- MH statistic combine information of y_{11k}'s from K tables:

y_{11+} ← Does it actually use only marginal 2×2 table? Ans. No. Check

$$\frac{\left(\left|\sum_k [y_{11k} - E(y_{11k})]\right| - 1/2\right)^2}{\sum_k Var(y_{11k})} \overset{a}{\sim} \chi_1^2$$

continuity correction (increasing p-value, more conservative, check LNp.5-7)

← dim(H_1) − dim(H_0)

∵ y_{11k}'s are independent.

If some $\Delta_k \gg 1$ & some $\Delta_k \ll 1$ then some $y_{11k} \gg E(y_{11k})$ & some $y_{11k} \ll E(y_{11k})$ they might cancel out & MH stat. becomes small (not rejected)

where $E(y_{11k})$ and $Var(y_{11k})$ are calculated under the H_0 → check (*) in LNp. 5-32

- can calculate an exact p-value for smaller dataset using hypergeometric distribution ← # of possible values of y_{11k}'s are not large (check LNp 5-12)

⇒ useful when data is sparse, under which the χ^2 approximations based on asymptotic thm is questionable ← many small y_{ijk}'s

➢ MH test is sometimes called Cochran-Mantel-Haenszel test because a version without the 1/2 is published earlier by Cochran (1954).

❖ **Reading**: Faraway (2006, 1st ed.), 4.4

5/14

Ordinal Variables

(LNp. 2-3 ~4) ① continuous (interval) ② discrete interval ③ ordinal ④ nominal

ⓐ In analysis, cannot arbitrarily rearrange the order of categories
ⓑ the order matters for the objective of analysis

- Some variables have a nature ordering between categories

Otherwise, we can find order(s) in many (almost any) variables, but some not matter

distance = ?

➢ e.g., education: HS, BA, MA; political ideology: VL, SL, M, SC, VC

cf. • ordinal response 隨 ← LNp.5-28
• ordinal covariate 規

➢ The ordinal structure not matter when # of categories = 2

no difference in the treatments for ordinal & nominal

(□) ➢ For ordinal variables, can use the methods for nominal variable

- But, more information can be extracted by taking advantage of the ordinal structure

some dummy var., e.g., Helmert, forward difference, ..., more meaningful for ordinal predictor

check ✱ in LNp. 2-4

➢ Treatments for ordinal response (future lecture) and ordinal covariates are different

could be wrong → assumption added; information lost

cont ← ordinal → nominal → dummy var.; linear, quadratic, ...

- Treatment for ordinal predictors: assign each category a *score* ← cf. → CA in LNp. 5-15

like adding distance btwn categories

➢ It kind of turns an ordinal variable into a continuous variable

➢ The choice of scores requires some judgment

• known values
(Δ) ⊙ should increase with order
• assignment of scores can be quite flexible
(*) ⊙ scores ← cf. → dummy var.

order: 1 2 ⋯ I → treat them as quantitative

⊙ If no particular preference, even spacing allows for the simplest interpretation

- For interval scales, midpoints of the intervals are often used

usually defined for continuous variables

➢ Should check whether the inference is robust to different assignments of scores

e.g., sign of β

- If *qualitative* conclusions are changed, this is an indication that you cannot make any strong finding based on scores

might ignore order & treat ordinal as nominal

- Poisson GLM with linear-by-linear association for 2-way tables:

➢ Consider table with ordinal row (X_1) and column (X_2) variables known values

- assign scores $u_1 \le u_2 \le \ldots \le u_I$ to rows, denoted by $u(X_1)$ i
- assign scores $v_1 \le v_2 \le \ldots \le v_J$ to columns, denoted by $v(X_2)$ j

check LNp.5-5

➢ Linear-by-linear association model: cf. interaction in Poisson GLM

$$\eta_{ij} \doteq \log(\mu_{ij}) \doteq \log(t\,\pi_{ij}) = \log(t) + \log(\pi_{i+}) + \log(\pi_{+j})$$

X_1 X_2

$$+\gamma \times u_i \times v_j$$

$= u_i(\gamma v_j)$, $\gamma_j = $ ★

only 1 parameter interaction

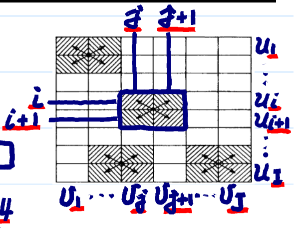
use codings for nominal X_1: I−1 d.f. X_2: J−1 d.f.

where u_i's, v_j's are known scores, and γ is an unknown parameter

- $Y \sim X_1 + X_2 + u(X_1)v(X_2) \equiv S_{O\times O}$

➢ Some notes about γ: only 1 d.f.

larger |γ|, stronger association (under same u_i's, v_j's)



