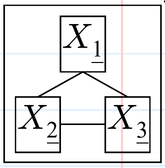
p. 5-25

➢ Uniform association (UA)

conditional independent ⇒ uniform association



▪ Consider a model with *all* two-factor interactions

$$Y \sim X_1 + X_2 + X_3 + X_1{:}X_2 + X_1{:}X_3 + X_2{:}X_3 \equiv S_4 \;(\supset S_3)$$

▪ S_4 has **no simple** interpretation **in terms** of independence

Also, fix X_1: $X_2 \xleftrightarrow{UA} X_3$; fix X_2: $X_1 \xleftrightarrow{UA} X_3$

◉ S_4 asserts that for every level of one variable, say X_3, we have the same association between X_1 and X_2

if all 3 2fi's are significant

i.e., $X_3=k$ is fixed

$$Y \sim X_1 + X_2 + X_3 + X_1{:}X_2 + X_1{:}X_3 + X_2{:}X_3$$

Recall. Interaction ↔ association (check ✱ in LNp 5-13)

might have different marginal dist.

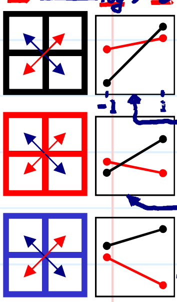
η	X_1	X_2	β_{X_1}	β_{X_2}	$\beta_{X_1X_2}$
η_{11}	1	1	-1	-1	+1
η_{12}	1	2	-1	+1	-1
η_{21}	2	1	+1	-1	-1
η_{22}	2	2	+1	+1	+1

$4\beta_{X_1X_2} = (\eta_{11}-\eta_{12}) - (\eta_{21}-\eta_{22})$

▫ For each levels of X_3, the *reduced models* of S_4 have *different* coefficients for the main effects of X_1 and X_2, but have the *same* coefficients for the interaction $X_1{:}X_2$

Under S_4, $\beta_{12k} = \beta_{12}$

$\pi \xrightarrow{\log} \eta$, $\pi \xleftarrow{\exp} \eta$



same difference in slopes

▫ e.g., $I=J=2$, same fitted odds-ratio between X_1 and X_2 for each category of X_3. Note that

← a measure of association

$$\hat{\Delta}_{12k} = \text{fitted odd-ratio} = \frac{\hat{y}_{11k}\hat{y}_{22k}}{\hat{y}_{12k}\hat{y}_{21k}} = \frac{\hat{\pi}_{11k}\hat{\pi}_{22k}}{\hat{\pi}_{12k}\hat{\pi}_{21k}} = e^{4\hat{\beta}_{12k}}$$

$\beta_{12k}=0 \Leftrightarrow \Delta_{12k} = e^{\beta_{12k}} = 1 \Leftrightarrow$ 2 parallel lines in the interaction plot of η

where β_{12k} is the coefficient of the $X_1{:}X_2$ term (under a coding $\propto \{+1, -1\}$) in the reduced model of $X_3=k$.

