STATISTICAL MODELLING Part IIC
Example Sheet 3 (of 4) RDS/Lent 2015

1. Look at the `cabbages` data in the `library(MASS)` package. Investigate whether the planting date has a significant effect on the weight of the cabbage head. Write out the models you have fitted and explain any conclusions you come to.

2. Let $Y \in \mathbb{R}$ be a random variable whose moment generating function is finite on an open interval containing zero. Show that the first two cumulants are equal to $\mathbb{E}(Y)$ and $\mathrm{Var}(Y)$ respectively.

3. Let Y have a model function from an exponential dispersion family. Compute the cumulant generating function of Y and deduce expressions for the mean and variance of Y.

4. We say Y has the inverse Gaussian distribution with parameters ϕ and λ, and write $Y \sim IG(\phi, \lambda)$ if its density is

$$f_Y(y; \phi, \lambda) = \frac{\sqrt{\lambda}}{\sqrt{2\pi} y^{3/2}} e^{\sqrt{\lambda\phi}} \exp\Big\{-\frac{1}{2}\Big(\frac{\lambda}{y} + \phi y\Big)\Big\},$$

   $y \in (0, \infty)$, $\lambda \in (0, \infty)$, $\phi \in (0, \infty)$. Compute the cumulant generating function of Y, and hence find its mean and variance.

   Show that the family of inverse Gaussian densities above is an exponential dispersion family, specifying the mean function μ, variance function $V(\mu)$, mean space $\mathcal{M}$, the range for the dispersion parameter Φ and the canonical link function. *Hint: First find σ^2 as a function of ϕ and λ by guessing that σ^2 is a function of λ alone.*

5. Let Y be a random variable with density $f(y; \theta)$ for $y \in \mathcal{Y} \subseteq \mathbb{R}^n$ and some $\theta \in \Theta \subseteq \mathbb{R}^d$, and write $\ell(\theta; Y)$ and $U(\theta; Y)$ for the corresponding log-likelihood and score functions. Assume that the order of differentiation with respect to a component of θ and integration over $\mathcal{Y}$ may be interchanged where necessary. Show that, for $r, s = 1, \ldots, d$,

$$\mathrm{Cov}_\theta\{U_r(\theta; Y), U_s(\theta; Y)\} = -\mathbb{E}_\theta\Big\{\frac{\partial^2}{\partial\theta_r \partial\theta_s}\ell(\theta; Y)\Big\}.$$

6. Let $Y_1, \ldots, Y_n$ be independent Poisson random variables with mean θ. Compute the maximum likelihood estimator $\hat{\theta}_n$. By considering $n\hat{\theta}_n$, write down the distribution of $\hat{\theta}_n$ and deduce its asymptotic distribution directly. Verify that this asymptotic distribution agrees with that predicted by the general asymptotic theory for maximum likelihood estimators.

7. Find the Fisher information matrix for the parameters (β, σ^2) in the normal linear model.

8. Let Y have a model function from the exponential dispersion family

$$f(y; \mu, \sigma^2) = \exp\Big[\frac{1}{\sigma^2}\{y\theta(\mu) - K(\theta(\mu))\}\Big] a(\sigma^2, y),$$

   $y \in \mathcal{Y}$, $\mu \in \mathcal{M}$, $\sigma^2 \in \Phi \subseteq (0, \infty)$, and variance function $V(\mu)$.

   (a) Use the identity $\mu = \mu(\theta(\mu))$ to show that $\theta'(\mu) = 1/V(\mu)$.

(b) Show that the maximum likelihood estimator for μ is Y.

9. Consider a generalised linear model for data $(y_1, x_1^T), \ldots, (y_n, x_n^T)$ and let the design matrix X have i^{th} row x_i^T for $i = 1, \ldots, n$.

   (a) Use the chain rule to show that the likelihood equations for β may be written as

   $$\sum_{i=1}^{n} \frac{(y_i - \mu_i)X_{ir}}{\sigma_i^2 V(\mu_i) g'(\mu_i)} = 0, \quad r = 1, \ldots, p,$$

   where $\mu_i = g^{-1}(x_i^T \beta)$.

   (b) Show that the Fisher information matrix for the parameters (β, σ^2) takes the form

   $$i(\beta, \sigma^2) = \begin{pmatrix} i(\beta) & 0 \\ 0 & i(\sigma^2) \end{pmatrix},$$

   where (with a slight abuse of notation) we have written $i(\beta)$ as the $p \times p$ block of the Fisher information matrix corresponding to β. Show that $i(\beta)$ can be expressed as $\sigma^{-2} X^T W X$ where W is a diagonal matrix with

   $$W_{ii} = \frac{1}{a_i V(\mu_i)\{g'(\mu_i)\}^2},$$

   (you need not specify $i(\sigma^2)$, and you may assume $\partial \ell / \partial \beta_j \partial \sigma^2 = \partial \ell / \partial \sigma^2 \partial \beta_j$ for all j).

   (c) How do the expressions in (a) and (b) simplify when $g(\mu_i)$ is the canonical link function? Show also that in this case, $i(\beta) = j(\beta)$, the $p \times p$ block of the observed information matrix corresponding to β.

10. Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim N(\mu_i, \sigma^2)$ and $\mu_i = x_i^T \beta$, for $i = 1, \ldots, n$.

    (a) Show that the deviance is equal to the residual sum of squares.

    (b) Assume now for simplicity that σ^2 is known. Show that only one iteration of the Fisher scoring method is required to attain the maximum likelihood estimator $\hat{\beta}$, regardless of the initial values for the algorithm. What feature of the log-likelihood function ensures that this is the case?

11*. Suppose that $Y = X\beta + \sigma\varepsilon$ where X is an $n \times p$ design matrix of full column rank, $\varepsilon \sim t_\nu$ and $\nu > 2$. Assume that ν and $\sigma^2 > 0$ are known and consider estimating β. Let $\hat{\beta}_{\text{OLS}} = (X^T X)^{-1} X^T Y$. Write down $\text{Var}(\hat{\beta}_{\text{OLS}})$. Show that the asymptotic variance of $\hat{\beta}$, the maximum likelihood estimator of β is

$$\frac{\nu + 3}{\nu + 1} \sigma^2 (X^T X)^{-1}.$$

Hint: The following facts may be of use. If $A \sim \chi_k^2$, then $\mathbb{E}(A^{-1}) = (k-2)^{-1}$ provided $k > 2$. Now if $B \sim \chi_l^2$ and A and B are independent, then

$$\frac{A}{A + B} \sim \textit{Beta}(k/2, l/2),$$

a Beta distribution with parameters $k/2$ and $l/2$, provided $k, l > 0$. If $Z \sim \textit{Beta}(a, b)$ then

$$\mathbb{E}(Z) = \frac{a}{a+b}, \qquad \text{Var}(Z) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Also the t_ν distribution has density proportional to

$$f(x) = (1 + x^2/\nu)^{-(\nu+1)/2}.$$