Log-Linear Approaches to Lag-Sequential Analysis When Consecutive Codes May and Cannot Repeat

Roger Bakeman  
Georgia State University

Vicenç Quera  
University of Barcelona

G. P. Sackett's (1979) lag-sequential analysis is recast in log-linear terms. Differing from articles by P. D. Allison and J. K. Liker (1982) and by D. Iacobucci and S. Wasserman (1988), sequences that, for logical reasons, do not allow consecutive events to be assigned the same code are considered. In addition, overlapped sampling, which the sequential analysis literature usually assumes, is contrasted with nonoverlapped sampling of sequences, which is also used and may seem more faithful to some statistical models. Several advantages of a log-linear approach to sequential problems are noted, including its ability to deal routinely with the structural zeros created when consecutive codes cannot repeat and its integration of sequential methods into an established and well supported statistical tradition.

Investigators as diverse as those who study animals, children, and married couples have found observational methods useful (e.g., Altman, 1965; Bakeman & Brown, 1977; Gottman, 1979). Such investigators systematically code selected aspects of the passing stream of behavior and then search for regularities or patterns in the data collected. Here we consider event sequences (Bakeman & Gottman, 1986; Bakeman & Quera, 1992, 1995), a basic and much used form of observational data that results when (mutually exclusive) sequential events in the passing stream of behavior are identified and assigned codes.

Event sequences are of two kinds: those that permit consecutive events to be assigned the same code (i.e., to repeat) and those that require consecutive codes to be different. For example, often conversation is transcribed and then segmented into stretches of relatively homogenous content (such segments are often called thought units; e.g., Gottman, 1979). Then content codes (e.g., Agree, Complain, Criticize, Compromise, and Disagree; see Hops, Wills, Weiss, & Patterson, 1972) are assigned to each thought unit. In such cases, not only may consecutive events be assigned the same code, but often it is also of considerable substantive interest, for example, whether Complain–Complain sequences occur frequently in the data.

In other cases, however, events are defined so that a repeated occurrence of a code is logically impossible. For example, often a list of activity categories for an animal or person is prepared (thus, e.g., Bakeman & Brownlee, 1980, following Parten, 1932, categorized preschoolers during free play as Unoccupied alone, Together with others, or engaged in Solitary, Parallel, or Group play). Then observers note each time the activity changes (perhaps recording its time as well) and record the code for the new activity. In such cases, a new event is necessarily different from the preceding, so consecutive events can never be coded the same (e.g., the sequence Solitary–Solitary can never appear in the data).

The analysis of event sequences has long interested students of animal and human behavior. Developed over the past few decades, one set of helpful techniques is called lag-sequential analysis (Sackett, 1979, 1987) or, more generally, sequential analysis (e.g., Bakeman, 1978; Bakeman & Dabbs, 1976; Bakeman & Gottman, 1986; Gottman & Roy, 1990; Sackett, 1978). Lag-sequential analysis was developed primarily by investigators, and the animal behavior literature in particular contains numerous examples (e.g., Lemon & Chatfield, 1971). Early articles stressed binomial $z$ scores for individual pairs of behaviors (e.g., Bakeman, 1978; Sackett, 1979). In an influential and often-cited article, Allison and Liker (1982) agreed that $z$ scores could be used to test sequential dependency between pairs of behaviors, but they noted that the binomial computation recommended in the early articles was technically incorrect and that a slightly different computation should be used instead (because expected cell frequencies were estimated from observed ones, not determined a priori). They also suggested that

We would like to thank M. T. Anguera, whose writings, invitations, and introductions made this work possible, and K. Kienapple, whose comments proved of considerable help when we first began thinking about the issues discussed here. We would also like to thank T. Wickers, whose writings helped our work immeasurably, and Georgia State and Emory University graduate students D. McArthur, B. Robinson, D. Reed, H. Mursalimi, and K. Baldwin for helpful comments on an earlier version.

Correspondence concerning this article should be addressed to Roger Bakeman, Department of Psychology, Georgia State University, University Plaza, Atlanta, Georgia 30303-3083, or Vicenç Quera, Departamento de Metodología, Facultad de Psicología, Universidad de Barcelona, Campus de la Dehesa, Paseo de la Salud, 171, 08035 Barcelona, Spain. Electronic mail may be sent via Internet to bakeman@gsu.edu (R. Bakeman) or vquer@ub.es (V. Quera).

Copyright 1995 by the American Psychological Association, Inc. 0033-2909/95/$3.00

272
log-linear modeling might be used but did not develop this idea in any detail.

Observational data of the sort appropriate for lag-sequential analysis can almost always be presented in the form of multidi

mensional contingency tables. Castellani (1979) may have been

the first to point this out and, as noted earlier, Allison and Liker (1982) explicitly suggested log-linear analyses for such tables.

Budescu (1984) developed this idea further, as did Iacobucci and Wasserman (1988); Iacobucci and Wasserman's informative article covers sequential analyses and focuses primarily on Lag I, dyadic relations. But only rarely have researchers with sequential data used log-linear analyses (exceptions are Cohn & Tronicck, 1987; Stevenson, Ver Hoeve, Roach, & Leavitt, 1986), even though such analyses seem well suited for such data (e.g., see Bakeman, 1991; Bakeman, Adamson, & Strisik, 1989; Gottman & Roy, 1990; cf. Farace & Dorman, 1987).

This article recasts traditional lag-sequential analysis in log-linear terms. It differs from Allison and Liker (1982) and Iacobucci and Wasserman (1988) in its consideration of sequences that, for logical reasons, do not allow consecutive events to be assigned the same code. Such situations are quite common, especially in the animal behavior literature, and have long troubled researchers (e.g., see Chaffin & Lemo, 1970). In addition, overlapped sampling, which the sequential analysis literature usually assumes, is contrasted with nonoverlapped sampling of sequences, which is also used and may seem more faithful to some statistical models. Throughout, we use the more general term ‘event sequences,’ although we recognize that the events of interest might be time intervals instead of bounded events; in such cases, some writers might use the more restrictive term ‘time sequences’ instead (e.g., see Bakeman & Gottman, 1986; Sackett, 1979).

Our discussion is divided into two major sections. In the first section, we consider log-linear analyses of Lag I associations, both when consecutive codes may and cannot repeat, and also consider how patterns involving specific behaviors might best be identified. In the second section, we consider similar problems but for legs greater than 1. As legs increase, so can demands for data; thus similar to traditional lag-sequential analysis, we emphasize analytic strategies that limit the amount of data required. Unlike lag-sequential analysis, however, we propose solutions that apply for nonoverlapped as well as overlapped sampling.

Analysis of Association at Lag 1

First, we define some basic notation and terminology: Assume that each event in a sequence is assigned one of K codes from a predefined set; that the codes are named A, B, C, and so forth; and that N is events are coded (the subscript i indicates that single events are counted; this distinction becomes important later). When consecutive codes may repeat, a sequence might look like A A B B C A C A . . . ; but when the codes cannot repeat, it might look like A C B C A . . . instead. If we were interested in Lag I effects, we might tally each two-event chain in the sequence—first b1b2, then b2b3, and so forth, where b represents a behavioral event and the subscript represents its sequential position—and organize the results in a table as follows (for simplicity, assume K = 3):

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
<tr>
<td>x1</td>
<td>x1</td>
<td>x1</td>
<td>x1</td>
</tr>
<tr>
<td>x2</td>
<td>x2</td>
<td>x2</td>
<td>x2</td>
</tr>
<tr>
<td>x3</td>
<td>x3</td>
<td>x3</td>
<td>x3</td>
</tr>
<tr>
<td>x4</td>
<td>x4</td>
<td>x4</td>
<td>x4</td>
</tr>
</tbody>
</table>

Rows represent the first event (Lag 0 or given) and columns the second (Lag 1 or target) event in the chain, x represents the frequency of chains that begin with the jth event and end with the jth event, and plus signs indicate summation over the subscripts indicated.

