Dipartimento di Statistica
"Giuseppe Parenti"

Independence in multi-way contingency tables: S.N. Roy's breakthroughs and later developments

Alan Agresti, Anna Gottard

Università degli Studi
di Firenze

Independence in Multi-Way Contingency Tables:
S. N. Roy's Breakthroughs and Later Developments

Alan Agresti[1] and Anna Gottard[2]

[1]Department of Statistics, University of Florida, Gainesville, Florida 32611, U.S.A

e-mail AA@STAT.UFL.EDU

phone (352) 392-1941, FAX (352) 392-5175

[2]Department of Statistics, University of Florence, 50134 Firenze, ITALY

e-mail GOTTARD@DS.UNIFI.IT

phone (Italy) 055-4237-267, FAX 055-4223-560

Independence in Multi-Way Contingency Tables: S. N. Roy's Breakthroughs and Later Developments

Abstract

In the mid 1950s S. N. Roy and his students contributed two landmark articles to the contingency table literature (Roy and Kastenbaum 1956, Roy and Mitra 1956). The first article generalized concepts of interaction from $2\times2\times2$ contingency tables to three-way tables of arbitrary size and to larger tables. In the second article, which is the source of our primary focus, various notions of independence were clarified for three-way contingency tables, Roy's union-intersection test was applied to construct chi-squared tests of hypotheses about the structure of such tables, and the chi-squared statistics were shown not to depend on the distinction between response and explanatory variables. This work pre-dates by many years later developments that expressed such results in the context of loglinear models. It pre-dates by a quarter century the development of graphical models. We summarize the main results in these key articles and discuss the connection between them and the later developments of loglinear modeling and of graphical modeling. We also mention ways in which these later developments have themselves been further generalized.

Keywords: Chi-squared tests; Conditional independence; Graphical models; Interaction; Loglinear models; Marginal models; Odds ratio; Union–intersection test.

1 Introduction

Through the mid 1950s, the literature on contingency table analysis had focused almost entirely on two-way tables. Then, in 1956 two revolutionary articles were published by S. N. Roy with two of his Ph.D. students – Marvin Kastenbaum and S. K. Mitra. These articles dealt mainly with the structure and analysis of three-way contingency tables. Let (X, Y, Z) denote the three categorical variables.

Bartlett (1935) had defined *no interaction* in a 2×2×2 table as meaning that the odds ratio between two of the variables is identical for the two strata of the third variable. This property is symmetric in the choice of variables, with XY common odds ratios being equivalent to XZ common odds ratios and to YZ common odds ratios. Roy and Kastenbaum (1956) generalized this definition of interaction to $r \times s \times t$ tables and presented chi-squared tests of the hypothesis of no interaction.

Roy and Mitra (1956) focused more generally on contingency table structure – not just for a lack of interaction but for the various ways that independence for a two-way table extends to three-way tables. In addition, this article exploited Roy's union-intersection principle to provide a heuristic justification for chi-squared goodness-of-fit tests for the various structures. It was not until the 1960s that the implications of this work were fully explored in the context of loglinear modeling of contingency tables. The development of graphical models starting with Darroch, Lauritzen, and Speed (1980) also put this paper and its key results in a broader context.

In this article, Section 2 surveys the most important results from Roy and Mitra (1956) and Roy and Kastenbaum (1956). We show their connections with these later developments of loglinear modeling in Section 3 and graphical modeling in Section 4. In Sections 5 and 6 we also mention ways in which these later developments have themselves been further generalized in work that, at least indirectly, has its genesis in

these papers by Roy and his students.

2 Summary of the Roy–Mitra–Kastenbaum Articles

In their introductory paragraph, Roy and Mitra stated their opinion that as of the time of their article (1956), the landmark articles in contingency table analysis were Pearson's introduction of the chi-squared test (in 1900), Fisher's correction of that test giving the proper degrees of freedom (in 1922), Cramér's classic book on mathematical statistics which derived the asymptotic distribution of the Pearson chi-squared statistic in general parametric cases (see Cramér (1946), pp. 424-434), Neyman's fundamental paper (see Neyman 1949) introducing the class of best asymptotic normal estimators and its attendant discussion of minimum chi-squared statistics (also discussed by Cramér), the discussion of association in Yule and Kendall's landmark book (see Yule and Kendall 1950), and articles in 1947 by Barnard and Pearson that applied unconditional approaches to inference. With few exceptions, nearly all methodological discussion of contingency table analysis up to that date focused on two-way tables.

As an alternative to R. A. Fisher's famous exact conditional test of independence for 2×2 tables, Barnard (1945) had proposed an unconditional approach for a small-sample analysis comparing two binomial parameters. Barnard's approach was strongly criticized by Fisher (in a subsequent letter to the editor of *Nature*) and not revived for about 35 years. The Pearson (1947) article, while not itself groundbreaking, gave support to the viewpoint that Fisher's conditional analysis was not always optimal or even appropriate. Roy and Mitra expressed agreement with this view. They emphasized throughout their article that the formulation of the probability distribution assumed for the contingency table should reflect the identification of the response variables (which they referred to as the 'variates') and the explanatory variables (which

3

they referred to as the 'ways of classification').

