Modern Regression: Examination 2018

29 June 2018

Instructions: The time allotted for the examination is 180 minutes. You may answer in either English or French. No written material may be brought into the examination, but a simple calculator may be used. Full marks may be obtained with complete answers to four questions. The final mark will be based on the best four solutions.

Notation: $a_+ = \max(a, 0)$ for $a \in \mathbb{R}$; $A_{r\times s}$ means that A is an $r \times s$ matrix; $X \sim \mathcal{N}_p(\mu, \Omega)$ means that X has a p-dimensional multivariate normal distribution with mean vector $\mu_{p\times 1}$ and variance matrix $\Omega_{p\times p}$; and $X_{p\times 1} \sim (\mu, \Omega)$ means that $\mathrm{E}(X) = \mu_{p\times 1}$ and $\mathrm{var}(X) = \Omega_{p\times p}$.

First name:

Last name:

SCIPER number:

Exercise	Points	Indicative marks
1		/10 points
2		/10 points
3		/10 points
4		/10 points
5		/10 points
Total:		/40 points

1. The log likelihood function for data $y_1, \ldots, y_n$ believed to come from a parametric statistical model that is regular for maximum likelihood estimation is of the form

$$\ell(\beta) = \sum_{j=1}^{n} \ell_j\{\eta_j(\beta)\},$$

where β is a $p \times 1$ vector of unknown real-valued parameters.

(a) Show that in many cases the maximum likelihood estimate $\hat\beta$ can be obtained starting from some β_0 by the iteration

$$\hat\beta_{i+1} = \{X^{\mathrm{T}}(\beta_i)W(\beta_i)X(\beta_i)\}^{-1}X^{\mathrm{T}}(\beta_i)W(\beta_i)z(\beta_i), \quad i = 0, 1, \ldots,$$

where $X(\beta)$, $W(\beta)$ and $z(\beta)$ should be given in terms of ℓ_j and η_j.

(b) Briefly discuss any potential problems with applying the algorithm derived in (a), and discuss how they might be overcome.

(c) An M-estimate of β is obtained by minimizing

$$g(\beta) = \sum_{j=1}^{n} \rho(y_j - x_j^{\mathrm{T}}\beta),$$

where the twice continuously differentiable function $\rho : \mathbb{R} \to \mathbb{R}_+$ is chosen to downweight outlying observations. Suggest how $\hat\beta$ may be obtained using the algorithm from (a). How would you find initial values?

2. (a) An observation Y has probability function of the form

$$f(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y;\phi)\right\}, \quad y \in \mathcal{Y}, \theta \in \Theta, \phi > 0,$$

where $\mathcal{Y}, \Theta \subset \mathbb{R}$. Find the cumulant-generating function of Y and hence show that $\mathrm{E}(Y) = b'(\theta)$ and $\mathrm{var}(Y) = \phi b''(\theta)$.

(b) Show that the binomial distribution can be written in the above form, and hence find the relation between its mean and variance function.

(c) What is the *deviance* of a fitted model?

(d) Independent Bernoulli variables $Y_1, \ldots, Y_n$ have means $\pi_1, \ldots, \pi_n$, where $\pi_j = 1/\{1+\exp(-x_j^{\mathrm{T}}\beta)\}$, with $x_1, \ldots, x_n$ being $p \times 1$ vectors of covariates. Show that the deviance for this model is a function of the data only through the maximum likelihood estimate $\hat\beta$, and explain the implications of this for model-checking.

3. Table 1 shows data on the chemical composition of turnip greens. Four plants were taken from a field at random, three leaves were randomly selected from each plant, and each of these leaves was sampled at two randomly-chosen sites; the measurements show the calcium concentration (%) at each site, as determined by a microchemical method.

(a) Explain the terms *components of variance*, *nested random effects* and *crossed random effects*.

(b) A linear model for the data in Table 1 is of the form

$$y_{pls} = \mu + \alpha_p + \beta_{pl} + \varepsilon_{pls}, \quad p = 1, \ldots, 4, \quad l = 1, 2, 3, \quad s = 1, 2. \qquad (1)$$

State distributional assumptions appropriate for (1), and use them to complete the analysis of variance by giving the values of A, B and C in Table 2.

Table 1: Calcium concentration (%, dry basis) of turnip greens, at two sites on each of three leaves randomly chosen on four plants.

Plant	Leaf	Site 1	Site 2	Plant	Leaf	Site 1	Site 2
1	1	3.28	3.09	3	1	2.77	2.66
	2	3.52	3.48		2	3.74	3.44
	3	2.88	2.80		3	2.55	2.55
2	1	2.46	2.44	4	1	3.78	3.87
	2	1.87	1.92		2	4.07	4.12
	3	2.19	2.19		3	3.31	3.31

Table 2: Analysis of variance for the data in Table 1.

Source	df	Mean square, M	$\mathrm{E}(M)$
Plant	A	3.52	$\sigma^2 + 2\sigma^2_{\text{leaf}} + 6\sigma^2_{\text{plant}}$
Leaf within plant	8	0.3288	B
Site within leaf	C	0.0067	σ^2

(c) Give estimates of the components of variance, and discuss the results of the analysis.

4. Consider a regression model

$$y_{n\times 1} = \mu_{n\times 1} + \varepsilon_{n\times 1}, \quad \varepsilon \sim (0, \sigma^2 I_n),$$

where the jth element of the vector μ equals $\mu(x_j)$, with $\mu(x)$ twice continuously differentiable for $x \in [a, b]$ and $a < x_1 < \cdots < x_n < b$.

(a) Explain why it may be desirable to estimate μ by minimising the penalised sum of squares

$$Q_\alpha(\mu) = \sum_{j=1}^{n} \{y_j - \mu(x_j)\}^2 + \alpha \int_a^b \mu''(x)^2 \,\mathrm{d}x.$$

Discuss the cases $\alpha \to 0, \infty$.

(b) If one can write

$$\int_a^b \mu''(x)^2 \,\mathrm{d}x = \mu^{\mathrm{T}} \Delta \mu,$$

for some symmetric positive semi-definite matrix $\Delta_{n\times n}$ of rank $n - 2$, show that the vector of fitted values $\hat{\mu}_{n\times 1}$ may be written as $S_\alpha y$, where S_α should be given. Show also that we may write

$$\mathrm{tr}(S_\alpha) = \sum_{j=1}^{n} \frac{1}{1 + \alpha d_j},$$

for some non-negative $d_1, \ldots, d_n$, and discuss how $\mathrm{tr}(S_\alpha)$ varies as a function of α.

(c) Show that

$$\mathrm{E}\left\{(\hat{\mu} - \mu)^{\mathrm{T}}(\hat{\mu} - \mu)\right\} = \sigma^2 \mathrm{tr}(S_\alpha^{\mathrm{T}} S_\alpha) + \|(I - S_\alpha)\mu\|^2,$$

and discuss the interpretation of the terms on the right-hand side of this expression. Show that $\mathrm{tr}(S_\alpha^{\mathrm{T}} S_\alpha) < \mathrm{tr}(S_\alpha)$ for any $\alpha > 0$. What does this imply about the corresponding fitted models?

5. (a) Give two definitions of residuals in a general regression model, and discuss how they might be used to assess the fit of the model to data.

   (b) Figure 1 shows residuals for fits of generalized linear models to four sets of $n = 200$ observations. In each case (i) say what you think the model is, giving reasons, (ii) say whether the fit seems adequate, giving reasons, and (iii) suggest what steps you would take to deal with any model failure suggested by the plot.

Figure 1: Diagnostic plots for generalized linear model fits.