When two-event chains are tallied, x (the number of tallies in the table) is symbolized N. Assuming overlapped sampling as described in the previous paragraph, N would be N - 1 if the sequence represented a single unbroken observation session and N - S if breaks occurred (S is the number of separate sessions, as occurs, e.g., when the child wanders out of view briefly or recording is scheduled over different days). Nonoverlapped sampling also occurs. For example, imagine that a researcher is interested only in tallying Speaker 1 to Speaker 2 transitions. Then N would be N + 2 (tallying first b1b2, then b3b4, etc.) if the speakers alternate turns, or some smaller number if a speaker's turn may contain more than one thought unit (which could result in tallying b1b2, b3b4, b5b6, b7b8, etc.; see Bakeman & Casey, 1995).

No matter how sampling is effected, if consecutive events cannot be assigned the same code, then the diagonal elements (x) when i = j are zero. Such zeros are called structural, as opposed to empirical, because their value follows from necessity, not from a happenstance of data collection.

To answer the question: Are events at Lag 0 systematically related to events at Lag 1, a contingency table such as the one depicted earlier invites a chi-square computation. But the sampling gives us pause. In the case of overlapped sampling, the next two-event chain is constrained to begin with the code that ended the previous chain (i.e., if a two-event chain ends in B, adding a tally to the second column, the next must add a tally to the second row), and this violates sampling independence. A Pearson or likelihood-ratio chi-square (defined later) could be computed and would be an index of the extent to which observed frequencies in this table tend to deviate from their expected ones. But we would probably have even less confidence than usual that the test statistic is distributed as χ² and so, quite properly, would be reluctant to apply a p value.

Nonoverlapped sampling does not pose the same threat to sampling independence. Moreover, the consequences of overlapped sampling may not be as severe as they at first seem. Bake

man and Dorval (1989) found that when sequences were generated randomly, distributions of a test statistic assumed their theoretically expected form equally for the overlapped and nonoverlapped procedure and concluded that the apparent violation of sampling independence associated with overlapped sampling was not consequential. Thus, for the remainder of this article conventional .05 critical values are assumed, but questions concerning the accuracy of probabilities assigned test statistics may arise. Moreover, limitations not just of multiple tests, but of multiple interrelated ones as arise in contingency table analysis, are noted.
Log-Linear Analyses When Consecutive Codes May Repeat

It is useful to compare a log-linear analysis of a Lag 0–Lag 1 table with a traditional lag-sequential analysis. For example, assume \( K = 3 \) and tallies for two-event chains are as follows:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>23</td>
<td>5</td>
<td>15</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>1</td>
<td>7</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>14</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>20</td>
<td>38</td>
<td>100</td>
</tr>
</tbody>
</table>

Given a two-dimensional table such as this one, two log-linear models are of interest: Using bracket notation, (Fienberg, 1980) the saturated model is symbolized as \([01]\) and the model of independence is symbolized as \([0][1]\). Each model generates a set of expected frequencies. For the \([01]\) model, the expected values are identical to the observed ones; whereas for the \([0][1]\) model, they are computed using the familiar formula

\[
m_{ij} = \frac{x_{ij} x_{..}}{x_{..}},
\]

where \( m_{ij} \) is an estimate of the expected frequency (\( m \) because often expected values are means; later Equation 1 is referred to as MM for marginal model because \( m \) is computed from the row and column totals or marginals). Note, however, that Equation 1 assumes a complete table (i.e., one with no structural zeros).

A model's goodness of fit can be evaluated with the familiar Pearson chi-square statistic,

\[
X^2 = \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{(x_{ij} - m_{ij})^2}{m_{ij}},
\]

where \( X^2 \) represents the estimated value for \( \chi^2 \). Or the less familiar likelihood-ratio chi-square,

\[
G^2 = 2 \sum_{i=1}^{K} \sum_{j=1}^{K} x_{ij} \log \frac{x_{ij}}{m_{ij}},
\]

can be used to estimate \( \chi^2 \) instead (here and elsewhere, \( \log \) represents the natural logarithm, i.e., the logarithm to the base \( e \); for a discussion of the differences between \( X^2 \) and \( G^2 \), see Wickers, 1989). If \( X^2 \) or \( G^2 \) is large when \( m_{ij} \) is computed according to the model of independence, then the \([0][1]\) model does not fit the data. Consequently, the \([01]\) model is accepted, which assumes that row and column dimensions (in this case, Lag 0 and Lag 1) are associated.

Alternatively, because the models are related hierarchically (i.e., terms in the \([0][1]\) model are a subset of those in the \([01]\) model; here, terms included in the \([0][1]\) model represent a row and column effect, whereas terms included in the \([01]\) model include, in addition to these two, a term representing their interaction), one could ask whether the partial chi-square,

\[
\Delta G^2_{[0]} = G^2_{[0][1]} - G^2_{[01]},
\]

is large. Differences between likelihood-ratio chi-squares (but not Pearson chi-squares) associated with hierarchically related models are themselves distributed approximately as \( \chi^2 \), with degrees of freedom equal to the difference between the degrees of freedom for the two models (Fienberg, 1980), which explains why \( G^2 \) is often preferred to \( X^2 \) in a log-linear context.

\( G^2_{[01]} \) is always zero (because the expected frequencies for the saturated model are identical with the observed ones); thus, when \( m_{ij} \) is generated by the \([0][1]\) model, Equations 3 and 4 yield identical results. Still, these equations embody slightly different points of view, which proves useful for tables with more than two dimensions. For the tallies, given earlier, the result is 16.43 with 4 degrees of freedom [for a complete two-dimensional table, \( df = (K - 1)^2 \), here \((3 - 1)^2 = 4\)], which exceeds the critical value of 9.49 and so suggests that Lag 0 and Lag 1 are associated.

The preceding discussion of \( X^2 \) and \( G^2 \) suggests one immediate advantage of a log-linear view. Traditional lag-sequential analysis rarely considers whether Lag 0 and Lag 1 behaviors are associated generally, which is what the tablewise \( X^2 \) or \( G^2 \) indicates, but instead focuses on cellwise statistics, which identify the particular chains (such as C-B, A-A, etc.) whose occurrence differs substantially from expected. When many cells are examined exploratively, the traditional approach invites Type I errors. As a corrective, absent specific predictions, we recommend examining cell statistics only when the tablewise \( X^2 \) or \( G^2 \) is large (just as post hoc tests are pursued in analysis of variance only when the omnibus F ratio is significant). Yet, even then the Type I error rate of subsequent pairwise tests may not be adequately controlled when the partial null hypothesis is true.

Usually investigators are drawn to traditional lag-sequential analysis because they want to determine whether some behaviors occur quite frequently (or only rarely) after other behaviors. A statistic is needed that gauges the extent to which an observed frequency, \( x_{ij} \), differs from its expected value, \( m_{ij} \). The log-linear literature provides such an index. It is called the adjusted residual, and for two-dimensional complete tables is defined as

\[
z_{ij} = \frac{x_{ij} - m_{ij}}{\sqrt{m_{ij}(1 - p_{ii})(1 - p_{jj})}},
\]

where \( p_{ii} = x_{ii} / x_{..} \) and \( p_{jj} = x_{..} / x_{ij} \) (Haberman, 1978, p. 111). Haberman (1978, 1979) symbolized the adjusted residual with \( r \); however, to avoid confusion with the Pearson correlation coefficient and to emphasize correspondence with similar statistics in the lag-sequential literature, \( z \) is used (Equation 5 is later referred to with the mnemonic ZCT because it requires complete tables).

A similar but not identical statistic appears in the lag-sequential literature. The standard Sackett (1979) formula, corrected as noted by Allison and Liker (1982; symbolized by them as \( z_1 \)) and expressed in terms of expected frequencies (Bakeman & Gottman, 1986), which emphasizes its similarity with Equation 5, is

\[
z^{(1)}_{ij} = \frac{x_{ij} - m_{ij}}{\sqrt{m_{ij}(1 - p_{i1})(1 - p_{j})}},
\]

where \( m_{ij} = x_{i} x_{j} / N_{ij} \), \( x_{i} \) is the number of events in the se-
quence assigned the \( i \)th code, and \( p_i = x_i / N_i \) (our mnemonic for Equation 6 is ALZ for Allison & Liker). Note that \( m_{ij} \), \( p_i \), and \( p_j \) are based on occurrences of single events not two-event chains.