$S_{O\times O}$ reduced to $Y \sim X_1 + X_2$ main-effect model ⟺ independent

- values of γ represents the amount of association
- γ=0 ⟺ independence
- positive and negative γ

$\gamma > 0$, $X_1 \uparrow, X_2 \uparrow$; $X_1 \downarrow, X_2 \downarrow$

$\gamma < 0$, $X_1 \uparrow, X_2 \downarrow$; $X_1 \downarrow, X_2 \uparrow$

check (*) in LNp.5-34

$$\log\left(\frac{\pi_{i,j}\ \pi_{i+k,j+l}}{\pi_{i,j+l}\ \pi_{i+k,j}}\right) = \gamma\,(u_{i+k} - u_i) \times (v_{j+l} - v_j)$$ (exercise)

e.g., $u_{i+1} - u_i = v_{j+1} - v_j = 1$ or $u_{i+k} - u_i = k$, $v_{j+l} - v_j = l$

- Interpretation of γ by log-odds-ratio:

$$\log\left(\frac{\pi_{i,j}\ \pi_{i+1,j+1}}{\pi_{i,j+1}\ \pi_{i+1,j}}\right) = \log\left(\frac{\mu_{i,j}\ \mu_{i+1,j+1}}{\mu_{i,j+1}\ \mu_{i+1,j}}\right) = \gamma u_i v_j + \gamma u_{i+1} v_{j+1} - \gamma u_i v_{j+1} - \gamma u_{i+1} v_j$$

scores affect the interpretation of γ

$$= (\eta_{i,j} + \eta_{i+1,j+1}) - (\eta_{i,j+1} + \eta_{i+1,j}) \doteq \gamma(u_{i+1} - u_i)(v_{j+1} - v_j)$$

uniform association for 3-way table (LNp.5-25~26)

- for evenly spaced scores, these log-odds-ratios are equal ← fixed k, l, any i, j
- cf. ⇒ called uniform association in Goodman (1979)

- Latent (continuous) variable Z motivation for γ:

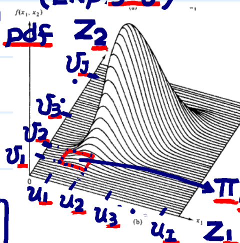
SS1. Poisson (LNp.5-4~5) or SS2. multinomial (LNp.5-8)

Z can not be directly observed, but, a function of Z, X=f(Z), can be observed to gain information of Z

- Assume π_{ij}'s are obtained by putting a grid on an approximately bi-variate Normal (Z_1, Z_2) for latent variables and u_i's and v_j's are cutpoints

This explains under what conditions a model like $S_{O\times O}$ is appropriate for a 2-way table

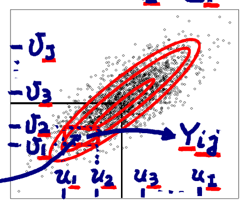
$N\left(\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$



- γ can then be identified with the correlation coefficient ρ of the latent variables (cf., positive and negative ρ)

$\gamma \approx \rho/(1-\rho^2)$ if u_i's & v_j's are standardized (Agresti, 2013, 10.4.1)

γ	ρ
positive ↔	positive
negative ↔	negative



$X_1 = i$, $X_2 = j$ ⟺ $Z_1 \in (u_{i-1}, u_i)$ & $Z_2 \in (v_{j-1}, v_j)$

➢ **Q**: for the tests of independence or goodness-of-fit, what is the benefit of using $S_{O\times O}$ over the nominal approach, i.e., fitting a nominal-by-nominal model $S_{N\times N}$: $Y \sim X_1 + X_2 + X_1{:}X_2$?

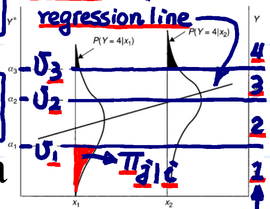
Note. indep $\subset S_{O\times O} \subset S_{N\times N}$

saturated model

X_1: 7 levels, X_2: 7 levels

As shown in a lab example, $(I-1)(J-1) = 6\times 6 = 36$ d.f.

SS3. product multinomial (LNp.5-10)



$Z_2 | Z_1 = z_1$ ~ simple regression model

ordinal

H0: indep, H1: $S_{N\times N}$, not significant

- in the $N\times N$ approach, interaction effects reduce a deviance of 40.743 on 36 degrees of freedom, but

cf. H0: indep, H1: $S_{O\times O}$, significant

- the $O\times O$ interaction effect reduces a deviance of 10.175 on *one* degrees of freedom, i.e., the other 35 interaction effects only reduce a deviance of 30.568

deviance & its d.f. of $S_{O\times O}$

◉ Ordinal-by-nominal model (or nominal-by-ordinal model) → for interactions in Poisson GLM association

row — column (check (□) in LNp.5-34) — row — column

Note. row & column are ordinal variables

➢ Rows (or columns) assigned scores, but column known (or row) variable treated as a nominal variable — use dummy variables

