tot punt 85 ⇒ $\frac{\text{punt}+10}{9.5}$

Exam Statistical Modelling

Thursday April 5, 2018, 18:30 - 21:30 A. Jacobshal 01

RULES AND REMARKS:

- The use of a normal, non-graphical calculator is permitted. In this exam you can use the usual significance cut-off $\alpha = 0.05$.
- This is a CLOSED-BOOK exam consisting of 5 exercises in total.
- Explain in all cases the reasoning leading to your answer and provide each page with your name and student number.
- The number of points per question are indicated by a box. We wish you success with the completion!
- Clearly indicate your type of curriculum.

1. **Poisson Modelling**. Suppose that $Y_1, \cdots, Y_n$ are independently Poisson distributed with mean $\theta_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$, for $i = 1, \cdots, n$, with density

$$f(y_i; \theta_i) = \frac{\theta_i^{y_i} e^{-\theta_i}}{y_i!}, \quad \text{where } y_i > 0.$$

(a) $\boxed{5}$ Assume $\theta_i = \theta$, for all i and show that the Poisson belongs to the exponential family of distributions and use the general expectation and variance formulas for exponential families to prove that $E[Y] = \text{Var}[Y] = \theta$.

(b) $\boxed{5}$ Assume $\theta_i = \theta$ for all i and derive the Fisher information for a sample $Y_1, \cdots, Y_n$ and give a simple condition under which the information tends to infinity as the size of the sample increases.

(c) $\boxed{10}$ Assume that $\theta_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$ for all i and derive the following expression for the deviance between this model and the saturated model

$$D = 2\left\{\sum_{i=1}^n y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - \sum_{i=1}^n (y_i - \hat{y}_i)\right\}.$$

Use the 2nd order Taylor approximation $o\log(\frac{o}{e}) = o - e + \frac{1}{2}\left(\frac{(o-e)^2}{e}\right)$ to show that the deviance is asymptotically chi-squared distributed.

(d) $\boxed{5}$ Find explicit expressions for the terms $\boldsymbol{W}^{(m-1)}$ and $\boldsymbol{z}^{(m-1)}$ in the iteratively reweighted least squares updating scheme

$$\boldsymbol{b}^{(m)} = (\boldsymbol{X}^T\boldsymbol{W}^{(m-1)}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}^{(m-1)}\boldsymbol{z}^{(m-1)}, \quad \text{where}$$

$$w_{ii} = \frac{1}{\text{Var}Y_i}\left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2, \quad \text{and} \quad z_i = \boldsymbol{x}_i^T\boldsymbol{b} + (y_i - \mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right)$$

Table 1: 2×2 table for a prospective study of exposure and disease outcome.

		Disease	
		Yes	No
Exposure	Yes	π_1	$1-\pi_1$
	No	π_2	$1-\pi_2$

2. **Logistic modelling**. Consider a 2×2 contingency table from a prospective study in which people who were or were not exposed to some pollutant are followed up and, after several years, categorized according to the presence or absence of a disease. Table 1 shows the probabilities for each cell. The odds of disease for either exposure group is $O_i = \pi_i/(1-\pi_i)$, for $i = 1, 2$, and the odds ratio $\phi = O_1/O_2$.

   (a) $\boxed{7}$ For the simple logistic model $\pi_i = e^{\beta_i}/(1+e^{\beta_i})$, show that $\phi = 1$ if and only if there is no difference between the exposed and not exposed groups (i.e., $\beta_1 = \beta_2$).

   (b) $\boxed{8}$ Consider J 2×2 tables like Table 1, one for each level x_j of a factor, such as age group, with $j = 1, ..., J$. For the logistic model

$$\pi_{ij} = \frac{\exp(\alpha_i + \beta_i x_j)}{1 + \exp(\alpha_i + \beta_i x_j)}, \quad i = 1, 2, \; j = 1, ..., J.$$

   Show that $\log \phi$ is constant over all tables if and only if $\beta_1 = \beta_2$.

3. $\boxed{10}$ **Independent Poisson**. Suppose that Y_1, Y_2 are independent Poisson random variables with means μ and $\rho\mu$ respectively. Show that

$$Y_1 | Y_1 + Y_2 = m \quad \sim \quad \text{Bin}\left(m, \frac{1}{1+\rho}\right)$$

4. $\boxed{10}$ **Chi-square**. Suppose that the n-dimensional vector $Y \sim MVN(0, I)$ and suppose that the matrix A is symmetric with rank a, $a < n$. Show that $A = A^2$ is a necessary and sufficient condition for $Y^T AY \sim \chi^2(a)$.
Hint: The moment generating function of the $\chi^2(1)$ distribution is $(1-2t)^{-1/2}$.

