Sociological Methods & Research
http://smr.sagepub.com

# Linked Markov Sources: Modeling Outcome-Dependent Social Processes

Cees H. Elzinga, Adriaan W. Hoogendoorn and Wil Dijkstra
*Sociological Methods Research* 2007; 36; 26
DOI: 10.1177/0049124106295975

**Sociological Methods & Research**
Volume 36 Number 1
August 2007 26-47

10.1177/0049124106295975
http://smr.sagepub.com
hosted at
http://online.sagepub.com

# Linked Markov Sources

## Modeling Outcome-Dependent Social Processes

Cees H. Elzinga
Adriaan W. Hoogendoorn
Wil Dijkstra
*Vrije Universiteit Amsterdam*

Many social processes are adaptive in the sense that the process changes as a result of previous outcomes. Data on such processes may come in the form of categorical time series. First, the authors propose a class of Markov Source models that embody such adaptation. Second, the authors discuss new methods to evaluate the fit of such models. Third, the authors apply these models and methods to data on a social process that is a preeminent example of an adaptive process: (encoded) conversation as arises in structured interviews.

***Keywords:*** *Markov models; model selection; Markov model fit; survey interview conversations*

Categorical time series are a quite common kind of data in many branches of the social sciences: They come in the form of encodings of life histories, interview conversations, children's play activities, job histories, records of criminal recidivism, negotiation protocols, and so forth. Such data basically consist of sequences of the form $\mathbf{o}_\ell = o_1 \dots o_\ell$ of encoded events of length $\ell$. Each event is encoded by a category from a set of observable categories $O = \{u, v, w, \dots\}$, that is, $o_t \in O$ for each $1 \leq t \leq \ell$. Throughout this article, we will assume that $O$ contains $m$ distinct categories, that is, $|O| = m$. A simple model for such time series is

$$\begin{cases} P(o_1 = u) = \delta_{1,u} \\ P(o_{t+1} = v | o_t = u) = \tau_{v|u}. \end{cases} \tag{1}$$

Such a model is called a first-order, time-homogeneous Markov process or Markov Chain with parameter set $\lambda = \{\boldsymbol{\delta}_1, \mathbf{T}\}$ consisting of an initial probability distribution $\boldsymbol{\delta}_1 = (\boldsymbol{\delta}_{1,u}, \boldsymbol{\delta}_{1,v}, \dots)$ of observing elements of $O$ at

$t = 1$ and a $(m \times m)$-matrix $\mathbf{T} = \{\tau_{u|v}\}$ of conditional transition probabilities. For $t > 1$, we have the marginal distribution $\boldsymbol{\delta}_t = \boldsymbol{\delta}_1 \mathbf{T}^{t-1}$. The likelihood of observing a particular sequence $\mathbf{o}_\ell \in O^\ell$ is given by

$$P(\mathbf{o}_\ell|\lambda) = \boldsymbol{\delta}_{o_1} \cdot \prod_{t=2}^{\ell} \boldsymbol{\tau}_{o_t|o_{t-1}}. \tag{2}$$

The chain is called first-order because events before $t$ do not condition the probabilities at $t+1$ and time-homogeneous because the probabilities do not change with time. Because each row of $\mathbf{T}$ is a conditional distribution, they must sum to 1 so that equation (1) contains $(m-1) + m(m-1)$ free parameters. At least since the 1950s, social scientists have used Markov models to account for a wide variety of social phenomena, such as voting, social mobility, and migration (e.g. Miller 1952; Prais 1955).

Whatever substantive theory leads to the hypothesis of a Markovian time series of observables and regardless of the fit of such a model, a Markov chain cannot formalize the mechanism that generates the data. Markov chains merely summarize data. Oftentimes, social scientists try to understand the observable time series in terms of an underlying mechanism that is not (directly) observable and that often has no physical correlate. A quite natural extension of a Markov Chain then arises by assuming a system that occupies either one of a set $S = \{i, j, k, \ldots\}$ of not directly observable states and that the system, being in a state $s_t \in S$ at $t$, "emits" an observable $o_t$. Assuming that the system switches states according to a Markov Chain, this amounts to the model

$$\begin{cases} P(s_1 = i) = \delta_{1,i} \\ P(s_{t+1} = j|s_t = i) = \tau_{j|i} \\ P(o_t = u|s_t = i) = \rho_{u|i} \ . \end{cases} \tag{3}$$

The model embodied in equation (3) is, depending on the substantive area of application, called a Markov Source, a Hidden Markov Model, or a Latent Markov Model. As stated previously, the model is first-order and time homogeneous. The model has the parameter set $\lambda = \{\boldsymbol{\delta}_1, \mathbf{T}, \mathbf{P}\}$, where $\mathbf{P} = \{\boldsymbol{\rho}_{u|i}\}$ is a $(n \times m)$ matrix of emission probabilities. If we take $|S| = n$, the model has $(n-1) + n(n-1) + n(m-1)$ free parameters. Analogous to the equations for the Markov Chain, we have $\boldsymbol{\delta}_t = \boldsymbol{\delta}_1 \mathbf{T}^{t-1}$, the marginal distribution of observables equals $\boldsymbol{\delta}_t \mathbf{P}$, and

$$P(\mathbf{o}_\ell|\lambda) = \boldsymbol{\delta}_{s_1} \cdot \prod_{t=2}^{\ell} \boldsymbol{\tau}_{s_t|s_{t-1}} \cdot \prod_{t=1}^{\ell} \rho_{o_t|s_t}. \tag{4}$$

Markov Sources have been successfully applied in fields as diverse as bio-informatics, 3-D pattern recognition, and the social sciences. In the social sciences, Markov Sources have found their place within the framework of Latent Class Analysis (e.g., Langeheine and Van de Pol 1990, 2002), where it has been generalized to accommodate time- and population-heterogeneity, and it has found wide application in the study of learning and forgetting (e.g. Visser, Raijmakers, and Molenaar 2002). A political science application is provided by Schrodt (2000), who modelled event sequences in negotiations. Bijlevelt and Mooijaart (2003) modelled criminal recidivism data.

