Sol'ns

Exam 1 for Stat 7249 - Spring 2009

Feb. 24, 2009

Instructions

- budget your time wisely
- make sure to show your work for partial credit
- problems are worth 25 points each

1. Consider the inverse Gaussian density given below,

$$f(y) = \begin{cases} 0, & y \le 0 \\ (\frac{\lambda}{2\pi y^2})^{1/2}\exp\{-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\}, & y > 0. \end{cases}$$

   (a) Show that the density can be written in exponential dispersion form. Identify the components. Also, what is the canonical link and the variance function, $V(\mu)$ [where $Var(Y) = a(\phi)V(\mu)$]?

   (b) Derive the cumulant generating function and use it to derive the first two cumulants.

(a) $f(y) = \exp\left\{\left[-\frac{y}{2\mu^2} - \left(-\frac{1}{\mu}\right)\right]/(1/\lambda) + \underbrace{\frac{1}{2}\log\lambda - \frac{1}{2}\log 2\pi y^2 - \frac{\lambda}{2y}}_{c(y,\lambda)}\right\}$

$\theta = -\frac{1}{2\mu^2}$, $\mu = (-2\theta)^{-\frac{1}{2}}$

$b(\theta) = -(-2\theta)^{\frac{1}{2}} = -\frac{1}{\mu}$

canonical link $\theta = -\frac{1}{2\mu^2}$, $V(\mu) = \mu^3$

(b) from class/hw

$K_Y(t) = \frac{1}{a(\phi)}\left[b(\theta + a(\phi)t) - b(\theta)\right]$

$E[Y] = K_Y'(t)|_{t=0} = b'(\theta) = (-2\theta)^{-\frac{1}{2}} = \mu$

$Var(Y) = K_Y''(t)|_{t=0} = a(\phi)b''(\theta) = \frac{1}{\lambda}(-2\theta)^{-3/2} = \mu^3/\lambda$

2. In toxicological experiments in which the probability killed, $\pi(x)$ is modelled as a function of dose, $g(\pi(x)) = \alpha + \beta x$, the x value at which $\pi(x) = \frac{1}{2}$ is called the median lethal dose (denoted by LD50).

   (a) Derive the LD50 for the complementary log-log link.

   (b) Show that for the probit and logit link, the $LD50 = -\alpha/\beta$.

   (c) Let's now derive a confidence interval for the LD50 in the logit model. Suppose the sample is large enough such that the mle's $(\hat{\alpha}, \hat{\beta})$ are bivariate normal with covariance matrix Σ. Let $\theta = -\alpha/\beta$.

      i. Derive the distribution of $\hat{\alpha} + \theta\hat{\beta}$.

      ii. Use this to derive a confidence interval for θ (the LD50).

(a) $\frac{1}{2} = \pi(x) = 1 - e^{-e^{\alpha+\beta x}}$

$\Rightarrow x = LD50 = \frac{\log\log 2 - \alpha}{\beta}$

(b) $\frac{1}{2} = \pi(x) = \frac{e^{\alpha+\beta x}}{1+e^{\alpha+\beta x}}$ logit

$\Rightarrow x = LD50 = -\alpha/\beta$

$\frac{1}{2} = \pi(x) = \Phi(\alpha+\beta x)$ probit

$\Rightarrow x = LD50 = -\alpha/\beta$

(c) (i) $(\hat{\alpha} + \theta\hat{\beta}) \sim N\left(0, \underbrace{(1,\theta)\Sigma\binom{1}{\theta}}_{\sigma^2(\theta)}\right)$

(ii) $P\left(\frac{(\hat{\alpha}+\theta\hat{\beta})^2}{\sigma^2(\theta)} < Z^2_{\alpha/2}\right) = \alpha$

$P\left(\frac{(\hat{\alpha}+\theta\hat{\beta})^2}{Var(\hat{\alpha}) + \theta^2 Var(\hat{\beta}) + 2\theta Cov(\hat{\alpha},\hat{\beta})} < Z^2_{\alpha/2}\right) = \alpha$

- quadratic in θ
- solve for θ to obtain exact confidence limits

3. Let $Y_1, \dots Y_n \sim Ber(\pi_i)$, where $\pi_i = \Phi(\beta_0 + \sum_{j=1}^p x_{ij}\beta_j)$. Define a latent variable $Z_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \epsilon_i$, where $\epsilon_i \sim N(0, \sigma^2)$. The observed data is connected to this latent variable by $Y_i = I\{Z_i > c\}$.

