§10.4 - Contingency tables

We can model a

2×2 contingency table as 2 binomials:

$r_0 \sim Bin(m_0, \pi_0)$; $r_1 \sim Bin(m_1, \pi_1)$

$$\psi = \log\left\{\frac{\pi_1/(1-\pi_1)}{\pi_0\ (1-\pi_0)}\right\}$$ canonical par. in exp. family model

$\hat{\psi} \sim N(\psi, \hat{\sigma}_\psi)$; approx.$^{\underline{n}}$ is poor for small m_1, m_2

We can also find the exact distribution:

$$f(r_1, r_0) = \binom{m_1}{r_1}\binom{m_0}{r_0}\pi_1^{r_1}(1-\pi_1)^{m_1-r_1}\pi_0^{r_0}(1-\pi_0)^{m_0-r_0} \ ; \ \pi_1 = e^{\psi+\lambda}/(1+e^{\psi+\lambda}) \ ; \ \pi_0 = e^{\lambda}/(1+e^{\lambda})$$

$$= e^{(r_1+r_0)\lambda + r_1\psi - c(\psi,\lambda)}(\ )(\ )$$

$$f(r_1 \mid r_1+r_0=a;\psi) = \frac{\binom{m_1}{r_r}\binom{m-m_1}{a-r_r}e^{\psi r_1}}{\sum_{u=r_-}^{r_+}\binom{m_1}{u}\binom{m-m_1}{a-u}e^{\psi u}}$$

$$r_- = \max(0, a-m_0)$$
$$r_+ = \min(m_1, a)$$

sum over all tables with
compatible margins

Example 10.20 ; line 1

	(response) bleeding	not	
new	$8 = r_1$	7	$15 = m_1$
old	$2 = r_0$	11	$13 = m_0$
	10	18	

If $\psi = 0$,

$$P_r(\text{table}) = \binom{15}{8}\binom{13}{2} \Big/ \binom{28}{10}$$

p-value is $P(r_1 \geq 8 \; ; \psi = 0)$

$= P(r_1 = 8) + P(r_1 = 9) + P(r_1 = 10)$

$= 0.0453$ (typo in text)

Homework - Ex. 10.20
- read & program in R

§10.5 - loglinear models
& Poisson regression

10.5.2 $R \times C$ contingency tables

	1	$\cdots$	C
1	y_{11}	$\cdots$	y_{1C}
$\vdots$	$\vdots$		
R	y_{R1}	$\cdots$	y_{RC}

Three sampling models
1. Poisson
2. Multinomial
3. Product multinomial

<u>Model 1</u> : $y_{rc} \sim Po(\mu_{rc})$, ind't

$$f(\underset{\sim}{y}; \underset{\sim}{\mu}) = \prod_{r,c} \mu_{rc}^{y_{rc}} e^{-\mu_{rc}} / y_{rc}!$$

$$\mu_{rc} = e^{\gamma_r + x_{rc}^T \beta}$$

$$\ell(\underset{\sim}{\gamma}, \underset{\sim}{\beta}) = \sum_{r,c} \left\{ y_{rc}(\gamma_r + x_{rc}^T \beta) - e^{\gamma_r + x_{rc}^T \beta} \right\}$$

Model 2: fixed sample size

$\underset{\sim}{y} \sim Mult(n; \underset{\sim}{\pi})$

$$f(\underset{\sim}{y}) = \frac{n!}{\prod_{rc} y_{rc}!} \prod_{r,c} \pi_{rc}^{y_{rc}}$$

$$\pi_{rc} = \mu_{rc} / \sum_{st} \mu_{st}$$

Model 3: Fix row totals m_r

Product multinomial, one in each row

$$f(\underset{\sim}{y}) = \prod_{r=1}^{R} \frac{m_r!}{y_{r1}! \cdots y_{rC}!} \prod_{c} \pi_{rc}^{y_{rc}}$$

$$\sum_{c=1}^{C} \pi_{rc} = 1 \quad \forall r$$

Ex. Butterflies:

48 trinomials

GLM: $\mu_{rc} = e^{\gamma_r + x_{rc}^T \beta}$

$$\pi_{rc} = \frac{e^{\gamma_r + x_{rc}^T \beta}}{\sum e^{\gamma_r + x_{rt}^T \beta}}$$

sum over all r, c for model 2;
over c only for model 3
<u>In</u> either case π_{rc} is free of γ_r.

e.g., for 3;

$$\pi_{rc} = \frac{\mu_{rc}}{\sum_t \mu_{rt}} = \frac{e^{\gamma_r + x_{rc}^T \beta}}{\sum_t e^{\gamma_r + x_{rt}^T \beta}}$$

$$= e^{x_{rc}^T \beta} \Big/ \sum_t e^{x_{rt}^T \beta}$$

Poisson likelihood can be factored:

$$\ell_{Pois}(\beta) = \sum_{r,c} \mu_{rc} \log y_{rc} - \mu_{rc}$$

$$= \sum_{r,c} \left\{ y_{rc}(\gamma_r + x_{rc}^T \beta) - e^{\gamma_r + x_{rc}^T \beta} \right\}$$

$$= \sum_r m_r \gamma_r + \sum_{rc} y_{rc} x_{rc}^T \beta - \sum_{rc} e^{\gamma_r} e^{x_{rc}^T \beta}$$

Let $\tau_r = e^{\gamma_r} \sum_c e^{x_{rc}^T \beta}$

$$\log \tau_r = \gamma_r + \log \sum_c e^{x_{rc}^T \beta}$$

$$\ell = \sum_r (m_r \log \tau_r - \tau_r) + \sum_{rc} \left\{ y_{rc} x_{rc}^T \beta - m_r e^{x_{rc}^T \beta} \right\}$$

This shows that the log-likelihood function for the Poisson model, Model 1, factors into 2 terms — the first is a Poisson likelihood for m_r, the row totals, and the second is a multinomial log-lik function for β.

This means that Poisson glm's can be used to fit multinomial models. We'll do this with Jacamar data next week.