However, we know that many social systems are adaptive, that is, their behavior changes as a result of previous outcomes. For example, we know that criminal careers may drastically change after a spell of imprisonment and that job careers may be seriously affected—"scarred"—by a spell of unemployment (e.g., Arulampulan and Gregg 2001). Similarly, the course of a conversation or negotiation may take quite a turn as a result of what either participant says. Such output dependencies cannot be adequately modelled with a Markov Source. So these examples illustrate the need for a class of simple models that do incorporate such output dependencies. If one wants to incorporate output-dependencies in a Markov Source, this can be done in several ways. One way is to have transition matrices depend on the previous observable. This would, for example, formalize the idea that (aspects of) a life history are "caused" by a plan or design of the subject (e.g., Giddens 1991) and the subject is then supposed to adapt this design according to his evaluation of events so far. Such an adaptation would be formalized by stating that $P(s_{t+1} = j | s_t = i, o_t = u) = \tau_{j|i,u}$. On the other hand, consider negotiating parties who decide to change their style of expressing but not to adapt their goals or strategy. This would probably be formalized as $P(o_{t+1} = v | s_{t+1} = i, o_t = u) = \rho_{v|i,u}$. A general model where both the underlying Markov Chain and the emitting mechanism is linked to previous outputs is given by

$$\begin{cases} P(s_1 = i) = \delta_{1,i} \\ P(s_{t+1} = j | s_t = i, o_t = u) = \tau_{j|i,u} \\ P(o_1 = v | s_1 = i) = \rho^*_{v|i} \\ P(o_{t+1} = v | s_{t+1} = i, o_t = u) = \rho_{v|i,u}. \end{cases} \tag{5}$$

We call equation (5) a Linked Markov Source (LMS), as transitions and emissions are linked to previous outputs. The LMS has quite a lot more parameters than the Markov Source, as we need a $(n \times m)$-matrix $\mathbf{P}^* = \{\rho^*_{u|i}\}$ of initial emission probabilities and separate transition and emission matrices for each observable $u \in O$. The LMS has parameter set $\lambda = \{\boldsymbol{\delta}_1, \{\mathbf{T}_u\}, \{\mathbf{P}^*, \mathbf{P}_u\}\}$ contains $(n-1) + mn(n-1) + n(m-1) + mn(m-1)$ free parameters. Because of the output dependencies, expressions for the marginal distributions of states and observables at $t$ must now be conditioned on the sequence $\mathbf{o}_{t-1}$ and the likelihood of a particular sequence of observables is provided by

$$P(\mathbf{o}_\ell|\lambda) = \delta_{s_1} \cdot \rho^*_{o_1|s_1} \cdot \prod_{t=2}^{\ell} \tau_{s_t|s_{t-1},o_{t-1}} \cdot \prod_{t=2}^{\ell} \rho_{o_t|s_t,o_{t-1}}. \tag{6}$$

Attempts to model output dependies are certainly not new. Matras (1967) already suggested a Markov migration model in which state transition probabilities depend on population density. Conlisk (1976, 1978) formalized this model, and Bartholomew (1982) discussed several variants thereof. But to the best of our knowledge, no such attempts have been undertaken within the realm of Markov Source models.

The parameters of an LMS can be estimated using the EM-algorithm (Dempster, Laird, and Rubin 1977). We implemented slightly adapted forward and backward algorithms (Baum, Petrie, Soules, and Weiss, 1970) in combination with Devijver's (1985) scaling algorithm for the E-step. By setting up the appropriate Lagrangians and differentiating, one obtains the maximizing expressions for the M-step. These methods are explained in, for example, Rabiner (1989) and Clote and Backofen (2000).

## An Application: Cognitive Models of Survey Interviews

Conversations can be seen as categorical event sequences and a relatively simple kind of conversation arises in the context of survey interviewing. It is not surprising that attempts to model encoded interviews with a Markov Chain were already published in the 1970s (Hawes and Foley 1972; Jaffe and Feldstein 1970). Furthermore, a conversation is a preeminent example of a social process in which present events are codetermined by previous events. Therefore, encoded transcripts of structured survey interviews are an obvious type of data to fit with an LMS. Ideally such sequences have a very simple structure: The interviewer (I) poses a question and explains the format of the admissible response alternatives (IQ), and the respondent (R) answers by picking one of these (RA) on which the I ends the conversation explicitly or implicitly (by posing the next question from the survey) through accepting the answer given (IE). So the resulting encoded sequence should have the form of IQ, RA, and IE, and such a sequence is called a paradigmatic sequence (Schaeffer and Maynard 1996).

However, in practice and even when the interviewers are well trained and the questions have been thoroughly tested, many sequences depart from the paradigmatic one and are much longer because of misunderstandings, clarifications, social talk, inadequate behaviour, and so forth. This fact has lead to an abundance of research into questioning and answering behaviour with the objective of understanding the functioning of the interview as a measurement device. This research, in turn, has given rise to a variety of conceptual frameworks, inspired by cognitive psychology, that presume different cognitive processes involved in answering a survey question (i.e., comprehension of question, retrieval of information, judging the retrieved information and formatting the response; e.g., Tourangeau, Rips, and Rasinsky 2000). Bradburn (2004) provides for a concise review of the current state of affairs in this research. What these theories roughly amount to is that the switching between distinct mental states or cognitive processes—processing and formatting—determines the utterances of I and R and thus may account for the nonparadigmatic sequences occurring so frequently. Because such cognitive processes are naturally represented as the unobservable states of a Markov Source and because previous events in an interview are believed to determine events that follow, fitting an LMS to such data may be considered as a rough empirical test of the cognitive theoretical framework as employed in research on survey questioning methodology.

Because survey conversations end by an act of the I after a varying number of events, it seems logical that this (perhaps nonverbal) event is generated from an absorbing state of the Markov Source that models the cognitive processes going on. Furthermore, because it is plausible that an utterance from, say I, is the exclusive result of a mental state or process of I and not of R, an emission matrix $\mathbf{P}_u$of the mimicking LMS should have the structure

$$\mathbf{P}_u = \begin{pmatrix} \mathbf{P}_{u_I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{u_R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{pmatrix},$$

that is, I-utterances stem from I-states and R-utterances stem from R-states exclusively. Imposing these constraints on the emission matrices of the LMS has the consequence that some of the rows of the transition matrices $\mathbf{T}_u$ will be undefined. For example, consider the event $o_t = u$ to be an R-utterance. Then, $s_t = i$ must be an R-state, hence the conditional probability

**Table 1**
**Encoding Scheme of Transcripts of Structured Interviews**

| Code | Description | Code | Description |
|---|---|---|---|
| Ia | Correctly posing question, corrective comment | Ra | Admissible answer |
| In | Suggestive question or accepting inadmissible answer | Rn | Inadmissible answer |
| Ip | Clarifies question, supports or stimulates thinking | Rp | Asks for clarification, motivates choice, qualifies question, etc. |
| Ii | Talks about irrelevant subjects | Ri | Talks about irrelevant subjects |
| Ie | (Implicitly or explicitly) accepts answer given, that is, ends sequence | | |

Note: The data are those described in Smit and Neyens (2000). I = interviewer; R = respondent.

of the system switching from an I-state to any other state given the R-utterance $u$ is undefined: $\boldsymbol{\tau}_{j|i,u}$ is undefined because the conditioning event simply cannot take place. Hence, if the lower rows of $\mathbf{T}_u$ pertain to R-states, the structure of $\mathbf{T}_u$ must be of the form

$$\mathbf{T}_u = \begin{pmatrix} \mathbf{X}_{u_R} \\ \mathbf{T}_{u_R} \\ \mathbf{0} \quad 1 \end{pmatrix},$$

wherein $\mathbf{X}_{u_I}$ is undefined. Similarly, the same holds if $u$ is an I-utterance.