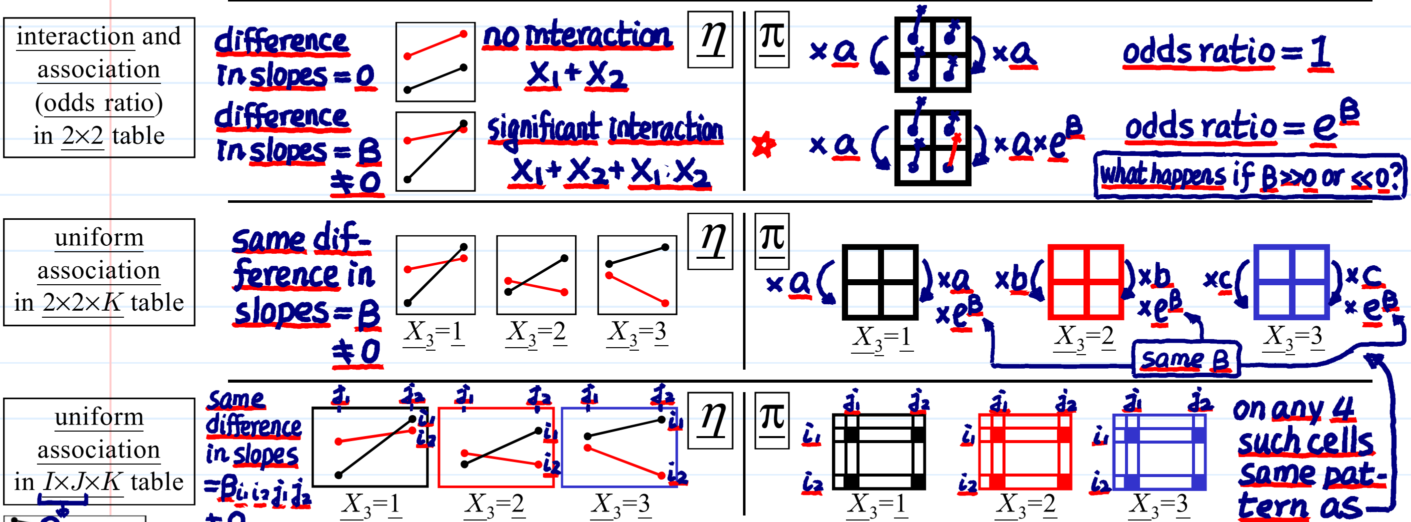
by ✱ in LNp.5-23

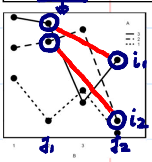
p. 5-26

▪ **Q**: What does uniform association mean? How to interpret the association? How does it connect with interaction terms?

(additive effects of $\underline{X}$) $\eta \underset{\exp}{\overset{\log}{\rightleftarrows}} \mu \propto \pi$ (multiplicative effects of $\underline{X}$)

		η	π	
interaction and association (odds ratio) in 2×2 table	difference in slopes = 0	no interaction X_1+X_2	$\times a$ ( ) $\times a$	odds ratio = 1
	difference in slopes = $\beta \neq 0$	significant interaction $X_1+X_2+X_1{:}X_2$	✱ $\times a$ ( ) $\times a \times e^{\beta}$	odds ratio = e^{β} what happens if $\beta \gg 0$ or $\ll 0$?
uniform association in 2×2×K table	same difference in slopes = $\beta \neq 0$	$X_3=1$, $X_3=2$, $X_3=3$	$\times a$ ( ) $\times a \times e^{\beta}$ ($X_3=1$); $\times b$ ( ) $\times b \times e^{\beta}$ ($X_3=2$); $\times c$ ( ) $\times c \times e^{\beta}$ ($X_3=3$) same β	
uniform association in $I\times J\times K$ table	same difference in slopes $=\beta_{i_1 i_2 j_1 j_2} \neq 0$	$X_3=1$, $X_3=2$, $X_3=3$	$X_3=1$, $X_3=2$, $X_3=3$	on any 4 such cells same pattern as



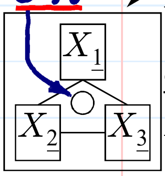


▪ S_4 is not saturated ⇒ some degrees of freedoms left for goodness-of-fit test

d.f. for 3-factor interactions (3fi's) = $(I-1)(J-1)(K-1)$

different odds ratios

3-fi



➢ A saturated model corresponds to a 3-way table with *different association* between, say X_1 and X_2, across K levels of X_3

IJK parameters

whereas $Y \sim 1$ corresponds to a 3-way table with constant π

intercept-only model

$\pi_{ijk} = 1/IJK$

for testing individual property

➢ **Q**: how to examine whether X_1, X_2, X_3 in a 3-way table are mutually independent (S_1), jointly independent (S_2), conditionally independent (S_3), or uniformly associated (S_4), indivudually?

- **Ans**: Perform deviance-based/Pearson's X^2 goodness-of-fit (GoF) tests for S_1, S_2, S_3, S_4 (as H_0), respectively.

use Poisson GLM

might find several S's are good enough

- However, be careful of zero or small y_{ijk} (rule of thumb: 20% of cells less than 5) in the table ⇒ there will be some doubt about the accuracy of chi-square approximation in GoF test

e.g. $H_0: S_1, H_1: S_2 \setminus S_1$; $H_0: S_2, H_1: S_3 \setminus S_2$; $H_0: S_3, H_1: S_4 \setminus S_3$

- The chi-square approximation is better in comparing models than assessing GoF

S_1, S_2, S_3, S_4 ↔ saturated

Then, report the condition the reduced model corresponding to

➢ Analysis strategy: start with complex Poisson GLM (e.g., saturated one) and see how far the model can be reduced (e.g., using model selection or sequential deviance-based tests as in ANOVA to compare models).