As of 1956, there had been little in the way of systematic presentation of the types of association structure that could exist in a three-way contingency table. Among the exceptions were Bartlett's (1935) article defining no interaction in 2×2×2 tables, Kendall's (1945) explanation for such tables that conditional independence does not imply marginal independence (and the related discussion about analyzing partial association in three-way tables in Chapter 2 of the 14th and final edition of Yule and Kendall (1950)), Simpson's (1951) study about the dangers of collapsibility when there is no interaction (which, like Yule and Kendall, also mentioned briefly the generalization from independence in a two-way table to mutual independence in a three-way table), and Lancaster's (1951) attempt to partition the Pearson chi-squared statistic for testing mutual independence in three-way tables. These articles all focused on 2×2×2 tables. Although Lancaster mentioned the possibility of generalization to larger tables, it is amusing in retrospect to read his comment that "Doubtless little use will ever be made of more than a three-dimensional classification."

2.1 Sampling schemes

Using the notation from Roy and Kastenbaum (1956) and Roy and Mitra (1956), let $\{n_{ijk}\}$ denote cell counts in the $r \times s \times t$ cross-classification, with total sample size n. As Lancaster (1951) had done, Roy and Mitra considered three sampling schemes for a three-way contingency table:

1. A single multinomial sample over the entire table. This case treats the overall sample size n alone as fixed.

2. Multinomial sampling over the $r \times s$ combinations of categories of (X, Y) at a fixed category of Z, with independent samples at the t categories of Z. This case treats $\{n_{++k}\}$ as fixed (a + subscript denotes summation over that index).

3. Multinomial sampling over the s categories of Y, with independent samples at the $r \times t$ categories of (X, Z). This case treats $\{n_{i+k}\}$ as fixed.

We will summarize the Roy – Mitra results for these three cases. We let $\{p_{ijk}\}$ denote the cell probabilities. In case (1) $\sum_i \sum_j \sum_k p_{ijk} = 1$. In case (2), $\sum_i \sum_j p_{ijk} = 1$ for each fixed k. In case (3), $\sum_j p_{ijk} = 1$ for each fixed combination of i and k.

2.2 A single multinomial sample

For a single multinomial sample, Roy and Mitra considered the hypotheses:

- X and Y are *conditionally independent*, given Z. That is,
  $H_0 : \frac{p_{ijk}}{p_{++k}} \frac{p_{i+k}}{p_{++k}} \frac{p_{+jk}}{p_{++k}}$ for all i, j, and k.

- X and Y are *jointly independent* of Z. That is,
  $H_0 : p_{ijk} = p_{ij+}p_{++k}$ for all i, j, and k.

Roy and Mitra showed that:

- If there is conditional independence between X and Y and if also X and Z are marginally independent and Y and Z are marginally independent, then X, Y, and Z are *mutually independent*; that is,

$$p_{ijk} = p_{i++}p_{+j+}p_{++k}$$

  for all i, j, and k.

- If X and Z are marginally independent and if Y and Z are marginally independent, then if there is no interaction (as defined in the Bartlett and Roy–Kastenbaum sense of equality of odds ratios between two variables at each category of the third variable, as developed below), then X and Y are jointly independent of Z.

Roy and Kastenbaum also noted that no interaction is the condition which, when added to marginal independence of X and Y and marginal independence of Y and Z, implies joint independence of Y from X and Z. They observed that for a multinomial sample with only the overall sample size n fixed, this condition of no interaction means that the cell probabilities $\{p_{ijk}\}$ have the form

$$p_{ijk} = a_{ij}b_{ik}c_{jk},$$

but there is no closed-form expression in terms of the two-way marginal probabilities of $\{p_{ijk}\}$.

For an $r \times s \times t$ table, Roy and Kastenbaum showed that the hypothesis of no interaction corresponds to $(r-1)(s-1)(t-1)$ constraint equations involving odds ratios,

$$\frac{p_{rst}p_{ijt}}{p_{ist}p_{rjt}} = \frac{p_{rsk}p_{ijk}}{p_{rsk}p_{rjk}} \quad \text{for} \quad i = 1, ..., (r-1), j = 1, ..., (s-1), k = 1, ..., (t-1).$$

Their concluding remarks explained how to generalize the concept of no interaction to higher dimensions and pointed out that the no interaction constraint is itself a set of contrasts in the logarithms of the probabilities. This is an important article that previews what became the natural definition of no interaction implied by loglinear models and logistic regression models with categorical predictors.

2.3   Multinomial sampling with Z fixed or with X and Z fixed

When Z is fixed, Roy and Mitra considered the hypothesis that X and Y are independent at each level of Z. This is equivalent in structure to the hypothesis of conditional independence mentioned in the previous subsection. They noted that if X and Y are conditionally independent and if Y and Z are marginally independent, then X and Y are also marginally independent. This result relates to later work on conditions

for *collapsibility* of a contingency table in terms of identical conditional and marginal associations (e.g., Shapiro 1982).

When X and Z are fixed and Y is a single response variable, Roy and Mitra discussed how hypotheses of interest have analogs in the analysis of variance for a *quantitative* response variable with two fixed categorical factors. For example, a relevant hypothesis is that for a fixed value of one factor, for any particular response outcome the probability is the same at each category of the other factor.

2.4   Union–Intersection Tests

Roy (1953) had introduced the union–intersection principle for conducting hypothesis tests, described briefly as follows: Suppose the null hypothesis of interest can be expressed as an intersection of several component hypotheses; then, the rejection region for the union–intersection test is the union of the rejection regions for the tests for the component hypotheses. Roy and Mitra (1956) showed that application of this principle to testing the simple null hypothesis that a set of multinomial probabilities equal particular fixed values leads to a test statistic that is the Pearson chi-squared statistic for this hypothesis.

