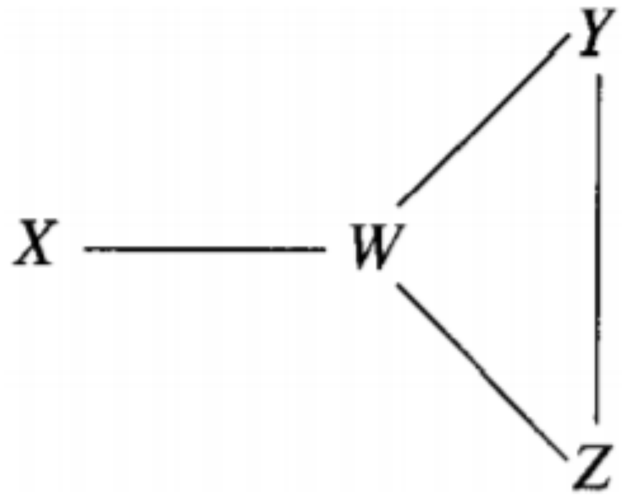
Ch10. Building and Extending Loglinear Models

Jaehyoung Hong

Conditional independence graphs

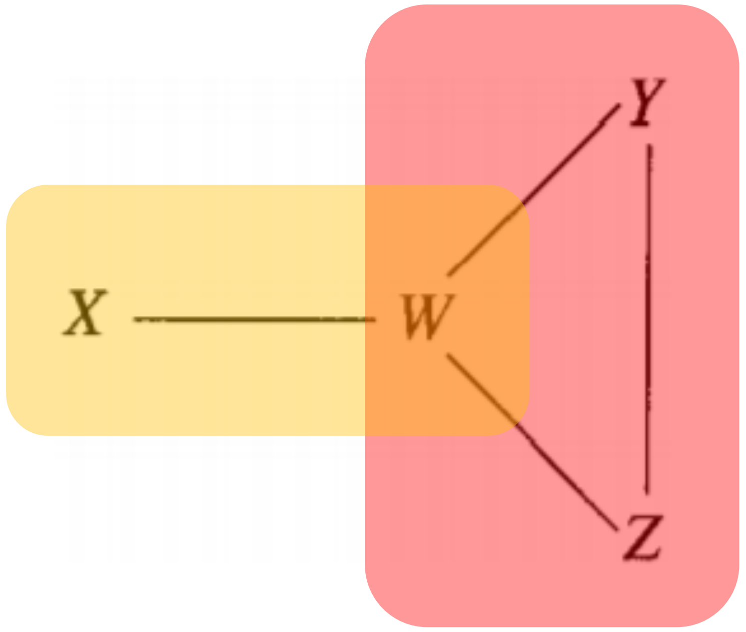


<Graphs of (WX,WY,WZ,YZ) and (WX, WYZ)>

- ✓ X and Y, X and Z are independent conditional on the remaining two variables
- ✓ Same pairwise associations have the same conditional independence graph
- ✓ X and Y are conditionally independent given W or given {W,Z}

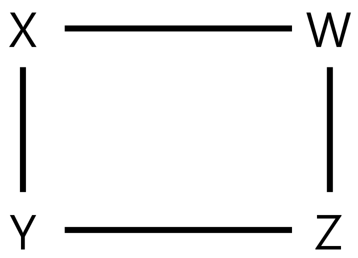
Graphical loglinear models

Clique : maximally connected subset



<Graphs of (WX,WY,WZ,YZ) and (WX, WYZ)>

- In graphical models, clique is sufficient statics
- Graphical models ⊇ Decomposable models
- Not all graphical models are decomposable



Example of non-decomposable graphical model

Collapsibility in three-way contingency tables

- Under the collapsibility condition, conditional associations usually differ from marginal associations

- Recall : XY marginal and conditional odds are identical
  if either Z and X or Z and Y (or both) are conditionally independent

➡ (XY,YZ) , (XY,XZ) and (XY,Z)

- **Model : (XY, XZ)**

$$\mu_{ij+} = \sum_k \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ})$$

$$= \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}) \sum_k \exp(\lambda_k^Z + \lambda_{ik}^{XZ})$$

Collapsibility in three-way contingency tables

- Recall : XY marginal and conditional odds are identical
  if either Z and X or Z and Y (or both) are conditionally independent

