Advanced Regression: Examination 2017

25 January 2017

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 $\mathrm{var}(X) = \Omega_{p\times p}$.

---

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) If the maximum likelihood estimate $\hat\beta$ is known to satisfy the score equation

$$\frac{\partial \ell(\hat\beta)}{\partial \beta} = 0,$$

derive the iterative weighted least squares algorithm to obtain $\hat\beta$.

(b) Briefly describe how a 95% confidence interval for a component β_r of β might be constructed, and how the hypothesis that $\beta_1 = 0$ would be tested.

(c) The Michaelis–Menten equation links the rate v of an enzymatic reaction to the concentration c of a substrate through the equation

$$v = \frac{Uc}{K + c},$$

where U and K are the maximum rate and the Michaelis constant, respectively. Derive an algorithm to obtain maximum likelihood estimates of U and K based on independent observation pairs $(v_1, c_1), \ldots, (v_n, c_n)$, where the c_j are regarded as constants but the v_j are measured with Gaussian errors having variance σ^2.

2. (a) Under what circumstances would you represent the expectation of a response variable Y corresponding to a covariate x by the truncated power series

$$\beta_0 + \beta_1 x + \cdots + \beta_p x^p + b_1(x - \kappa_1)_+^p + \cdots + b_K(x - \kappa_K)_+^p, \quad x \in \mathbb{R}?$$

Explain the meanings of the symbols in the expression

$$(y - B\gamma)^{\mathrm{T}}(y - B\gamma) + \alpha\gamma^{\mathrm{T}} D\gamma. \tag{1}$$

(b) Use (1) to obtain an estimator of γ. Hence show that the fitted values based on independent observations $(y_1, x_1), \ldots, (y_n, x_n)$ may be written as $S_\alpha y$, and specify S_α.

(c) If η is an eigenvalue of $N^{-1/2}MN^{-1/2}$, where M and N are $q \times q$ matrices, with M positive semi-definite, N invertible with matrix square root $N^{1/2}$ and $\alpha \geq 0$, show that $(1 + \alpha\eta)^{-1}$ is an eigenvalue of $(N + \alpha M)^{-1}N$.

(d) Show that $\mathrm{tr}(S_\alpha)$ is a bounded monotone function of α, and give an interpretation of the function and its bounds in terms of the truncated power series.

3. (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) Explain the terms in italics in the phrase 'the *link function* and *variance function* are key ingredients of a *generalized linear model*'.

(c) Independent responses $Y_1, \ldots, Y_n$ with corresponding $p \times 1$ covariate vectors $x_1, \ldots, x_n$, with $p < n$, follow a logistic regression model. Unfortunately, instead of the Y_j, the noisy responses $Z_j = (1 - I_j)Y_j + I_j(1 - Y_j)$ are observed, where $\Pr(I_j = 1) = 1 - \Pr(I_j = 0) = \gamma \in (0, 1)$, and the I_j are independent, and independent of the Y_j. The probability γ is known from another study.

Formulate a generalised linear model with responses $Z_1, \ldots, Z_n$, and give its link and variance functions. What if $\gamma = 1/2$?

How would you estimate an unknown γ?

4. Consider a regression model

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + Z_{n\times q}b_{q\times 1} + \varepsilon_{n\times 1}, \quad p, q < n,$$

where X and Z have respective ranks p and q, $b \sim \mathcal{N}\{0, \sigma^2 Q(\psi)\}$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ are independent, and β and σ^2 are unknown.

(a) Show that if ψ is known, then the maximum likelihood estimator of β is

$$\hat{\beta}_\psi = \{X^{\mathrm{T}}(I_n + ZQZ^{\mathrm{T}})^{-1}X\}^{-1}X^{\mathrm{T}}(I_n + ZQZ^{\mathrm{T}})^{-1}y.$$

(b) Show that we can factorise the probability density of y as

$$f(y; \beta, \sigma^2, \psi) = f(\hat{\beta}_\psi; \beta, \sigma^2, \psi) f(y \mid \hat{\beta}_\psi; \sigma^2, \psi),$$

and deduce that estimators for σ^2 and ψ can be obtained from

$$\log f(y; \hat{\beta}_\psi, \sigma^2, \psi) + \frac{p}{2}\log\sigma^2 - \frac{1}{2}\log|X^{\mathrm{T}}(I_n + ZQZ^{\mathrm{T}})^{-1}X|.$$

(c) When Z is absent from the model, compute the estimator of σ^2 obtained in (b) and compare it with the usual maximum likelihood estimator. Discuss.

5. (a) Let $Y_{1,1}, \dots, Y_{R,C}$ be independent Poisson variables with means

$$\mu_{r,c} = \exp\left(\gamma_r + x_{rc}^{\mathrm{T}}\beta\right), \quad r = 1, \dots, R, c = 1, \dots, C,$$

where x_{rc} are vectors of covariates and $\gamma_1, \dots, \gamma_C, \beta$ are unknown parameters. By writing $\tau_r = \sum_c \mu_{rc}$, or otherwise, show that we can write

$$\ell_{\text{Poiss}}(\beta, \tau; y) = \ell_{\text{Poiss}}(\tau; m) + \ell_{\text{Mult}}(\beta; y \mid m),$$

where ℓ_{Poiss} and ℓ_{Mult} are respectively Poisson and multinomial log likelihoods. Explain how this result is useful when fitting log-linear models to count data.

(b) The table below shows data from a study of preferences for two brands of detergent, A and B, with the respondents also cross-classified by the temperature and softness of their washing water and whether or not they were previous users of brand B.

Previous B user	No				Yes			
Temperature	Low		High		Low		High	
Preference	A	B	A	B	A	B	A	B
Water softness								
Hard	68	42	42	30	37	52	24	43
Medium	66	50	33	23	47	55	23	47
Soft	63	53	29	27	57	49	19	29

If the classifying factors are denoted by `B.User`, `Temp`, `Brand` and `Soft` for previous B user, temperature, brand preference and water softness respectively, and, for example, `B.User*Temp` denotes the interaction of B user and temperature, use the calculations in (a) to say what model(s) it would be sensible to fit to the data. Explain your reasoning carefully.

(c) What do you learn from the `R` output below?

```
> anova(fit0,fit1,fit2,fit3)
Analysis of Deviance Table

Model 1: y ~ Temp * B.User * Soft + Brand
Model 2: y ~ Temp * B.User * Soft + Brand + Brand * B.User
Model 3: y ~ Temp * B.User * Soft + Brand + Brand * B.User + Brand * Temp
Model 4: y ~ Temp * B.User * Soft + Brand + Brand * B.User + Brand * Temp + Brand * Soft
  Resid. Df Resid. Dev Df Deviance
1        11     32.826
2        10     12.244  1  20.5815
3         9      8.444  1   3.8002
4         7      8.228  2   0.2160
```