For testing a composite null hypothesis, such as one of the independence conditions or the no interaction condition specified above, they showed that the union–intersection principle yields the minimum chi-squared statistic as the test statistic. That test statistic has Pearson chi-squared form

$$\sum (n_{ijk} - n\hat{p}_{ijk})^2 / (n\hat{p}_{ijk}),$$

but the parameter estimates $\{\hat{p}_{ijk}\}$ for it are those based on minimizing this Pearson statistic instead of the usual maximum likelihood (ML) estimates. From Neyman (1949), the minimum chi-squared estimates share with ML estimates the best asymp-

totic normal property, and Roy and Mitra noted that for computational simplicity one could instead insert the ML estimates in the test statistic. They applied this result to several cases for two-way and three-way tables. They noted that for a particular hypothesis, the same Pearson test statistic and the same asymptotic chi-squared distribution results regardless of the multinomial sampling scheme (and thus, regardless of the distinction between response and explanatory variables).

The asymptotic chi-squared distribution has degrees of freedom equal to the number of multinomial parameters minus the number of parameters estimated from the data. In particular, they showed that $df = rst - r - s - t + 2$ for testing mutual independence, $df = t(r - 1)(s - 1)$ for testing conditional independence between X and Y (given Z), and $df = (rt - 1)(s - 1)$ for testing that Y is jointly independent of X and Z.

Finally, they noted that when a hypothesis with an associated chi-squared statistic is an intersection of several hypotheses with their own chi-squared statistics, then the summary chi-squared statistic partitions into a sum of the component chi-squared statistics asymptotically. (This was later noted to happen *exactly* for any n when using the likelihood-ratio statistic instead of the Pearson statistic.) For example, they noted that the hypothesis of mutual independence of X, Y, and Z is the intersection of (1) conditional independence between X and Y given Z, (2) X and Z marginal independence, and (3) Y and Z marginal independence. Under the null hypothesis of mutual independence, they noted that the three component Pearson statistics are asymptotically independent and converge in probability to the overall chi-squared statistic.

3   Related Connections with Loglinear Models

The results just described from Roy and Mitra (1956) and Roy and Kastenbaum (1956) relate to the extensive literature that evolved in the 1960s on *loglinear models*. Let

$\{\mu_{ijk}\}$ denote expected frequencies in the cells of a three-way contingency table. For instance, for a single multinomial sample over the entire table, $\{\mu_{ijk} = np_{ijk}\}$, but these means could also refer to independent Poisson observations or to other sorts of multinomial sampling. See Bishop et al. (1975) and Agresti (2002) for introductions to loglinear models.

3.1 Independence structure expressed as a loglinear model

The hypothesis of no interaction can be expressed as the loglinear model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

The hypothesis of conditional independence between X and Y, given Z, can be expressed as the loglinear model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}.$$

The hypothesis of X and Y being jointly independent of Z is the loglinear model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}.$$

The hypothesis of mutual independence among X, Y, and Z is the loglinear model

$$\log \mu_{ijk} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z.$$

Below we will use the common shorthand for loglinear models that specifies the marginal distributions that are the minimal sufficient statistics for the models. The four models just stated are denoted (XY, XZ, YZ), (XZ, YZ), (XY, Z), and (X, Y, Z).

In terms of such notation, the Roy and Mitra results stated in the previous section for a single multinomial sample were:

9

- Suppose the loglinear model (XZ, YZ) holds and also (X, Z) holds (that is, X and Z are marginally independent) and (Y, Z) holds. Then, the loglinear model (X, Y, Z) holds.

- Suppose loglinear models (X, Z) and (Y, Z) hold and suppose there is no interaction (that is, loglinear model (XY, XZ, YZ) holds). Then, loglinear model (XY, Z) holds.

For related results on collapsibility for loglinear models, see Agresti (2002, pp. 358-360, 398). In the 1960s the work of Roy and Kastenbaum and of Roy and Mitra was extended by many. Important contributions included those from L. Goodman, J. Darroch, I. J. Good, M. Birch, N. Mantel, and R. Plackett (see, for instance, the summary in Chapter 16 of Agresti 2002). In particular, Birch (1963) noted how various margins of a contingency table are sufficient statistics for loglinear models, the likelihood equations equate those margins to their expected values, ML inference is the same for Poisson sampling and the various types of multinomial sampling, and the ML estimates are unique.

3.2 Chi-squared statistics

As mentioned above, for testing hypotheses about various types of independence in three-way tables, Roy and Mitra showed that the union–intersection principle yields the minimum chi-squared test statistic. For about 25 years following the publication of their paper, the minimum chi-squared approach and the related but simpler minimum modified chi-squared approach in which the cell probability estimates are based on minimizing

$$\sum (n_{ijk} - n\hat{p}_{ijk})^2 / (n_{ijk})$$

received considerable attention. For example, when the estimating equations for the minimum modified chi-squared statistic are linear, Bhapkar (1966) showed that the estimators are exactly those obtained using weighted least squares (WLS). Interestingly, Bhapkar was himself one of Roy's students. Soon after Bhapkar's article, with the publication of Grizzle, Starmer and Koch (1969) WLS became a popular competitor for ML for fitting models to contingency tables, because of its computational simplicity.

Minimum chi-squared statistics are now much less commonly used for contingency table analysis. However, the connection between the minimum modified chi-squared statistic and the Pearson chi-squared statistic using ML estimates was strengthened by the observation by Cressie and Read (1984) that these are both special cases of the power divergence statistic. For a three-way table, this is

$$\frac{2}{\lambda(\lambda+1)} \sum n_{ijk} \Big[(n_{ijk}/n\hat{p}_{ijk})^{\lambda} - 1 \Big].$$

The Pearson statistic results when $\lambda = 1$ and the modified chi-squared statistic results when $\lambda = -2$.

