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

The inferences in the multinomial model would coincide with that in Poisson model, i.e.,

- same estimates (MLE)
- same test statistics and p-values

1. link function: ① $\log(y_{++}\pi_{ij})=\eta_{ij}$ • $\log(\mu_{ij})=\eta_{ij}$; logit ↔ $\log(\pi_{ij}/\pi_{i'j'})=\eta_{ij}-\eta_{i'j'}$

2. MLE
- $0=\sum_{ij}\frac{\partial}{\partial\pi_{ij}}\left[\ell+\lambda(\sum_{ij}\pi_{ij}-1)\right]\frac{\partial\pi_{ij}}{\partial\eta_{ij}}\frac{\partial\eta_{ij}}{\partial\beta}$ ↳ $=y_{ij}/\pi_{ij}-y_{++}$ (∵ $\hat\lambda=-y_{++}$) $\propto \frac{y_{ij}}{y_{++}\pi_{ij}}-1$
- $0=\sum_{ij}\frac{\partial\ell}{\partial\mu_{ij}}\frac{\partial\mu_{ij}}{\partial\eta_{ij}}\frac{\partial\eta_{ij}}{\partial\beta}$ ↳ $=y_{ij}/\mu_{ij}-1$ (LNp 4-5)

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

3. deviance of S:
- $\sum_{ij}y_{ij}\log(y_{++}\hat\pi_{ij}/y_{ij})$
- $\sum_{ij}y_{ij}\log(\hat\mu_{ij}/y_{ij})+\sum_{ij}(y_{ij}-\hat\mu_{ij})$, (LNp 4-6) $=0$ if S contain intercept

➢ 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_{++}, π_{ij}'s) $i=1,\dots,I$, $j=1,\dots,J$, $\sum_{ij}\pi_{ij}=1$

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

➢ Model: for a random sample, can assume SRS

$Y_{1j}\sim$ 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)

to be consistent with log link in Poisson GLM

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 (規)

	X_2: 1	2	
X_1: 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:

$E(Y_{1+})=y_{+1}\pi_{1|1}+y_{+2}\pi_{1|2}=(y_{+1}+y_{+2})\pi_{1+}$ if indep. (*)
& $Y_{1+}\sim B(y_{++},\pi_{1+})$ ⇒ Y_{1+}/y_{++} can be used to estimate π_{1+}
similar result for Y_{2+}

◉ When $\pi_{ij}=\pi_{i+}\pi_{+j}$ (X_1 and X_2 independent),

Q Why can study independence in scheme 3? Ans. ∵ indep can be defined in terms of conditional dist. (check LNp 5-2)

$$\mathrm{logit}(\pi_{i=1|j})=\log\left[(\pi_{1j}/\cancel{\pi_{+j}})/(\pi_{2j}/\cancel{\pi_{+j}})\right]$$

$1-\pi_{i=1|j}$

∵ indep.

$$=\log\left[(\pi_{1+}\cancel{\pi_{+j}})/(\pi_{2+}\cancel{\pi_{+j}})\right]=\log(\pi_{1+}/\pi_{2+})$$

→ irrelevant to j (i.e., X_2)

⇒ corresponds to a constant-effect model, i.e., $Y_{1j}\sim 1\equiv S$

- When $\pi_{ij}\neq\pi_{i+}\pi_{+j}$ (X_1 and X_2 not independent) → The 2 conditional dist. are different, i.e., $\pi_{i=1|j}$ changes with j (i.e., X_2)

⇒ $Y_{1j}\sim X_2\equiv L$ (saturated model) → $\hat\pi_{i=1|j}=\frac{Y_{1j}}{y_{+j}}$

➢ To test whether $\pi_{ij}=\pi_{i+}\pi_{+j}$ (H_0) ⇒ H_0: S vs. H_1: $L\backslash S$

→ The test results would be identical to that obtained by applying Poisson GLM (LNp 5-6~7) on the scheme 3 data (why? check ★ in LNp 5-8) But, the main effects of X_1 & X_2 in Poisson fit could be meaningless

- Deviance based: $D_S-D_L=D_S$ ← a goodness-of-fit test for S
  $D_L=0$ ∵ saturated
- Perason X^2 under S

➢ Can be generalized to X_2 with J (>2) levels in the same manner →

X_2	Y_x	n_x	P_x
1	Y_{11}	y_{+1}	$\pi_{1\|1}$
2	Y_{12}	y_{+2}	$\pi_{1\|2}$
⋮	⋮	⋮	⋮
J	Y_{1J}	y_{+J}	$\pi_{1\|J}$

Binomial GLM

future lecture

◉ For the case that X_1 has I (>2) levels

$\sum_i\pi_{i|j}=1$, $j=1,\dots,J$, dim(parameters) $=(I-1)J$

indep.

