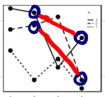


- When $\pi_{ij} \neq \pi_{i+}\pi_{+j}$ (X_1 and X_2 not independent) → check LNp.5-3: at least one $\Delta_{ij} \neq 1$ for a pair of (i, j)
  - ⇒ add interaction $X_1{:}X_2$ — $(I-1)(J-1)$ dummy var. — $\hat{\mu}_{ij} = Y_{ij}$
  - ⇒ may consider $Y_{ij} \sim X_1 + X_2 + X_1{:}X_2 \equiv L$ (saturated model)

$\eta_{ij} = \gamma + \alpha_i + \beta_j + (\alpha\beta)_{ij}$

$\alpha_I = 0$, $\beta_J = 0$, $(\alpha\beta)_{iJ} = 0, \forall i$, $(\alpha\beta)_{Ij} = 0, \forall j$ → $IJ - I - J + 1$

- **Q**: what type of π's corresponds to the following models?
  - without or with interactions ⇔ independent or dependent
  - without or with main effects ⇔? uniform or non-uniform marginal dist.

$Y_{ij} \sim 1$ $\qquad Y_{ij} \sim X_1 + X_1{:}X_2$ $\qquad Y_{ij} \sim X_2 + X_1{:}X_2$ $\qquad$ 4/21

➤ Recall. For a Poisson GLM with log link (canonical link), $X^T Y = X^T \hat{\mu}$ — exp $\hat{\eta} = X\hat{\beta}_{MLE}$

For models without interactions,

Sufficient statistics of β

⇒ $X^T Y$ is only related to marginal totals (Y_{i+}'s, Y_{+j}'s)

e.g., 2×2 table $X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$, $Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix}$, $X^T Y = \begin{bmatrix} Y_{++} \\ Y_{1+} \\ Y_{+1} \end{bmatrix}$ — $= Y_{1+} + Y_{2+}$, $= Y_{+1} + Y_{+2}$

In general, consider 2-way ANOVA model in DOE

⇒ the fitted values $\hat{\mu}$ is a function of marginal totals → $\hat{\mu}_{ij} = \exp(\hat{\eta}_{ij}) = \exp(\hat{\gamma} + \hat{\alpha}_i + \hat{\beta}_j)$ — Y_{++}, Y_{i+}'s, Y_{+j}'s

$\mu_{ij} = t\,\pi_{ij}$

$= t\,\pi_{i+}\pi_{+j}$

⇒ for example, for main-effect model $Y_{ij} \sim X_1 + X_2$

$= t\,\frac{\mu_{i+}}{\mu_{++}} \times \frac{\mu_{+j}}{\mu_{++}} = \frac{\mu_{i+} \times \mu_{+j}}{\mu_{++}}$ → $\hat{\mu}_{ij} = Y_{++}\,\hat{\pi}_{i+}\,\hat{\pi}_{+j} \overset{\Delta}{=} Y_{i+}Y_{+j}/Y_{++}$ (cf. $\square$) ← $\hat{\pi}_{i+} = Y_{i+}/Y_{++}$, $\hat{\pi}_{+j} = Y_{+j}/Y_{++}$

$t = \mu_{++}$

➤ To test whether $\pi_{ij} = \pi_{i+}\pi_{+j}$ (H_0) ⇒ H_0: S vs. H_1: $L \backslash S$ ★(LNp.5-5) — It's the goodness-of-fit test for S

using Poisson GLM

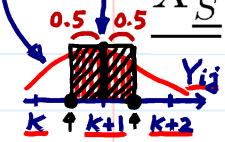
- Deviance based (also called G^2_S): $D_S - D_L \overset{a}{\sim} \chi^2_{(I-1)(J-1)}$ ← ∵ L is saturated ⇒ $D_L = 0$

- Pearson's X^2 (goodness-of-fit measure) under S: — estimate of expected count under S (independent)

continuous pdf, true pmf

$$X_S^2 = \sum_{ij} \frac{(Y_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \overset{a}{\sim} \chi^2_{(I-1)(J-1)}$$

Δ in LNp.5-6



□ *Yate's continuity correction*: (better for ...)