3.3   The sampling model with all margins fixed

In a two-way table, when both margins are fixed, under independence the relevant sampling distribution is the hypergeometric. Fisher's exact test is a well-known inference that uses this structure. Roy and Mitra briefly described the fixed margins case for two-way contingency tables. They noted that this case is less common in practice, and they did not consider it at all for three-way tables.

As of 1956, the hypergeometric model seems only to have been considered for 2×2 tables, by Fisher for his exact test in 1935 and by Cornfield (1956) for interval estimation of the odds ratio in case-control studies. Roy and Mitra argued that a disadvantage of this setting is the lack of an obvious distribution to use under the alternative hypoth-

esis. However, Cornfield's result about the relevance of the odds ratio to retrospective studies and his use of a noncentral hypergeometric distribution to obtain a confidence interval for the odds ratio showed that this case has much greater scope than believed prior to 1956. Although the fixed-margins sampling scheme is less common in practice, the hypergeometric and related sampling models received more attention starting in the 1980s as attention focused on exact, small-sample inference. Using the Fisherian conditional approach, unknown nuisance parameters are eliminated by conditioning on their sufficient statistics.

For example, consider the loglinear model (XZ, YZ) of conditional independence between X and Y, given Z. Suppose we would like to construct a small-sample test of the null hypothesis that this model holds. To obtain a conditional distribution free of the unknown nuisance parameters $\{\lambda_{ik}^{XZ}\}$ and $\{\lambda_{jk}^{YZ}\}$, we condition on their sufficient statistics $\{n_{i+k}\}$ and $\{n_{+jk}\}$. The conditional probability mass function of $\{n_{ijk}\}$ corresponds to a product of hypergeometric probability mass functions relating X and Y at the various strata of Z. The computations for the conditional approach for hypotheses in contingency tables or more generally for inference about parameters in logistic regression models have been developed in several articles over the past 20 years by Cyrus Mehta and Nitin Patel. See Agresti (1992) for a survey of the primary small-sample methods.

Likewise, in considering interaction, Roy and Kastenbaum (1956) thought if natural to have only the overall sample size n fixed. They concluded their article by stating their belief that the structure of no interaction is not especially meaningful if any of the margins are fixed. This was not borne out by later events. For example, this structure of no interaction is the standard one for logistic regression models containing only main effect terms. Yet, for such models it is standard to treat only Y as random and to treat the cross classification of the explanatory variables as fixed counts.

12

4 Related Connections with Graphical Models

About 25 years ago and yet 25 years after Roy and Mitra (1956), conditional independence began to receive special attention as its own focus of research, in connection with a class of probabilistic models called *graphical Markov models*. This is a class of multivariate statistical models for which the conditional independence structure of the joint distribution can be read directly off a graph. In this context, the graph $\mathcal{G}$ is a pair (V, E), where V is a finite and nonempty set of *nodes* and E is a set of ordered pairs of distinct nodes, representing *edges* between nodes. Edges can be *undirected* or *directed* (with arrows). Two nodes $u, v \in V$ connected by an undirected edge are called *neighbors*. If there is an arrow pointing from u to v, then u is called a *parent* of v and v is called a *child* of u. A graph is said to be *complete* if all the nodes are connected.

In a (conditional) *independence graph*, each node represents a univariate random variable, while the absence of an edge between two nodes represents a type of conditional independence. A set of properties, called *Markov properties*, provides the criteria to read off conditional independence restrictions from a graph. The importance of independence graphs relates to their ability of providing with a graph an intuitive visual display of complex association structures in terms of conditional independence. For extended presentations of various types of independence graphs, see Lauritzen (1996) or Cox and Wermuth (1996).

4.1 Graphical models for a single multinomial sample

As mentioned before, the first attempt to give a graphical representation of the pattern of association in a contingency table is due to Darroch *et al.* (1980). They defined *graphical models* as a subclass of hierarchical loglinear models that can be interpreted in terms of conditional independence using an *undirected graph*, which is a graph having

only undirected edges. The graph associated with a hierarchical loglinear model has as many nodes as the dimension of the contingency table: Two nodes are not connected whenever in the hierarchical loglinear model their two-factor interaction is absent.

We illustrate with the first sampling scheme discussed by Roy and Mitra (see also Section 2.2) in which the three categorical random variables X, Y and Z are all response variables with a single multinomial sample. Undirected graphs having three nodes can represent this case. Such undirected graphs are appropriate when we wish to treat the variables on an equal footing (all as response variables), with interest in modeling the associations among them.
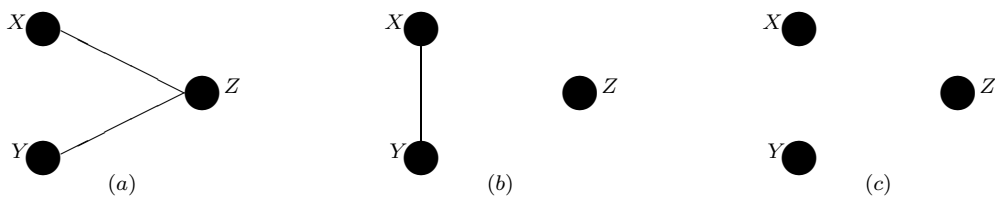

Figure 1: *Examples of independence graphs involving three categorical response variables.*

Let's consider now the independence hypotheses proposed by Roy and Mitra and their independence graphs for this case in which each variable is a response. First, suppose X and Y are conditionally independent, given Z, that is, using Dawid's notation (Dawid, 1979), $X \perp\!\!\!\perp Y \mid Z$. In this case, X and Y are not neighbors and their nodes are separated by Z, as shown in Figure 1(a). As Roy and Mitra noted, this hypothesis is the analogue of no partial XY correlation in a three-variate normal population, which has a similar graph. Using a loglinear model specification, this hypothesis corresponds to model (XZ, YZ) reported in Section 3.1, that is, to the absence of terms λ_{ijk}^{XYZ} and λ_{ij}^{XY}.