TABLE 8.5 Estimated Odds Ratios for Loglinear Models in Table 8.5

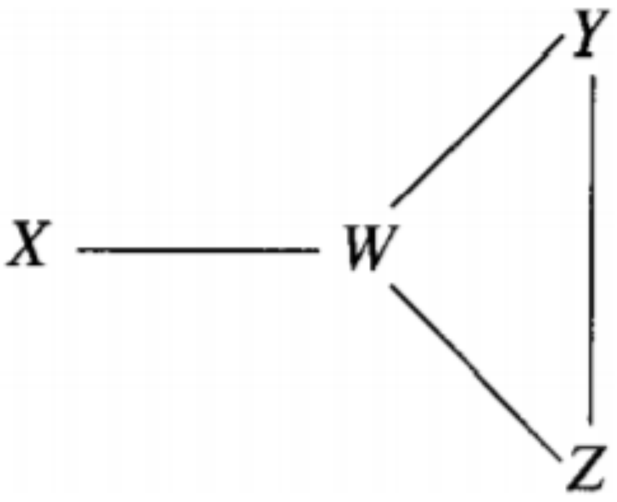
	Conditional Association			Marginal Association		
Model	AC	AM	CM	AC	AM	CM
(A, C, M)	1.0	1.0	1.0	1.0	1.0	1.0
(AC, M)	17.7	1.0	1.0	17.7	1.0	1.0
(AM, CM)	1.0	61.9	25.1	2.7	61.9	25.1
(AC, AM, CM)	7.8	19.8	17.3	17.7	61.9	25.1
(ACM) level 1	13.8	24.3	17.5	17.7	61.9	25.1
(ACM) level 2	7.7	13.5	9.7			

$$A \text{ ——— } M \text{ ——— } C.$$

Collapsibility in three-way contingency tables

- Bishop et al. (1975)

Suppose that a model for a multiway table partitions variables into three mutually exclusive subsets, A, B, C, such that B separates A and C. After collapsing the table over the variables in C, parameters relating variables in A and parameters relating variables in A to variables in B are unchanged.



- $A = \{X\}, B = \{W\}, C = \{Y, Z\}$, collapsing over C does change A and B association

✓ If separating B contains more than one variable, ML estimates of parameters may differ slightly when collapsing over C

Considerations in Model selection

✓ The model should contain the most general interaction term relating the explanatory variables

➡ $\{\hat{\mu}_{g+l+} = n_{g+l+}\}$ when G and L is explanatory variables

- **Gender / Race is explanatory → Include GR**

TABLE 9.1 Alcohol, Cigarette, and Marijuana Use for High School Seniors

		Marijuana Use							
		Race = White				Race = Other			
		Female		Male		Female		Male	
Alcohol Use	Cigarette Use	Yes	No	Yes	No	Yes	No	Yes	No
Yes	Yes	405	268	453	228	23	23	30	19
	No	13	218	28	201	2	19	1	18
No	Yes	1	17	1	17	0	1	1	8
	No	1	117	1	133	0	12	0	17

Considerations in Model selection

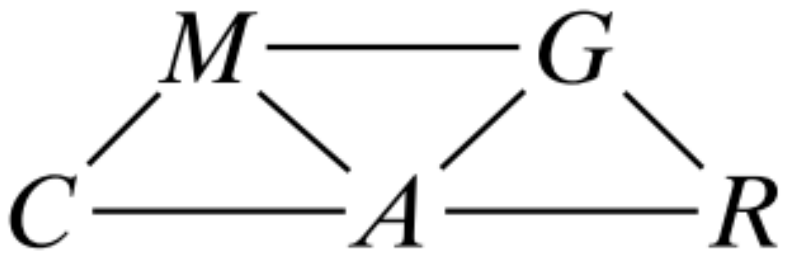
Backward elimination

TABLE 9.2 Goodness-of-Fit Tests for Loglinear Models for Table 9.1

Model[a]	G^2	df
1. Mutual independence + GR	1325.1	25
2. Homogeneous association	15.3	16
3. All three-factor terms	5.3	6
4a. (2)−AC	201.2	17
4b. (2)−AM	107.0	17
4c. (2)−CM	513.5	17
4d. (2)−AG	18.7	17
4e. (2)−AR	20.3	17
4f. (2)−CG	16.3	17
4g. (2)−CR	15.8	17
4h. (2)−GM	25.2	17
4i. (2)−MR	18.9	17
5. (AC, AM, CM, AG, AR, GM, GR, MR)	16.7	18
6. (AC, AM, CM, AG, AR, GM, GR)	19.9	19
7. (AC, AM, CM, AG, AR, GR)	28.8	20

[a] G, gender; R, race; A, alcohol use; C, cigarette use; M, marijuana use.