We used the transcripts of 6,635 structured interviews on TV watching and commercials as amply described in Smit and Neyens (2000). These interviews were encoded using the methods explained in Dijkstra (1999) in combination with the scheme of Table 1. With this encoding method, a sequence may look like

Ia Rp Ia Rn In Rn.

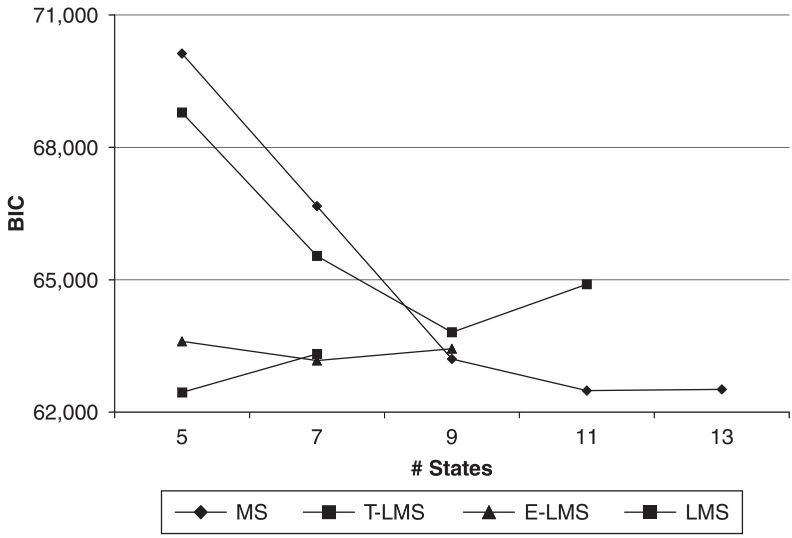
Because all sequences appeared to start with the code “Ia,” we removed this code from all sequences, as it does not make sense to probabilistically model an event that is certain to happen. Because sequences of the form “Ia Ra Ie” (paradigmatic sequences) do not provide any information about cognitive processes, we removed all such sequences from the data. This left us with 5,825 sequences, constructed with 9 codes. Some details about these data are presented in Appendix A.

To these data, an MS and an LMS were fitted plus two restricted versions of an LMS. In the restricted versions of the LMS, linkage of emissions at $t$ was constrained to either the transmissions or the emissions. This amounts to a transmission-linked LMS (T-LMS) wherein $\boldsymbol{\rho}_{u|i,v} = \boldsymbol{\rho}_{u|i}$ for all $v \in O$ and an emission-linked version (E-LMS) wherein $\boldsymbol{\tau}_{j|i,v} = \boldsymbol{\tau}_{j|i}$ for all $v \in O$. Because it is well known (Wu 1983) that an EM-algorithm does not necessarily converge to the global maximum of the log-likelihood function, we ran each model with at least 50 initial configurations $\boldsymbol{\lambda}^0$ and took the optimal solution for a close approximation of the global solution. At all models, this same optimum was attained many times.

In Figure 1, we show the Bayesian Information Criterion (BIC; e.g., Weakliem 2004) as obtained for various models (MS, T-LMS, E-LMS, and LMS with different numbers of hidden states restricted to the models where the number of I-states equals the number of R-states). If, for example, the number of states equals 7, this implies that the model has 3 I-states and 3 R-states plus an absorbing state. We did investigate models with unequal numbers of I- and R-states, but these produced a worse fit than the models shown in Figure 1. As soon as, for each particular model, the BIC starts to increase again with an increasing number of states, we stopped, so the graphs in Figure 1 do not all cover the same range. The picture shows that LMS with 2 I- and 2 R-states fits best. In Table 2, the details of the results of the parameter estimation are shown for the best fitting model in each of the classes.

Evidently, the complete LMS model fits best; therefore we present the estimated parameters of that model in Appendices B and C. The full LMS model only provides a rough rendering of the sophisticated cognitive frameworks that are presently used to describe the question-answering process. Moreover, the way of encoding the transcripts of course affects the fit of the models. A more detailed investigation of the models, both from a statistical and a substantive point of view, is beyond the scope and purpose of the present article. Therefore, we do not attempt to test the hypothesis that the LMS as fitted is the true model in a statistical sense. This would, given the extreme sparseness of the data, require a fair amount of Monte Carlo bootstrapping (e.g., Langeheine and Van de Pol 1996). Thus, the variance estimation is problematic because of its computational burden. Instead, in the next section, we will discuss methods that allow us to investigate how well the fitted model is capable of predicting several relevant aspects of the data.

**Figure 1**
**Graphs of BIC Against the Total Number of States (# States) for Four Different Models**




Note: BIC = Bayesian information criterion; MS = Markov Source; LMS = Linked Markov Source; T-LMS = transmission-linked LMS; E-LMS = emission-linked LMS. Details of models and parameter estimation are in the main text. Detailed results of the fit of the best fitting LMS models are in Table 2. The authors thank Rob de Vries and his staff for providing a PC lab to obtain these results.

Although it is not our intent to present a model that fully renders modern concepts of cognitive processes in question-answering behavior, a few remarks on an interpretation of some states of the model are in place. Typically, as a starting point for such an interpretation, one begins looking at the emission matrices because these parameters connect the unobservable states to the observable world. These matrices are shown in Appendix C. The reader should be aware that turn-taking is quite pronounced, as can be seen from Appendix A. So if the conditioning event is an I-utterance, the next event will probably be an R-utterance and vice versa. With this in mind, we look at the difference between the emission matrices that are conditional on Ia and In. The rightmost parts of these matrices show the emission probabilities of R-utterances from the two R-states. Let us call the upper one of these states R1 and the lower one R2. The reader observes that

**Table 2**
**Fit Statistics for LMS Models Fitted to Interview Sequences**

| Model | # States | *df* | log − *L* | $G^2$ | BIC | AIC | Runs |
|---|---|---|---|---|---|---|---|
| MS | 11 | 139 | –30,535.8 | 61,071.69 | 62,488.98 | 61,349.69 | 100 |
| T-LMS | 9 | 287 | –30,442.0 | 60,883.98 | 63,810.33 | 61,457.98 | 100 |
| E-LMS | 7 | 203 | –30,551.0 | 61,101.92 | 63,171.78 | 61,507.92 | 100 |
| LMS | 5 | 175 | –30,333.3 | 60,666.64 | 62,451.00 | 61,016.64 | 100 |

Note: MS = Markov Source; T-LMS = transmission-linked LMS; E-LMS = emission-linked LMS; LMS = Linked Markov Source. BIC = Bayesian information criterion; AIC = Akaike information criterion. For details about the models, data, and parameter estimation, we refer to the text. $G^2$ denotes the log-likelihood chi-square.

if the conditioning event is Ia, the likelihood of the next observable being Ra is much higher if R is in state R2 that when R is in state R1. But this pattern dramatically changes when the conditioning event is In: Then suddenly, Ra is much more likely to stem from R1. This invites for interpreting R1 as a “formatting state” and R2 as a “retrieving/processing state.” This interpretation is corroborated by the fact that the likelihood of Rp is quite pronounced if the conditioning event is Ip. Because of lack of space, we will not dwell on the issue of providing for a much more thorough and complete interpretation of the model and its many parameters. Instead, in the next sections, we will turn our attention to some methods of model evaluation.