In the Darroch *et al.* (1980) graphical loglinear models, the maximal interaction parameters of the hierarchical loglinear model correspond to the *cliques* of the graph.

A clique is a subset of nodes forming a maximal complete subgraph; that is, it forms an undirected subgraph such that all the nodes are neighbors and such that with the inclusion of an additional node the resulting subgraph would no longer be complete. The graph in Figure 1(a) consists of two cliques, $\{X, Z\}$ and $\{Y, Z\}$, because X and Y are not neighbors. Therefore, the graphical loglinear model is (XZ, YZ).

Roy and Mitra next considered the hypothesis that X and Y are jointly independent of Z, denoted by $(X, Y) \perp\!\!\!\perp Z$. Figure 1(b) shows the corresponding undirected graph, where both X and Y are not connected to Z. The cliques of this graph are $\{X, Y\}$ and $\{Z\}$. Therefore, in the graphical loglinear specification denoted by (XY, Z), the maximal interaction term is λ_{ij}^{XY}. The graph suggests also that $X \perp\!\!\!\perp Z \mid Y$ and $Y \perp\!\!\!\perp Z \mid X$.

The independence graph in Figure 1(c) consists of all singletons, implying conditional independence for each pair of variables, given the third one. Equivalently, there is mutual independence among the variables. This was the third and last hypothesis considered in this scheme by Roy and Mitra. This graph has cliques $\{X\}$, $\{Y\}$ and $\{Z\}$, and the corresponding loglinear model denoted by (X, Y, Z) has no interaction terms.

Figure 1 does not show the graph with an edge between each pair of variables. Its single clique is $\{X, Y, Z\}$. So, it corresponds to the loglinear model with maximal interaction parameter λ_{ijk}^{XYZ}, which is the saturated model for a three-way table. That is, whenever a loglinear model contains all the pairwise associations, only the case with the highest-order interaction present is a graphical model. The model (XY, XZ, YZ) of no interaction that was the basis of the Roy and Kastenbaum paper is not a graphical model.

4.2 Graphical models for multinomial sampling with Z fixed or with X and Z fixed

Roy and Mitra emphasized taking into account the sampling design, and hence distinguishing between response and explanatory variables. Undirected graphs cannot make this distinction. Other types of graphs have to be used that allow edges to be directed. Graphs admitting only directed edges and containing no directed cycles are called *directed acyclic graphs*. They treat variables in an asymmetric way, so that, for example, an arrow from X to Y means that X is explanatory to Y as a response. More general graphs admitting both directed and undirected nodes are called *chain graphs* (Lauritzen and Wermuth, 1989). They can incorporate symmetric and asymmetric dependence. Both undirected graphs and directed acyclic graphs are special cases of chain graphs.

In chain graphs, nodes can be partitioned into an ordered sequence of subsets, called *blocks*, so that the variables within a block are classified as pure explanatory, intermediate, or pure response variables. Within each block nodes can be connected only by undirected edges, forming an undirected subgraph, while nodes in different blocks can be connected only by arrows. The variables in a block from which arrows emanate are considered explanatory of the ones in the successive blocks.

Consider now the case in which X and Y are response variables, while Z is explanatory. The appropriate graph in this case is a *joint-response graph*, which is a type of chain graph. The nodes are partitioned into two blocks: The block on the right contains the node Z as a pure explanatory variable and the block on the left contains Y and X as a joint response. Figure 2(a) illustrates a joint-response graph, showing the case of no conditional independence for any pair of the variables. The graph is complete.
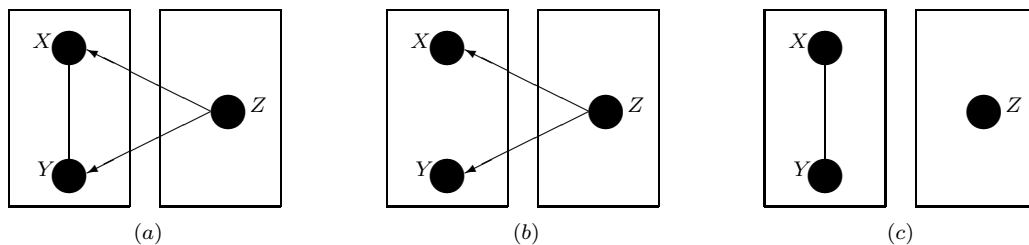
Figure 2: *Examples of independence graphs for two response variables (X and Y) and an explanatory variable (Z).*

Figure 2(b) portrays the Roy and Mitra hypothesis stating that the response variables are conditionally independent given the explanatory variable Z. In this case the chain graph reduces to a directed acyclic graph. This graph is Markov equivalent to Figure 1(a); that is, it encodes the same conditional independence statements. Figure 2(c) depicts the Roy and Mitra hypothesis stating that X and Y are jointly independent of Z. The graph is equivalent to Figure 1(b), since if $(X, Y) \perp\!\!\!\perp Z$ then Z is not a useful explanatory variable for X and Y in that it explains nothing about them.