Considerations in Model selection



- ✓ C is conditionally independent with {G,R}
- ✓ Collapsing over the {G,R}, the conditional association between C and A and between C and M are the same as with model (AC,AM,CM)
- ✓ When removing MG (model 7), collapsing over G and R then, all pairwise conditional associations among A, C and M in model 7 are identical to (AC,AM,CM)

Loglinear model comparison statistics

- **Likelihood-ratio statistic (M_0 is special case of M_1 : M_0 is simple model)**

$$G^2(M_0|M_1) = 2\sum_i n_i \log(\frac{\hat{\mu}_{1i}}{\hat{\mu}_{0i}})$$

➕

$$\log \boldsymbol{\mu_0} = \boldsymbol{X_0\beta_0} = \boldsymbol{X_1\beta_1^*} \text{ and } \log \boldsymbol{\mu_1} = \boldsymbol{X_1\beta_1}$$

$$G^2(M_0|M_1) = 2\boldsymbol{n^T}(\log\boldsymbol{\hat{\mu}_1} - \log\boldsymbol{\hat{\mu}_0}) = 2\boldsymbol{n^T}\left[\boldsymbol{X_1\hat{\beta}_1} - \boldsymbol{X_1\hat{\beta}_1^*}\right]$$

$$= 2\boldsymbol{\hat{\mu}_1^T}\left[\boldsymbol{X_1\hat{\beta}_1} - \boldsymbol{X_1\hat{\beta}_1^*}\right] \text{ (Likelihood equation } \boldsymbol{n^T X_1} = \boldsymbol{\hat{\mu}_1^T X_1})$$

$$= 2\boldsymbol{\hat{\mu}_1^T}(\log\boldsymbol{\hat{\mu}_1} - \log\boldsymbol{\hat{\mu}_0}) = 2\sum_i \boldsymbol{\hat{\mu}_{1i}} \log(\boldsymbol{\hat{\mu}_{1i}}/\boldsymbol{\hat{\mu}_{0i}})$$

Partitioning Chi-squared with model comparison

- **Partitioning**

$$\begin{aligned} G^2(M_J) &= G^2(M_J|M_{J-1}) + G^2(M_{J-1}) \\ &= G^2(M_J|M_{J-1}) + G^2(M_{J-1}|M_{J-2}) + G^2(M_{J-2}) \\ &= \cdots = G^2(M_J|M_{J-1}) + \cdots + G^2(M_3|M_2) + G^2(M_2). \end{aligned}$$

- **Adjusted significance for multiple comparison**

Use $1-(1-\alpha)^{1/s}$ for s independent tests $\rightarrow P(type\ I\ error) \approx \alpha$

e.g.) 3 comparison for $\alpha = 0.05$, use $1-(0.95)^{\frac{1}{3}} = 0.01695$ for each

Modeling ordinal associations

TABLE 9.3 Opinions about Premarital Sex and Availability of Teenage Birth Control

	Teenage Birth Control[a]			
Premarital Sex	Strongly Disagree	Disagree	Agree	Strongly Agree
Always wrong	81 (42.4)[1] 7.6[2] (80.9)[3]	68 (51.2) 3.1 (67.6)	60 (86.4) −4.1 (69.4)	38 (67.0) −4.8 (29.1)
Almost always wrong	24 (16.0) 2.3 (20.8)	26 (19.3) 1.8 (23.1)	29 (32.5) −0.8 (31.5)	14 (25.2) −2.8 (17.6)
Wrong only sometimes	18 (30.1) −2.7 (24.4)	41 (36.3) 1.0 (36.1)	74 (61.2) 2.2 (65.7)	42 (47.4) −1.0 (48.8)
Not wrong at all	36 (70.6) −6.1 (33.0)	57 (85.2) −4.6 (65.1)	161 (143.8) 2.4 (157.4)	157 (111.4) 6.8 (155.5)

[a1] Independence model fit; [2] standardized Pearson residuals for the independence model fit; [3] linear-by-linear association model fit.