$(Y_{1j},\dots,Y_{Ij})\sim$ multinomial$(y_{+j},\ \pi_{i=1|j},\dots,\pi_{i=I|j})$, $j=1,\dots,J$ ---(□)

a column in table; X_1; X_2

⇒ called product multinomial model (cf., unrestricted multinomial model in scheme 2) → only one multinomial (LNp 5-8)

$\sum_{ij}\pi_{ij}=1$, dim(parameters) $=IJ-1$

- Scheme 4: ← LNp.5-4 SRS from the y_{++} subjects

  ➢ Model: if $\pi_{ij}=\pi_{i+}\pi_{+j}$ (H_0), for a random sample, test statistic → Y_{11} ~ hypergeometric(y_{1+}, y_{+1}, y_{+2}), i.e.,

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_{++}

LNp.2-27, sampling without replacement

$$P(Y_{11}=y_{11}) = \binom{y_{+1}}{y_{11}}\binom{y_{+2}}{y_{12}} \Big/ \binom{y_{++}}{y_{1+}} = \frac{y_{1+}!\,y_{2+}!\,y_{+1}!\,y_{+2}!}{y_{11}!\,y_{12}!\,y_{21}!\,y_{22}!\,y_{++}!}$$

\# of black balls drawn (Y_{11}); \# of black balls (y_{+1}); \# of white balls (y_{+2}); total \# of balls (y_{++}); \# of balls drawn (y_{1+})

✱ $\boxed{y_{11} \le \min\{y_{+1}, y_{1+}\}}$

null distribution

- Under Scheme 3 and H_0, the joint pmf of (Y_{11}, Y_{12}, Y_{21}, Y_{22}) is: (LNp.5-9; $\pi_{1|1}=\pi_{1|2}=\pi_{1+}$, $\pi_{2|1}=\pi_{2|2}=\pi_{2+}$)

$$\binom{y_{+1}}{y_{11}}\pi_{1|1}^{y_{11}}\pi_{2|1}^{y_{21}} \times \binom{y_{+2}}{y_{12}}\pi_{1|2}^{y_{12}}\pi_{2|2}^{y_{22}} \overset{\because\text{ indep}}{=} \frac{y_{+1}!\,y_{+2}!}{y_{11}!\,y_{21}!\,y_{12}!\,y_{22}!}\,\pi_{1+}^{\overbrace{y_{11}+y_{12}}^{y_{1+}=}}\,\pi_{2+}^{\overbrace{y_{21}+y_{22}}^{y_{2+}}}$$

- Under Scheme 3 and H_0, the sufficient statistics of π_{1+} (P(success)) and π_{2+} (P(failure)) are Y_{1+} ($\sim B(y_{++}, \pi_{1+})$) and Y_{2+} ($=y_{++}-Y_{1+}$), respectively, and their joint pmf is: (check (✱) in LNp.5-10)

$$\frac{y_{++}!}{y_{1+}!\,y_{2+}!}\,\pi_{1+}^{y_{1+}}\,\pi_{2+}^{y_{2+}}$$

$$P(Y_{11}, Y_{12}, Y_{21}, Y_{22} \mid y_{1+}, y_{2+}, y_{+1}, y_{+2}) \overset{\Delta}{=} \frac{P(Y_{11}, Y_{12}, Y_{21}, Y_{22}, \cancel{Y_{1+}}, \cancel{Y_{2+}} \mid y_{+1}, y_{+2})}{P(Y_{1+}, Y_{2+} \mid y_{+1}, y_{+2})}$$

  ➢ When $\pi_{ij}\neq\pi_{i+}\pi_{+j}$ (H_1), the probability a black ball is drawn is different from the probability a white ball is drawn

➢ To test whether $\pi_{ij}=\pi_{i+}\pi_{+j}$ (*Fisher's exact test*) ← ∵ exact null dist. ← check ✱ in LNp.5-11

- Because Y_{11} can only take a limited number of values, can compute the probability of all these outcomes under H_0 (use hypergeometric)

  ⇒ can compute the total probability (p-value) of all outcomes that are *more extreme* than the one observed (y_{11})

- **Q**: what outcomes are more extreme? Some options:
  - The outcomes with probability ≤ $P(Y_{11}=y_{11})$ (more extreme, less possible to appear under H_0)
  - Outcomes y_{11}' s.t. $|y_{11}'-E(Y_{11})| \ge |y_{11}-E(Y_{11})|$ ($E(Y_{11}) = y_{1+}\,y_{+1}/y_{++}$ under H_0)
  - Others (see Agresti, 2013, 3.5)

➢ Generalization to $I\times J$ table for testing H_0: $\pi_{ij}=\pi_{i+}\pi_{+j}$

⇒ use multiple hypergeometric as null distribution, whose probability mass function is:

$$\Big(\prod_i y_{i+}!\Big)\Big(\prod_j y_{+j}!\Big)\Big/\Big(y_{++}! \times \prod_{ij} y_{ij}!\Big)$$