As usual in the lag-sequential literature, Equation 6 assumes overlapped sampling. When an unbroken sequence is tallied using overlapped sampling, row and column marginals differ from the corresponding simple frequencies by at most 1 (i.e., \( x_{i+} = x_i + 1 \), or \( x_i - 1 \); so from the tallies given earlier, it is inferred that \( A, B, \) and \( C \) occurred 43, 20, and 38 times, respectively, and that the sequence of 101 events began with \( A \) and ended with \( B \)). In such cases, the estimate for the probability that a two-event chain ends in \( A \), for example, might reasonably be the probability of a simple event being coded \( A \), which is \( p_i \) or \( x_{i+} / N_{i+} \); thus, \( p_i \) appears in Equation 6. However, in other cases and especially when sampling is not overlapped (in which case row and column margins need not be similar), a better estimate for the probability that a two-event chain ends in an \( A \) is \( p_{i+} \) or \( x_{i+} / N_{i+} \); thus, \( p_{ij} \) appears in Equation 5.

When a single, unbroken sequence is tallied using overlapped sampling, the results of Equations 5 and 6 are usually quite similar. For example, for the 101 events tallied in the table given previously, \( z_{ij}^{(1)} = 1.91 \), whereas \( z_{ij} = 2.02 \) (if the decision rule were 1.96 absolute, however, this would matter). Nonetheless—and this is a second advantage of the log-linear view—Equation 5 is recommended in all cases. Unlike Equation 6, it is firmly based in the log-linear literature and is appropriate for both overlapped and nonoverlapped sampling.

For the present example, the adjusted residuals (per Equation 5) are

\[
\begin{align*}
1: & \ A \quad B \quad C \\
0: & \ A \quad 2.02 \quad -1.82 \quad -0.56 \\
& \ B \quad 1.56 \quad -1.78 \quad -0.12 \\
& \ C \quad -3.32 \quad 3.30 \quad 0.66
\end{align*}
\]

which suggests that the two-event chains A-A and C-B occurred more often than expected, whereas C-A occurred less often than expected. This conclusion is provisional and may overstate the case; we return to it later.

**Log-Linear Analyses When Consecutive Codes Cannot Repeat**

A third advantage of the log-linear view becomes evident when tables are incomplete (i.e., contain structural zeros). For example, imagine we decide that only transitions from one content code to a different one are of interest. In that case, we would replace autotransitions with structural zeros; thus, the previous tallies become

\[
\begin{align*}
1: & \ A \quad B \quad C \\
0: & \ A \quad 5 \quad 15 \quad 20 \\
& \ B \quad 11 \quad 7 \quad 18 \\
& \ C \quad 8 \quad 14 \quad 22 \\
19 & \overline{19} \ 22 \ 60
\end{align*}
\]

where a dash represents a structural zero. The authors argued that the probability for two-event chains that end in code \( j \) should be the number of such chains \( (x_{ij} \) relative to the number of chains that could end in code \( j \) which is the total number of two-event chains \( (N_j) \) reduced by chains that begin with code \( j \) \( (x_{ij}) \) and therefore cannot end in code \( j \). Thus, compared with Equation 6 (GS1), Equation 8 has the merit of applying no matter whether overlapped or separate sequences are sampled.

In fact, both Equations 7 and 8 are problematic. Neither generates expected frequencies whose marginals match the marginals for the observed ones as they should. For example, expected frequencies for the data given earlier per GS1 (Equation 7) are

\[
\begin{align*}
1: & \ A \quad B \quad C \\
0: & \ A \quad 9.50 \quad 11.00 \quad 20.50 \\
& \ B \quad 8.14 \quad 9.43 \quad 17.57 \\
& \ C \quad 11.00 \quad 11.00 \quad 22.00 \\
19.14 & \overline{20.50} \ 20.43 \ 60.07
\end{align*}
\]

Lemon and Chatfield (1971) understood that the lack of a match was a problem; referring to their Equation 8 as approximate (see also Van Hooff, 1982), they noted that an exact solution requires an iterative procedure as presented by Goodman (1968). Iterative procedures can handle a variety of models, as well as tables of various dimensions with or without structural zeros, so these procedures are used routinely in log-linear analysis.

Thus, when consecutive codes cannot repeat, resulting in structural zeros on the diagonal, expected frequencies for the [0][1] model should be computed with an iterative procedure. Two of the most widely used are iterative proportional fitting.
(IPF, also called the Deming–Stephan algorithm) and the Newton–Raphson algorithm; for descriptions see Bishop, Fienberg, and Holland (1975) and Fienberg (1980). With either method, expected frequencies for the example would be

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>1: A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:</td>
<td>A</td>
<td>B</td>
<td>C</td>
<td>20.00</td>
</tr>
<tr>
<td></td>
<td>8.30</td>
<td>11.70</td>
<td>20.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7.70</td>
<td>10.30</td>
<td>18.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.30</td>
<td>10.70</td>
<td>22.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.00</td>
<td>19.00</td>
<td>22.00</td>
<td>60.00</td>
</tr>
</tbody>
</table>

As a practical matter, statistics based on these two sets of expected frequencies are not dramatically different: \( G^2 = 6.99 \) using GS1 and 6.81 using an iterative procedure (if one tries to reproduce these values for \( G^2 \) using the expected frequencies shown in the tables, one gets slightly different results because the latter are rounded). Degrees of freedom are 1 instead of the usual 4 for a \( 3 \times 3 \) table because 3 are lost to the structurally zero cells.

Just as MM (Equation 1) does not produce the correct expected frequencies for incomplete tables, so too ZCT (Equation 5) does not yield the correct adjusted residuals. The formula (actually several formulas, many of which are used by the Newton–Raphson algorithm) is given in Haberman (1979, p. 454).\(^1\)

The most practical way for an investigator to compute these values is to use a computer program (such as SPSS's old LOGLINEAR or new GENLOG) that produces the values (programs such as SPSS's HILOGLINEAR and Bakeman & Robinson's, 1994, ILOG that rely on IPF do not).

### Untangling Patterns of Adjusted Residuals

Earlier we noted that lag-sequential analysis is prone to Type I errors. Given \( K \) codes, the number of possible two-event chains is \( K^2 \) when consecutive codes may repeat and \( K(K - 1) \) when they cannot. Thus, the number of cells not structurally zero increases exponentially with \( K \); when many adjusted residuals are examined, some may be large just by chance. The problem, however, is more severe. Adjusted residuals in a table form an interrelated web. If some are large, others necessarily must be small, and the investigator needs to determine which should be emphasized. For example, three of the nine adjusted residuals in the complete table examined earlier were greater than 1.96 absolute, but some of these adjusted residuals might be induced by others.

This is easy to test using methods for incomplete tables. Assume the C-B chain, whose adjusted residual is 3.30, is of primary theoretical interest. To test its importance, we declare the C-B cell structurally zero—note that now the [0][1][1] model fits the remaining data (\( G^2[3] = 5.79, \) critical value for \( 3 \) df = 7.81; \( df = 3 \) because one is lost to the structurally zero cell). We conclude that interpretation should emphasize the C-B chain as the other two effects (decreased occurrences for C-A and increased occurrences for A-A) disappear when the C-B chain is removed.

Had the model not fit the reduced data, we would have declared another structural zero and tested the data reduced by two cells. Proceeding stepwise (but letting theoretical considerations not raw empiricism determine the next chain to delete, else one risks capitalizing on chance as with backward elimination in multiple regression and compromising Type I error control), we would identify those chains that prevent the [0][1] model from fitting. A logically similar suggestion, not in a log-linear context, was made by Rechten and Fernald (1978; see also Wickens, 1989, pp. 251–253). The ability to winnow results in this way is yet another advantage of the log-linear view.