- Positive trend

Linear-by-Linear association model for two-way tables

- **Model with row scores $\boldsymbol{u_1 \le \cdots \le u_I}$ and column scores $\boldsymbol{v_1 \le \cdots \le v_J}$**

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j$$

✓ $\lambda_{ij}^{XY} = \beta u_i v_j$ but uses only one parameter to describe association

✓ $\beta > 0$ Y tends to increase as X increase

- **Rows a and c with columns b and d**

$$\log \frac{\mu_{ab}\mu_{cd}}{\mu_{ad}\mu_{cb}} = \beta(u_c - u_a)(v_d - v_b)$$

Linear-by-Linear association model for two-way tables

- **Standardizing row scores $\boldsymbol{u_1} \le \cdots \le \boldsymbol{u_I}$ and column scores $\boldsymbol{\nu_1} \le \cdots \le \boldsymbol{\nu_J}$**

$$\sum u_i \pi_{i+} = \sum v_j \pi_{+j} = 0$$
$$\sum u_i^2 \pi_{i+} = \sum v_j^2 \pi_{+j} = 1.$$

✓ $L \times L$ model tends to fit well when underlying distribution is approximately bivariate normal

✓ For standardized scores, β is comparable to $\rho/(1-\rho^2)$ where ρ is the underlying correlation

✓ For weak associations, $\beta \approx \rho$

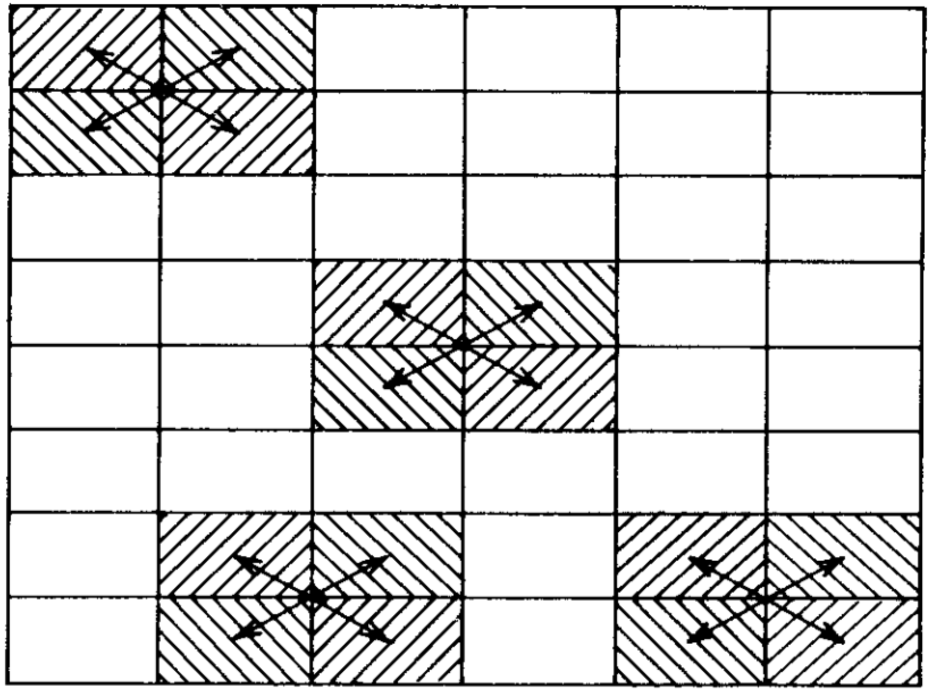
Corresponding logistic model for adjacent responses

- **Let $\boldsymbol{\pi_{j|i} = P(Y = j|X = i)}$**

$$\log \frac{\pi_{j+1|i}}{\pi_{j|i}} = \log \frac{\mu_{i,j+1}}{\mu_{ij}} = \left(\lambda^Y_{j+1} - \lambda^Y_j\right) + \beta(v_{j+1} - v_j)u_i.$$

- **Unit-spaced $\{\boldsymbol{\nu_j}\}$**

$$\log \frac{\pi_{j+1|i}}{\pi_{j|i}} = \alpha_j + \beta u_i$$



✓ Same linear logit effect β applies for all pairs of a

Likelihood equations and model fitting

- **Poisson log likelihood $L(\mu) = \sum_i \sum_j n_{ij} \log \mu_{ij} - \sum_i \sum_j \mu_{ij}$ for $L \times L$ model**

$$L(\boldsymbol{\mu}) = n\lambda + \sum_i n_{i+}\lambda_i^X + \sum_j n_{+j}\lambda_j^Y + \beta \sum_i \sum_j u_i v_j n_{ij}$$

$$- \sum_i \sum_j \exp\left(\lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j\right).$$

- **Differentiating $L(\mu)$ w.r.t ($\lambda_i^X, \lambda_j^Y, \beta$)**

$$\hat{\mu}_{i+} = n_{i+}, \; i = 1, \ldots, I, \qquad \hat{\mu}_{+j} = n_{+j}, \; j = 1, \ldots, J,$$

$$\sum_i \sum_j u_i v_j \hat{\mu}_{ij} = \sum_i \sum_j u_i v_j n_{ij}.$$

Likelihood equations and model fitting

TABLE 9.4 Output for Fitting Linear-by-Linear Association Model to Table 9.3

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value
Deviance	8	11.5337
Pearson Chi-Square	8	11.5085