➢ called *column* (or *row*) *effects model* because the columns (or rows) are not assigned scores; instead, their effects are estimated

Recall. coding for an interaction: componentwise product of a coding of X_1 & a coding of X_2

▪ alternative viewpoint: the scores of the ordinal columns (or rows) regarded as parameters (unknown) — i.e., coefficients of dummy var

Under Poisson GLM. Check LNp.5-5

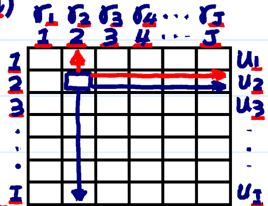
➢ Column effects model:

Under ✱ in LNp 5-35 (△) $\gamma_1:\gamma_2:\cdots:\gamma_J$ — proportion unknown (estimated by data) in $O\times N$; $=v_1:v_2:\cdots:v_J$ — proportion known in $O\times O$

$$\eta_{ij} \doteq \log(\mu_{ij}) \doteq \log(t\,\pi_{ij})$$

$$= \log(t) + \log(\pi_{i+}) + \log(\pi_{+j}) + u_i \times \gamma_j$$

X_1 only — X_2 only — interactions — X_1 as continuous (scores) — X_2 as nominal (coefficients of dummy var.)



$$\begin{array}{c|cccccc|c}
 & 1 & 2 & 3 & 4 & \cdots & J & \\
\gamma_1 & 1 & 0 & 0 & 0 & \cdots & 0 & \\
\gamma_2 & 0 & 1 & 0 & 0 & \cdots & 0 & \\
\gamma_3 & 0 & 0 & 1 & 0 & \cdots & 0 & \\
\gamma_4 & 0 & 0 & 0 & 1 & \cdots & 0 & \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \\
\gamma_J & 0 & 0 & 0 & 0 & \cdots & 1 &
\end{array}$$

dummy var. — reference level

where u_i's, i=1,…, I, are known scores, and

each column (jth) a parameter γ_j

γ_j's, j=1,…, J, are unknown parameters (over-parameterized; only requires J–1 parameters), (✱)

adding 1 constraint. e.g., $\gamma_1=0$ or $\sum_j \gamma_j=0$

use codings for nominal: X_1: I-1 d.f., X_2: J-1 d.f.

▪ $Y \sim X_1 + X_2 + u(X_1){:}X_2 \equiv S_{O\times N}$ ($\supseteq S_{O\times O}$)

(J-1) d.f.

∵ [v_j] in LNp.5-35 ∈ the space of [γ_j]

denoted by $S_{N\times O}$ for row effects model

➢ Some notes about γ_j's, called the *column effects*:

Under the constraint (✱) in LNp 5-37, it becomes $\gamma_1=\cdots=\gamma_J=0$ ⇒ no interaction model ⇔ independent

▪ Equality of the γ_j's (then, $u_i\times\gamma_j=u_i\times\gamma$) corresponds to the hypothesis of independence between X_1 and X_2

It becomes a main effect of X_1 & is merged into the X_1 term in $S_{O\times N}$ in LNp 5-37

◉ For ordinal column variable, if the model $S_{O\times O}$ were a good fit, we would expect the estimates of the γ_j's in $S_{O\times N}$ to be roughly proportional to v_j's (e.g., for evenly spaced v_j's, estimates of γ_j's should follow a linear trend)

check (△) in LNp 5-37

data determine — of order (by (△) in LNp.5-37)

Note. u_i's are fixed, i.e., under the assumption that u_i's are suitable scores for rows

◉ We can use the estimates of γ_j's in $S_{O\times N}$ (1) to examine whether the chosen scores for columns in $S_{O\times O}$ (i.e., v_j's) are appropriate, or (2) to possibly suggest better scores (see an example in lab)

Note. In $S_{O\times N}$, the estimated $\hat{\gamma}_j$'s might not increase or decrease with the order of X_2 (cf. (△) in LNp.5-34)

- Some advantages of using scores for ordinal variables

➢ helpful in reducing the complexity of models for categorical data with ordinal variables

identifying few but meaningful signif-icant interactions

cf. nominal

	d.f.
main effect,	X_1: I-1, X_2: J-1, X_3: K-1
2fi,	$X_1{\cdot}X_2$: $(I-1)(J-1)$
	$X_1{\cdot}X_3$: $(I-1)(K-1)$
	$X_2{\cdot}X_3$: $(J-1)(K-1)$
3fi,	$X_1{\cdot}X_2{\cdot}X_3$: $(I-1)(J-1)(K-1)$

➢ especially useful in higher dimensional table where a reduction in the # of parameters is particularly welcome

➢ can also sharpen our ability to detect associations

❖ **Reading**: Faraway (2006, 1st ed.), 4.5 — CA (LNp.5-13~17) cf. — interactions in Poisson GLM