## Expected Distributions of Sequence Lengths

In rendering complex theoretical frameworks through relatively simple Markov models, one can hardly expect that the rendering is perfect, even if the data were perfect. Therefore, testing the hypothesis that the fitted model is the true model may not be very productive, for if that hypothesis is rejected indeed, it may not be so clear what the alternative is. So perhaps a better question is whether the best-fitting model is a useful model. A useful model should do a nice job in predicting certain aspects of the data that are substantially relevant (too). Such an aspect then must be an aggregate of the raw data. For example, in the present context of long, nonparadigmatic sequences being modelled, it is interesting to see how good or bad the fitted model predicts the observed sequence lengths. So sequence length could be a relevant data aggregate. In the present section, we will provide an algorithm that is useful to derive expected distributions of many more different data aggregates than just sequence lengths. In the next sections, we will demonstrate how this type of algorithm can be used to evaluate expected distributions of other data aggregates as well.

The algorithm that we present here is an extension of the famous forward algorithm of Baum et al. (1970). It allows for recursive computation of the extended forward variable $\alpha_t(i,u) = P(s_t = i, o_t = u)$, that is, the probability that the system will be in state $i$ and emit observable $j$ at some point in time $t \geq 1$. Obviously, we have for $t = 1$, $\alpha_1(i,u) = P(s_1 = i, o_1 = u) = \delta_{1,i}\rho^*_{u|i}$, and this quantity is directly computable from the estimated parameters $\boldsymbol{\lambda}$. By using the rules of conditional probability, it is not difficult to derive the recursive formula

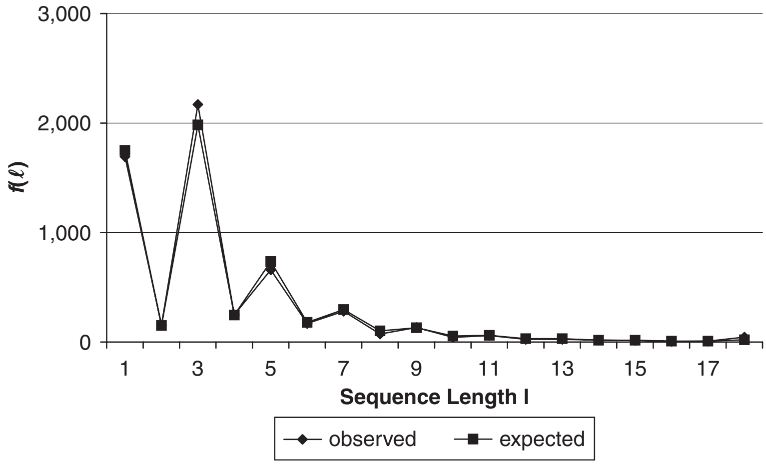
$$\alpha_t(i,u) = \sum_{v \in O} \rho_{v|i,u} \sum_{j \in S} \alpha_{t-1}(j,v) \cdot a_{i|j,v}. \tag{7}$$

The above algorithm is quite fast since it requires only $t(nm)$ additions of estimated parameters $\rho_{u|i,v}$ and $a_{i|j,v}$ and it is accurate enough, since by the time $\alpha_t(i,u)$ approximates the limits of computational precision, the event $(s_t = i, o_t = u)$ is extremely rare and therefore of no practical relevance. Next, we use the extended forward variable to compute the distribution of sequence lengths. We will assume that we have more than just one absorbing state and more than just one observable that can be emitted when the system is in an absorbing state. Therefore, we separate the set of states S into two disjoint sets $S_{trans}$ and $S_{abs}$ containing the transient states and the absorbing states, respectively. Analogously, we define $O_{abs}$ and $O_{trans}$. Now let $\ell$ denote the length of a sequence of observables and $P(L = \ell)$ its probability distribution. A sequence having length $\ell$ is a sequence with $s_\ell \in S_{abs}$ so it must have been in a nonabsorbing state at time $\ell - 1$: $s_\ell \in S_{trans}$, hence

$$\begin{aligned} P(L = \ell) &= P(s_l \in S_{abs} | s_{l-1} \in S_{trans}, o_{\ell-1} \in O_{trans}) \\ &= \sum_{j \in S_{abs}} \sum_{v \in O_{trans}} \sum_{i \in S_{trans}} \alpha_{\ell-1}(i,v) \cdot \tau_{j|i,v}. \end{aligned} \tag{8}$$

So once the matrix of the extended forward variables $\{\boldsymbol{\alpha}_t(i,u)\}$ has been calculated, the predicted distribution of sequence lengths is easily computed with the estimated $\tau_{i|j,v}$. The above expression is general, also because it is readily adapted to the more restricted models. In the most restricted case of a pure Markov source, the emission probabilities will not play a role, except at $\ell - 1$.

**Figure 2**
**Frequencies $f(\ell)$ of Observed and Expected Sequence Lengths**




Note: Expectations were calculated from the fitted parameters as shown in Appendix B of the appendix. Lengths more than 18 have been taken together because of low frequencies. For details, see Table 3.

We applied the above machinery to the five-state LMS with the parameter estimates as shown in the appendix and compared this expected distribution to the one observed. A graphical representation of these probability distributions is shown in Figure 2 and some details in Table 3.

Visually, the observations and predictions are quite close. The jagged shape of the graphs reflects the turn-taking in structured interviewing: Sequences wherein either party says something that requires more than one code are relatively rare, so sequences of odd lengths (remember that we removed the first code that always reflects the posing of the question) are rare, and this effect should fade away when the sequences become longer, as it does. On the basis of Figure 2 and Table 3, we are inclined to think that the LMS as fitted is quite useful.

