Psychology Science, Volume 47, 2005 (3/4), p. 304-314

Paradoxes in multidimensional contingency tables: What does this mean for CFA?

JOACHIM KRAUTH[1]

Abstract

Lienert (1969, 1971a) justified his invention of Configural Frequency Analysis (CFA) by the observation that higher-order interactions may exist in three- and higher-dimensional data even if no correlations between any pairs of variables can be observed. In Lienert (1971b) Hierarchical CFA (HCFA) was proposed in order to find that subset of variables which shows the highest degree of interdependence. In HCFA, the original contingency table is collapsed in a hierarchical way until two-dimensional tables result and CFA is applied to all tables in this hierarchy.

Now, it has been known for more than a hundred years that collapsing contingency tables may lead to strange results which are nowadays often denoted as Simpson's paradox (Simpson, 1951). Another paradox studied first by Lord (1967) occurs when groups are compared which may differ in their baseline values. Here, a contingency table version of this latter paradox is considered. Implications of these possible paradoxes for the interpretation of the results of CFA are discussed.

Key words: Configural Frequency Analysis, Simpson's paradox, Lord's paradox

[1] Prof. Dr. J. Krauth, Department of Experimental Psychology, University of Düsseldorf, D-40225 Düsseldorf, Germany; email: J.Krauth@uni-duesseldorf.de

Introduction

G.A. Lienert introduced CFA in a talk given 1968 at the Congress of the German Society of Psychology (Lienert, 1969). His main argument for proposing a new method of classification was that with this method it was possible to allow not only for correlations for pairs of variables as this is the case for factor and configural analysis but also for higher-order associations between more than two variables. In Lienert (1971a) it is argued that the invention of CFA was caused by a Figure in an article by Olmstead and Tukey (1947) illustrating a second-order interaction between three binary variables where no first-order interactions are present. This Figure could explain according to Lienert the structure of his LSD data presented in Lienert (1971a). According to Lienert the detection of the Figure in the Olmstead-Tukey article would not have been necessary if he had understood the algebraic formulation of – in Lienert's terminology – Meehl's paradox (Meehl, 1950). Perhaps some little doubt is permitted with respect to this nice narrative by which Lienert tried to unveil the origins of CFA: Both articles by Meehl and by Olmstead and Tukey were cited already in Lienert (1969) where the newly invented CFA is applied to five items of the Hamburg Depression Scale and in this contribution neither Olmstead-Tukey's Figure nor Meehl's paradox are mentioned anywhere.

1. Simpson's paradox

The denotion Simpson's paradox was chosen after Simpson (1951) gave a simple example of a paradoxical situation in a 2×2×2 table. Nowadays it is known that Simpson's finding was a rediscovery because already Yule (1903) discussed problems of this kind. Meanwhile many authors have studied the effects which may occur if higher-dimensional contingency tables are collapsed as this is done e.g. in HCFA (Lienert, 1971b). The essence of Simpson's paradox is that statistical inference in a low-dimensional table can distort seriously the reality in a high-dimensional table from which it has been derived by collapsing. We cite here only the article by Guo, Geng and Shi (2001) as an example for the many papers which studied the conditions (here, the term collapsibility is used) which higher-dimensional tables must fulfill in order to guarantee that the statistical information in low-dimensional tables obtained by collapsing is not distorted. That Simpson's paradox is by no means a theoretical phenomenon which can be ignored in practice is shown in the articles by Wainer (1986) and Wainer and Brown (2004) where just this paradox is exhibited in real data sets.

In the following we restrict our attention to the most simple case where Simpson's paradox can occur. This is the case of three binary variables X, Y, and Z constituting a 2×2×2 table. The corresponding scheme is given in Table 1.

