한국통계학회논문집 제2권 2호
1995년 12월 pp. 310-319

Graphical Descriptions for Hierarchical Log Linear Models

Hyun Jip Choi[1] and Chong Sun Hong[2]

Abstract

We represent graphically the relationship of hierarchical log linear models by regarding the values of the likelihood ratio statistics as the squared norm of the corresponding vectors. Right angled triangles, tetrahedrons, and modified polyhedrons are used for graphical description. We find that the angle between the two vectors depends on the coefficient of determination and the partial coefficient of determination. These graphical descriptions could be applied to the model selection method.

1. Introduction

There have been several attempts to plot the properties of contingency tables. Fienberg and Gilbert (1970) developed the geometry of measure for 2×2 contingency tables, which allows to visualize the properties of the various models in terms of the loci of the tetrahedron. Goodman (1991) reviewed several plots for identifying associations in log linear models. For multi-way contingency tables, Friendly (1994) developed Mosaic display. On the other hand, for the interpretation of the relationships among the effects of the terms in a given log-linear model, a simple and undirected graph which is called an association graph has been explored by Darroch, Lauritzen and Speed (1980), Goodman (1971, 1973) and others. By this association graph, they also defined the class of graphical models. And Edwards and Kreiner (1983) gave an overview of the use of graphical models which are generated by the graphs. They, especially, suggested strategies for model selection based on this class of models. These previous works have been studied to explain interactions within certain log linear models with undirected graphical representation. For a given model, the association graph plays an important role in describing the relationship among main factors including interactions by using points and lines which connect points. However, we have not stressed the measured degrees of associations and the magnitudes of the goodness of fit statistics.

In this paper, we propose some graphical descriptions to examine the relationships of the likelihood ratio statistics, G^2, corresponding to several hierarchical log linear models, so that

1) Department of Applied Statistics, Kyunggi University, Changan Gu, Suwon, 440-760, KOREA
2) Department of Statistics, Sung Kyun Kwan University, Seoul, 110-745, KOREA

we can evaluate visually the relationships among log linear models. In order to plot the relationship of two hierarchical log linear models, right angled triangle is used. We adapt the shape of tetrahedron for three log linear models. And for more than three log linear models, some polyhedrons are modified to plot. The partial coefficient of determination for hierarchical log linear model is suggested in this paper. Both values of the well known coefficient of determination and the proposed partial coefficient of determination are found to depend on the magnitudes of angles which consist in the proposed plotting shapes. Those are discussed in the following section.

2. Graphical descriptions

Consider models (a), (b) and (c) such that model (a) is the special case of model (b) which is the special case of model (c). And define $G^2(a)$, $G^2(b)$ and $G^2(c)$ as the likelihood ratio statistics for models (a), (b) and (c), respectively. In this paper, our attention is restricted to hierarchical log linear models for complete tables. Under this hierarchical structure, it satisfies that $G^2(a) \geq G^2(b) \geq G^2(c)$. First of all, we consider only models (a) and (b). For model (a) and (b), we can have an equation

$$\begin{aligned} G^2(a) &= [G^2(a) - G^2(b)] + G^2(b) \\ &\equiv G^2(a|b) + G^2(b) \ . \end{aligned} \tag{1}$$

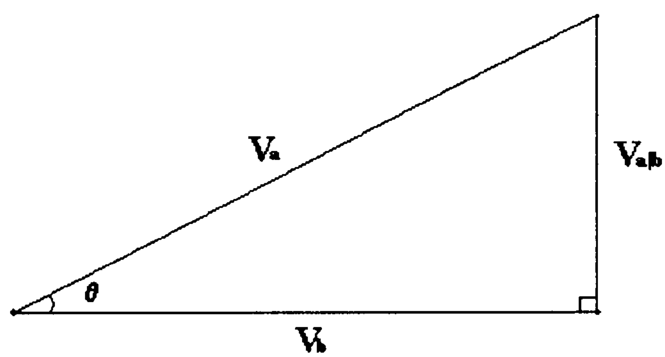
From this partitioning, $G^2(a)$ and $G^2(b)$ are the variation for model (a) and model (b), respectively. And the first component of the RHS in equation (1), $G^2(a|b)$, could be regarded as the improved variation explained by model (b) from model (a). Now we can define three components in equation (1) as the squared norms of the corresponding vectors V_a, $V_{a|b}$ and V_b, respectively. Let us denote

$$\begin{aligned} |V_a|^2 &= G^2(a) \, , \\ |V_{a|b}|^2 &= G^2(a|b) \, , \\ |V_b|^2 &= G^2(b) \, . \end{aligned}$$