### Analysis of Association at Lags Greater Than 1

Investigating Lag \( L \) effects, where \( L \) represents the lag under investigation, involves consideration of \( (L + 1) \) event chains. The first event in the chain occupies Lag Position 0, the second Lag Position 1, and so forth. The \( (L + 1) \) event chains can be tallied in a \( K^{L+1} \) table; so when investigating Lag 2 effects, for example, three-event sequences would be tallied in a \( K^3 \), Lag 0 by Lag 1 by Lag 2 table. Earlier, to test for a Lag 1 effect, we asked whether events at Lag 0 were independent of events at Lag 1. Such unconditional independence can be symbolized \( 0 \| 1 \) (Wickens, 1989; Bishop et al., 1975, symbolize this \( 0 \odot 1 \) and viewed as analogous to the usual bivariate correlation. Now, we ask whether events at Lag 0 are independent of events at Lag \( L \); but more than two lag positions are involved, so we can phrase this conditionally and, analogous to a partial correlation, ask whether events at Lag 0 and Lag \( L \) are associated after taking intermediate effects into account (symbolized \( 0 \| 1 \) \( L \) \) \( L \) \( L \) – 1).

In general terms, the test of conditional independence, taking all intermediate effects into account, is

\[
\Delta G^2_{011 \cdots L-1} = G^2_{0 \cdots (L-1)1 \cdots L} - G^2_{0 \cdots (L-1)1 \cdots L \| 0L2} \tag{9}
\]

(see Wickens, 1989, pp. 61–64). For example, when considering Lag 2 effects, we would test whether \( 0 \| 1 \) \( L \),

\[
\Delta G^2_{011} = G^2_{011} - G^2_{0112} \tag{10}
\]

The more complete model includes the 02 term; otherwise the two models are the same. If the [01][12] model fits and fits only a little worse than the more complete model, then the partial chi-square is small. If simpler models do not fit, then we would conclude that events at Lag 0 are independent of those at Lag 2 after taking Lag 1 events into account (technically, this assumes that the 012 term is not required for a fitting model either). Such information is not available from unconditional tests. Just as a large bivariate correlation may become a small partial regression coefficient when a mediating variable is taken into account (Baron & Kenny, 1986), so too a large unconditional Lag 2 effect may become a small Lag 2 effect when conditioned on Lag 1 events (as we demonstrate later).

The full, conditional test (Equation 9)—which assesses whether events at Lag 0 and Lag \( L \) are associated after taking effects at intermediate lags into account—has considerable appeal, but it does require \( (L + 1) \) dimensional tables. The number of cells not structurally zero in such tables is \( K^{L+1} \) when

\(^1\) In Haberman's (1979) volume (first printing), a term is missing in the formula for \( c_{ij} \) on p. 454. An \( m \) should be inserted after the minus sign and before the two summation signs.
consecutive codes may repeat and $K(K-1)^L$ when they cannot, hence the number of such cells increases exponentially with increases in $K$ and $L$ (although the increase is somewhat less pronounced when consecutive codes cannot repeat). Especially for larger values of $L$, unless the number of events observed is almost astronomically large, expected frequencies for many of the cells may be near zero, which is problematic for log-linear analysis. In the next two sections, we consider strategies for reducing the size of the tables analyzed and the amount of data required.

**Analyses When Consecutive Codes May Repeat**

Although details vary somewhat when consecutive codes may and cannot repeat, in both cases the overall sequential search strategy we propose is the same (such generality results from the log-linear view). We begin by looking for complex Lag 2 effects in the 012 table. If no complex Lag 2 effects are found (what constitutes a complex effect is defined in the next paragraph), the three-dimensional table is collapsed to a 02 table, and the Lag 3 dimension is added; then we look for Lag 3 effects in the 023 table, and so forth. In general terms, beginning with Lag $L$, we test whether the three-dimensional 0($L-1$)$L$ table can be collapsed over the $L-1$ dimension. If so, we collapse to the 0$L$ table, add the $L+1$ dimension, thereby creating a new three-dimensional table, increment $L$, and repeat the procedure, continuing until we find a table that does not permit collapsing over the $L-1$ dimension. Once such a table is found, we explicate the Lag $L$ effects in this three-dimensional table. If data are sufficient, next we might analyze the four-dimensional 0($L-1$)$L$($L+1$) table, and so forth; but further collapsing is unwarranted due to the Lag $L$ effects just found. Nonetheless, this strategy may let us examine lags longer than 2 without requiring tables larger than $K^3$.

At each lag, complex effects are defined as effects that implicate Lags 0 and $L$ with Lag $L-1$; if present, collapsing over the $L-1$ dimension is unwarranted. For example, if $L = 2$, then complex effects are present if the simplest model that fits the 012 table includes any of the following:

1. [012] because then Lag 0 and Lag 2 are associated and interact with Lag 1 (three-way associations).
2. [01][12][02] because then Lag 0 and Lag 2 are associated with each other and Lag 1 but do not interact with Lag 1 (homogeneous associations).
3. [01][12] because then Lag 0 and Lag 2 are independent conditional on Lag 1.

However, if any other model fits, then collapsing over the $L-1$ dimension is justified (Wickens, 1989, pp. 79–81, 142–143). After collapsing, we would test whether 0 12 in the 02 table. If not, we would examine adjusted residuals per ZCT (Equation 5) to explicate the Lag 0–Lag $L$ effect identified.

Consider the following example. We generated a sequence of 250 codes (the computer program is described later); $K = 3$, and consecutive codes were allowed to repeat. Tallys for the overlapping two-event chains constituting this sequence are

\[
\begin{array}{ccc}
1: & A & B & C \\
0: & 21 & 25 & 49 \\
B & 23 & 26 & 21 \\
C & 50 & 19 & 15 \\
\end{array}
\]

As a preliminary, we test whether 0 1. $G^2_{10[1]} = 35.2$ (critical value for 4 df = 9.49), so we conclude that events at Lag 0 and Lag 1 are associated. Next, beginning the sequential search strategy described in the previous paragraph, we consider Lag 2. The tallies for three-event chains, derived from the same sequence, are

\[
\begin{array}{ccc}
2: & A & B & C \\
0: & A & 7 & 6 & 8 \\
& B & 10 & 9 & 6 \\
& C & 31 & 8 & 10 \\
B & 1: & 3 & 6 & 14 \\
& B & 6 & 11 & 9 \\
& C & 12 & 4 & 4 \\
C & 1: & 11 & 13 & 26 \\
& B & 7 & 6 & 6 \\
& C & 7 & 7 & 1 \\
\end{array}
\]

Consistent with traditional lag-sequential practice, which considers only the collapsed 0$L$ tables, we could proceed immediately to test whether 0 2 in the collapsed 02 table (i.e., the table consisting of the $x_{i+}$ marginals),

\[
\begin{array}{ccc}
2: & A & B & C \\
0: & 48 & 23 & 24 \\
B & 21 & 21 & 27 \\
C & 25 & 26 & 33 \\
\end{array}
\]

without first using the strategy described in the previous paragraph to verify that collapsing is justified. Then, however, we would fail to detect any of the effects that implicate Lag 1 in Lag 0–Lag 2 effects, which is the gain of the log-linear approach. For the unconditional test, $G^2_{11[2]} = 10.70$, we would conclude that events at Lag 0 and Lag 2 (like events at Lags 0 and 1) are associated. But perhaps 0 2 1. Indeed, for these tallies, the [01][12] model fits the 012 table ($G^2_{12} = 11.52$ and $\Delta G^2_{10[2]} = 3.85$, both less than their critical value) and simpler ones do not, so we should conclude that events at Lag 0 and Lag 2 are in fact independent, once events at Lag 1 are taken into account. Such a circumstance would not be detected with traditional lag-sequential methods.

As demonstrated in the previous section, log-linear analysis can detect complex effects that would elude simpler approaches, but an insight of traditional lag-sequential analysis has always been that effects at higher lags should be interpreted in light of effects at lower lags because the former may be induced in part by the latter. We conclude this section with a demonstration, using a computer program we wrote for this article. The program generated sequences of a specified length and a value for $K$. A matrix of transitional probabilities, $t_{ij}$, is defined; thus,

\[
\Sigma t_{ij} = 1
\]
for all values of \( i \). Let \( b_s \) represent the \( s \)th code in the sequence (where \( s = 1 \) through \( N_s \)). To specify a Lag 1 effect, we set \( i = b_{s-1} \) (the numeric value of the preceding code, assuming \( A = 1 \), \( B = 2 \), etc.) where \( i \) specifies a row in the matrix and hence a vector of transitional probabilities. Then the value for \( b_s \) is a random number 1 through \( K \) whose probabilities reflect those in the \( i \)th row of the transitional probability matrix. Generally, if we want a Lag \( L \) effect, then \( i = b_{s-L} \) for each successive \( b_s \) (\( s = L + 1 \) through \( N_s \)). For clarity, we limited effects to a single lag, although effects at several lags are possible and may be more realistic. In addition, if we wish no effect, then \( i = b_{s-1} \), but all rows are defined the same (i.e., \( t_{ij} = t_{i2} \), etc.), or a single vector of transitional probabilities is used.