Table 1:
Scheme for a 2×2×2 table with probabilities π_{ijk} and frequencies N_{ijk} for the 8 configurations (n denotes the sample size)

X	Y	Z	π_{ijk}	N_{ijk}
1	1	1	π_{111}	N_{111}
1	1	0	π_{110}	N_{110}
1	0	1	π_{101}	N_{101}
1	0	0	π_{100}	N_{100}
0	1	1	π_{011}	N_{011}
0	1	0	π_{010}	N_{010}
0	0	1	π_{001}	N_{001}
0	0	0	π_{000}	N_{000}
Sums:			1	n

In the following a dot indicates that a sum is formed over the two possible values (1 and 0) of the corresponding index, e.g. $\pi_{1\bullet 0} = \pi_{100} + \pi_{110}$, $\pi_{1\bullet\bullet} = \pi_{111} + \pi_{110} + \pi_{101} + \pi_{100}$ etc.

Independence of X and Y is given if $\pi_{11\bullet} = \pi_{1\bullet\bullet}\pi_{\bullet 1\bullet}$, independence of X and Z holds if $\pi_{1\bullet 1} = \pi_{1\bullet\bullet}\pi_{\bullet\bullet 1}$, independence of Y and Z means that $\pi_{\bullet 11} = \pi_{\bullet 1\bullet}\pi_{\bullet\bullet 1}$. The independence of X, Y, and Z is assumed if $\pi_{ijk} = \pi_{i\bullet\bullet}\pi_{\bullet j\bullet}\pi_{\bullet\bullet k}$ for all possible selections of $i, j, k \in \{0, 1\}$.

From the independence of X, Y, and Z it follows that no second-order interaction is present and that the 3 pairs of variables (X, Y; X, Z; Y, Z) are independent. The reverse is also true.

If the 3 pairs of variables are independent and if X, Y, and Z are dependent (i.e. not independent) then a second-order interaction is present.

Several different definitions of second-order interaction in 2×2×2 tables have been proposed. According to Bartlett (1935) no second-order interaction is assumed if $\pi_{111}\pi_{100}\pi_{010}\pi_{001} = \pi_{110}\pi_{101}\pi_{011}\pi_{000}$. A different definition due to Lancaster (1969) is given for this situation by

$$\pi_{111} = \pi_{1\bullet\bullet}\pi_{\bullet 1\bullet}\pi_{\bullet\bullet 1} + \pi_{1\bullet\bullet}(\pi_{\bullet 11} - \pi_{\bullet 1\bullet}\pi_{\bullet\bullet 1}) + \pi_{\bullet 1\bullet}(\pi_{1\bullet 1} - \pi_{1\bullet\bullet}\pi_{\bullet\bullet 1}) + \pi_{\bullet\bullet 1}(\pi_{11\bullet} - \pi_{1\bullet\bullet}\ \pi_{\bullet 1\bullet}).$$

An idealized version of Lienert's LSD data (Lienert 1971a) is given in Table 2.

Table 2:
Idealized version of Lienert's LSD data

X	Y	Z	π_{ijk}	N_{ijk}
1	1	1	.25	16
1	1	0	0	0
1	0	1	0	0
1	0	0	.25	16
0	1	1	0	0
0	1	0	.25	16
0	0	1	.25	16
0	0	0	0	0
	Sums:		1.00	64

Here, X and Y are independent $(.25 = .5 \times .5)$, X and Z are independent $(.25 = .5 \times .5)$, and Y and Z are independent $(.25 = .5 \times .5)$. However, X, Y, and Z are dependent $(.25 \neq .5 \times .5 \times .5 \text{ for } i = j = k = 1)$ and therefore a second-order interaction exists. Bartlett's definition gives $.25 \times .25 \times .25 \times .25 \neq 0 \times 0 \times 0 \times 0$ and Lancaster's definition $.25 \neq .5 \times .5 \times .5 + .5(.25 - .5 \times .5) + .5(.25 - .5 \times .5) + .5(.25 - .5 \times .5)$.

In order to show more clearly the possible dramatic loss of information which can occur if a 2×2×2 contingency table is collapsed in the presence of a second-order interaction, we consider the following example which is an adaptation of an example by Kendall and Stuart (1961, pp. 544-545). Here (cf. Table 3), we assume that X denotes gender (1 = male, 0 = female), Y denotes treatment (1 = treatment, 0 = no treatment), and Z denotes recovery (1 = recovery, 0 = no recovery). We observe for the males (first 4 configurations) a positive association between treatment and recovery (.12×.12 – .50×.01 > 0), while for the females (second 4 configurations) a negative association between treatment and recovery results $(.04 \times .04 - .14 \times .03 < 0)$. In other words: The treatment is good for males but bad for females.