- Subtracts 0.5 from $Y_{ij} - \hat{\mu}_{ij}$ when it is positive
- Add 0.5 to $Y_{ij} - \hat{\mu}_{ij}$ when it is negative

$Y_{ij} \pm 0.5$ often closer to $\hat{\mu}_{ij}$ than Y_{ij}

often reduce X^2 ∵ $|Y - \hat{\mu}| - 0.5 < |Y - \hat{\mu}|$ ⇒ increase p-value ⇒ more conservative (sometimes, too conservative)

this give superior results for small samples — often suggest if some $Y_{ij} < 5$

➤ To test H_0: $\pi_{1+} = \cdots = \pi_{I+}$ (or $\pi_{+1} = \cdots = \pi_{+J}$),

compare models S^* and L^* (H_0: S^* vs. H_1: $L^* \backslash S^*$), where

- S^*: $Y_{ij} \sim X_2 + X_1{:}X_2$ and L^*: $Y_{ij} \sim X_1 + X_2 + X_1{:}X_2$ ✗
- S^*: $Y_{ij} \sim X_1{:}X_2$ and L^*: $Y_{ij} \sim X_1 + X_1{:}X_2$
- S^*: $Y_{ij} \sim X_2$ and L^*: $Y_{ij} \sim X_1 + X_2$ ○
- S^*: $Y_{ij} \sim 1$ and L^*: $Y_{ij} \sim X_1$

analysis strategy: Find a Poisson fitted GLM that
1. pass goodness-of-fit
2. pass diagnostics
3. effects are significant.

Then, explain the type of π's according to this Poisson fit

they might have different test results, be wise to choose a right one.

- Deviance-based test: $D_{S^*} - D_{L^*} \overset{a}{\sim} \chi^2_{df_{S^*} - df_{L^*}}$

$Y_{ij} \sim$ Poisson(μ_{ij}) indep., $i = 1, \cdots, I$, $j = 1, \cdots, J$

➤ Can be generalized to X_1 with I levels and X_2 with J levels

- Scheme 2: ← LNp.5-4 SRS

Consider 1 draw from the population. It follows multinomial$(1, \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$. For y_{++} independent draws, it then follows (LNp.2-23)

  - Model: for a random sample, we can assume

$\Sigma_{ij} Y_{ij} = y_{++}$ fixed

$(Y_{11}, Y_{12}, Y_{21}, Y_{22})$ ← LNp.2-22~23

The y_x's under different settings of $X=(X_1, X_2)$ are not independent

$\sim$ multinomial$(y_{++}, \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$

$\lambda_i/\Sigma_i\lambda_i \leftrightarrow \mu_{ij}/t = t\pi_{ij}/t$ ⇒ same π_{ij}'s as in the Scheme 1

as a function g^{-1} of η_{ij}, $\eta_{x_1,x_2} = X\beta$

where π_{ij} $(i=1, 2; j=1, 2)$ is linked to X_1 and X_2 according to the model we choose

Interpreted as conditional on $Y_{++} = y_{++}$

X_1 \ X_2	1	2	
1	Y_{11}	Y_{12}	Y_{1+}
2	Y_{21}	Y_{22}	Y_{2+}
	Y_{+1}	Y_{+2}	y_{++}

★ Connection between Poisson and multinomial: ← Scheme 1 ← LNp.2-26

$Y_i \leftrightarrow Y_{ij}$ | $\lambda_i \leftrightarrow \mu_{ij}$ | $\Sigma_i Y_i \leftrightarrow Y_{++}$ | $\Sigma_i\lambda_i \leftrightarrow t$ | $n \leftrightarrow y_{++}$

Let $Y_i \sim$ Poisson(λ_i), $i=1,\ldots,k$, and independent,

$\sim$Poisson$(\Sigma_i\lambda_i)$

$(Y_1,\ldots,Y_k | \Sigma_i Y_i = n) \sim$ multinomial$(n, \lambda_1/\Sigma_i\lambda_i, \ldots, \lambda_k/\Sigma_i\lambda_i)$ ↔ π_{ij}'s

check △ in LNp.5-5

Note. Analysis methods for scheme 2 can be regarded as conditional approaches for data of scheme 1

⇒ the parameter t (value of size variable) in Poisson is removed, but π_{ij}'s are not affected

After conditional on Y_{++} the information of t is gone (i.e., the fixed y_{++} not carry information of t)

⇒ would expect there is a lot of similarity between the inferences for Poisson and multinomial models