## Expected Conditional Frequency Distributions

Of course, in many applications, sequence length is not the only data property that is of interest. For example, in studying job careers, the number

**Table 3**
**Observed and Expected Sequence Lengths According to the Fitted LMS**

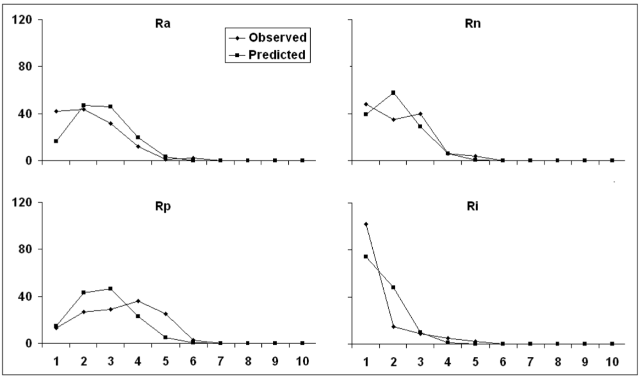
| $\ell$ | Observed | Expected | $\chi^2$ | $G^2$ |
|---|---|---|---|---|
| 1 | 1,692 | 1,751.1 | 2.0 | 120.2 |
| 2 | 156 | 151.3 | 0.1 | −9.3 |
| 3 | 2,169 | 1,983.0 | 17.4 | −355.6 |
| 4 | 250 | 247.3 | 0.0 | −5.4 |
| 5 | 659 | 736.1 | 8.1 | 162.9 |
| 6 | 169 | 179.0 | 0.6 | 20.6 |
| 7 | 281 | 297.1 | 0.9 | 33.1 |
| 8 | 74 | 102.7 | 8.0 | 67.3 |
| 9 | 133 | 130.9 | 0.0 | −4.2 |
| 10 | 43 | 55.7 | 2.9 | 28.8 |
| 11 | 59 | 61.5 | 0.1 | 5.1 |
| 12 | 24 | 29.9 | 1.2 | 13.1 |
| 13 | 24 | 30.3 | 1.3 | 14.1 |
| 14 | 18 | 16.1 | 0.2 | −3.6 |
| 15 | 15 | 15.5 | 0.0 | 1.0 |
| 16 | 6 | 8.8 | 0.9 | 6.7 |
| 17 | 8 | 8.1 | 0.0 | 0.2 |
| 18 | 45 | 20.5 | 29.3 | −32.2 |
| Σ | 5,825 | 5,825 | 73.1 | 63.1 |

Note: LMS = Linked Markov Source. The last row shows the fit-statistics Pearson's $\chi^2$ and log-likelihood chi-square $G^2$. Sequences with lengths of 18 and longer have been taken together.

of occurrences of distinct, short spells of employment, say shorter than six months, may be an interesting feature. Similarly, within the context of theories on the survey interview, the frequency of occurrence of certain types of utterances is of importance. In Figure 3, we plotted the observed and expected frequencies of different R-utterances, given a sequence length of 10.

Because there are only 133 sequences of precisely this length in the data, that is, less than 2.5% of the sequences, we consider these results quite satisfying. Of course, similar predictions for shorter sequences are very much closer to the observations. From a substantial point of view, it is, at least in the context of the survey interview, interesting to study the change of such distributions over different sequence lengths. However, this is beyond the purpose of the present article. Instead, we will focus here on the algorithmic aspects of such predictions. Calculating these predictions

**Figure 3**
**Observed and Predicted Frequency Distributions of the Utterances of Respondents in Sequences of Length 10**




Note: Ra = admissible answer; Rn = inadmissible answer; Rp = asks for clarification, motivates choice, qualifies question, etcetera; Ri = talks about irrelevant subjects.

should not be naively attempted, for the amount of computation would certainly be prohibitive except for the shortest sequences. Therefore, we present an algorithm that allows for efficient computation.

Let the quantity $n_t(w) = h$ denote the observation that an observable $w$ occurred $h$ times before or at $t$. Because there are $\binom{t}{h}$ ways of distributing $h$ occurrences of $w$ over $t$ events, naively evaluating the distribution of $P(n_t(w) = h)$ is not feasible except for small $t$. So we set out to evaluate this type of frequency distributions and show the details of a recursive algorithm to allow for easy adaptation to other needs. For $t > 1$, we have

$$P(n_t(w) = h) = P(o_t = w, n_{t-1}(w) = h - 1) + P(o_t \neq w, n_{t-1}(w) = h) \quad (9)$$

with

$$P(n_1(w) = 1) = \sum_{i \in S} \alpha_1(i, w) \text{ and } P(n_1(w) = 0) = 1 - P(n_1(w) = 1). \quad (10)$$

A general expression for the quantities in the right-hand side of equation (9) is the recursive

$$\begin{aligned} P(o_t, n_{t-1}(w) = h) = {} & P(o_t | o_{t-1} = w) \cdot P(o_{t-1} = w, n_{t-2}(w) = h - 1) \\ & + P(o_t | o_{t-1} \neq w) \cdot P(o_{t-1} \neq w, n_{t-2}(w) = h) \end{aligned} \quad (11)$$

for $t > 2$ and for $t = 2$ we have

$$P(o_2, n_1(w) = h) = \begin{cases} P(o_2|o_1 \neq w) \cdot P(o_1 \neq w) & \text{if } h = 0 \\ P(o_2|o_1 = w) \cdot P(o_1 = w) & \text{if } h = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (12)$$

In order that equation (11) is valid for all $h$, we additionally define $P(o_t, n_{t-1}(w) = h) = 0$ if $m < 0$ and if $m > t - 1$. To compute the quantity $P(n_t(w) = m)$, we use the recursion equation (11) with the initials as in equation (12) and substitute the results in equation (9). The conditional probabilities as appear in the right-hand side of equations (11) and (12) are computable through the extended forward algorithm as discussed in the previous section. For example, we have, for $t > 1$,

$$P(o_t = w|o_{t-1} \neq w) = \frac{\sum_{i,j \in S_{trans}} \sum_{v \in O \setminus \{w\}} \boldsymbol{\alpha}_{t-1}(i, v) \cdot \boldsymbol{\tau}_{j|i,v} \cdot \boldsymbol{\rho}_{w|j,v}}{P(o_{t-1} \neq w)} \quad (13)$$

with

$$P(o_{t-1} \neq w) = \sum_{v \in O \setminus \{w\}} \sum_{i \in S_{trans}} \boldsymbol{\alpha}_{t-1}(i, v). \quad (14)$$

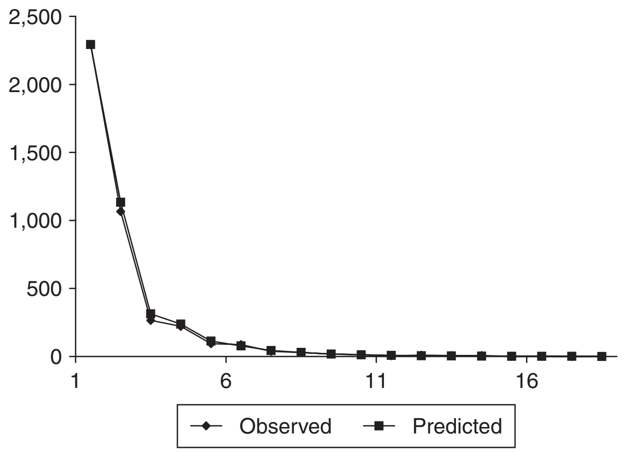
The interested reader will now easily derive expressions analogous to equations (13) and (14) for the other quantities required. It is not difficult now to compute the expected frequency distributions, conditional on some observable property $x$. For example, conditional on the sequence length, that is, $P(n_t(w) = h|L = \ell)$, where $L$ denotes the sequence length. We already know how to compute the expected distribution of sequence lengths $P(L = \ell)$, so we concentrate on the quantity $P(n_t(w) = h, L = \ell)$. Like in equation (9), we decompose the event and use the fact that conditioning on an event $n_{\ell-1}(w)$ is irrelevant for an event at $\ell$. This yields, for $\ell > 2$,

$$\begin{aligned} P(n_{\ell-1}(w) = h, L = \ell) = {} & P(s_\ell \in S_{abs}|o_{\ell-1} = w) \cdot P(o_{\ell-1} = w, n_{\ell-2}(w) = h-1) \\ & + P(s_\ell \in S_{abs}|o_{\ell-1} \neq w) \cdot P(o_{\ell-1} \neq w, n_{\ell-2}(w) = h) \end{aligned} \quad (15)$$