\# of free Y_{ij}'s = $IJ-I-J+1$ $=(I-1)(J-1)$

exercise: similar derivation as △ in LNp.5-11, but changing binomial into multinomial (check (□) in LNp.5-10)

y_{i+}: \# of balls drawn at the ith time; y_{+j}: \# of jth-type balls in the urn

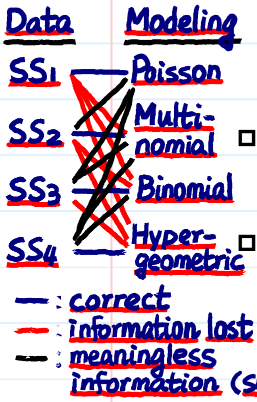
➢ Some notes:

- The situation that both marginal totals are fixed is rather less common in practical sampling applications

true \ answer	A	B	
A			y_{1+}
B			y_{2+}
	y_{+1}	y_{+2}	y_{++}

$y_{+1}=y_{1+}$, $y_{+2}=y_{2+}$ known

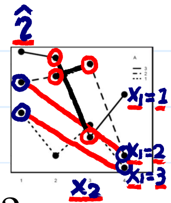
- It can arise when classifying objects into one of 2 types when the true proportions of each type are known → a close situation: know π_{1+}, π_{2+} & y_{+1}, y_{+2} are fixed (scheme 3)
  - Example: the lady tasting tea ← Table 3.9 in LNp.2-11
- It suggests a more accurate test for independence ← cf. "conditional" approaches in LNp.3-31 & LNp.3-45~46
  - Fisher's exact test is attractive because the null distribution for deviance-based and Pearson's X^2 test statistics is only *approximately* χ^2 distributed.
  - For tables with small counts, this χ^2 approximation is suspicious, which makes the exact method valuable. → I, J large, too many possible outcomes
  - Fisher's exact test becomes more difficult to compute for larger tables. However, the χ^2 approximation will tend to be accurate for larger tables. → usually, y_{++} large



❖ **Reading**: Faraway (2006, 1st ed.), 4.1, 4.2

Correspondence Analysis (CA)

cf. Recall. In DOE, if ANOVA (if difference exists?) rejected then multiple comparison (what different?)



- **Q**: when independence of a 2-way contingency table is rejected, how to know where the dependence is coming from?

★ — Interaction terms in a Poisson GLM contain dependence ⟺ association information; however, interpretation of them could be difficult.

intuitive

➢ CA: a visual residual analysis for contingency table ← Recall Pearson, Deviance residuals (LNp 3-34)

In a model corresp. to independent ($\pi_{ij}=\pi_{i+}\pi_{+j}$), its residuals contain the dependence information
- Poisson main effect model
- Multinomial main effect model
- Binomial intercept model

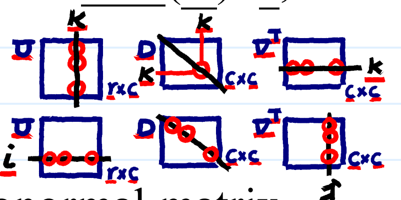
- Singular value decomposition ← cf. → eigen-decomposition for square matrix

➢ $R=[R_{ij}]$: an $r\times c$ matrix. W.l.o.g, assume $r\geq c$ and rank$(R)=c$, then

Fix k. $[U_{ik}V_{jk}] = U_k V_k^T$

$$R = U_{r\times c} D_{c\times c} V_{c\times c}^T = \Sigma_k\, d_k U_k V_k^T \text{, i.e.,} \quad (U_k V_k^T: r\times c \text{ matrix})$$

$$R_{ij} = \sum_{k=1}^{c} U_{ik}\, d_k\, V_{jk}\text{, where}$$



- $U=[U_{ij}]=[U_1, \ldots, U_c]$: an $r\times c$ column orthonormal matrix, i.e., $U^TU=I_{c\times c}$; its columns are called *left singular vectors*

(d_k, U_k, V_k): a triplet

- $V=[V_{ij}]=[V_1, \ldots, V_c]$: a $c\times c$ column orthonormal matrix, i.e., $V^TV=I_{c\times c}$; its columns called *right singular vectors*
- $D=\text{diag}(d_1, \ldots, d_c)$, $d_1 \geq \ldots \geq d_c > 0$, called *singular values*

➢ Some properties $RR^T=(UDV^T)(UDV^T)^T=UDV^TVDU^T=UD^2U^T$ ($V^TV=I$). Similarly, $R^TR=VD^2V^T$

Note $r\geq c$, the rest $r-c$ eigen-values are zeros

- Columns of $U_{r\times c}$ are eigenvectors of $(RR^T)_{r\times r}$
- Columns of $V_{c\times c}$ are eigenvectors of $(R^TR)_{c\times c}$
- $\{d_1^2, \ldots, d_c^2\}$ are eigenvalues of RR^T and R^TR

(LNp.3-34) $\dfrac{O_x - E_x}{\sqrt{E_x}} = \dfrac{y_{ij}-\hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}}}$

- Procedure of correspondence analysis on Pearson residuals