Using this program (written in Borland’s Pascal 6.0), we generated six sets of 1,000 sequences, each 100 events in length. Each set reflected a different effect (no effect, a Lag 1 effect, etc. through a Lag 5 effect). The length was set at 100, because with overlapped sampling and \( K = 3 \), this assures a reasonable number of tallies per cell (e.g., 95 tallies for 9 cells when \( L = 5 \); slightly more for shorter lags). Initially, values for \( t_{ij} \) were set at

\[
\begin{array}{ccc}
  j: & A & B & C \\
  i: & A & .17 & .17 & .67 \\
    & B & .33 & .33 & .33 \\
    & C & .33 & .33 & .33 \\
\end{array}
\]

We tested each sequence generated using these parameters for an unconditional Lag 1 through Lag 5 effect (i.e., we computed \( G^2 \) for the \([0][L]\) model where \( L = 1 \) through 5) and noted the proportion (for each effect at each lag) that exceeded the .05 critical value. These proportions were

\[
\begin{array}{cccccc}
  L: & 1 & 2 & 3 & 4 & 5 \\
  E: 0 & .07 & .05 & .06 & .06 & .05 \\
    1 & .69 & .08 & .07 & .06 & .06 \\
    2 & .06 & .69 & .07 & .07 & .05 \\
    3 & .07 & .05 & .72 & .06 & .05 \\
    4 & .06 & .04 & .06 & .67 & .06 \\
    5 & .06 & .06 & .06 & .67 & .06 \\
\end{array}
\]

where \( L \) represents the lag used to compute \( G^2 \) and \( E \) represents the lag at which the program was set to produce an effect. Proportions greater than .10 are bold, here and in later tables. When no effect was specified (\( E = 0 \)) and also when an effect was specified but at another lag (\( E \neq L \)), proportions represent Type I error rates and were close to the theoretically expected .05. Values for \( X^2 \) were even closer. As others have reported—for example, Fienberg, 1980, Appendix IV; Larnitz, 1978—the Pearson is the statisticist as far as null hypothesis behavior is concerned. Small counts in cells with moderate expected values tend to inflate the likelihood-ratio chi-square, so more values of .06–.08 than .04–.06 are seen. Otherwise, when an effect was specified (\( E = L \)), proportions represent power; for the values of \( N_s \) and \( t_{ij} \) used here, Type II error rates, or 1 minus power, were approximately .30.

Next, to demonstrate how lower order effects can induce higher order ones, we specified a second set of transitional probability parameters,

\[
\begin{array}{ccc}
  j: & A & B & C \\
  i: & A & .20 & .20 & .60 \\
    & B & .33 & .33 & .33 \\
    & C & .60 & .20 & .20 \\
\end{array}
\]

Before, only \( p(C_2 | A_0) \) was elevated (.67, the parameter value for \( C \) at Lag 1 given \( A \) at Lag 0 or \( t_{13} \)). Here, \( p(C_2 | A_0) \) and \( p(A_2 | C_0) \) are both elevated. When a Lag 1 effect is specified for these two transitional probabilities, a Lag 2 effect is likely (because A-C and C-A chains tend to produce A-C-A chains), a Lag 4 effect is more likely when a Lag 2 effect is specified, and so forth. Indeed, proportions of \( G^2 \) exceeding their critical values for the second set of transitional probability parameters were

\[
\begin{array}{cccccc}
  L: & 1 & 2 & 3 & 4 & 5 \\
  E: 0 & .06 & .05 & .06 & .05 & .05 \\
    1 & .95 & .22 & .09 & .08 & .08 \\
    2 & .04 & .93 & .06 & .21 & .06 \\
    3 & .08 & .08 & .95 & .07 & .06 \\
    4 & .07 & .05 & .08 & .93 & .07 \\
    5 & .09 & .08 & .07 & .08 & .94 \\
\end{array}
\]

As anticipated, Lag 2 effects were noted more than expected when a Lag 1 effect was specified (proportion exceeding critical value was .22); likewise Lag 4 effects were noted more than expected when a Lag 2 effect was specified (proportion exceeding critical value was .21). Later, we return to the issue of induced effects and discuss how the issue becomes more complex when consecutive codes cannot repeat.

**Analyses When Consecutive Codes Cannot Repeat**

The sequential search strategy described earlier applies when consecutive codes cannot repeat with one modification. When consecutive codes may repeat and no complex Lag \( L \) effects are found (i.e., each table examined sequentially permits collapsing), then the test series becomes \( 0 \| 1 \| 2 \), \( 0 \| 1 \| 3 \), and so forth (i.e., \( 0 \| L \) is tested in the \( 0 \| L \) table). When consecutive codes cannot repeat, the unconditional test makes no sense because it does not reflect the constraints imposed when consecutive codes cannot repeat. Then, when no complex Lag \( L \) effects are found, the analogous series becomes \( 0 \| 1 \| 2 \), \( 0 \| 1 \| 3 \), and so forth (i.e., \( 0 \| L \) is tested in the \( 0 \| L \) table). Models associated with these tests include the \( (L - 1) \) L term. The corresponding marginal table has structural zeros on the diagonal, which reflect the cannot-repeat constraint.

As an example of a Lag 2 inquiry, assume that \( K = 3 \) and 120 three-event sequences are tallied with the following result:
When consecutive codes cannot repeat, the 012 table always contains the pattern of structural zeros exemplified here. Moreover, when \( K = 3 \) and only when \( K = 3 \), the [01][12][02] model is completely determined; its degrees of freedom are zero and expected frequencies duplicate the observed ones (like the [012] model when consecutive codes may repeat). For these tallies, expected frequencies for the [01][12] model are

\[
\begin{array}{ccc}
2: & A & B & C \\
0: A & \begin{array}{ccc}
 & - & - & - \\
 & 15 & - & 6 \\
 & 10 & 9 & - \\
\end{array} \\
B & \begin{array}{ccc}
 & - & 12 & 8 \\
 & - & - & - \\
 & 9 & 12 & - \\
\end{array} \\
C & \begin{array}{ccc}
 & - & 8 & 11 \\
 & 5 & - & 15 \\
 & - & - & - \\
\end{array}
\end{array}
\]

For these data, \( G^2 = 10.8 \) (\( df = 3 \), critical value for \( 3 \) \( df \) = 7.81); thus, the test of conditional independence fails. We accept the [01][12][02] model and conclude that events at Lag 0 and Lag 2 are associated (and both are associated with Lag 1).

Now consider the traditional lag-sequential approach. Even if we were willing to ignore intermediate lags as traditional lag-sequential analysis does, we could not simply collapse to the 0L table and compute expected frequencies per Equation 1 because MM does not take into account that consecutive codes cannot repeat. In such cases, Sackett (1979) suggested that

\[
m_{ij} = \frac{N_{ij} - x_i x_j}{N_i x_i}
\]

(11)

where \( m_{ij} \) represents the expected frequency for the target behavior at Lag \( L \) when preceded by the given behavior at Lag 0, and \( x_{ij} \) is the observed frequency for the target at Lag \( L - 1 \) preceded by the given behavior at Lag 0 (our mnemonic for this equation is GS2 because it applies to lags of 2 or greater). As usual, \( x_i \) and \( x_{ij} \) are the simple frequencies for the given and target behaviors, respectively, and \( N_i \) is the number of events in the sequence.

