Anthony Davison

Problem 1 When the list of standard link functions is insufficiently rich, it may be useful to introduce more flexible, parametrised, link functions.

(a) Show that the parametric link functions

$$g(\pi;\gamma) = \log\left[\gamma^{-1}\left\{(1-\pi)^{-\gamma}-1\right\}\right], \quad g(\pi;\gamma) = -\log\left\{\gamma^{-1}\left(\pi^{-\gamma}-1\right)\right\}, \quad \gamma \neq 0,$$

give respectively the logit and complementary log-log link functions, and the logit and log-log link functions, when $\gamma = 1$ and when $\gamma \to 0$. Give formulae for π in terms of η and γ in each case.

(b) A link function $\eta = g(\pi;\gamma)$ is called symmetric if $-\eta = g(1-\pi;\gamma)$. Show that

$$g(\pi;\gamma) = 2\gamma^{-1}\frac{\pi^\gamma - (1-\pi)^\gamma}{\pi^\gamma + (1-\pi)^\gamma}, \quad \gamma \neq 0,$$

is symmetric for all γ and gives the logit and identity functions when $\gamma \to 0$ and when $\gamma = 1$.

(c) How would you fit such models and choose a suitable value of γ?

Problem 2

(a) If $Y_1, \ldots, Y_D$ are independent Poisson variables with means $\mu_1, \ldots, \mu_D$, find the distribution of $S = \sum_{d=1}^D Y_d$, and show that their joint distribution conditional on $S = m$ is multinomial with probabilities $\pi_d = \mu_d / \sum_{d'=1}^D \mu_{d'}$ and denominator m, i.e.,

$$\mathrm{P}(Y_1 = y_1, \ldots, Y_D = y_D \mid S = m) = \frac{m!}{\prod_{d=1}^D y_d!}\prod_{d=1}^D \pi_d^{y_d}, \quad y_1, \ldots, y_D \in \{0, \ldots, m\}, \sum_{d=1}^D y_d = m.$$

(b) When m independent individuals are put into categories $1, \ldots, D$ with respective probabilities $\pi_1, \ldots, \pi_D$, let I_{jd} denote the indicator variable that the jth individual is placed into category d; note that $I_{j1} + \cdots + I_{jD} = 1$, as each individual can only be placed into one category. Find the joint density of $R_1, \ldots, R_D$, where $R_d = \sum_{j=1}^m I_{jd}$.

(c) Given that $(R_1, \ldots, R_D) \sim \mathrm{Mult}(m; \pi_1, \ldots, \pi_D)$, for some $D \geq 5$, find the joint distribution of $(R_1 + R_2, R_3 + R_4, R_5, \ldots, R_D)$ and the joint distribution of $(R_1, R_2, R_3 + R_4)$ conditional on $R_5 + \cdots + R_D = n$. Give the general version of this result.

Problem 3 For a 2×2 contingency table with probabilities

$$\begin{array}{cc} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{array}$$

and corresponding Poisson variables Y_{00} etc., the maximal log-linear model may be written as

$$\begin{aligned} \eta_{00} &= \alpha - \beta - \gamma + (\beta\gamma), & \eta_{01} &= \alpha - \beta + \gamma - (\beta\gamma), \\ \eta_{10} &= \alpha + \beta - \gamma - (\beta\gamma), & \eta_{11} &= \alpha + \beta + \gamma + (\beta\gamma), \end{aligned}$$

where $\eta_{jk} = \log \mathrm{E}(Y_{jk}) = \log(m\pi_{jk})$ and $m = \sum_{j,k} y_{jk}$ is taken as fixed.

(a) Show that the 'interaction' term $(\beta\gamma)$ may be written $(\beta\gamma) = \frac{1}{4}\log\Delta$, where Δ is the odds ratio $(\pi_{11}\pi_{00})/(\pi_{01}\pi_{10})$, so that $(\beta\gamma) = 0$ is equivalent to $\Delta = 1$.

(b) When $(\beta\gamma) = 0$ give interpretations of the parameters α, β and γ in terms of row, column and other effects.

Problem 4 One standard model for over-dispersed binomial data assumes that R is binomial with denominator m and probability π, where π has the beta density

$$f(\pi; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\pi^{a-1}(1-\pi)^{b-1}, \quad 0 < \pi < 1, a, b > 0,$$

with $\Gamma(a)$ the gamma function: if $n \in \mathbb{N}$, then $n! = \Gamma(n+1)$ and that $\Gamma(a+1) = a\Gamma(a)$ for $a > 0$.

(a) Show that the beta model for π yields the *beta-binomial* density

$$\mathrm{P}(R = r; a, b) = \frac{\Gamma(m+1)\Gamma(r+a)\Gamma(m-r+b)\Gamma(a+b)}{\Gamma(r+1)\Gamma(m-r+1)\Gamma(a)\Gamma(b)\Gamma(m+a+b)}, \quad r = 0, \ldots, m.$$

(b) Let μ and σ^2 denote the mean and variance of π. Show that in general,

$$\mathrm{E}(R) = m\mu, \quad \mathrm{var}(R) = m\mu(1-\mu) + m(m-1)\sigma^2,$$

and that the beta density has $\mu = a/(a+b)$ and $\sigma^2 = ab/\{(a+b)(a+b+1)\}$. Deduce that the beta-binomial density has mean and variance

$$\mathrm{E}(R) = ma/(a+b), \quad \mathrm{var}(R) = m\mu(1-\mu)\{1 + (m-1)\delta\}, \quad \delta = (a+b+1)^{-1}.$$

Can you explain the overdispersion is undetectable when $m = 1$? What is the condition for uniform overdispersion?