STATISTICAL MODELLING Part IIC
Example Sheet 3 (of 4) RDS/Lent 2014

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 fit and explain any conclusions you come to.

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

3. Show that the family of Poisson probability mass functions

$$f(y;\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \qquad y \in \{0,1,\ldots\},\ \lambda \in (0,\infty)$$

is an exponential dispersion family. Specify the mean function μ, variance function $V(\mu)$, mean space $\mathcal{M}$, and the canonical link function (you may take the dispersion parameter to be equal to 1).

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. Let $Y_1,\ldots,Y_n$ be independent Poisson(θ) random variables. Show that both $\bar{Y} = n^{-1}\sum Y_i$ and $S^2 = (n-1)^{-1}\sum(Y_i - \bar{Y})^2$ are unbiased estimators of θ. Without calculating $\mathrm{Var}_\theta(S^2)$, argue that $\bar{Y}$ is at least as good an estimator as S^2.

8. The asymptotic distribution theory for maximum likelihood estimators was valid under regularity conditions. Here is a situation where those conditions are not met. Let $Y_1, \ldots, Y_n$ be independent $U[0, \theta]$ random variables, for some $\theta \in \Theta = (0, \infty)$. Find the maximum likelihood estimator $\hat{\theta}_n$, as well as its distribution function, mean and variance. What is the asymptotic distribution of $n(\theta - \hat{\theta}_n)/\theta$?

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

10. Let Y have the exponential dispersion model function

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

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

(a) Use the identity $\mu = \mu(\theta(\mu))$ to show that

$$\frac{d\theta}{d\mu} = \frac{1}{V(\mu)}.$$

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

11. 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 $X^T W X$ where W is a diagonal matrix with

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

(you need not specify $i(\sigma^2)$). *Hint: Use the definition of the Fisher information in terms of products of first derivatives of the likelihood function.*

(c) How do the expressions in (a) and (b) simplify when $g(\mu_i)$ is the canonical link function?

12. 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 $i = 1, \ldots, n$. Show that the deviance is equal to the residual sum of squares.

13. Let $Y_1, \ldots, Y_n$ be independent with $Y_i \sim N(\mu_i, \sigma^2)$ for $i = 1, \ldots, n$, where $\mu_i = \alpha + \beta x_i$, and assume 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 $(\hat{\alpha}, \hat{\beta})^T$, regardless of the initial values for the algorithm. What feature of the log-likelihood function ensures that this is the case?

14. 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. It can be shown that $\mathrm{Var}(\varepsilon_i) = \nu/(\nu - 2)$. Let $\hat{\beta}_{\mathrm{OLS}} = (X^T X)^{-1} X^T Y$. Write down $\mathrm{Var}(\hat{\beta}_{\mathrm{OLS}})$. According to the theory of general maximum likelihood estimators, what is the asymptotic variance of $\hat{\beta}$, the maximum likelihood estimator of β? *Hint: You will need to use the fact that* $\Gamma(t+1) = t\Gamma(t)$ *for* $t > 0$.

15*. (a) A set $S \subseteq \mathbb{R}^d$ is convex if for all $x, y \in S$, $tx + (1-t)y \in S$ for all $t \in [0,1]$ (so S contains the line segment joining x and y). Show that the set of values $\mathcal{B}$ that β can take in a generalised linear model with canonical link function is an open convex set. *Hint: Recall that the natural parameter space* Θ *of the underlying exponential dispersion family must be an open interval.*

(b) Let $S \subseteq \mathbb{R}^d$ be an open convex set. Show that if $g : S \to \mathbb{R}$ is a twice differentiable function for which the Hessian matrix is negative definite and $x^* \in S$ satisfies $\frac{\partial g(x^*)}{\partial x} = 0$, then x^* is the unique maximiser of g on S.

(c) Conclude that if $\hat{\beta}$ satisfies $\frac{\partial \ell(\hat{\beta}, \sigma^2)}{\partial \beta} = 0$ where $\ell(\beta, \sigma^2)$ is the likelihood of a generalised linear model with canonical link and the matrix of predictors X with i^{th} row x_i^T for $i = 1, \ldots, n$ has full column rank, then $\hat{\beta}$ is the unique maximiser of $\ell(\beta, \sigma^2)$ over β.

16*. Suppose that the maximum likelihood estimator $\hat{\theta}^{(n)}$ when n observations are available for estimating a vector of parameters $\theta \in \mathbb{R}^d$ satisfies

$$\sqrt{n}(\hat{\theta}^{(n)} - \theta) \xrightarrow{d} N_d(0, I^{-1}(\theta)),$$

where $I(\theta) = \lim_{n\to\infty} i^{(n)}(\theta)/n$ is a positive definite matrix. Suppose further that $i^{(n)}(\hat{\theta}^{(n)})/n \xrightarrow{p} I(\theta)$.

A more general version of Slutsky's lemma states that if $Z_1, Z_2, \ldots$ is a sequence of random vectors in $\mathbb{R}^{k_1}$ and $Y_1, Y_2, \ldots$ is a sequence of random vectors (or matrices) in $\mathbb{R}^{k_2}$, if $Z_n \xrightarrow{d} Z$ and $Y_n \xrightarrow{p} c$ and $g : \mathbb{R}^{k_1} \times \mathbb{R}^{k_2} \to \mathbb{R}^{k_3}$ is continuous, then $g(Z_n, Y_n) \xrightarrow{d} g(Z, c)$.

Prove that under the assumptions above,

$$(\hat{\theta}^{(n)} - \theta)^T i^{(n)}(\theta)(\hat{\theta}^{(n)} - \theta) \xrightarrow{d} \chi^2_d.$$