SEE NEXT PAGE

5. **Survival Analysis**. Suppose that Y is a continuous random variable indicating the survival time with distribution function $F(y)$ and survivor function $S(y) = P(Y \geq y)$.

(a) $\boxed{7}$ The hazard function $h(y)$ is defined as the probability of death in $[y, y + \delta y]$ given survival up to y, i.e. $Y > y$, relative to an infinitely small interval ($\delta y \to 0$). Show that

$$h(y) = -\frac{d}{dy}\log[S(y)].$$

(b) $\boxed{8}$ Suppose that Y has a Weibull density $f(y; \lambda, \phi) = \lambda\phi y^{\lambda-1}\exp(-\phi y^{\lambda})$. Derive the median survival time, the survivor function $S(y) = \exp(-\phi y^{\lambda})$, and its corresponding hazard function $h(y; \lambda, \phi) = \lambda\phi y^{\lambda-1}$.

(c) $\boxed{2}$ Proportional hazards are defined as those having the property $h_i(y) = \eta_i h_0(y)$, where $\eta_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$ and $h_0(y)$ the baseline hazard. Let $\phi_i = \eta_i\phi$ for the i-th subject. Show that the Weibull survival model has proportional hazards.

(d) $\boxed{8}$ The odds of survival past time y is defined as

$$O(y) = \frac{S(y)}{1 - S(y)}.$$

Give for the Weibull survival model an expression for $O(y)$ and show that the model does not act multiplicatively on the odds of survival beyond time y in the sense that $O_i = \eta_i O_0$, where O_0 is the baseline odds, $\phi_i = \eta_i\phi$, and $\eta_i = \exp(\boldsymbol{x}_i^T\boldsymbol{\beta})$.

1(a) $f(y;\theta) = \exp\{y \cdot \log(\theta) - \theta - \log y!\}$

$a(y) = y$

$b(\theta) = \log\theta, \; b'(\theta) = \theta^{-1}, \; b''(\theta) = -\theta^{-2}$ 1

$c(\theta) = -\theta \;, \; c'(\theta) = -1 \;, \; c''(\theta) = 0$ 1

(5) $E[Y] = -\frac{c'(\theta)}{b'(\theta)} = -\frac{-1}{\theta^{-1}} = \theta$ 1

$Var[Y] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{[b'(\theta)]^3} = \frac{-\theta^{-2}\cdot -1 - 0}{[\theta^{-1}]^3} = \theta$ 2

(b) $\mathfrak{J} = [b'(\theta)]^2 \cdot Var[a(Y)] = (\theta^{-1})^2 \cdot \theta = \theta^{-1}$ 2

(5) $0 < \theta \Rightarrow 0 < \frac{1}{\theta} \Rightarrow \mathfrak{J}_n = n\mathfrak{J} = n\theta^{-1} \longrightarrow \infty \;, \; n \to \infty$ 3

(c) $\ell(y_1,\dots,y_n;\theta_1,\dots,\theta_n) = \log\prod_{i=1}^{n} \frac{\theta_i^{y_i} e^{-\theta_i}}{y_i!} = \sum_{i=1}^{n} y_i \log\theta_i - \theta_i - \log(y_i!)$

$\frac{\partial}{\partial\theta_i}\ell = \frac{y_i}{\theta_i} - 1 = 0 \Rightarrow \hat{\theta}_i = y_i \;;\; \frac{\partial^2}{\partial\theta_i^2}\ell = -y_i\theta_i^{-2} < 0 \Rightarrow$ max at $\hat{\theta}_i = y_i$ 2

$\ell(b_{max}, y) = \sum_{i=1}^{n} y_i \log(y_i) - y_i - \log(y_i!)$

(10) IRLS $\Rightarrow \hat{b} \Rightarrow \hat{\theta}_i = \exp(x_i^T b) \;;\; EY_i = \theta_i$ we take $\hat{y}_i = \hat{\theta}_i$

$\ell(\hat{b}; y) = \sum_{i=1}^{n} y_i \log\hat{y}_i - \hat{y}_i - \log y_i!$ 2

saturated model

$D = 2[\ell(b_{max}; y) - \ell(\hat{b}; y)] = 2\left[\sum_{i=1}^{n} y_i \log\left(\frac{y_i}{\hat{y}_i}\right) - \sum_{i=1}^{n}(y_i - \hat{y}_i)\right]$ 3

$= 2\left\{\sum_{i=1}^{n} o_i \log\left(\frac{o_i}{e_i}\right) - \sum(o_i - e_i)\right\} \underset{\text{Taylor}}{\approx} \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(n-p)$ 3

