Solution 1

(a) This is easily checked for the first link when $\gamma = 1$, and for $\gamma \to 0$ we write

$$\gamma^{-1}\left\{(1-\pi)^{-\gamma}-1\right\} = \gamma^{-1}\left[\exp\{-\gamma\log(1-\pi)\}-1\right] = -\log(1-\pi)+O(\gamma) \to -\log(1-\pi),$$

as required. Setting $\eta = g(\pi;\gamma)$ and solving for π gives $\pi(\eta;\gamma) = 1-(1+\gamma e^{\eta})^{-1/\gamma}$, which indeed equals $1-1/(1+e^{\eta})$ when $\gamma = 1$ and tends to $1-\exp(-e^{\eta})$ when $\gamma \to 0$. Similar computations apply for the second link function.

(b) Symmetry is easily verified for all γ. When $\gamma \to 0$ we have $p^{\gamma} = 1+\gamma\log p + O(\gamma^2)$, so

$$g(\pi;\gamma) = 2\gamma^{-1}\frac{\pi^{\gamma}-(1-\pi)^{\gamma}}{\pi^{\gamma}+(1-\pi)^{\gamma}} = \frac{2}{\gamma}\frac{1+\gamma\log\pi-1-\gamma\log(1-\pi)+O(\gamma^2)}{1+\gamma\log\pi+1+\gamma\log(1-\pi)+O(\gamma^2)} \to \log\pi - \log(1-\pi),$$

as required, and $g(\pi;1) = 4\pi-2$, which is linear (though with a location shift).

(c) To fit such a model we would need code for fitting with γ fixed, and then compute a profile log likelihood for γ, which would then form the basis for confidence intervals on γ. Rather large amounts of data would be needed for such intervals to be at all useful.

Solution 2

(a) The distribution of S is Poisson with mean $\mu_1+\cdots+\mu_D$, so

$$\mathrm{P}(Y_1 = y_1,\ldots,Y_D = y_D \mid S = m) = \frac{\prod_{d=1}^{D}\frac{\mu_d^{y_d}}{y_d!}e^{-\mu_d}}{\frac{(\mu_1+\cdots+\mu_D)^m}{m!}\exp\{-(\mu_1+\cdots+\mu_D)\}},$$

which reduces to the given form on cancelling the exponential terms and writing $\mu_d = \pi_d(\mu_1+\cdots+\mu_D)$.

(b) R_d is the number of independent individuals in category d, and clearly $R_1+\cdots+R_D = m$. This is a generalisation of the binomial distribution from two categories (success/failure) to D categories, giving the multinomial distribution found in part (a), with $R_d \equiv Y_d$.

(c) We can use the indicator functions from (b) to see that the new variables $R_1' = R_1+R_2$ and $R_2' = R_3+R_4$ are sums of indicator variables for (partially merged) categories $1', 2', 5, \ldots, D$ with respective probabilities $\pi_1+\pi_2, \pi_3+\pi_4, \pi_5, \ldots, \pi_D$. Hence $(R_1+R_2, R_3+R_4, R_5, \ldots, R_D)$ is multinomial with these probabilities and denominator m.

For the second part, the joint distribution of $(R_1, R_2, R_3+R_4, R_5+\cdots+R_D)$ is multinomial with probabilities $(\pi_1, \pi_2, \pi_3+\pi_4, \pi_5+\cdots+\pi_D)$ and denominator m, and $R_5+\cdots+R_D$ is binomial with denominator m and probability $\pi_5+\cdots+\pi_D$. Now

$$\mathrm{P}(R_1 = r_1, R_2 = r_2, R_3+R_4 = r_3' \mid R_5+\cdots+R_D = n)$$

equals

$$\frac{\mathrm{P}(R_1 = r_1, R_2 = r_2, R_3+R_4 = r_3, R_5+\cdots+R_D = n)}{\mathrm{P}(R_5+\cdots+R_D = n)}$$

and after substituting in the corresponding multinomial and binomial probabilities and lots of cancellations we find that the required conditional distribution is multinomial with probability vector (π_1', π_2', π_3') and denominator $m - n$, where

$$\pi_1' = \frac{\pi_1}{\pi_1 + \cdots + \pi_4}, \quad \pi_2' = \frac{\pi_2}{\pi_1 + \cdots + \pi_4}, \quad \pi_3' = \frac{\pi_3 + \pi_4}{\pi_1 + \cdots + \pi_4}.$$

The general results are messy to state, but the broad conclusion is that the multinomial distribution is closed under marginalisation by summing variables or by conditioning on sums of variables. These properties are analogous to those of the multivariate normal distribution.