In support of GS2 (Equation 11), Sackett (1979) reasoned that when adjacent codes cannot repeat, the expected probability for a particular target code at Lag \( L \) (assuming a particular given code at Lag 0) is the frequency for that target code, diminished by the number of times it appears in the Lag \( L - 1 \) position (because then it could not appear in the \( L \) position, after itself) divided by the number of events that may occur at Lag \( L \) (which is the sum of the Lag \( L \) minus the Lag \( L - 1 \) frequencies summed across all \( K \) target codes). Simply, this sum is the number of all events less the number of events assigned the given code. As with similar equations (ALZ and GS1), GS2 assumes overlapped sampling; like GS1, marginals for expected frequencies per GS2 do not match the observed marginals. For these reasons, we sought to develop an alternative that would apply no matter what the sampling design was.

For data such as those used in the previous example, once an overall Lag 0–Lag 2 association is established, the usual next question concerns the specific behavioral chains involved. In this case, traditional lag-sequential analysis would have computed expected frequencies for the 0L table per GS2 and then computed \( z \) scores per ALZ. Cellwise statistics can also be derived from the log-linear analysis, albeit not directly. The expected frequencies shown for the [01][12] model can be collapsed into a 02 table (0L generally) as follows:

\[
m_{jk} = \sum_{i=0}^{K} m_{ijk}
\]

(12)

(our mnemonic for Equation 12 is CIS because expected frequencies generated by the conditional independence model are summed). Then (assuming the [0, L - 1][L - 1, L] model does not fit, suggesting a Lag 0–Lag \( L \) association), \( z \) scores based on expected frequencies computed per CIS and themselves computed per ZCT could be used to describe particular Lag 0–Lag \( L \) associations.

The CIS + ZCT (Equation 12 + 5) approach seems more mathematically based than the lag-sequential GS2 + ALZ one (Equation 11 + 6) and does not assume overlapped sampling, but both are presented as simple descriptive devices. Neither has a firm statistical base, and we do not claim that these \( z \) scores necessarily reflect a standard normal distribution. Later, we compare them using simulated data, but a comparison based on the tallies given earlier in this section is instructive. For these data, the \( z \) scores derived per CIS + ZCT are

\[
\begin{array}{ccc}
0: & A & B & C \\
2: & -0.59 & -1.97 & \\
B & -0.70 & 0.63 & -0.94 \\
C & -2.44 & -1.06 & 2.19 \\
\end{array}
\]

\( ^2 \)One way to explain the 3 degrees of freedom is as follows. When \( K = 3 \) and \( L = 2 \), total degrees of freedom are 11 (27 cells minus 15 that are structurally zero minus 1). Each one-way term is associated with 2 \( df \) (\( K - 1 \)). The 01 and 12 terms are each associated with 1 \( df \), not 4, due to the structural zeros. This leaves 3 \( df \) for the 01 term (11 - 2 - 2 - 2 - 1 - 1), not the (\( K - 1 \))^2 or 4 expected, and none for the 012 term.
and per GS2 + ALZ are

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0: A</td>
<td>1.95</td>
<td>-0.47</td>
<td>-2.18</td>
</tr>
<tr>
<td>B</td>
<td>-0.46</td>
<td>0.90</td>
<td>-0.84</td>
</tr>
<tr>
<td>C</td>
<td>-2.26</td>
<td>-0.84</td>
<td>2.19</td>
</tr>
</tbody>
</table>

which are not markedly different. Both suggest that chains C-B-A (because only B can occur between C and A) and A-B-C (because only B can occur between A and C) are less frequent than expected, whereas chains that begin and end with C (i.e., C-A-C and C-B-C) are more frequent than expected. However, as noted earlier, z scores in such tables are interrelated. If cells associated with any of these three effects are replaced with structural zeros (i.e., C-B-A, A-B-C, or C-A-C and C-B-C), the modified data fit the [01][12] model. Hence, only one of these effects should be interpreted; the other two should be regarded as induced by it. This demonstrates how descriptive z scores (i.e., CIS + ZCT and GS2 + ALZ) can suggest formal log-linear tests, which may well be one of their better uses.

When codes in a sequence are unconstrained by the previous code, sequences can be generated that manifest an effect only at the lag specified, as demonstrated earlier. However, when codes are constrained by the previous code, lagged effects might be expected, by definition. To examine this possibility and compare the relative merits of CIS + ZCT and GS2 + ALZ, we again made use of our simulation program. When no effect or only a Lag 1 effect is desired, the program described previously can be used without modification. For no effect, when \( K = 3 \), transitional probability parameters are set at

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>i: A</td>
<td>0</td>
<td>.50</td>
<td>.50</td>
</tr>
<tr>
<td>B</td>
<td>.50</td>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>C</td>
<td>.50</td>
<td>.50</td>
<td>0</td>
</tr>
</tbody>
</table>

In addition, for a Lag 1 effect, the first set of values we used was

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>i: A</td>
<td>0</td>
<td>.25</td>
<td>.75</td>
</tr>
<tr>
<td>B</td>
<td>.50</td>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>C</td>
<td>.50</td>
<td>.50</td>
<td>0</td>
</tr>
</tbody>
</table>

which specifies an elevated value for \( p(C_1|A_0) \). Zeros on the diagonal signal structural zeros.

For effects at higher lags, a three-dimensional matrix of transitional probabilities, \( t_{ijk} \) is required. The \( i \) indicates Lag 0, the \( j \) Lag \( L - 1 \), and the \( k \) Lag \( L \). Thus,

\[
\sum_j t_{ijk} = 1
\]

for all values of \( i \) and \( j \) (except at Lag 2, in which case \( t_{ijk} = 0 \) when \( i = j \) because Lag 0 and Lag 1 codes cannot be the same; similarly, at all lags \( t_{ijk} = 0 \) when \( j = k \) because Lag \( L - 1 \) and Lag \( L \) codes cannot be the same). Accordingly, the first set of transitional probability parameters specified for Lag 2 was

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>i: A</td>
<td>( j = A )</td>
<td>0</td>
<td>.25</td>
</tr>
<tr>
<td>B</td>
<td>.25</td>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>C</td>
<td>.50</td>
<td>.50</td>
<td>0</td>
</tr>
</tbody>
</table>

and for lags greater than 2 was

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>i: A</td>
<td>( j = A )</td>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>B</td>
<td>.50</td>
<td>0</td>
<td>.50</td>
</tr>
<tr>
<td>C</td>
<td>.50</td>
<td>.50</td>
<td>0</td>
</tr>
</tbody>
</table>

Thus, the last three sets (Lag 1, Lag 2, and greater than Lag 2) specify an elevated value for \( p(C_L|A_0) \). Here, transitional probabilities that were not structurally zero were .75 when associated with \( p(C_L|A_0) \), .25 for other nonzero entries in a row containing a .75, and .50 otherwise.

Again, we generated six sets of 1,000 sequences, each set reflecting a different effect (no effect, a Lag 1 effect, etc. through a Lag 5 effect). This time length was set at 200 events because with overlapped sampling, \( K = 3 \) and three-dimensional tables are required for the \( 0 \mid 1 \) \( L - 1 \) test; this assures a reasonable number of tallies per cell (e.g., 195 tallies when \( L = 5 \) for the 18 cells not structurally zero that occur when lag is greater than 2). Then we tested for a Lag 1 through Lag 5 effect (0 \( \| \) 1, 0 \( \| \) 2 \( \| \) 1, 0 \( \| \) 3 \( \| \) 2, etc.) with the appropriate partial \( G^2 \). Proportions exceeding their critical value were

<table>
<thead>
<tr>
<th>L</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>E: 0</td>
<td>.05</td>
<td>.06</td>
<td>.07</td>
<td>.06</td>
<td>.06</td>
</tr>
<tr>
<td>1</td>
<td>.70</td>
<td>.06</td>
<td>.06</td>
<td>.08</td>
<td>.06</td>
</tr>
<tr>
<td>2</td>
<td>.22</td>
<td>.40</td>
<td>.13</td>
<td>.06</td>
<td>.05</td>
</tr>
<tr>
<td>3</td>
<td>.06</td>
<td>.14</td>
<td>.59</td>
<td>.13</td>
<td>.10</td>
</tr>
<tr>
<td>4</td>
<td>.07</td>
<td>.06</td>
<td>.18</td>
<td>.58</td>
<td>.14</td>
</tr>
<tr>
<td>5</td>
<td>.04</td>
<td>.09</td>
<td>.08</td>
<td>.12</td>
<td>.60</td>
</tr>
</tbody>
</table>
As with the consecutive-codes-may-repeat simulation, when no effect was specified \((E = 0)\), values were close to the theoretically expected .05; when an effect was specified, the null hypothesis was often rejected at the appropriate lag \((E = L; \text{different values for } \alpha \text{ would, of course, result in different proportions exceeding the critical value}).