Table 3:
Adaptation of an example by Kendall and Stuart

X	Y	Z	π_{ijk}	N_{ijk}
1	1	1	.12	12
1	1	0	.50	50
1	0	1	.01	1
1	0	0	.12	12
0	1	1	.04	4
0	1	0	.14	14
0	0	1	.03	3
0	0	0	.04	4
	Sums:		1.00	100

If we collapse the 2×2×2 table in Table 3 with respect to gender (*X*) we obtain Table 4. In this table no association between treatment and recovery is observed $(.16 \times .16 - .64 \times .04 = 0)$. This means that treatment has no effect at all.

Table 4:
Table 3 collapsed with respect to variable *X*

Y	*Z*	$\pi_{\bullet jk}$	$N_{\bullet jk}$
1	1	.16	16
1	0	.64	64
0	1	.04	4
0	0	.16	16
Sums:		1.00	100

In the example depicted in Table 3 a second-order interaction is present according to Bartlett's definition $(.12 \times .12 \times .14 \times .03 > .50 \times .01 \times .04 \times .04)$ and according to Lancaster's definition

$$(.12 > .75 \times .80 \times .20 + .75(.16 - .80 \times .20) + .80(.13 - .75 \times .20) + .20(.62 - .75 \times .80)).$$

Simpson (1951) argued that it is obvious which strategy should be followed if a second-order interaction is present: In this case, the six possible 2×2 tables ($X \times Y$ for $Z = 1$, $X \times Y$ for $Z = 0$, $X \times Z$ for $Y = 1$, $X \times Z$ for $Y = 0$, $Y \times Z$ for $X = 1$, $Y \times Z$ for $X = 0$) have to be studied with respect to possible associations. However, paradoxical situations can arise also if no second-order interactions are present and this is what is called nowadays Simpson's paradox. An adaptation of Simpson's example is given in Table 5, where again we identify *X* with gender, *Y* with treatment, and *Z* with recovery.

Table 5:
Adaptation of Simpson's example

X	*Y*	*Z*	π_{ijk}	N_{ijk}
1	1	1	.16	8
1	1	0	.08	4
1	0	1	.08	4
1	0	0	.06	3
0	1	1	.24	12
0	1	0	.32	16
0	0	1	.02	1
0	0	0	.04	2
Sums:			1.00	50

According to Bartlett's definition, no second-order interaction is present: .16×.06×.32×.02 = .08×.08×.24×.04. The same holds for Lancaster's definition: .16 = .38×.80×.50 + .38(.40 – .80×.50) + .80(.24 – .38×.50) + .50(.24 – .38×.80).

For the males we observe a positive association between treatment and recovery (.16×.06 – .08×.08 > 0) and this holds also for the females (.24×.04 – .32×.02 > 0). In other words: The treatment is benign for the males as well as for the females! In Table 6 we have collapsed Table 5 with respect to gender (*X*). Contrary to our expectation there is no association between treatment and recovery (.40×.10 – .40×.10 = 0) though we observed positive associations for both subpopulations. Above we mentioned that Wainer (1986) and Wainer and Brown (2004) presented real data sets where this kind of paradox occurred.

Table 6:
Table 5 collapsed with respect to variable *X*

Y	Z	$\pi_{\bullet jk}$	$N_{\bullet jk}$
1	1	.40	20
1	0	.40	20
0	1	.10	5
0	0	.10	5
Sums:		1.00	50

3. Lord's paradox

Lord's paradox was first described by Lord (1967) and has been discussed by many authors (e.g. Lord, 1969, 1976; Holland and Rubin, 1983; Wainer, 1991). In Wainer and Brown (2004) real data are presented where not only Simpson's paradox but also Lord's paradox are observed.