Parameter		Estimate	Standard Error	Wald 95% Conf. Limits		Chi-Square	Pr > ChiSq
Intercept		0.4735	0.4339	−0.3769	1.3239	1.19	0.2751
premar	1	1.7537	0.2343	1.2944	2.2129	56.01	<.0001
premar	2	0.1077	0.1988	−0.2820	0.4974	0.29	0.5880
premar	3	−0.0163	0.1264	−0.2641	0.2314	0.02	0.8972
premar	4	0.0000	0.0000	0.0000	0.0000	.	.
birth	1	1.8797	0.2491	1.3914	2.3679	56.94	<.0001
birth	2	1.4156	0.1996	1.0243	1.8068	50.29	<.0001
birth	3	1.1551	0.1291	0.9021	1.4082	80.07	<.0001
birth	4	0.0000	0.0000	0.0000	0.0000	.	.
linlin		0.2858	0.0282	0.2305	0.3412	102.46	<.0001

LR Statistics

Source	DF	Chi-Square	Pr > ChiSq
linlin	1	116.12	>.0001

- Positive trend
- Estimated local odds ratio $= \exp(\hat{\beta}) = \exp(0.286) = 1.33$
- $\exp[\hat{\beta}(u_4 - u_1)(v_4 - v_1)]$
  $= \exp[0.286(4-1)(4-1)] = 13.1$
- Standardized scores give $\hat{\beta} = 0.374$
- $\hat{\beta} = \hat{\rho}/(1-\hat{\rho}^2)$ yields $\hat{\rho} = 0.333$

Row effects and column effects association models

- **Scores are parameters rather than fixed : ordered row values $\{\boldsymbol{\beta u_i}\} \rightarrow$ unordered parameters $\{\boldsymbol{\mu_i}\}$**

$$\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \mu_i \upsilon_j .$$

✓ Treat X : nominal and Y : ordinal

- **Adjacent logit for equal-column score (row-effect model)**

$$\log \frac{P(Y = j + 1 | X = i)}{P(Y = j | X = i)} = \alpha_j + \mu_i .$$

✓ Row effects model shows Table 10.3 have large conditional distributions on the column variable between row 2 and 3 than between row 1 and 2

Generalized loglinear model

- **Generalized loglinear model**

$$\boldsymbol{C}\log(\boldsymbol{A}\boldsymbol{\mu}) = \boldsymbol{X}\boldsymbol{\beta}$$

Multiplicative row and column effects model

- **Row and column effect model (RC model) : Row / Column scores are parameter**

$$\log\ \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta\mu_i\nu_j.$$

✓ Not linear

✓ Both row and column treated as nominal

$$\log\theta_{ij} = \beta(\mu_{i+1} - \mu_i)(\nu_{j+1} - \nu_j).$$

✓ Likelihood may not be concave / may have local maxima

Multiplicative row and column effects model

TABLE 9.9 Cross-Classification of Mental Health Status and Socioeconomic Status

Parents' Socioeconomic Status	Mental Health Status: Well	Mild Symptom Formation	Moderate Symptom Formation	Impaired
A (high)	64	94	58	46
B	57	94	54	40
C	57	105	65	60
D	72	141	77	94
E	36	97	54	78
F (low)	21	71	54	71

Source: Reprinted with permission from L. Srole et al. *Mental Health in the Metropolis: The Midtown Manhattan Study*, (New York: NYU Press, 1978), p. 289.

- ✓ $G^2(RC) = 3.57\ with\ df = 8$ (fits well)
- ✓ ML estimates : $(-1.11, -1.12, -0.37, 0.03, 1.01, 1.82)$ / $(-1.68, -0.14, 0.14, 1.41)$ / $\hat{\beta} = 0.17$
- ✓ Positive trend

Correlation models

- **Correlation model for one-dimensional version (all scores are parameters)**

$$\pi_{ij} = \pi_{i+}\pi_{+j}(1 + \lambda\mu_i\nu_j),$$

$$\sum \mu_i \pi_{i+} = \sum \nu_j \pi_{+j} = 0 \quad \text{and} \quad \sum \mu_i^2 \pi_{i+} = \sum \nu_j^2 \pi_{+j} = 1.$$

✓ Standardized scores

✓ λ is the correlation between the scores for joint distribution

✓ ML estimates of the scores maximize the correlation

✓ ML estimates of λ and the score parameters are similar to those of β and the score parameters of RC model (when λ is close to zero)