We already know that the unconditional probabilities in equation (15) are computable through the recursion in equations (11) and (12). The conditional likelihoods in equation (15) are easily computed by using the extended forward algorithm again. For example,

$$P(s_l \in S_{abs}|o_{\ell-1} \neq w) = \frac{\sum_{i \in S_{abs}} \sum_{v \in O \setminus \{w\}} \sum_{j \in S_{trans}} \alpha_{\ell-1}(j, v) \cdot a_{i|j,v}}{P(o_{\ell-1} \neq w)}. \quad (16)$$

**Figure 4**
**Frequencies of First Occurrences of Inadequate Behavior of Interviewer or Respondent, as Observed and as Predicted by the Fitted LMS**




Note: Estimated parameters of the Linked Markov Source (LMS) model with five states and nine observables.

The derivation of the other conditional probability and the special case for $\ell = 2$ is left to the reader. So to evaluate$P(n_t(w) = h|L = \ell)$, one evaluates equation (12) and divides by $P(L = \ell)$, the computation of which was already dealt with.

## Expected Conditional Waiting Time Distributions

In negotiations, certain moves may have far-reaching consequences for the duration of the process and the likelihood of success. In general, the course of output-linked processes may be seriously affected by certain first-time events. In the present context of the survey interview, it is known that inadequate behaviour of either I or R is likely to be followed by a series of other inadequate or irrelevant behavior. Therefore, it is interesting to be able to generate expected waiting time distributions, given a fitted model. In Figure 4, we present the unconditional waiting time distributions, as observed and as expected, of inadequate behavior of I or R. From this figure, we conclude that the LMS as fitted quite adequately predicts this kind of behavior, although there seems to be a slight underestimation of moderate waiting times. To generate such a distribution, one needs a different kind of extended forward algorithm. Furthermore, to evaluate distributions that are conditional on waiting time, one even needs a conditional forward variable. In this section, we define these variables and demonstrate how to use them in generating the expected distributions.

Let $f_w = t$ denote the observation that the first occurrence of $w \in O$ occurs at $t \geq 1$. So we wonder about the expected distribution

$$P(L = \ell | f_w = t) = \frac{P(L = \ell, f_w = t)}{P(f_w = t)}. \tag{17}$$

The quantity $P(f_w = t)$ cannot be computed with the extended forward algorithm because $\boldsymbol{\alpha}_t(i, w)$ takes all previous occurrences of $w$ into account. We solve this by creating a "negative sweep" of the event space $S \times O$ as follows. We write $\mathbf{o}_t \neq w$ to denote that the observable $w$ did not occur in the sequence of observations $\mathbf{o}_t$, and we define, for $t > 1$, the quantity

$$\boldsymbol{\alpha}_{t,\bar{w}}(i, u) = P(s_t = i \in S_{trans}, o_t = u, \mathbf{o}_{t-1} \neq w)$$

and $\boldsymbol{\alpha}_{1,\bar{w}}(i, u) = \boldsymbol{\alpha}_1(i, u)$.

The reader now easily verifies the recurrence

$$\boldsymbol{\alpha}_{t,\bar{w}}(i, u) = \sum_{v \in O \setminus \{w\}} \boldsymbol{\rho}_{u|i,v} \sum_{j \in S_{trans}} \boldsymbol{\alpha}_{t-1,\bar{w}}(j, v) \cdot \boldsymbol{\tau}_{i|j,v}. \tag{18}$$

With this recurrence, it is immediately clear that

$$P(f_w = t) = \sum_{i \in S} \boldsymbol{\alpha}_{t,\bar{w}}(j, v) \cdot \boldsymbol{\tau}_{i|j,v} \tag{19}$$

which is the expected unconditional waiting time distribution for the first occurrence of $w$. So the problem of evaluating equation (17) is reduced to calculating its numerator:

$$\begin{aligned} P(L = \ell, f_w = t) = P(o_\ell \in O_{abs} | s_{\ell-1} \in S_{trans}) \times P(s_{\ell-1} \in S_{trans} | o_t = w) \\ \times P(o_t = w | \mathbf{o}_{t-1} \neq w). \end{aligned} \tag{20}$$

Now we already know how to evaluate the first and last multiplicands of equation (20), so we have to deal with $P(s_{\ell-1} \in S_{trans} | o_t = w)$. Thereto, we define, for $t \geq 1$ and $d \geq 1$, the conditional forward variable

$$\boldsymbol{\alpha}_{t|i,u}(j, v) = P(s_{t+d} = j, o_{t+d} = v | s_t = i, o_t = u)$$

and $\boldsymbol{\alpha}_{1|i,u}(j, v) = 1$ if $(i, u) = (j, v)$ and $\boldsymbol{\alpha}_{1|i,u}(j, v) = 0$ otherwise. This variable also satisfies the now familiar recursive structure

$$\boldsymbol{\alpha}_{t|i,u}(j, v) = \sum_{k \in S} \sum_{w \in O} \boldsymbol{\alpha}_{t-1|i,u}(k, w) \cdot \boldsymbol{\tau}_{j|k,w} \cdot \boldsymbol{\rho}_{v|k,w}. \tag{21}$$

Note that because $\boldsymbol{\alpha}_{1|i,u}(j, v)$ is independent from $d$, the quantities $\boldsymbol{\alpha}_{t|i,u}(j, v)$ are also independent from $d$. Now we can use the multiplicative structure of equation (20) to write

$$P(L = \ell, f_w = t) = \sum_{i \in S_{abs}} \sum_{j,k \in S_{trans}} \sum_{v \in O_{trans}} \alpha_{t,\bar{w}}(j, w) \cdot \boldsymbol{\alpha}_{\ell - t - 1|j,w}(k, v) \cdot \boldsymbol{\rho}_{i|k,v} \tag{22}$$

where, for the sake of simplicity of notation, we presume that $|O_{abs}| = 1$. So, equations (22) and (19) together allow for the computation of the distribution $P(L = \ell | f_w = t)$.