Lord (1967) considered the following example: In a large university the administration is interested in studying the effects which the diet provided in the university dining halls has on the students and whether any gender difference can be observed for these effects. Beside other variables the weight (*Y*) of each student at the time of arrival in September and the corresponding weight (*Z*) in the following June are recorded. The two scatterplots for boys and girls can be described by intersecting ellipses which are both symmetric with respect to the 45° line through the origin. From this follows that the centers of the ellipses lie on the 45° line through the origin. Naturally, the center of the boys' ellipse is shifted to the right in comparison with that of the girls because, in general, the weight of boys is higher than that of girls. By projecting the ellipses to the *Y* and *Z* axes we find that there is no difference between the distributions of the September and June weights and this is true also if the weights for boys and girls are considered separately. As a consequence we can conclude that the diet had no effect on the weight of the boys and girls.

Predicting the June weights by the September weights using linear regressions separately for boys and girls it is found that the slopes of the two regression lines are identical but that the intercepts differ. In fact, the boys exhibit a larger intercept. If an analysis of covariance would be performed with the September weight as a covariate it would be found that the boys show more gain in weight than the girls.

These two contradictory interpretations of the same data constitute Lord's paradox. It is very difficult to decide which answer is the correct one in an empirical situation and many authors have discussed Lord's and other examples.

According to Lienert one important application of CFA is the analysis of change (e.g. Lienert, 1971b; Lienert and Krauth, 1973a, b). In Lienert (1971b) it is proposed to code an increase of a variable between two points of time by a „+“, a decrease by a „–“, and no change by a „0“ and to use this change score with 3 categories in CFA. An obvious question is whether Lord's paradox is also important for contingency table analyses where we do not have metric variables and cannot use regression and covariance analysis. The answer is yes, if we consider the example in Table 7 which is an adaptation of Lord's example for contingency tables. Here, X denotes again gender (1 = boys, 0 = girls), Y denotes the September weight (0 = at most 55 kg, 1 = more than 55 kg and less than 70 kg, 2 = at least 70 kg), and Z denotes the June weight with the same categories.

Table 7:
Adaptation of Lord's example for a 2×3×3 table

X	Y	Z	π_{ijk}	N_{ijk}	X	Y	Z	π_{ijk}	N_{ijk}
1	0	0	.00	0	0	0	0	.08	4
1	0	1	.12	6	0	0	1	.00	0
1	0	2	.00	0	0	0	2	.12	6
1	1	0	.00	0	0	1	0	.12	6
1	1	1	.06	3	0	1	1	.06	3
1	1	2	.12	6	0	1	2	.00	0
1	2	0	.12	6	0	2	0	.00	0
1	2	1	.00	0	0	2	1	.12	6
1	2	2	.08	4	0	2	2	.00	0
Sums:			.50	25	Sums:			.50	25

By collapsing the subtable with $X = 1$ (boys) with respect to Z (June weights) we derive Table 8 (distribution of September weights for the boys). By collapsing the subtable with $X = 1$ (boys) with respect to Y (September weights) we get Table 9 (distribution of June weights for the boys). By collapsing the subtable with $X = 0$ (girls) with respect to Z or Y, respectively, we obtain Tables 10 and 11. We see at once that the distributions of June and September weights do not differ, and this is true for boys (Tables 8, 9) and girls (Tables 10 and 11), though, naturally, boys and girls exhibit different distributions. Thus, we may conclude that the diet had no effect and that this non-effect is observed as well for the boys as for the girls.

Table 8:
Distribution of September weights for the boys

X	Y	$\pi_{ij\bullet}$	$N_{ij\bullet}$
1	0	.24	6
1	1	.36	9
1	2	.40	10
Sums:		1.00	25

Table 9:
Distribution of June weights for the boys

X	Z	$\pi_{i\bullet k}$	$N_{i\bullet k}$
1	0	.24	6
1	1	.36	9
1	2	.40	10
Sums:		1.00	25

Table 10:
Distribution of September weights for the girls

X	Y	$\pi_{ij\bullet}$	$N_{ij\bullet}$
0	0	.40	10
0	1	.36	9
0	2	.24	6
Sums:		1.00	25

Table 11:
Distribution of June weights for the girls

X	Y	$\pi_{i\bullet k}$	$N_{i\bullet k}$
0	0	.40	10
0	1	.36	9
0	2	.24	6
Sums:		1.00	25

Now, we define for each subject a change score (C) as this was proposed in Lienert (1971b). This means that we set C = „–“ for ($Y = 1, Z = 0$), ($Y = 2, Z = 0$), and ($Y = 2, Z = 1$), we set C = „0“ for ($Y = 0, Z = 0$), ($Y = 1, Z = 1$), and ($Y = 2, Z = 2$), and we set C = „+“ for ($Y = 0, Z = 1$), ($Y = 0, Z = 2$), and ($Y = 1, Z = 2$). This yields Table 12.