With these vectors, we may draw a right angled triangle like the one in Figure 1. All information concerning the goodness of fits for models (a) and (b) can be obtained simply by comparing the lengths of the vectors V_a, V_b, and $V_{a|b}$. We note that the angle θ between two vectors V_a and V_b satisfies

$$\cos\theta = \frac{|V_b|}{|V_a|} \quad . \tag{2}$$

From the right angled triangle in Figure 1, we can see that the shorter the length of the vector V_b is, the longer the length of vector $V_{a|b}$, and the larger the value of angle θ. The large angle θ means that the length of the vector V_b is relatively shorter than that of vector V_a, so that we can say that model (b) may fit better than model (a) does, and the variation due to the difference between model (a) and (b), $G^2(a|b)$, may be significant. Therefore through the right angled triangle, we can identify not only the goodness of fits for each log linear model but also the difference between goodness of fits of model (a) and (b).



(Figure 1) Right angled triangle for two hierarchical models

Now, we regard model (a) as the smallest log linear model, e.g. complete independence model. Then the coefficient of determination R^2 which is defined by Christensen (1990) for log linear models could be considered with angle θ of this triangle such that

$$\begin{aligned} R^2(b) &= 1 - \frac{G^2(b)}{G^2(a)} \\ &= 1 - \frac{|V_b|^2}{|V_a|^2} = 1-\cos^2\theta \\ &= \sin^2\theta \quad . \end{aligned} \tag{3}$$

If we regard model (a) and model (b) in (1) and (3) as reduced model and full model, respectively, $R^2(b)$ could be the generalized coefficient of determination suggested by Anderson-Sprecher (1994) in regression analysis. This form emphasizes that R^2 is a model comparison like the right angled triangle.

As in the regression, the value of R^2 gets large as number of parameters in a log linear model increases. So we can adjust R^2 by the number of degree of freedom for the models of interest as

$$R^2_{adj.}(b) = 1 - \frac{d_b}{d_a}\,[1 - R^2(b)\,]$$

$$= 1 - \frac{G^2(b)/d_b}{G^2(a)/d_a} \quad , \tag{4}$$

where d_a and d_b are the degrees of freedom for model (a) and model (b), respectively. The shape of right angled triangle does not considered the degrees of freedom of the corresponding log linear models. But with the adjusted coefficient of determination, we can make a better interpretation of the relationship of the two log linear models.

Now examine the relationship among model (a), (b), and (c). With equation (1), we also obtain the following :

$$G^2(a) = G^2(a|c) + G^2(c) \quad ,$$

$$G^2(b) = G^2(b|c) + G^2(c) \quad . \tag{5}$$

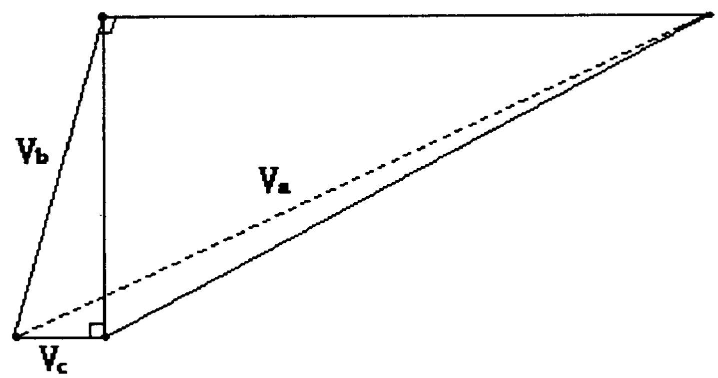
Each equation brings up the image of a right-angled triangle, so that three equations in (1) and (5) could be made up to a tetrahedron like the one in Figure 2, where we denote

$$G^2(a|c) = |V_{a|c}|^2,\ G^2(b|c) = |V_{b|c}|^2 \text{ and } G^2(c) = |V_c|^2 .$$