However, unlike the first consecutive-codes-may-repeat simulation, whose transitional probabilities were likewise intended to elevate \(p(C_L | A_0)\), effects intended to create a Lag \(L\) effect also induced effects at adjacent lags (e.g., at \(L = 1, 2, 3\) and when \(E = 2\)). This phenomenon is apparent in the previous table (and was evident in other simulations run with different values for the transitional probabilities). It is a consequence of the constraint that consecutive events cannot be assigned the same code and should be taken into account when interpreting results of analyses such as those described here.

To understand why efforts to induce a Lag \(L\) effect cannot be contained to Lag \(L\), consider Lag 2. We set transitional probabilities to encourage a C at Lag 2 after an A at Lag 0. This can only be accomplished, however, by the A-B-C chain. By making A-B-C more likely, we thereby make B-C more likely, thus inducing a (albeit weaker) Lag 1 effect. Similarly, after A-B-C only an A or B can follow, so a (likewise weaker) Lag 3 effect is also induced. Similar arguments apply to other lags.

For the simulations described earlier, \(K = 3\), which can be viewed as a limiting value when consecutive codes cannot repeat. When \(K = 2\), the sequence is totally determined (it can only be A B A B . . . ). When \(K = 3\), the sequence is maximally constrained in the sense that one third of the possible codes are not permitted to the following event: the limitation is one fourth when \(K = 4\), one fifth when \(K = 5\), and so forth, ever decreasing as \(K\) increases. The examples presented in this article have assumed \(K = 3\) for two reasons. First and most obviously, when \(K = 3\) and not some larger value, examples are simpler and can be presented with fewer numbers. Also, because 3 usually represents a limiting value, problems that might be obscured with larger values of \(K\) should be more apparent when \(K = 3\). Thus, if matters appear reasonable when \(K = 3\), one can be reasonably confident that results should generalize to higher values—values that are far more frequent in the literature than the 3.

Similarly, problems that appear when \(K = 3\) (such as the contamination of adjacent lags noted earlier) may diminish or disappear for higher values. To examine this possibility, we generated an additional six sets of 1,000 sequences. This time \(K = 4\), and transitional probabilities that were not structurally zero were .50 when associated with \(p(C_L | A_0)\), .25 for other nonzero entries in a row containing a .50, and .33 otherwise. Length was set to 500 events because this assures a reasonable number of tallies per cell (e.g., 495 tallies when \(L = 5\) for the 48 cells not structurally zero that occur when lag is greater than 2). This time, proportions exceeding their critical value were

\[
\begin{array}{c|ccccc}
L: & 1 & 2 & 3 & 4 & 5 \\
E: 0 & .06 & .06 & .05 & .06 & .07 \\
1 & .62 & .06 & .05 & .06 & .05 \\
2 & .08 & .36 & .07 & .06 & .05 \\
3 & .05 & .08 & .46 & .08 & .06 \\
4 & .05 & .06 & .10 & .44 & .08 \\
5 & .05 & .07 & .06 & .09 & .47 \\
\end{array}
\]

which suggests some slight induction of effects at adjacent lags but not much more than was seen when consecutive codes may repeat (and, as noted earlier, values using \(X^2\) were smaller). Moreover, a similar run with \(K = 5\) showed somewhat less evidence of induction. We conclude that the contamination of adjacent lags when attempting to induce an effect at Lag \(L\) occurs mainly when \(K = 3\) and otherwise is probably not problematic.

Finally, to evaluate the ad hoc descriptive procedures described earlier, for each sequence at each lag two sets of \(z\) scores evaluating \(p(C_L | A_0)\) were computed. For the first set, \(z\) scores at Lag 1 were based on expected frequencies per GS1 computed per ALZ and at lags greater than 1, GS2 + ALZ (these procedures were described earlier). Proportions of these scores exceeding 1.96 absolute were

\[
\begin{array}{c|ccccc}
L: & 1 & 2 & 3 & 4 & 5 \\
E: 0 & .01 & .03 & .02 & .03 & .03 \\
1 & .88 & .68 & .04 & .08 & .03 \\
2 & .02 & .35 & .13 & .03 & .05 \\
3 & .13 & .44 & .73 & .47 & .06 \\
4 & .03 & .02 & .15 & .48 & .24 \\
5 & .06 & .08 & .07 & .36 & .59 \\
\end{array}
\]

For the second set, \(z\) scores at Lag 1 were adjusted residuals computed per Haberman (1979, p. 454) and at lags greater than 1 were based on CIS + ZCT. Similar proportions for these scores were

\[
\begin{array}{c|ccccc}
L: & 1 & 2 & 3 & 4 & 5 \\
E: 0 & .04 & .01 & .01 & .03 & .02 \\
1 & .70 & .05 & .00 & .01 & .01 \\
2 & .21 & .20 & .06 & .02 & .03 \\
3 & .04 & .06 & .52 & .14 & .02 \\
4 & .06 & .02 & .14 & .56 & .10 \\
5 & .04 & .03 & .02 & .16 & .56 \\
\end{array}
\]

When no effect was specified \((E = 0)\), the observed proportion for the Lag 1 adjusted residual was .04, almost the expected .05. Proportions for other computations when \(E = 0\) (based on GS1 at Lag 1 and GS2 and CIS at other lags), which as mentioned lack the statistical grounding of adjusted residuals, were all less than .05 and so conservative, with CIS + ZCT appearing somewhat more conservative than GS2 + ALZ.

In general, conservatism is manifested by a lower Type I error rate when no effect is specified (fewer false claims) but by a higher Type II rate when effects are specified (more missed effects). Here CIS + ZCT was more conservative than GS2 + ALZ when no effect was specified and remained so when effects were specified. When a Lag 1 effect was specified, for example, the Type II error rate was .30 for CIS + ZCT and .12 for GS2 + ALZ. Not surprisingly, the greater liberality of GS2 + ALZ has a price. It claimed A to C effects at adjacent lags more often than CIS + ZCT and more often than might be expected, given the proportion of adjacent (induced) effects noted earlier for
the overall tests based on the appropriate partial $G^2$ (similar patterns were noted when $K = 3$ and 5).

**Number of Events Required for Tests of Association**

In previous sections, we mentioned the importance of sufficient data but did not quantify sufficient exactly. Several considerations play a role, so absolute guidelines are as difficult to define as they are desired. Summarizing current advice, Wickens (1989) noted that (a) expected frequencies for two-dimensional tables with 1 degree of freedom should exceed 2 or 3 but that with more degrees of freedom some expected frequencies may be near 1 and with large tables up to 20% may be less than 1, (b) the total sample should be at least 4 or 5 times the number of cells (more if marginal categories are not equally likely), and (c) similar rules apply when testing whether a model fits a three- or larger dimensional table (p. 30). Consider implications for $N_e$ (the total sample of chains tallied). As noted earlier, the number of cells is $K^{L+1}$ when consecutive codes may repeat and $K(K - 1)^L$ when they cannot. As $L$ increases, these numbers can become discouragingly large. However, if attention is limited to three-dimensional $0(L - 1)1$ tables, as suggested earlier, then the number of cells is no greater than $K^3$ when consecutive codes may repeat and $K^3(K - 1)$ when they cannot.