(d) $\mu_i = e^{\eta_i} \Rightarrow \frac{\partial\eta_i}{\partial\mu_i} = \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^{-1} = (e^{\eta_i})^{-1} = \mu_i^{-1} \;;\; \eta_i = x_i^T\beta$

$w_{ii} = \frac{1}{Var[Y_i]} \cdot \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2 = \frac{1}{\mu_i} \cdot \mu_i^2 = \mu_i = e^{\eta_i} = e^{x_i^T b} \;;\; w_{ii}^{(m-1)} = \exp(x_i^T b^{(m-1)})$ 2

(5) $z_i^{(m-1)} = x_i^T b^{m-1} + (y_i - \mu_i)\left(\frac{\partial\eta_i}{\partial\mu_i}\right) = x_i^T b^{(m-1)} + (y_i - \mu_i)\mu_i^{-1}$

$= x_i^T b^{(m-1)} + (y_i - \exp(x_i^T b^{(m-1)}) / \exp(x_i^T b^{(m-1)})$ 3

2(a) $1-\pi_i = \frac{1+e^{\beta_i}}{1+e^{\beta_i}} - \frac{e^{\beta_i}}{1+e^{\beta_i}} = (1+e^{\beta_i})^{-1}$

$O_i = \frac{\pi_i}{1-\pi_i} = \frac{e^{\beta_i}}{1+e^{\beta_i}} \cdot 1+e^{\beta_i} = e^{\beta_i} \quad , \; i=1,2$ 3

7

$\phi = \frac{O_1}{O_2} = \frac{e^{\beta_1}}{e^{\beta_2}} = \exp(\beta_1-\beta_2) = 1 \iff \beta_1=\beta_2$ 4

(b) $1-\pi_{ij} = (1+e^{\alpha_i+\beta_i x_j})^{-1}$

$O_{ij} = \frac{\pi_{ij}}{1-\pi_{ij}} = e^{\alpha_i+\beta_i x_j} \quad , \; i=1,2 , \; j=1,\dots,9$ 2

8

$\frac{O_{1j}}{O_{2j}} = \frac{e^{\alpha_1+\beta_1 x_j}}{e^{\alpha_2+\beta_2 x_j}} = \exp(\alpha_1-\alpha_2)\cdot \exp((\beta_1-\beta_2)x_j)$ constant odds ratio over j 4

$\iff \beta_1-\beta_2=0$ 2

Table j is

		Disease Yes	Disease no
Exposure	Yes	π_{1j}	$1-\pi_{1j}$
	no	π_{2j}	$1-\pi_{2j}$

3) $Y_1 = \text{Pois}(\mu)$, $Y_2 = \text{Pois}(p\mu)$, $Y_1 \perp Y_2 \Rightarrow$

$$P(Y_1+Y_2=m) = \sum_{y=0}^{m} P(Y_1=y)\cdot P(Y_2=m-y)$$ 2

$$= \sum_{y=0}^{m} \frac{\mu^y e^{-\mu}}{y!} \cdot \frac{(p\mu)^{m-y} e^{-p\mu}}{(m-y)!}$$

$$= e^{-\mu(1+p)} \cdot \mu^m p^m \cdot \frac{1}{m!} \sum_{y=0}^{m} \frac{m!}{y!(m-y)!} \cdot p^{-y}$$

$$= e^{-\mu(1+p)} \mu^m p^m \frac{1}{m!} \left(\frac{p}{1+p}\right)^{-m} \underbrace{\sum_{y=0}^{m} \binom{m}{y} \left(\frac{1}{1+p}\right)^y \left(\frac{p}{1+p}\right)^{m-y}}_{=1}$$ using that 2

$$\left\{ \left(\frac{p}{1+p}\right)^m p^{-y} = \left(\frac{1}{1+p}\right)^y \cdot \left(\frac{p}{1+p}\right)^{m-y} \right.$$

$$P(Y_1=y_1 \mid Y_1+Y_2=m) = \frac{P(Y_1=y_1 \cap Y_1+Y_2=m)}{P(Y_1+Y_2=m)} = \frac{P(Y_1=y_1, Y_2=m-y_1)}{P(Y_1+Y_2=m)}$$ 2

$$= \frac{\frac{\mu^{y_1} e^{-\mu}}{y_1!} \cdot \frac{(p\mu)^{m-y_1} e^{-p\mu}}{(m-y_1)!}}{e^{-\mu(1+p)} \mu^m p^m \frac{1}{m!} \left(\frac{p}{1+p}\right)^{-m}}$$

$$= \frac{m!}{y_1!(m-y_1)!} \cdot \left(\frac{p}{1+p}\right)^m p^{-y_1}$$ 2

$$= \binom{m}{y_1} \cdot \left(\frac{1}{1+p}\right)^{y_1} \cdot \left(\frac{p}{1+p}\right)^{m-y_1}$$

$$= P(Y=y_1), \quad Y \sim \text{Bin}\left(m, \frac{1}{1+p}\right)$$ 2

$\Leftarrow$: $A^2 = A = A^T$ w log.