$1 \subset S_1 \subset S_2$ "$\subset$" $S_3 \subset S_4$ $\subset$ saturated

SS3 (LNp. 5-9)

a multinomial with y_{+++} trials

their sum y_{ij+} is fixed or conditioned

X_1 (i), X_2 (j), X_3 (k), $\pi_{k|ij}$

indep

a multinomial with y_{IJ+} trials

cf.

- Binomial (or multinomial) GLM approach for 3-way table

① fixed originally ② random but conditioned

What if only y_{++k}'s are fixed? (cf., SS2 in LNp. 5-8)

➢ If y_{ij+}'s regarded as fixed, can treat Y_{X_3} as response and X_1, X_2 as covariates

response, covariates

	X_1	X_2	X_3	1	2	3
1st subject	1	3	2	0	1	0
2nd subject	2	5	3	0	0	1

count data

➢ **Q**$_1$: what information been gone? **Q**$_2$: what still attainable?

∵ y_{ij+}'s do not carry information about π's (i.e., π_{ij+}'s)

- **Ans** for **Q**$_1$: information about π_{ij+}

Note. $\pi_{ijk} = \pi_{k|ij} \times \pi_{ij+}$

- **Ans** for **Q**$_2$: information about $\pi_{k|ij}$

➢ Statistical Modeling

interested in how $\pi_{k|ij}$, $\forall k$, changes with $X_1(i)$ & $X_2(j)$

$Y_{ij2} = y_{ij+} - Y_{ij1}$

- $Y_{X_3} = Y_{ij1} \sim$ binomial(y_{ij+}, $\pi_{k=1|ij}$) when K=2

$Y_{ij1} \overset{\text{indep}}{\sim} B(y_{ij+}, \pi_{k=1|ij})$; link; function of X_1, X_2 (covariates); $X\beta = \eta_{ij}$

future lecture

- $Y_{X_3} = (Y_{ij1}, \ldots, Y_{ijK})$ $\sim$ multinomial(y_{ij+}, $\pi_{k=1|ij}$, …, $\pi_{k=K|ij}$) when $K > 2$

Note. $X_3(k)$ appears in response 隨, not in η 規 (cf., SS3, (*) in LNp. 5-9)

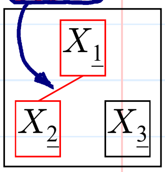
$\sum_k \pi_{k|ij} = 1$, but $\sum_{i,j} \pi_{k|ij} \neq 1$ in general

use logit link

no information

➢ **Q**: how is a binomial GLM connected to a Poisson GLM in 3-way tables?

covariates X_1, X_2 | covariates X_1, X_2, X_3



- $Y_{X_3} \sim 1 \Leftrightarrow S_2$ (joint independence)

LNp. 5-23~24

irrelevant to $X_1(i)$ & $X_2(j)$, e.g., when $k=1$, π_{++1} is a constant over i, j

binomial GLM		Poisson GLM
1	⟷	$1{:}X_3 = X_3$
X_1	⟷	$X_1{:}X_3$
X_2	⟷	$X_2{:}X_3$
$X_1{\cdot}X_2$	⟷	$X_1{\cdot}X_2{:}X_3$

$\pi_{k|ij} = \dfrac{\pi_{ijk}}{\pi_{ij+}}$

- $\pi_{ijk} = \pi_{ij+} \times \pi_{++k} \Leftrightarrow \pi_{k|ij} = \pi_{++k}$

cf.

corresponds to the main effect X_3 in Poisson GLM

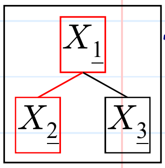
- The binomial GLM implicitly assumes an association between X_1 and X_2 (**Q**: why?)

∵ y_{ij+}'s are fixed ⇒ data contain no information about π_{ij}'s ⇒ true π_{ij+}'s can be any forms ⇒ corresponds to a saturated model of X_1 & X_2, i.e., $1 + X_1 + X_2 + X_1{:}X_2$

unable to use this model to exam whether X_1, X_2, X_3 are mutually indep. (S_1, LNp. 5-22~23)

- Poisson GLM allows us to drop the $X_1{:}X_2$ term, but binomial GLM does not

no information to exam whether X_1 & X_2 are independent or associated

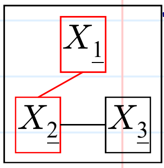


- $Y_{X_3} \sim 1 + X_1 \Leftrightarrow X_2, X_3$ are independent given X_1 ← conditional independence (LNp.5-24)