Still, considerable data may be required for a log-linear analysis. Following the strategy that limits attention to three-dimensional tables, $K$ is the determining factor. For example, numbers of cells when $K = 3, 5, 7, 9$ and consecutive codes may repeat are

<table>
<thead>
<tr>
<th>$L$:</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$:</td>
<td>3</td>
<td>9 27</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>25 125</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>49 343</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>81 729</td>
</tr>
</tbody>
</table>

(values for higher lags are the same as when $L = 2$; the general formula, derived from those given earlier, is $K^3$ when $L = 1$ and $K^3$ when $L = 2$ or higher) and when consecutive codes cannot repeat, numbers of cells (excluding structural zeros) are

<table>
<thead>
<tr>
<th>$L$:</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$:</td>
<td>3</td>
<td>6 12 18</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>20 80 100</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>42 252 294</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>72 576 648</td>
<td></td>
</tr>
</tbody>
</table>

(values for higher lags are the same as when $L = 3$; the general formula, derived from those given earlier, is $K(K - 1)$ when $L = 1$, $K(K - 1)^2$ when $L = 2$, and $K^2(K - 1)$ when $L = 3$ or higher). Taking into account only the times-the-number-of-cells rule, to compute $N_e$ these values should be multiplied by 4, 5, or whatever factor seems justified; although, as Bakeman and Robinson (1994) noted, such guidelines should not be regarded as minimal goals to be satisfied but as troubled frontiers from which as much distance as possible is desired. If the numbers still seem onerous, reconsider expected frequencies and note that Haberman (1977; cited in Wickens, 1989, p. 30) suggested that requirements for tests based on the difference between two chi-squares—such as the tests of $0 \mid L \mid L - 1$—may depend more on frequency per degree of freedom than on the minimum expected cell size. Still, when doubts arise, it may be best to consult local experts.

**Summary and Recommendations**

A log-linear approach to lag-sequential problems has many advantages. Primarily, it integrates methods that researchers who focus on sequential aspects of behavior have found useful into an established and well supported statistical tradition. Moreover, it provides a number of solutions to problems that have long troubled investigators and, at the same time, yields a more differentiated understanding of the sequences studied than the usual lag-sequential analysis.

Traditionally, lag-sequential analysis has focused on bivariate relations between particular behaviors at Lag 0 and Lag L and has gauged those relations with an Allison and Liker (1982) z score. A log-linear approach offers two useful refinements. First, it emphasizes a preliminary omnibus test (Is the tablewise association large?) which, absent specific predictions, should be required before any exploration of cellwise associations takes place. Second, it offers a statistic—the adjusted residual—that is similar but not identical to the usual lag-sequential z score yet has the advantage of a firm statistical base and broader application (e.g., it applies no matter if sampling is overlapped or not).

Moreover, structural zeros are routine to log-linear analysis and not troublesome as in the lag-sequential tradition, which proves useful when consecutive codes both may and cannot repeat. When consecutive events cannot be coded identically, iterative procedures provide better estimates for expected frequencies than ad hoc equations such as GS1; moreover, the log-linear literature provides estimates of adjusted residuals, which in this case essentially elude lag-sequential analysis (if GS1 is ad hoc, then GS1 + ALZ is doubly ad hoc). In addition, no matter whether consecutive codes may or cannot repeat, declaring cells that contain large adjusted residuals structurally zero and then testing whether the appropriate log-linear model now fits the modified data provides a bulwark against Type 1 errors. Indeed, the ability to reduce the number of cellwise associations caught in our initial web to those fundamentally responsible for the effect may be the log-linear approach’s single most important contribution to the analysis of event sequences.

The web of interconnected effects may extend both within and across time: within in the sense noted earlier as some Lag L effects may be induced by others and across in the sense that some Lag L effects may induce others at subsequent lags (a phenomenon that reaches its zenith when consecutive codes cannot repeat and $K = 3$). Here, too, a log-linear analysis yields a more textured view than traditional methods. Particularly advantageous is the ability to test conditional independence; lag-sequential analyses typically test whether events at Lag 0 are associated with Lag L unconditionally, ignoring any effect intermediate lags may have, whereas log-linear analyses may test whether events at Lag 0 are associated with Lag L over and above any effects due to intermediate lags (analogous to hierarchical
LOG-LINEAR APPROACHES TO LAG-SEQUENTIAL ANALYSIS

multiple regression). Thus, as demonstrated earlier, researchers may learn that what appeared to traditional lag-sequential analysis as a (unconditional) Lag 0–Lag L dependence may be in fact a Lag 0–Lag L independence conditioned on Lag L = 1.

Additionally, log-linear methods may be used to test the depth of association (Are current events affected only by the immediately previous one, the previous two, etc.?), as Altmann (1965) did several decades ago with information theory measures (Attneave, 1959) and as Cohn and Tronick (1987) did more recently using log-linear analysis. When tallies result from overlapped sampling, such analyses can be simplified somewhat; in such cases usually if the [01] term is required for a fitting model, so too is the [12] term and any terms that likewise indicate a Lag 1 effect. This reduces the number of models that reasonably need be considered.

When log-linear analyses are used, they differ from traditional lag-sequential approaches in requiring more data at lags greater than 1. If the sequential search strategy described earlier is followed, however, which limits consideration to tables no larger than three dimensions, then cell size need not exceed \( K^3 \) when consecutive codes may repeat (\( L > 1 \)) and \( K^4 (K - 1) \) when they cannot (\( L > 2 \)). The limiting factor is \( K \), and clearly the approach described here discourages analyses for which \( K \) is large. Note, however, it does not discourage studies but only analyses for which \( K \) is large. An earlier sequential analysis program (Bakeman, 1983) limited \( K \) to 20, which we might argue is appropriate. Often the codes observers record make finer distinctions than are required for meaningful analyses (Bakeman & Gottman, 1986); moreover, codes may represent several different domains. Thus, separate initial analyses for different domains and lumping of similar codes within domains—both of which limit \( K \)—usually are appropriate analytic strategies.

It is never wise to ignore the need for sufficient data. Even when \( p \) values are not an issue, there is little reason to place much confidence in a descriptive statistic based on only a few data points; this is as true for lag-sequential as for log-linear analysis. Still, the amount of data required by these two approaches, as seen through a log-linear lens, highlights their differences. First, even the two-dimensional tables of lag-sequential analysis contain \( K^2 \) cells, which suggests that, given reasonably equally distributed marginals, sufficient data are required so that at least 4 to 5 times this number of chains are tallied. Second, without the three-dimensional tables required for the suggested log-sequential investigations of lags greater than 1, the unconditional conclusions of the usual lag-sequential approach ignore effects associated with intermediate lags—and it is the data required by this extra dimension that permit the more complex understanding log-sequential analyses provide.

From this point of view, earlier lag-sequential studies are possibly limited but probably not wrong. The comparisons made here between adjusted residuals based on iteratively derived expected frequencies and GS1 + ALZ and between CIS + ZCT and GS2 + ALZ do not suggest that traditional lag-sequential procedures have led investigators seriously into Type I errors, at least at the level of individual tests (studies like Type I error rates are another matter, but problems associated with multiple tests are hardly unique to lag-sequential studies). If anything, formulas such as GS2 are overly conservative. True, our attempt at an alternative to GS2 was even more conservative but would nonetheless be used given nonoverlapped sampling. In any event, descriptive devices like GS1 + ALZ and CIS + ZCT (and adjusted residuals generally) are best regarded as hints, suggesting where structural zeros should be placed for subsequent follow-up log-linear analyses.

Throughout this article, we attempted to convey the generality of the log-linear approach (which easily accommodates structural zeros and applies whether or not sampling is overlapped) and to recast traditional lag-sequential practice in log-linear terms. However, often the best approach is to cast a particular analytic endeavor in log-linear terms in the first place. Iacobucci and Wasserman (1988) showed how this can be done for a number of dyadic problems, whereas Wickens (1989) suggested a wide range of more general strategies. One matter left unsettled concerns the trustworthiness of \( p \) values associated with various results. Randomization (or sampled permutation) tests as suggested by Rechten and Fernald (1978) yield exact \( p \) values; consequently, they might provide a welcome alternative to \( p \) values based on the asymptotic assumptions of the usual applications of chi-square, especially when samples are small or other questions (such as those associated with overlapped sampling) arise. In subsequent work, we plan to explore this possibility further.

References


Bakeman, R., & Gottman, J. M. (1986). Observing interaction: An in-
Introduction to sequential analysis. New York: Cambridge University Press.


Received November 10, 1993
Revision received March 5, 1995
Accepted March 17, 1995