We can state the graphical interpretation about the shape in Figure 2 as we did in Figure 1. With this tetrahedron, the relationship between model (a) and (b), and model (a) and (c) can be described by examining vectors $V_{a|b}$ and $V_{a|c}$. Moreover, the two angles between vectors V_a and V_b, and V_a and V_c could be explained by extending equations (2) and (3) as the following :

$$\cos\theta_1 = \frac{|V_b|}{|V_a|} \quad , \quad \cos\theta_2 = \frac{|V_c|}{|V_a|} \quad ,$$

$$R^2(b) = \sin^2\theta_1 \quad , \quad R^2(c) = \sin^2\theta_2 \quad . \tag{6}$$



(Figure 2) Tetrahedron for three hierarchical models

Now we can set up the following equation from the relationship in (5) and (6),

$$R^2(c) - R^2(b) = \frac{G^2(b|c)}{G^2(a)} \quad . \tag{7}$$

Since $G^2(b|c)$ is the improved goodness of fit by model (c) from model (b), $R^2(c) - R^2(b)$ means the proportion of the variations explained by the difference of the two models (b) and (c). Therefore $R^2(c) - R^2(b)$ could be regarded as the partial coefficient of determination for models (b) and (c), and we will denote $R^2(b|c)$.

Definition 1

The partial coefficient of determination for model (b) and model (c) is defined as

$$R^2(b|c) = \frac{G^2(b|c)}{G^2(a)} \quad .$$

If the value of $R^2(b|c)$ is too large, then the difference between the goodness of fits of models (b) and (c) would be significant. And we might say that model (c) fits better than model (b) does. We note that the partial coefficient of determination can be also obtained by the difference between $R^2(b)$ and $R^2(c)$:

$$R^2(b|c) = R^2(c) - R^2(b) \quad . \tag{8}$$

And we could also adjust the partial R^2 for models (b) and (c), which we call the adjusted partial coefficient of determination, $R^2_{adj.}(b|c)$.

Definition 2

The adjusted partial coefficient of determination for model (b) and model (c) is defined as

$$R^2_{adj.}(b|c) = \frac{G^2(b)/d_b - G^2(c)/d_c}{G^2(a)/d_a} .$$

Also, the adjusted one could be obtained by

$$R^2_{adj.}(b|c) = [1 - R^2_{adj.}(b)] - [1 - R^2_{adj.}(c)]$$
$$= R^2_{adj.}(c) - R^2_{adj.}(b) . \qquad (10)$$

A large value of the adjusted partial R^2 indicates that model (c) might fit better than model (b) does.

For the $2\times2\times2$ structural habitat data for Lizards of Bimini(Schoener (1968)), we consider the following three hierarchical log linear models, and their results are listed in Table 1.

(Table 1) Goodness of fits for structural habitat data for Lizards of Bimini.

ID	MODEL	d.f.	G^2	Difference	d.f.	G^2
(a)	[1][23]	3	12.43*			
(b)	[13][23]	2	2.03	(a) and (b)	1	10.4*
(c)	[12][13][23]	1	0.15	(b) and (c)	1	1.88

* indicates that the p-value of the statistic is less than 5% significant level.

An exact graphical description for comparing model (a) with (b) is shown in Figure 1. The angle, θ_1, between two vectors V_a and V_b is 66.16°. Since the magnitude of this angle and the length of $V_{a|b}$ are both large, we might say that the difference between goodness of fits of models (a) and (b) is significant. Note that the value of $R^2(b)$ is 0.834 and $R^2_{adj.}(b)$ is 0.755, and these values are highly informative for assuring that model (b) is better fitted than model (a).

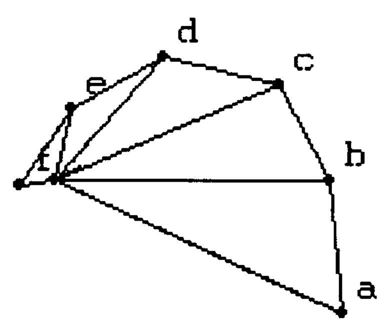
Furthermore, the relationships among the above three models are presented in Figure 2. Figure 2 indicates that $|V_a|$ is relatively longer than both $|V_b|$ and $|V_c|$, and that $|V_{a|b}|$ is longer than $|V_{b|c}|$. Also one obtains $R^2(b|c) = 0.151$ and $R^2_{adj.}(b|c) = 0.209$, which are irrelevant. Hence we may choose model (b) as the best.