Suppose next that Y is the sole response variable and X and Z are explanatory variables. In the usual modeling, one would not assume structure between the explanatory variables X and Z, regarding them as fixed rather than random by virtue of the response–explanatory distinction or the sampling design. Likewise, Roy and Mitra did not specify the relationship between X and Z, focusing instead on the conditional distribution of Y given the explanatory variables.
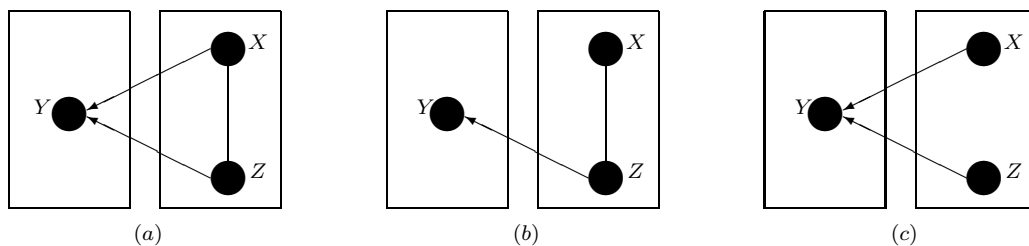

Figure 3: *Examples of independence graphs for one response variable (Y) and two explanatory variables (X, Z).*

With Y as the sole response, Figure 3(a) presents a graph without any independence

17

structure. Figure 3(b) depicts the graph of conditional independence of X and Y given Z, which is illustrated in the graph by the missing arrow from X to Y. This graph is Markov equivalent to those in Figure 1(a) and Figure 2(b). As mentioned above, the independence graph in Figure 3(c) would usually be of less interest. However, we mention in passing that for such graphs, variables (such as X and Z here) that are unconnected parents of a common child are *marginally* independent but conditionally dependent given the child. (For explanation, see the discussion of chain graph Markov properties in Lauritzen, 1996, Section 3.2.3.)

5 Other Types of Independence Graphical Models

Not all types of independence structures can be captured by undirected graphs, directed graphs, or chain graphs. As a consequence, new types of graphs have been proposed in recent years to encode different independence structures. For example, one may be interested in modeling *marginal* independence between pairs of variables. In this framework, Cox and Wermuth (1993) defined a new class of graphs, called *covariance graphs*, in which the presence of a dashed edge denotes a marginal dependence in a set of jointly Gaussian variables. In the same direction, as an alternative to graphical loglinear models, Drton and Richardson (2005) defined a class of graphical models for binary variables whose marginal independence structures can be read off a graph. Luparelli and Marchetti (2005) generalized their result by showing how such graphs apply to marginal loglinear models for which the log link function applies to marginal probabilities.

Another type of independence that has been recently investigated for categorical variables is conditional independence for specific values of a given variable: $X \perp\!\!\!\perp Y \mid Z = k^*$. In this context, Fienberg and Kim (1999) explored the possibility of using a

class of graphical loglinear models for combining conditional loglinear structures. The approach of considering conditional independence at specific values of a given variable has also been discussed by Højsgaard (2003, 2004). He introduced a new type of graph, called a *split graph*, for contingency tables. A split graph consists of a collection of graphs arranged in a hierarchy displaying the different association structures for the conditional joint distributions at each value of the given variable.

Finally, Bayesian approaches have also been used with graphical models. In a simple nonhierarchical approach, the prior distribution for the joint probabilities factors into prior distributions for marginal and conditional probabilities. With independent Dirichlet form for those distributions, one obtains independent Dirichlet posterior distributions as well. Dawid and Lauritzen (1993) introduced the notion of a probability distribution defined over probability measures on a multivariate space that concentrate on a set of such graphs. A special case includes a *hyper Dirichlet* distribution that is conjugate for multinomial sampling and that implies that certain marginal probabilities have a Dirichlet distribution.

6 Marginal Homogeneity: A Type of Independence for Marginal Distributions

The Roy and Mitra paper focused on independence structure when there are independent observations on a set of variables. With the increased frequency of longitudinal studies in practice and consequent *dependent* observations, another type of independence that has increasingly received attention is independence between a response and the time of measurement. In the contingency table literature, this is often referred to as *marginal homogeneity*, because it corresponds to a joint distribution for a multivariate response having identical marginal distributions.

Relevant literature here includes McNemar (1947) for matched pairs with binary data, Stuart (1955), Madansky (1963), Bhapkar (1966), and Caussinus (1966) for square contingency tables with a multiple-category response, Agresti (2002, pp. 420-421, 429-431, and Exercises 10.37, 10.38, and 12.35) for various approaches for ordered categories, and Cochran (1950), Bhapkar (1973), and Darroch (1981) for multi-way tables. See Bishop et al. (1975, Ch. 8) and Agresti (2002, Chapters 10 and 11) for surveys of such methods.

In the spirit of Roy's multivariate researches, one could extend formulations of marginal homogeneity and marginal inhomogeneity to multivariate response data when a *vector* of categorical response variables is observed repeatedly. For example, in analyzing safety in clinical trials for a new drug, it is common to measure the presence or absence of a large number of adverse side effects. In crossover trials, the same subjects are observed under two or more doses of the drug. For two doses with c potential adverse events (AEs), the data then take the form of paired multivariate binary data, with c binary variables measured for the subjects under two conditions.