Table 12:
Rescaling of Table 7 by replacing (Y, Z) by C

X	C	π_{ij}	N_{ij}
1	–	.12	6
1	0	.14	7
1	+	.24	12
0	–	.24	12
0	0	.14	7
0	+	.12	6
Sums:		1.00	50

From Table 12 we can conclude that for the boys the probability for an increase in weight (.24) is twice the probability of a decrease in weight (.12) while for the girls just the opposite is true. This indicates that the diet has a considerable effect on weight and that this effect is different for boys and girls. This observation contradicts the results of our analysis above where no weight changes were found for boys and girls.

4. Results of CFA

For getting more distinct results we double all frequencies in Tables 5, 6, 7, and 12. Of course, this does not alter the structure of these tables. Then we perform a CFA using exact hypergeometric tests (Krauth, 1993, pp. 30-34).

For Table 5 (Simpson's paradox) we find types for ($X = 0$, $Y = 1$, $Z = 0$) i.e. treated females without recovery (p = .006131) and ($X = 1$, $Y = 0$, $Z = 1$) i.e. untreated males with recovery (p = .016806). In Table 6 (Table 5 collapsed with respect to gender) we do not detect any type.

For Table 7 (Lord's paradox) we find types for ($X = 0$, $Y = 0$, $Z = 2$) i.e. girls with low September and high June weight (p = .000764), ($X = 0$, $Y = 1$, $Z = 0$) i.e. girls with normal September and low June weight (p = .002770), ($X = 0$, $Y = 2$, $Z = 1$) i.e. girls with high September and normal June weight (p = .002770), ($X = 1$, $Y = 0$, $Z = 1$), i.e. boys with low September and normal June weight (p = .002770), ($X = 1$, $Y = 1$, $Z = 2$) i.e. boys with normal September and high June weight (p = .002770), ($X = 1$, $Y = 2$, $Z = 0$) i.e. boys with high September and low June weight (p = .000746). For Table 12 (Table 7 with (Y, Z) replaced by change scores) we detect types for ($X = 0$, C = „–“) i.e. girls with a decrease in weight (p = .010648) and for ($X = 1$, C = „+“) i.e. boys with an increase in weight (p = .010648).

5. Discussion

What can we learn from the examples above which illustrate Simpson's and Lord's paradoxes? Our conclusions are as follows:

An important argument for introducing CFA was that higher-order interactions are no longer a danger for the interpretation of the data resulting from collapsing contingency tables (Lienert, 1969, 1971a). However, Simpson's paradox (Simpson, 1951) does not require the existence of such interactions and this means that the interpretation of CFA results for collapsed tables may be distorted with respect to the original table even if no higher-order interactions are present. Therefore, Lienert's proposal to find better interpretable types by means of HCFA (Lienert, 1971b) may sometimes yield interpretations of the data which are not really justified. This can be seen e.g. by comparing the CFA results for Tables 5 and 6. A further problem is that each multivariate contingency table which we analyze by means of CFA can be considered as a table which resulted from collapsing a table with a higher dimension because we can consider always only a restricted set of variables and infinitely many possible other variables are not considered by us.

The measurement of change is an important problem in psychology and other fields. In Lienert (1971b) and other articles a CFA based on measures of change is proposed. As we have seen, this kind of analysis can yield results which are quite different from those which we get by applying a CFA to the original data. A reason for this can be Lord's paradox (Lord, 1967). It does not exist a simple solution for this paradox and sometimes the first kind of analysis sems to be more appropriate and sometimes the second kind (cf. Holland and Rubin, 1983; Wainer, 1991). In our example, we found that the results of CFA for Table 7 (original data) are far more difficult to interpret than the CFA results for Table 12 where change scores were used. But we are not sure that this a valid argument to prefer a CFA based on change scores.