Solution 3

(a) On substituting the given expressions for the ηs we have

$$\log \Delta = \log\left(\frac{\pi_{11}\pi_{00}}{\pi_{01}\pi_{10}}\right) = \eta_{11} - \eta_{10} - (\eta_{01} - \eta_{00}) = \cdots = 4(\beta\gamma),$$

as required, so zero interaction, $\Delta = 1$, means that $\eta_{11} - \eta_{10} = \eta_{01} - \eta_{00}$, i.e, $\mu_{11}/\mu_{10} = \mu_{01}/\mu_{00}$ and $(\beta\gamma) = 0$.

(b) As $(\pi_{11}\pi_{10}\pi_{01}\pi_{00})^{1/4} = \exp(\alpha)$, α is the logarithm of the geometric mean of the probabilities for the cells.

We have $\eta_{11} - \eta_{10} = \eta_{01} - \eta_{00} = 2\gamma$, so

$$\mu_{11} = \mu_{10}e^{2\gamma}, \quad \mu_{01} = \mu_{00}e^{2\gamma},$$

i.e., the mean increases by a factor $\exp(2\gamma)$ when moving from column 1 to column 2.

A similar calculation shows that $\exp(2\beta)$ is the row effect.

Solution 4

(a) We have

$$\mathrm{P}(R = r \mid \pi) = \binom{m}{r}\pi^r(1-\pi)^{m-r}, \quad r \in \{0, 1, \ldots, m\},$$

and with the given beta density for π we therefore have

$$\begin{aligned}\mathrm{P}(R = r; a, b) &= \int_0^1 \mathrm{P}(R = r \mid \pi) f(\pi; a, b)\,\mathrm{d}\pi \\ &= \frac{\Gamma(m+1)}{\Gamma(r+1)\Gamma(m-r+1)} \times \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \times \frac{\Gamma(r+a)\Gamma(b+m-r)}{\Gamma(a+b+m)}, \quad r \in \{0, 1, \ldots, m\},\end{aligned}$$

where $a, b > 0$.

(b) We have $\mathrm{E}(R \mid \pi) = m\pi$ and $\mathrm{var}(R \mid \pi) = m\pi(1-\pi)$, so if $\mu = \mathrm{E}_\pi(\pi)$ and $\sigma^2 = \mathrm{var}_\pi(\pi)$ we have $\mathrm{E}(R) = m\mathrm{E}_\pi(\pi) = m\mu$ and

$$\mathrm{var}(R) = \mathrm{E}_\pi\{\mathrm{var}(R \mid \pi)\} + \mathrm{var}_\pi\{\mathrm{E}(R \mid \pi)\} = m(\mu - \sigma^2 - \mu^2) + m^2\sigma^2,$$

which reduces to the given formula. The mean and variance formulae for the beta density follow easily from computing $\mathrm{E}_\pi(\pi^n)$ for $n = 1, 2$ and using properties of the gamma function, and the rest is algebra.

When $m = 1$ we have $r \in \{0, 1\}$, so we could not see overdispersion, whereas if (say) $m = 2$ then $r \in \{0, 1, 2\}$, and in repeated sampling we might observe that the extremes 0 and 2 appeared more often than with the binomial distribution, corresponding to overdispersion. Likewise with $m = 3, \dots$, but with a two-point distribution any overdispersion is invisible, because the only possible values are 'extreme'.

The condition for uniform overdispersion would be that $1 + (m - 1)\delta$ was constant, i.e., that $m - 1 \propto (a + b + 1)$, so if the ms varied for different responses, there would need to be corresponding variation of $a + b$.