Consider paired multivariate measurement for n subjects of c binary response variables. For observation i of variable j on a subject, let $y_{ij} = 1$ for a "success" and $y_{ij} = 0$ for a "failure," $j = 1, \ldots, c$, $i = 1, 2$. Let $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)'(y_{11}, \ldots, y_{1c}, y_{21}, \ldots, y_{2c})'$ denote the $2c$-dimensional binary responses for a randomly selected subject. A 2^{2c} contingency table summarizes all possible outcomes for $\boldsymbol{y}$. One could assume a multinomial distribution for the counts in this 2^{2c} table. Let $\pi_i(j) = P(y_{ij} = 1)$ denote a first-order marginal probability. The hypothesis

$$H_0 : \pi_1(j) = \pi_2(j), \ j = 1, 2, \ldots, c.$$

corresponds to *simultaneous marginal homogeneity* (SMH) in terms of the c binary variables. This is a special case of a generalized loglinear model of form $\mathbf{C} \log \mathbf{A} \boldsymbol{\pi} =$

$\mathbf{X}\boldsymbol{\beta}$, where $\mathbf{A}$ is a binary matrix that sums the joint probabilities to give the first-order marginal probabilities.

Although the hypothesis is quite simplistic, even in this modern era of computers when c is large it is computationally difficult to maximize the likelihood subject to this constraint, for example in order to conduct a likelihood-ratio (LR) test. One must maximize a multinomial likelihood with $2^{2c} - 1$ joint probabilities, subject to equality constraints relating two sets of c marginal probabilities. A corresponding Pearson statistic compares the 2^{2c} observed and fitted counts for the SMH model, using the usual $X^2 = \sum$ (observed - fitted)^2/fitted. This is the score statistic for testing SMH. Like the LR test, it is computationally infeasible with current software when c is large.

For computationally simpler alternatives, in ongoing work Klingenberg and Agresti (2005) considered Wald and score-type tests for SMH. For the sample proportions, let $\boldsymbol{d} = (d_1, \ldots, d_c)'$, where $d_j = \hat{\pi}_1(j) - \hat{\pi}_2(j)$. The covariance matrix Σ of $\boldsymbol{d}$ is

$$\text{Var}(d_j) = \left[\pi_1(j) + \pi_2(j) - 2\pi(j,j) - \{\pi_1(j) - \pi_2(j)\}^2\right]\Big/n$$

$$\text{Cov}(d_j, d_k) = \left[\pi_1(j,k) + \pi_2(j,k) - \{\pi(j,k) + \pi(k,j)\}\right.$$
$$\left. -\{\pi_1(j) - \pi_2(j)\}\{\pi_1(k) - \pi_2(k)\}\right]\Big/n,$$

where $\pi_i(j,k) = P(y_{ij} = 1, y_{ik} = 1)$ and $\pi(j,k) = P(y_{1j}1, y_{2k} = 1)$. The statistic $W = \boldsymbol{d}'\hat{\Sigma}^{-1}\boldsymbol{d}$ with sample proportions in this matrix is a Wald statistic for testing SMH, having asymptotic chi-squared distribution with $df = c$.

An alternative statistic uses the pooled estimate of the covariance matrix under the null hypothesis of SMH. The pooled estimate for the common proportion $\pi_0(j)$ in each dose group is

$$\hat{\pi}_0(j) = \{\hat{\pi}_1(j) + \hat{\pi}_2(j)\}/2.$$

The variance of d_j and the covariance of d_j and d_k then simplify to

$$\text{Var}_0(d_j) = \left[\pi_1(j) + \pi_2(j) - 2\pi(j,j)\right]\Big/n = 2\{\pi_0(j) - \pi(j,j)\}/n$$

21

$$\text{Cov}_0(d_j, d_k) \;\; = \;\; \Big[\pi_1(j,k) + \pi_2(j,k) - \{\pi(j,k) + \pi(k,j)\}\Big]\Big/n.$$

Klingenberg and Agresti (2005) showed that the test statistic $W_0 d'\hat{\Sigma}_0^{-1}d$ converges much more quickly to the asymptotic chi-squared distribution that W does. Also, they showed that the two estimates of the covariance matrix are linked through $\hat{\Sigma} = \hat{\Sigma}_0 - dd'/n$, from which it can be shown that

$$W = W_0\Big/(1 - W_0/n).$$

Ireland et al. (1969) showed the same type of result for paired multicategorical responses in the univariate case.

Analogous hypotheses occur for the simultaneous comparisons of marginal distributions for multivariate responses with *independent* samples. See Agresti and Klingenberg (2005). For the independent and dependent samples cases, their articles noted that it is inappropriate to rely on large-sample chi-squared distributions when n is small or c is large or when some of the true marginal probabilities are near 0. For such cases, they recommended small-sample exact permutation tests.

7 Summary

The articles by Roy with Kastenbaum and with Mitra had major influences in increasing the attention focused on the analysis of multi-way contingency tables. Their work was followed by other seminal articles on this subject by students of Roy and by other students and faculty at the University of North Carolina following his death. Good examples of this work are the important contributions by Bhapkar (1966) and by Grizzle, Starmer, and Koch (1969). Directly or indirectly, such work also had a major impact on later developments in categorical data analysis, such as loglinear modeling and graphical models. As we've seen, these areas continue to be generalized today. We

in the statistical community owe a great debt to S. N. Roy for this aspect of his many contributions to the current body of statistical knowledge.

Acknowledgments

References

Agresti, A., 1992. A survey of exact inference for contingency tables. *Statist. Sci.* 7, 131-153.

Agresti, A., 2002. *Categorical Data Analysis, 2nd edition.* Wiley, New York.

Agresti, A., Klingenberg, B., 2005. Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *Applied Statistics*, in press.

Barnard, G. A., 1945. A new test for 2×2 tables. *Nature* 156, 177.