## Comments and Conclusions

The primary purpose of this article is to demonstrate the potential use of an LMS as a tool in modeling complicated social processes. To be a useful tool, such a model should describe the data adequately and be able to generate derived characteristics of the data that are in accordance with the observed characteristics. We feel that we demonstrated the LMS to be quite appropriate in these respects: It mimics the data and/or some of its aggregates sufficiently accurate to justify further research, the nature of which will depend on the substantive application. In our case, the modeling of survey interview conversation, it will need fine-tuning with respect to different kinds of questions, different kinds of interviewing, and different populations of Rs. Furthermore, the model only roughly represents current ideas in survey interview research: According to most current ideas, the model is primitive in that it only has two unobservable states, whereas most conceptual frameworks would require at least four of such states to properly render the cognitive processes. Furthermore, the model as discussed is only first order in its dependencies on previous outcomes, and this is, of course, a gross simplification. Therefore, we feel that more sophisticated elaboration and rigorous test of the model is beyond the scope and purpose of the present article; appropriate strategies and techniques are amply discussed in, for example, Hagenaars and McCutcheon (2002). We also believe that the methods and algorithms to generate model predictions on the level of data aggregates are useful; they permit model evaluation that is a bit more sophisticated than just acceptance or rejection on the basis of fit statistics, and they may lead to valuable clues about what and how to adapt a model. The four basic extensions of the forward algorithm are easily implemented, and if appropriately combined, they allow for generating almost any expected distribution once the model has been fitted.

## Appendix A1

| | Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| Ia | 47 | 22 | 9 | 2 | 2,269 | 393 | 626 | 60 | 9 | 3,437 |
| In | 42 | 42 | 5 | 1 | 860 | 552 | 379 | 33 | 16 | 1,930 |
| Ip | 32 | 22 | 11 | 0 | 257 | 171 | 961 | 128 | 7 | 1,589 |
| Ii | 8 | 1 | 0 | 2 | 11 | 5 | 29 | 384 | 58 | 498 |
| Ra | 1,901 | 135 | 24 | 14 | 9 | 1 | 178 | 15 | 3,614 | 5,891 |
| Rn | 821 | 1,231 | 60 | 6 | 16 | 8 | 188 | 10 | 1,077 | 3,417 |
| Rp | 537 | 400 | 1,429 | 31 | 74 | 33 | 92 | 19 | 921 | 3,536 |
| Ri | 48 | 36 | 14 | 442 | 4 | 2 | 1 | 12 | 123 | 682 |
| Ie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5,825 | 5,825 |

Note: See Table 1 for an explanation of the utterance types. The cells in the table denote the frequencies of observing the row-type utterance at $t$ and the column-type utterance at $t+1$. The last column denotes the frequency of each of the utterance types. The average sequence length equals 4.6 with a standard deviation of 3.15; the median sequence length equals 4.

## Appendix B

| | | | | | |
|---|---|---|---|---|---|
| $\pi$ | .00 | .01 | .98 | .01 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| Ia | .02 | .00 | .00 | .97 | .00 |
| | .00 | .03 | .91 | .06 | .00 |
| | — | — | — | — | — |
| | — | — | — | — | — |
| | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| In | .02 | .04 | .19 | .74 | .01 |
| | .00 | .01 | .46 | .52 | .01 |
| | — | — | — | — | — |
| | — | — | — | — | — |
| | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Ip | .00 | .01 | .00 | .98 | .01 |
| | .03 | .03 | .94 | .00 | .00 |
| | — | — | — | — | — |
| | — | — | — | — | — |
| | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Ii | .02 | .01 | .00 | .84 | .14 |
| | .00 | .01 | .83 | .09 | .07 |
| | — | — | — | — | — |
| | — | — | — | — | — |
| | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Ra | — | — | — | — | — |
| | — | — | — | — | — |
| | .49 | .04 | .01 | .04 | .43 |
| | .01 | .01 | .00 | .01 | .88 |
| | 0 | 0 | 0 | 0 | 1 |

| | | | | | |
|---|---|---|---|---|---|
| Rn | — | — | — | — | — |
| | — | — | — | — | — |
| | .49 | .16 | .06 | .01 | .28 |
| | .38 | .09 | .01 | .05 | .48 |
| | 0 | 0 | 0 | 0 | 1 |

*(continued)*

## Appendix B (continued)

**Rp**

| | | | | |
|---|---|---|---|---|
| — | — | — | — | — |
| — | — | — | — | — |
| .09 | .74 | .08 | .02 | .08 |
| .43 | .05 | .00 | .02 | .50 |
| 0 | 0 | 0 | 0 | 1 |

**Ri**

| | | | | |
|---|---|---|---|---|
| — | — | — | — | — |
| — | — | — | — | — |
| .00 | .95 | .03 | .02 | .00 |
| .68 | .04 | .00 | .02 | .27 |
| 0 | 0 | 0 | 0 | 1 |

**Ie**

| | | | | |
|---|---|---|---|---|
| — | — | — | — | — |
| — | — | — | — | — |
| — | — | — | — | — |
| — | — | — | — | — |
| 0 | 0 | 0 | 0 | 1 |

Note: See Table 1 for an explanation of the utterance types. Estimated transition probabilities $\{\boldsymbol{\delta}, \mathbf{T}_u\}$ of the Linked Markov Source model with five states and nine observables. The top vector shows the initial transition vector $\boldsymbol{\delta}$, the matrices are the conditional transition matrices $\mathbf{T}_u$, where $u$ denotes the pertaining encoded observable. Parameters in the nonstochastic rows (those filled with "–") could not estimated because of the restrictions imposed on the emission matrices $\{\mathbf{P}^*, \mathbf{P}_u\}$. The parameters of the last row in each matrix were fixed to the values shown. Because of rounding, the rows may not sum to 1.

## Appendix C

| Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie |
|---|---|---|---|---|---|---|---|---|
| .00 | .91 | .09 | .00 | | | | | |
| .01 | .47 | .52 | .00 | | | | | |
| | | | | .42 | .40 | .18 | .00 | |
| | | | | .00 | .00 | 1.0 | .00 | |
| | | | | | | | | 1 |

**Ia**

| Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie |
|---|---|---|---|---|---|---|---|---|
| .90 | .05 | .00 | .04 | | | | | |
| .17 | .57 | .26 | .00 | | | | | |
| | | | | .43 | .33 | .23 | .02 | |
| | | | | .77 | .04 | .17 | .02 | |
| | | | | | | | | 1 |

**In**

| Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie |
|---|---|---|---|---|---|---|---|---|
| .86 | .00 | .10 | .03 | | | | | |
| .27 | .70 | .03 | .00 | | | | | |
| | | | | .70 | .07 | .20 | .00 | |
| | | | | .38 | .40 | .21 | .01 | |
| | | | | | | | | 1 |

**Ip**

| Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie |
|---|---|---|---|---|---|---|---|---|
| .17 | .81 | .02 | .00 | | | | | |
| .72 | .00 | .28 | .00 | | | | | |
| | | | | .26 | .18 | .52 | .04 | |
| | | | | .04 | .01 | .80 | .15 | |
| | | | | | | | | 1 |