$A = K \Lambda K^T \ (KK^T = K^T K = I)$

$Y \sim N(0, I) \Leftrightarrow Y_1, \ldots, Y_n \sim \text{IID } N(0,1)$ 3

10

$Y^T A Y = Y^T K \Lambda K^T Y = X^T \Lambda X \, ; \quad X = K^T Y \sim N(0, I)$ 3

$$\varphi_{Y^T A Y}(t) = \varphi_{\sum \lambda_j X_j^2}(t) = \prod_{j=1}^{n} \varphi_{X_j^2}(\lambda_j t)$$

$$= \prod_{j=1}^{n} (1 - 2\lambda_j t)^{-1/2}$$

$$= (1 - 2t)^{-a/2} \Leftrightarrow \Lambda = \begin{bmatrix} I_a & 0 \\ 0 & 0 \end{bmatrix}$$ 4

5.(a) $h(y) = \lim_{\delta_y \to 0} \dfrac{P(Y \in [y, y+\delta_y] \mid Y > y)}{\delta}$ 2

$= \lim_{\delta_y \to 0} \dfrac{P(y \le Y \le y+\delta_y \cap y < Y)}{\delta_y P(Y > y)} = \dfrac{1}{S(y)} \lim_{\delta_y \to 0} \dfrac{F(y+\delta_y) - F(y)}{\delta_y}$ 2

(7)

$= \dfrac{f(y)}{S(y)} = -\dfrac{1}{S(y)} \dfrac{d}{dy}(1 - F(y)) = -\dfrac{1}{S(y)} \dfrac{d}{dy} S(y) = -\dfrac{d}{dy} \log S(y)$ 3

(b) $S(y) = P(Y > y) = -\int_y^{\infty} -\phi t^{\lambda-1} \cdot \lambda \cdot \exp(-\phi t^{\lambda}) dt$ (by substitution)

(8)

$\left.\begin{array}{l} \\ = -\int_{-\phi y^{\lambda}}^{-\infty} \exp(u) du = -\left[e^u\right]_{-\phi y^{\lambda}}^{-\infty} = -(0 - e^{-\phi y^{\lambda}}) = \exp(-\phi y^{\lambda}) \end{array}\right\} 5$

$h(y) = -\dfrac{d}{dy} \log S(y) = -\dfrac{d}{dy} -\phi y^{\lambda} = \lambda \phi y^{\lambda-1}$ 1

$e^{-\phi y^{\lambda}} = 2^{-1} \Rightarrow y = (\log(2)/\phi)^{1/\lambda}$ 2

(c) $h_i(y) = \lambda \phi_i y^{\lambda-1} = \lambda \eta_i \phi y^{\lambda-1} = \eta_i \lambda \phi y^{\lambda-1} = \eta_i h_0(y)$ 2

(2)

(d) $S(y) = e^{-\phi y^{\lambda}}$

$O(y) = \dfrac{S(y)}{1 - S(y)} = \dfrac{e^{-\phi y^{\lambda}}}{1 - e^{-\phi y^{\lambda}}} \cdot \dfrac{e^{\phi y^{\lambda}}}{e^{\phi y^{\lambda}}} = (e^{\phi y^{\lambda}} - 1)^{-1}$ 3

(8)

$\left.\begin{array}{l} O_i(y) = \eta_i O_0(y) \Leftrightarrow \qquad \phi_i = \eta_i \phi \\ (e^{\phi_i y^{\lambda}} - 1)^{-1} = \eta_i (e^{\phi y^{\lambda}} - 1)^{-1} \quad \forall y > 0, \forall \eta_i > 0 \end{array}\right\} 2$

$\left.\begin{array}{l} e^{\phi y^{\lambda}} - 1 = \eta_i (e^{\eta_i \phi y^{\lambda}} - 1) \quad \forall \eta_i \\ \eta (e^{\phi y^{\lambda}})^{\eta} - e^{\phi y^{\lambda}} + 1 - \eta = 0 \quad \forall \eta \\ \eta x^{\eta} - x + 1 - \eta = 0 \quad \forall \eta > 0, \forall x > 0. \quad \text{contradiction.} \end{array}\right\} 3$