Bartlett, M. S., 1935. Contingency table interactions. *J. Roy. Statist. Soc.* Supplement 2, 248-252.

Bergsma, W. P., Rudas, T., 2002. Marginal models for categorical data. *Ann. Statist.* 30, 140-159.

Bhapkar, V. P., 1966. A note on the equivalence of two test criteria for hypotheses in categorical data. *J. Amer. Statist. Assoc.* 61, 228-235.

Bhapkar, V. P., 1973. On the comparison of proportions in matched samples. *Sankhya* A 35, 341-356.

Birch, M. W., 1963. Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc., Ser. B* 25, 220-233.

Bishop, Y. M. M., Fienberg, S.E., Holland, P.W., 1975. *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press.

Caussinus, H., 1966. Contribution à l'analyse statistique des tableaux de correlation. *Ann. Fac. Sci. Univ. Toulouse* 29, 77-182.

Cochran, W. G., 1950. The comparison of percentages in matched samples. *Biometrika* 37, 256-266.

Cornfield, J., 1956. A statistical problem arising from retrospective studies. In *Proc. 3rd Berkeley Symposium on Mathematics, Statistics and Probability*, ed. J. Neyman. 4, 135-148.

Cox, D. R., Wermuth, N., 1993. Linear dependencies represented by chain graphs. *Statistical Science* 8, 204-218.

Cox, D. R., Wermuth, N., 1996. *Multivariate Dependencies: Models, Analysis and Interpretation.* Chapman and Hall.

Cramér, H., 1946. *Mathematical Methods of Statistics.* Princeton: Princeton Univer-

sity Press.

Cressie, N., Read, T.R.C., 1984. Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc., Ser. B* 46, 440-464.

Darroch, J. N., 1981. The Mantel-Haenszel test and tests of marginal symmetry; fixed-effects and mixed models for a categorical response. *Intern. Statist. Rev.* 49, 285-307.

Darroch, J. N., Lauritzen, S. L., Speed, T. P., 1980. Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* 8, 522-539.

Dawid, A. P., 1979. Conditional independence in statistical theory (with discussion). *J. Roy. Statist. Soc., Series B* 41, 1-31.

Dawid, A. P., Lauritzen, S. L., 1993. Hyper Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* 21, 1272-1317.

Drton, M., Richardson, T. S., 2005. Binary models for marginal independencies. *Technical Report* 474, Department of Statistics, University of Washington.

Fienberg S. E., Kim S.-H., 1999. Combining conditional log-linear structures, *J. Amer. Statist. Assoc.*, 94, 229-239.

Grizzle, J. E., Starmer, C. F., Koch, G. G., 1969. Analysis of categorical data by linear models. *Biometrics*, 25, 489-504.

Højsgaard, S., 2004. Statistical inference in context specific interaction models for contingency tables. *Scand. J. Statist.* 31, 143-158.

Højsgaard, S., 2003. Split models for contingency tables. *Comput. Statist. Data Anal.* 42, 621-645.

Ireland, C. T., Ku, H. H., Kullback S., 1969. Symmetry and marginal homogeneity of an r × r contingency table. *J. Amer. Statist. Assoc.* 64, 1323-1341.

Kendall, M. G., 1945. *The Advanced Theory of Statistics*, vol. 1. Griffin.

Klingenberg, B., Agresti, A., 2005. Multivariate extensions of McNemar's test. Unpublished technical report, University of Florida Department of Statistics.

Lancaster, H. O., 1951. Complex contingency tables treated by partition of χ^2. *J. Roy. Statist. Soc., Ser. B* 13, 242-249.

Lauritzen, S. L., 1996. *Graphical Models.* Oxford University Press.

Lauritzen, S. L., Wermuth, N., 1989. Graphical models for the association between variables, some of which are qualitative and some quantitative. *Ann. Statist.* 17, 31-57.

Lupparelli, M., Marchetti, G.M., 2005. Graphical models of marginal independence for categorical variables. Submitted for publication.

Madansky, A., 1963. Tests of homogeneity for correlated samples. *J. Amer. Statist. Assoc.* 58, 97-119.

McNemar, Q., 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12, 153-157.

Neyman, J., 1949. Contributions to the theory of the χ^2 test. pp. 239-273 in *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman. Berkeley: Univ. California Press.

Pearson, E. S., 1947. The choice of a statistical test illustrated on the interpretation of data classified in 2×2 tables. *Biometrika* 34, 139-167.

Roy, S. N., 1953. On a heuristic method of test construction and its use in multivariate analysis. *Ann. Math. Statist.* 24, 22-238.

Roy, S. N., Kastenbaum, M. A., 1956. On the hypothesis of no "interaction" in a multiway contingency table. *Ann. Math. Statist.* 27, 749-757.

Roy, S. N., Mitra, S. K., 1956. An introduction to some nonparametric generalizations of analysis of variance and multivariate analysis. *Biometrika* 43, 361-376.

Shapiro, S. H., 1982. Collapsing contingency tables—a geometric approach. *Amer. Statist.* 36, 43-46.

Simpson, E. H., 1951. The interpretation of interaction in contingency tables. *J. Roy. Statist. Soc., Ser. B* 13, 238-241.

Stuart, A., 1955. A test for homogeneity of the marginal distributions of a two-way classification. *Biometrika* 42, 412-416.

Whittaker, J., 1990. *Graphical Models in Applied Multivariate Statistics*. Wiley.

Yule, G. U., Kendall. M. G., 1950. *An Introduction to the Theory of Statistics*, 14th ed. London: Griffin.