**Ii**

| Ia | In | Ip | Ii | Ra | Rn | Rp | Ri | Ie |
|---|---|---|---|---|---|---|---|---|
| .53 | .16 | .00 | .31 | | | | | |
| 1.0 | .00 | .00 | .00 | | | | | |
| | | | | .06 | .01 | .05 | .88 | |
| | | | | .01 | .01 | .08 | .90 | |
| | | | | | | | | 1 |

## Appendix C (continued)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ra | .99 | .90 | .01 | .00 | | | | | |
| | .00 | .90 | .10 | .00 | | | | | |
| | | | | | .16 | .03 | .81 | .00 | |
| | | | | | .03 | .00 | .89 | .09 | |
| | | | | | | | | | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Rn | .23 | .76 | .01 | .00 | | | | | |
| | .90 | .00 | .10 | .00 | | | | | |
| | | | | | .08 | .00 | .91 | .01 | |
| | | | | | .06 | .13 | .69 | .12 | |
| | | | | | | | | | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Rp | .13 | .10 | .74 | .03 | | | | | |
| | .28 | .20 | .52 | .00 | | | | | |
| | | | | | .38 | .17 | .38 | .07 | |
| | | | | | .23 | .11 | .52 | .14 | |
| | | | | | | | | | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ri | .04 | .00 | .02 | .95 | | | | | |
| | .16 | .16 | .03 | .64 | | | | | |
| | | | | | .00 | .00 | .00 | 1.0 | |
| | | | | | .32 | .16 | .08 | .44 | |
| | | | | | | | | | 1 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ie | — | — | — | — | — | — | — | — | — |
| | — | — | — | — | — | — | — | — | — |
| | — | — | — | — | — | — | — | — | — |
| | — | — | — | — | — | — | — | — | — |
| | | | | | | | | | 1 |

Note: See Table 1 for an explanation of the utterance types. Emission matrices as estimated for the Linked Markov Source model; fixed zero's omitted, values indicated with "—" could not be estimated. Numbers without decimals were fixed parameters. The top rows show the encoded observables. The top matrix is the matrix of unconditional emission probabilities **P**, the other matrices represent the conditional emission matrices $\mathbf{P}_u$ with $u$ the conditioning observable as indicated to the left of each matrix.
Because of rounding, the rows may not sum to 1.

# References


Arulampalan, W., and P. Gregg. 2001. Unemployment "scarring." *Economic Journal* 111(475):577-84.

Bartholomew, D. J. 1982. *Stochastic models for social processes*. 3rd ed. New York: Wiley.

Baum, Leonard E., Ted Petrie, George Soules, and Norman Weiss. 1970. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains." *Annals of Mathematical Statistics* 41:164-71.

Bijlevelt, Catrien B. J. H. and Ab Mooijaart. 2003. "Latent Markov Modelling of Recidivism Data." *Statistica Neerlandica* 57:305-20.

Bradburn, Norman M. 2004. "Understanding the Question-Answer Process." *Survey Methodology* 30:5-15.

Clote, Peter and Rolf Backofen. 2000. *Computational Microbiology, an Introduction*. New York: John Wiley.

Conlisk, John. 1976. "Interactive Markov Chains." *Journal of Mathematical Sociology* 4:157-85.

———. 1978. "A Stability Theorem for an Interactive Markov Chain." *Journal of Mathematical Sociology* 6:163-8.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B* 39:1-38.

Devijver, Pierre A. 1985. "Baum's Forward-Backward Algorithm Revisited." *Pattern Recognition Letters* 3:369-73.

Dijkstra, Wil. 1999. "A New Method for Studying Verbal Interactions in Survey Interviews." *Journal of Official Statistics* 15:67-85.

Giddens, Anthony. 1991. *Modernity and Self-Identity: Self and Society in the Late Modern Age*. Cambridge, UK: Polity.

Hagenaars, Jacques A. and Allan L. McCutcheon, eds. 2002. *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press.

Hawes, Leonard C. and Joseph M. Foley. 1972. "A Markov Analysis of Interview Communication." *Speech Monographs* 40:208-19.

Jaffe, Joseph and Stanley Feldstein. 1970. *Rhythms of Dialogue*. New York: Academic Press.

Langeheine, Rolf and Frank Van de Pol. 1996. "Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis." *Sociological Methods & Research* 24:492-516.

——— and Frank Van de Pol. 2002. "Latent Markov Chains." Pp. 304-41 in *Applied Latent Class Analysis*, edited by J. A. Hagenaars and A. L McCutcheon. Cambridge, UK: Cambridge University Press.

Matras, Judah. 1967. "Social Mobility and Social Structure: Some Insights from the Linear Model." *American Sociological Review* 32:608-14.

Miller, George A. 1952. "Finite Markov Processes in Psychology." *Psychometrica* 17:149-67.

Prais, Sig J. 1955. "The Formal Theory of Social Mobility." *Population Studies* 9:72-81.

Rabiner, Lawrence R. 1989. "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition." *Proceedings of IEEE* 77:267-95.

Schaeffer, Nora Cate and Douglas W. Maynard. 1996. "From Paradigm to Prototype and Back Again: Interactive Aspects of Cognitive Processing in Standardized Survey Interviews." Pp. 65-88 in *Answering Questions: Methodology for Determining Cognitive and Communicative Processes in Survey Research*, edited by N. Schwarz and S. Sudman. San Francisco: Jossey-Bass.

Schrodt, Philip A. 2000. "Pattern Recognition of International Crises Using Hidden Markov Models." Pp. 296-328 in *Political Complexity: Non-Linear Models of Politics*, edited by D. Richards. Ann Arbor: University of Michigan Press.

Smit, Edith G. and Peter B. Neyens. 2000. "Segmentation Based on Affinity for Advertising." *Journal of Advertising Research* 40:35-43.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Visser, Ingmar, Maartje E. J. Raijmakers, and Peter C. M. Molenaar. 2002, "Fitting Hidden Markov Models to Psychological Data." *Scientific Programming* 10:185-99.

Weakliem, David L. 2004. "Introduction to the Special Issue on Model Selection." *Sociological Methods & Research* 33:167-87.

Wu, C. F. Jeff. 1983. "On the Convergence Properties of the EM-Algorithm." *Annals of Statistics* 11:95-103.

**Cees H. Elzinga** is chair of the Department of Social Research Methodology at Vrije Universiteit Amsterdam. His main interest is the methodology of categorical time series. Recent publications are on sequence similarity (*Sociological Methods & Research*, 2003) and on metric representation of sequence data (*Journal of Classification*, 2005).

**Adriaan W. Hoogendoorn** is an assistant professor in the Department of Social Research Methodology at Vrije Universiteit Amsterdam. His main interest is in methodological issues regarding Internet panel surveys. A recent publication is on the use of dependent interviewing in a Web survey (*Journal of Official Statistics*, 2004).

**Wil Dijkstra** is a professor in the Department of Social Research Methodology at Vrije Universiteit Amsterdam. His main research interest is the analysis of question-answer sequences in survey interviews. More recently, he became involved in the Event History Calendar as a method for obtaining retrospective data.