It is impossible to represent the relationships among four hierarchical log linear models on three dimensional space. So we suggest an alternative method to plot these four models in this paper. One can consider the relationships sequentially. For a given hierarchical model, we could compare this one with the following two models : one is the smallest model of all, and the other is the special and just previous case of the given model. For example, there are six models labeled as (a) to (f), where model (a) is the smallest and the special case of model (b) which is the special of model (c), and so on. For model (d), we might obtain the following equations :

$$G^2(a) = G^2(a|d) + G^2(d) \quad ,$$
$$G^2(c) = G^2(c|d) + G^2(d) \quad .$$

Also for final model (f),

$$G^2(a) = G^2(a|f) + G^2(f) \quad ,$$
$$G^2(e) = G^2(e|f) + G^2(f) \quad .$$



(Figure 3) A polyhedron for hierarchical models

Under the hierarchical structure, we might consider a polyhedron like the one in Figure 3, where each phase is a right angled triangle. And vectors V_a, V_b, $\cdots$, V_f in Figure 3 have the similar shape with ''umbrella ribes'' where each length is different. With this polyhedron, we might apply the analogous arguments of Figure 2, i.e., through each right angled triangle on phases of the polyhedron, we could compare any given model with both the just previous one and the smallest one. These arguments could be applied to the well known forward model selection method based on the conditional likelihood ratio test statistics.

As an illustrative example, we take the $3\times2\times2\times2$ detergent preference data of Ries and

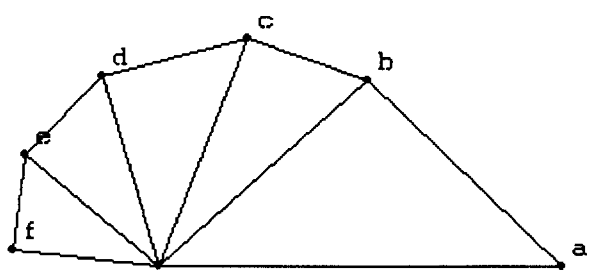
Smith (1963). Among others, Bishop et. al. (1975) and Fienberg (1983) considered the following hierarchical log linear models for model selection, so we consider the same hierarchy.

(Table 2) Goodness of fits for Detergent Preference data

ID	MODEL	d.f.	G^2	Difference	d.f.	G^2
(a)	[1][2][3][4]	18	42.93*			
(b)	[1][3][24]	17	22.35	(a) and (b)	1	20.58*
(c)	[1][24][34]	16	17.99	(b) and (c)	1	4.36*
(d)	[13][24][34]	14	11.89	(c) and (d)	2	6.10*
(e)	[1][234]	12	8.41	(d) and (e)	2	3.48
(f)	[123][234]	8	5.66	(e) and (f)	4	2.75

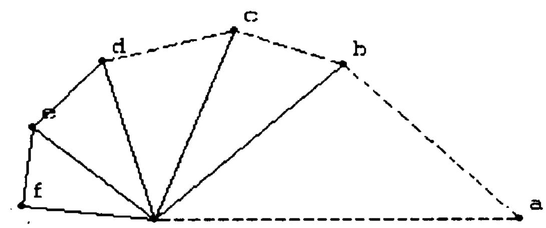
* indicates that the p-value of the statistic is less than 5% significant level.

An exact graphical description for the six models above is presented in Figure 3 which is based on the preceeding arguments. From this plot, model (d) might be regard as optimal. If we are interested only in the relationship between previous and posterior models for a certain model, we would regard a stretched-out shape like the one in Figure 4. This shape of the stretched polyhedron could be described graphically all informations for the models which include the relationship among the models in Table 2.



(Figure 4) Stretched polyhedron for hierarchical models

For example, if a goodness of fit statistic is significant, then the corresponding vector can be drawn as a dotted line. With this expression, results of Table 2 could be replotted in Figure 5. Then this kind of the plot helps us to explain clearly a certain hierarchical model structure and choose the best model. We may apply the forward and the backward selection method via Figure 5. As a result the best model could be obtained as model (d) quickly and clearly.