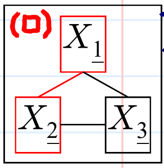
  check ✱ in LNp.5-28

  □ $\pi_{ijk} = \pi_{ij+}\pi_{i+k}/\pi_{i++} \Leftrightarrow \pi_{k|ij} = \pi_{k|i+}$ ← $\frac{\pi_{ijk}}{\pi_{ij+}} = \frac{\pi_{i+k}}{\pi_{i++}}$ ← a function of X_1 only

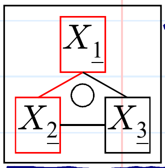
  need $X_1 \cdot X_3$ in Poisson GLM: $\pi_{i+k} \neq \pi_{i++} \times \pi_{++k}$ (not marginally indep) $\Leftrightarrow \pi_{k|i+} \neq \pi_{++k}$ → need X_1 in binomial GLM



- **Q**: how about $Y_{X_3} \sim 1 + X_2$? → X_1, X_3 indep given X_2
- **Q**: Can we exam whether X_1, X_2 are independent given X_3? → No check (□) No information to study $X_1 \cdot X_2$ in Poisson GLM

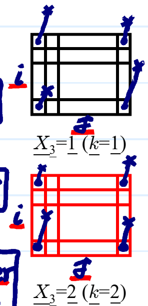


- $Y_{X_3} \sim 1 + X_1 + X_2 \Leftrightarrow S_4$ (uniform association) ← check ✱ in LNp. 5-28; LNp.5-25



- The saturated binomial GLM, $Y_{X_3} \sim 1 + X_1 + X_2 + X_1{:}X_2$, corresponds to a Poisson GLM for *different association* ← LNp 5-26
- Using binomial GLM loses little when we are interested in the relationship between the response X_3 and the two covariates X_1, X_2, and not interested in the association between X_1 and X_2

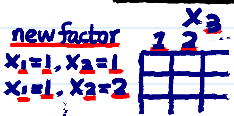
  Note. cannot study $X_1 \cdot X_2$ — e.g. y_{ij+}'s are fixed in advance — & the other 2 as covariates



- **Q**: Poisson or binomial GLM approach? Which to use?
  - ➢ Binomial if one variable is clearly identified as the response ← as in DOE; like in SSI (LNp 5-4)
  - ➢ Poisson if relationship between 3 variables is more symmetric

---

- Correspondence analysis ← LNp5-13~17, CA identifies association (i.e., interaction in Poisson GLM) between variables
  - ➢ Cannot directly apply to 3-way table ← ∵ CA is based on singular value decomposition, which applied on matrix ← 2-way
  - ➢ Can combine two of the factors, say X_1 and X_2, into a factor with $I\times J$ levels and apply correspondence analysis on the 2-way table formed by the new factor and X_3 (interested in $\pi_{k|ij}$'s in binomial GLM)



  - ➢ **Q**: which two factors should be chosen to merge?

    Ans: pick up the two whose association is least interesting to us ← ∵ after merging X_1 & X_2, becomes difficult to study $X_1 \cdot X_2$ in the main effects of the new factor (cf.)

- Simpson's paradox 5/12
  - ➢ example:

- X_1 (i): age
- X_2 (j): smoker
- X_3 (k): dead or alive

3-way table; there exists a significant interaction (association) btwn X_1 & X_2 ✱

$X_1(i)$	smoker $X_2(j)$	dead ($X_3(k)$=1)	alive (2)	y_{ij+}	$\hat{\pi}_{j\mid i+}$		smoker	dead	alive ($\hat{\pi}_{k\mid ij}$)	
age=35-44 (1, younger)	yes 1	14	95	✓ 109	(.47)	⇒	yes	.13	.87	1
	no 2	7	114	○ 121	(.53)		no	.06	.94 ≥	1

different base rates (cf.)

$X_1(i)$	smoker	dead	alive				smoker	dead	alive	
age=65-74 (2, older)	yes 1	29	7	✓ 36	(.22)	⇒	yes	.81	.19	1
	no 2	101	28	○ 129	(.78)		no	.78	.22 ≥	1

higher proportion of young people in smoker ($\hat{\pi}_{i|j+}$):
- 109/145 =.75
- 121/250 =.48

← $\hat{\pi}_{i|j+}$

2-way table

marginal total over age	smoker	dead	alive (y_{+jk})	y_{+j+}	$\hat{\pi}_{+j+}$		smoker	dead	alive ($\hat{\pi}_{k\mid +j}$)	
	yes	43	102	✓ 145	(.37)	⇒	yes	.30	.70	1
	no	108	142	○ 250	(.63)		no	.43	.57 ≤	1