References

1. Bartlett, M.S. (1935). Contingency table interactions. Journal of the Royal Statistical Society, Supplement, 2, 248 – 252.
2. Guo, J., Geng, Z. & Shi, N. (2001). On collapsibilities of Yule's measure. Science in China, Series A, 44, 829 – 836.
3. Holland, P.W. & Rubin, D.B. (1983). On Lord's paradox. In: Wainer, H. and Messick, S. (Eds.) Principles of modern psychological measurement, 3 – 25. Hillsdale, New Jersey, London.
4. Kendall, M.G. & Stuart, A. (1961). The advanced theory of statistics. Volume 2: Inference and relationship. London: Charles Griffin & Company Limited.
5. Krauth, J. (1993). Einführung in die Konfigurationsfrequenzanalyse (KFA). Ein multivariates nichtparametrisches Verfahren zum Nachweis und zur Interpretation von Typen und Syndromen. Weinheim/Basel: Beltz, Psychologie-Verlags-Union.
6. Lancaster, H.O. (1969). Contingency tables of higher dimensions. Bulletin of the International Statistical Institute, 43,1, 143 – 154.
7. Lienert, G.A. (1969). Die "Konfigurationsfrequenzanalyse" als Klassifikationsmethode in der klinischen Psychologie. In: Irle, M. (Ed.) Bericht über den 26. Kongreß der Deutschen Gesellschaft für Psychologie, Tübingen 16.9. – 19.9.1968, 244 – 253. Göttingen: Hogrefe.

8. Lienert, G.A. (1971a). Die Konfigurationsfrequenzanalyse. I. Ein neuer Weg zu Typen und Syndromen. Zeitschrift für Klinische Psychologie und Psychotherapie, 19, 99 – 115.
9. Lienert, G.A. (1971b). Die Konfigurationsfrequenzanalyse. II. Hierarchische und agglutinierende KFA in der klinischen Psychologie. Zeitschrift für Klinische Psychologie und Psychotherapie, 19, 207 – 220.
10. Lienert, G.A. & Krauth, J. (1973a). Die Konfigurationsfrequenzanalyse. VI. Profiländerungen und Symptomverschiebungen. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 100 – 109.
11. Lienert, G.A. & Krauth, J. (1973b). Die Konfigurationsfrequenzanalyse. VII. Konstellations-, Konstellationsänderungs- und Profilkonstellationstypen. Zeitschrift für Klinische Psychologie und Psychotherapie, 21, 197 – 209.
12. Lord, F.M. (1967). A paradox in the interpretation of group comparisons. Psychological Bulletin, 68, 304 – 305.
13. Lord, F.M. (1969). Statistical adjustments when comparing preexisting groups. Psychological Bulletin, 72, 336 – 337.
14. Lord, F.M. (1976). Lord's paradox. In: Anderson, S.B., Ball, S., Murphy, R.T. and Associates (Eds.) Encyclopedia of Educational Evaluation, 232 – 236. San Francisco, Washington, London: Jossey-Bass Publishers.
15. Meehl, P.E. (1950). Configural scoring. Journal of Consulting Psychology, 14, 165 – 171.
16. Olmstead, P.S. & Tukey, J.W. (1947). A corner test of association. Annals of Mathematical Statistics, 18, 495 – 513.
17. Simpson, E.H. (1951). The interpretation of interaction in contingency tables. Journal of the Royal Statistical Society, Series B, 13, 238 – 241.
18. Wainer, H. (1986). Minority contributions to the SAT score turnaround: An example of Simpson's paradox. Journal of Educational Statistics, 11, 239 – 244.
19. Wainer, H. (1991). Adjusting for differential base rates: Lord's paradox again. Psychological Bulletin, 109, 147 – 151.
20. Wainer, H. & Brown, L.M. (2004). Two statistical paradoxes in the interpretation of group differences. Illustrated with Medical School Admission and Licensing data. American Statistician, 58, 107 – 123.
21. Yule, G.U. (1903). Notes on the theory of association of attributes of statistics. Biometrika, 2, 121 – 134.