(Figure 5) A modified stretched polyhedron for hierarchical models

We can obtain the coefficients of determination for evaluating models in Table 2 by extending the equation (5) and the partial one by (6). All values about R^2 are listed in Table 3. Among partial R^2's, $R^2(c|d)$ and $R^2_{adj.}(c|d)$ have the largest values 0.142 and 0.116. This means that the model (d) gives a better fit than the model (c) does, and we could have exactly the same conclusion like that of graphical description. We may state that these information are stochastically supplement to the graphical descriptions.

(Table 3) Summarized results for (Table 2)

model	R^2	$R^2_{adj.}$		$R^2(\cdot\|\cdot)$	$R^2_{adj.}(\cdot\|\cdot)$
(b)	0.4800	0.4494			
(c)	0.5816	0.5293	(b) and (c)	0.1016	0.0799
(d)	0.7237	0.6448	(c) and (d)	0.1421	0.1155
(e)	0.8048	0.7072	(d) and (e)	0.0821	0.0624
(f)	0.8712	0.7102	(e) and (f)	0.0664	0.0030

3. Conclusion

We consider graphical descriptions to compare the relationship between the likelihood ratio statistics corresponding to several log linear models in a hierarchical structure. If we regard the values of the likelihood ratio statistics as a squared norm of the vectors, we could evaluate visually the relationship between two hierarchical models by using the shape of right angled triangles. And a tetrahedron could be described by the relationship of three hierarchical models graphically. For more than three hierarchical models, we might consider a polyhedron and a stretched polyhedron to explain the relationship sequentially. Moreover, we discuss that the angle on each phase of the plotting shapes relates with the coefficient of determinations, and propose the partial coefficient of determinations. The partial R^2 and the adjusted partial R^2 are also proposed in this paper. This graphical descriptions and several coefficients of determination help us to compare and evaluate the goodness of fits for the hierarchical log linear models, and could be applied to the model selection method.

References

[1] Anderson-Sprecher, Richard (1994). Model selection and R^2, *The American Statistician,* Vol. 48, 113-117.

[2] Bishop, Y. M. M, Fienberg, S. E., and Holland P. W. (1975). *Discrete Multivariate Analysis*, MIT Press.

[3] Christensen, R. (1990). *Log-Linear Model*, Springer-Velag.

[4] Darroch, J. N., Lauritzen, S. L. and Speed, T. P. (1980). Markov-fields and log-linear models for contingency tables, *Annals of Statistics,* Vol. 8, 522-39.

[5] Edwards, D. and Kreiner, S. (1983). The analysis of contingency tables by graphical models, *Biometrika,* Vol. 70, 553-565.

[6] Fienberg, S. E. (1983). *The Analysis of Cross-Classified Categorical Data*, MIT Press.

[7] Fienberg, S. E. and Gilbert, J. P. (1970). The geometry of a two by two contingency table, *Journal of the American Statistical Association,* Vol. 65, 694-701.

[8] Friendly, M. (1994). Mosaic displays for multi-way contingency tables, *Journal of American Statistical Association*, Vol. 65, 226-256.

[9] Goodman, L. A. (1971). Partitioning of chi-square, analysis of marginal contingency tables, and estimation of expected frequencies in multidimensional contingency tables, *Journal of American Statistical Association*, Vol. 66, 339-344.

[10] Goodman, L. A. (1973). The analysis of multidimensional contingency tables : stepwise procedures and direct methods for building models for multiple classifications," *Technometrics,* Vol. 13, 33-61.

[11] Goodman, L. A. (1991). Measures, model, and graphical displays in the analysis of cross-classified data, *Journal of the American Statistical Association,* Vol. 86, 1085-1111.

[12] Ries, P. N. and Smith, H. (1963). The use of chi-square for preference testing in multidimensional problems, *Chemical Engineering Progress,* Vol. 59, 39-43.

[13] Schoner, T. W. (1968). The anolis Lizards of Bimini : resource partitioning in a complex fauna, *Ecology*, Vol. 49, 704-726.