When $\mu_{ij} = t\pi_{ij}$

  - Log-likelihood of the multinomial: $\Sigma_{ij}\pi_{ij}=1$

$\Sigma_{ij} Y_{ij}[\log(t) + \log(\pi_{ij})]$

$= (\Sigma_{ij}Y_{ij})\log(t) + \Sigma_{ij} Y_{ij}\log(\pi_{ij})$, y_{++} (not random data any more)

pmf $\propto \prod_{ij} (\pi_{ij})^{y_{ij}}$ → $\ell = \log(\mathcal{L}) \propto \sum_{ij} Y_{ij} \log(\pi_{ij})$ ← LNp.2-22

(cf., log-likelihood for Poisson $\propto \sum_{ij} Y_{ij}\log(\mu_{ij}) - \mu_{ij}$) ← LNp.4-4

- The inferences in the multinomial model would coincide with that in Poisson model, i.e.,
  - same estimates (MLE)
  - same test statistics and p-values

∵ In Poisson GLM for scheme 1
- intercept ↔ t
- other effects ↔ π_{ij}'s

1. link function
   - $\log(y_{++}\pi_{ij}) = \eta_{ij}$ | $\log(\mu_{ij}) = \eta_{ij}$
   - logit ↔ $\log(\pi_{ij}/\pi_{i'j'}) = \eta_{ij} - \eta_{i'j'}$
2. MLE
   - $0 = \Sigma_{ij} \frac{\partial}{\partial \pi_{ij}}[\ell + \lambda(\Sigma_{ij}\pi_{ij} - 1)] \frac{\partial \pi_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \beta}$
     $= y_{ij}/\pi_{ij} - y_{++}$ (∵ $\hat\lambda = -y_{++}$) $\propto \frac{y_{ij}}{y_{++}\pi_{ij}} - 1$
   - $0 = \Sigma_{ij} \frac{\partial \ell}{\partial \mu_{ij}} \frac{\partial \mu_{ij}}{\partial \eta_{ij}} \frac{\partial \eta_{ij}}{\partial \beta}$
     $= y_{ij}/\mu_{ij} - 1$ (LNp.4-5)
3. deviance of S:
   - $\Sigma_{ij} y_{ij} \log(y_{++}\hat\pi_{ij}/y_{ij})$
   - $\Sigma_{ij} y_{ij}\log(\hat\mu_{ij}/y_{ij}) + \Sigma_{ij}(y_{ij} - \hat\mu_{ij})$, $= 0$ if S contain intercept (LNp.4-6)

Pretend y_{++} is random. But note that intercept not carry information of t

- The Poisson model is easier to execute in R, so we can fit a Poisson GLM for data from a multinomial sampling scheme
  - Can be generalized to $I \times J$ table in the same manner

4/23

- Scheme 3: binomial GLM ← LNp.5-4

X_2	Y_x	n_x
1	Y_{11}	y_{+1}
2	Y_{12}	y_{+2}

Y_{ij}'s ~ multinomial$(y_{++}, \pi_{ij}$'s$)$, $i=1,\ldots,I$, $j=1,\ldots,J$, $\Sigma_{ij}\pi_{ij}=1$

  - Model: for a random sample, can assume SRS

By ★ in LNp.5-8, $\pi_{i=1|j} = \frac{\mu_{1j}}{\mu_{1j}+\mu_{2j}}$

$Y_{1j} \sim$ binomial$(y_{+j}, \pi_{i=1|j} = \pi_{1j}/\pi_{+j})$, $j=1, 2$

indep.

as a function g^{-1} of η_j, $\eta_{x_2} = X\beta$

X_1 hidden in response (隱含)

where $\pi_{i=1|j}$ is linked to the covariate X_2 $(=j)$ only according to the model we choose (∗) cf. X_1 not a covariate (不是)

to be consistent with log link in Poisson GLM

X_1 \ X_2	1	2	
1	Y_{11}	Y_{12}	Y_{1+}
2	Y_{21}	Y_{22}	Y_{2+}
	y_{+1}	y_{+2}	y_{++}

  - Q: compared to schemes 1 and 2, what information has been gone/questionable in this scheme? → t ✗, π_{+j} ✗, π_{i+} ?, $\pi_{j|i}$ ✗

- Suppose fit the data with a Binomial GLM with logit link: