STATISTICAL MODELLING Part IIC / Michaelmas 2023
Example Sheet 3 (of 4)

1. Look at the `cabbages` data in the `library(MASS)` package (use `?cabbages` to find out about the dataset). 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, let K be its c.g.f. and let $\Theta = \{\theta : K(\theta) < \infty\}$.

   (a) Show that Θ is convex. *Hint: Use the Hölder inequality.*

   (b) Suppose that Θ contains an open interval containing zero. Show that the first two cumulants of Y 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

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

   $y \in (0, \infty)$, $\lambda \in (0, \infty)$, $\phi \in (0, \infty)$.

   (a) Compute the cumulant generating function of Y, and hence find its mean and variance.

   (b) 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 $\ell(\theta)$ and $U(\theta)$ for the corresponding log-likelihood and score functions.

   (a) 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 $r, s = 1, \ldots, d$,

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

   (b) Suppose the components Y_i of Y are i.i.d. and each Y_i has density $h(y_i; \theta)$. Let $i^{(n)}(\theta)$ be the Fisher information of Y. Show that $i^{(n)}(\theta) = n i^{(1)}(\theta)$.

6. Find the Fisher information for the parameters (β, σ^2) in the normal linear model $Y = X\beta + \varepsilon$, where $\varepsilon \sim N_n(0, \sigma^2 I)$.

7. Let $Y_1, \ldots, Y_n$ be independent Poisson random variables with mean θ.

   (a) Compute the maximum likelihood estimator $\hat{\theta}_n$.

   (b) 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.

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. 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.

10. Consider a generalised linear model for data $(Y_1, x_1), \ldots, (Y_n, x_n)$ 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) $i(\beta)$ is the $p \times p$ block of the Fisher information matrix corresponding to β. Show that $i(\beta)$ can be expressed as $\sigma^{-2}X^TWX$ 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 $\partial^2\ell/\partial\beta_j\partial\sigma^2 = \partial^2\ell/\partial\sigma^2\partial\beta_j$ for all j).

(c) Show for the canonical link function that the observed information and the Fisher information for β coincide, that is,

$$\mathbb{E}_{\beta,\sigma^2}\left(-\frac{\partial^2}{\partial\beta\partial\beta^T}\ell(\beta,\sigma^2)\right) = -\frac{\partial^2}{\partial\beta\partial\beta^T}\ell(\beta,\sigma^2).$$

11. 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?

12. Suppose that $Y = X\beta + \sigma\varepsilon$ where X is an $n \times p$ design matrix of full column rank, and where the components ε_i of ε are i.i.d. with $\varepsilon_i \sim t_\nu$ and degrees of freedom $\nu > 2$. Assume that ν and $\sigma^2 > 0$ are known and consider estimating β. Let $\hat{\beta}_{\text{OLS}} = (X^TX)^{-1}X^TY$.

(a) Write down $\text{Var}(\hat{\beta}_{\text{OLS}})$.

(b) Show that the asymptotic variance of $\hat{\beta}$, the maximum likelihood estimator of β, is

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

You may assume that the regularity conditions for the asymptotic theory of the maximum likelhood estimator are satisfied.

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

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

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

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

Also the t_ν *distribution has density proportional to*

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