(a) Show that the variance (σ^2) and threshold (c) are not identified by the data $\mathbf{Y}$.

(b) Suggest a way to extend this formulation to correlated binary data. Will there be unidentified parameters here? Explain.

(a)

$$\Phi(\beta_0 + \sum x_{ij}\beta_j) = \pi_i = P(Y_i = 1)$$
$$= P(Z_i > c)$$
$$= P\left(Z^* > \frac{c - (\beta_0 + \sum x_{ij}\beta_j)}{\sigma}\right) \quad \text{where } Z^* \sim N(0,1)$$
$$= 1 - \Phi\left(\frac{c - (\beta_0 + \sum x_{ij}\beta_j)}{\sigma}\right)$$
$$= \Phi\left(\frac{\beta_0 - c}{\sigma} + \sum x_{ij}\frac{\beta_j}{\sigma}\right)$$
$$= \Phi(\beta_0^* + \sum x_{ij}\beta_j^*) \quad \text{where } \beta_0^* = \frac{\beta_0 - c}{\sigma}, \ \beta_j^* = \beta_j/\sigma$$

(b) assume $\underset{\sim}{Y_i} = (Y_{i1}, \dots, Y_{iT}) \leftarrow$ correlated binary

$\underset{\sim}{\epsilon_i} \sim N(0, \Sigma)$ where $Y_{ij} = I\{Z_{ij} > c\}$

and variances (diagonal) elements of Σ are not identified (usually assume a correlation matrix)

- 'multivariate probit model'

4. Suppose $Y_1, \dots, Y_n$ are independently distributed as Poisson random variables with mean $\mu_i, i = 1, \dots, n$. Suppose μ_i is linked to covariates using a log link, $\log(\mu_i) = \boldsymbol{\beta}^T \boldsymbol{x}_i$.

(a) Derive the likelihood equations, the expected information matrix, and the deviance.

(b) Derive the likelihood ratio test for comparing two nested models, denoted by A and B (A nested within B) with parameters p_A and p_B respectively (and thus, $p_A < p_B$).

(c) Show that when the $\boldsymbol{x}_i$ vector contains an intercept (and the log link is used) that

$$D(\boldsymbol{y}; \hat{\boldsymbol{\mu}}_A) - D(\boldsymbol{y}; \hat{\boldsymbol{\mu}}_B) = 2\sum_i \hat{\mu}_{iB} \log \frac{\hat{\mu}_{iB}}{\hat{\mu}_{iA}}.$$

(a) see notes

(b) see notes

(c)

$$D(y; \hat{\mu}_A) - D(y, \hat{\mu}_B) = 2\sum\left[y_i \log \frac{\hat{\mu}_{iB}}{\hat{\mu}_{iA}} - \hat{\mu}_{iB} + \hat{\mu}_{iA}\right]$$

- from lik eqs w/ an intercept (and log link)

$\sum y_i = \sum \hat{\mu}_{iB} = \sum \hat{\mu}_{iA}$

$$\Rightarrow = 2\sum y_i \log \frac{\hat{\mu}_{iB}}{\hat{\mu}_{iA}}$$

log link

$$= 2\left[\sum_{i=1}^n y_i \sum_{j=1}^{p_B} x_{ij}\hat{\beta}_{jB} - \sum_{i=1}^n y_i \sum_{j=1}^{p_A} x_{ij}\hat{\beta}_{jA}\right]$$
$$= 2\left[\sum_{j=1}^{p_B} \hat{\beta}_{jB} \sum_{i=1}^n y_i x_{ij} - \sum_{j=1}^{p_A} \hat{\beta}_{jA} \sum_{i=1}^n y_i x_{ij}\right]$$
$$= 2\left[\sum_{j=1}^{p_B} \hat{\beta}_{jB} \sum \hat{\mu}_{iB} x_{ij} - \sum_{j=1}^{p_A} \hat{\beta}_{jA} \sum \hat{\mu}_{iB} x_{ij}\right]$$
$$= 2\left[\sum_{i=1}^n \hat{\mu}_{iB} \log \hat{\mu}_{iB} - \sum \hat{\mu}_{iB} \log \hat{\mu}_{iA}\right] \checkmark$$