MATH3012: Statistical Methods II–Problems Sheet

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

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

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

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

   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.

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.

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

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

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.

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.

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.