_1 + \beta_2 x.$$

Suppose that $y_1, ..., y_n$ are observations of the times to failure, $Y_1, ..., Y_n$ for n such components with corresponding resistances $x_1, ..., x_n$.

(a) Write down the likelihood in terms of β_1 and β_2 and hence derive a pair of simultaneous equations, the solutions of which are the maximum likelihood estimates.

(b) Calculate the observed and expected information matrices. Are the Newton-Raphson and the Fisher scoring method identical for this problem? Give reasons.

Solution 4. (a) Here the likelihood function is:

$$L(\mathbf{y}, \mathbf{x}; \beta_1, \beta_2) = \prod_{i=1}^{n} (\beta_1 + \beta_2 x_i) \times \exp\{-y_i(\beta_1 + \beta_2 x_i)\}.$$

Thus the log likelihood function is given by:

$$\log L(\mathbf{y}, \mathbf{x}; \beta_1, \beta_2) = \sum_{i=1}^{n} \log(\beta_1 + \beta_2 x_i) - \sum_{i=1}^{n} y_i(\beta_1 + \beta_2 x_i)$$

By differentiating and equating to zero we have:

$$\begin{aligned} \frac{\partial \log L}{\partial \beta_1} &= \textstyle\sum_{i=1}^{n} \frac{1}{\beta_1 + \beta_2 x_i} - \sum_{i=1}^{n} y_i = 0, \\ \frac{\partial \log L}{\partial \beta_2} &= \textstyle\sum_{i=1}^{n} \frac{x_i}{\beta_1 + \beta_2 x_i} - \sum_{i=1}^{n} y_i x_i = 0. \end{aligned}$$

The maximum likelihood estimates of β_1 and β_2 are found by solving the above two equations.
(b) To find the information matrix we first differenitiate the log-likelihood function twice. We obtain:

$$\begin{aligned} \frac{\partial^2 \log L}{\partial \beta_1^2} &= -\textstyle\sum_{i=1}^{n} \frac{1}{(\beta_1 + \beta_2 x_i)^2} \\ \frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_2} &= -\textstyle\sum_{i=1}^{n} \frac{x_i}{(\beta_1 + \beta_2 x_i)^2} \\ \frac{\partial^2 \log L}{\partial \beta_2^2} &= -\textstyle\sum_{i=1}^{n} \frac{x_i^2}{(\beta_1 + \beta_2 x_i)^2} \end{aligned}$$

Since none of the double derivatives depends on the observed data $\mathbf{y}$ the Fisher information matrix will be the negative of the matrix with the above elements, that is,

$$I(\boldsymbol{\beta}) = \begin{pmatrix} \sum_{i=1}^{n} \frac{1}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^{n} \frac{x_i}{(\beta_1 + \beta_2 x_i)^2} \\ \sum_{i=1}^{n} \frac{x_i}{(\beta_1 + \beta_2 x_i)^2} & \sum_{i=1}^{n} \frac{x_i^2}{(\beta_1 + \beta_2 x_i)^2} \end{pmatrix}.$$

Thus the Newton-Raphson and Fisher scoring methods are identical.

5. **Log-linear models** Suppose that $\mathbf{Y} = (Y_1, \ldots, Y_k)^T$ follows the multinomial distribution with parameters N and $\mathbf{p} = (p_1, \ldots, p_k)^T$ where its probability function is given by:

$$\begin{aligned} f_{\mathbf{Y}}(\mathbf{y}; \mathbf{p}) &= P(Y_1 = y_1, \ldots, Y_k = y_k) \\ &= \begin{cases} N! \frac{p_1^{y_1} \cdots p_k^{y_k}}{y_1! \cdots y_k!} & \text{if } \sum_{i=1}^{k} y_i = N \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

For given values of $x_1, \ldots, x_k$, consider the model

$$\log p_i = \beta_1 + \beta_2 x_i, \ 1 \leq i \leq k$$

where β_1 is such that $\sum_{i=1}^k p_i = 1$.

(a) Write down the log-likelihood function.

(b) Derive the equation for finding the maximum likelihood estimate of β_2.

Solution 5. Note that:

$$1 = \sum_{i=1}^{k} p_i = \sum_{i=1}^{k} e^{\beta_1 + \beta_2 x_i} = e^{\beta_1} \sum_{i=1}^{k} e^{\beta_2 x_i}.$$

Hence the log-likelihood function is given by:

$$\begin{aligned} l(\beta_2; \mathbf{y}) &= \textstyle\sum_{i=1}^k y_i \log(p_i) + \log(N!) - \sum_{i=1}^k \log(y_i!) \\ &= \textstyle\sum_{i=1}^k y_i(\beta_1 + \beta_2 x_i) + \log(N!) - \sum_{i=1}^k \log(y_i!) \\ &= \textstyle\beta_1 \sum_{i=1}^k y_i + \beta_2 \sum_{i=1}^k y_i x_i + \log(N!) - \sum_{i=1}^k \log(y_i!) \\ &= \textstyle -N \log(\sum_{i=1}^k e^{\beta_2 x_i}) + \beta_2 \sum_{i=1}^k y_i x_i + \log(N!) - \sum_{i=1}^k \log(y_i!). \end{aligned}$$

5. (b) Now

$$\frac{\partial l(\beta_2; \mathbf{y})}{\partial \beta_2} = \sum_{i=1}^{k} y_i x_i - N \sum_{j=1}^{k} x_j e^{\beta_2 x_j} \left(\sum_{i=1}^{k} e^{\beta_2 x_i} \right)^{-1} = 0.$$

6. **Log-linear models** Consider a two-way $r \times c$ contingency table with probability p_{jk}, $j = 1, \ldots, r$, $k = 1, \ldots, c$ and $\sum_{j,k} p_{jk} = 1$. Show that the absence of the row-column interaction effect in the log-linear model implies independence between the rows and columns.

Solution 6. Bookwork. See notes page 60.