Parameterization and estimation of path models for categorical data

Tamás Rudas[1], Wicher Bergsma[2], and Renáta Németh[3]

[1] Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University, rudas@tarki.hu
[2] Department of Statistics, London School of Economics and Political Science, W.P.Bergsma@lse.ac.uk
[3] Department of Statistics, Faculty of Social Sciences, Eötvös Loránd University, nmthrnt@freemail.hu

Summary. The paper discusses statistical models for categorical data based on directed acyclic graphs (DAGs) assuming that only effects associated with the arrows of the graph exist. Graphical models based on DAGs are similar, but allow the existence of effects not directly associated with any of the arrows. Graphical models based on DAGs are marginal models and are best parameterized by using hierarchical marginal log-linear parameters. Path models are defined here by assuming that all hierarchical marginal log-linear parameters not associated by an arrow are zero, providing a parameterization with straightforward interpretation. The paper gives a brief review of log-linear, graphical and marginal models, presents a method for the maximum likelihood estimation of path models and illustrates the use of path models, with special emphasis on the interpretation of estimated parameter values, to real data.

1 Introduction

This paper develops path models for categorical data and investigates their relationship with models associated with directed acyclic graphs, using marginal log-linear parameterizations of the distributions in the model. A path model is defined, intuitively, as a model associated with a directed acyclic graph, in the sense that the arrows of the graph represent direct effects between variables. A lack of an arrow between two variables means conditional independence between them, when conditioning on the parents of either one. Section 2 of the paper gives a general overview of models associated with directed acyclic graphs (DAGs) that possess the required conditional independence properties. Section 3 reviews DAG models as marginal models and Section 4 considers the implied marginal parameterization of DAG models.

Finding the appropriate parameterization in which the models can be defined and parameterized in an intuitive way is a central theme of the paper. In order to fully utilize the models considered, one would need to have a parameterization in

which the distributions in the model are parameterized by parameters measuring the strengths of the effect (arrows) allowed in the model.

As it will be illustrated in Section 4, DAG models, for categorical data, also allow effects that are not associated with any of the arrows in the model, thus further developments are needed to define a model class with the required properties. It is shown in Section 5 that in the marginal log-linear parameterization it is possible to identify the effects associated with the arrows in the graph and by assuming that the remaining parameters are zero, one obtains models that contain only effects related to arrows. Section 6 discusses algorithmic aspects of estimating the models and Section 7 presents an application.

2 Log-linear, graphical and DAG models

Let V_i, $i = 1, \ldots, q$ be categorical variables (classifications) with categories (or indices) $v_{i,1}, \ldots, v_{i,c(i)}$, $i = 1, \ldots, q$, respectively. The joint classification of N observations according to these form a frequency distribution on the Cartesian product $\Omega = \bigotimes_{i=1}^{q}(v_{i,1}, \ldots, v_{i,c(i)})$ which is called a contingency table. Such data are frequently observed in the social, behavioural or biological sciences. When analyzing such data, a question of primary interest is how the variables are related to each other. Simple structures are often formulated using a log-linear model [BFH75], [Agr02].

A log-linear model is based on a class of subsets of the variables Γ, the so-called generating class, and assumes that (in the strictly positive case)

$$\log P(\omega) = \sum_{\gamma \in \Gamma} f_\gamma(\omega_\gamma), \tag{1}$$

for all $\omega \in \Omega$, where $(.)_\gamma$ is a marginalization operator in the sense that it selects the indices from ω that belong to the variables in γ. The meaning of such a model depends on the subsets of variables that appear in Γ. One intuitive interpretation is that (1) means that there is no conditional order-$(|G|-1)$ association (that is, association involving all $|G|$ variables) within those groups of variables $G \subseteq \{V_1, \ldots, V_k\}$ that contain any of the maximal elements of Γ as a proper subset, when conditioned on all other variables $\{V_1, \ldots, V_k\} \setminus G$. Here, conditional association is measured by the conditional odds ratio [Rud98]. The elements of Γ are, therefore called interactions, because these groups of variables may be associated within each other. Another possible interpretation of (1) is that it is equivalent to a number of restrictions being valid for the joint distribution. The restrictions are either one of two types. The first type applies to subsets that are maximal with respect to the property that no interaction contains more than one of the variables from the subset. The first type of restriction is that the variables in any such subset, when conditioned on all other variables, are jointly independent. The second type is that those groups of variables of cardinality k that have the property that every subset of them of cardinality $k-1$ is an interaction, have no $(k-1)$st order association, conditioned on all other variables, when association is, again, measured by the odds ratio [Rud02].

Of particular interest are log-linear models based on generating classes with the property that the maximal interactions are the cliques (i.e., maximal complete subgraphs) of a graph G, where the nodes are the variables. Such models are called

graphical models [Lau96] and they can be interpreted using conditional independencies. In particular, in the characterization of log-linear models given in the previous paragraph, there are no subsets with the second property, that is, the first type of conditional independencies characterize the joint distribution. Another important characterization is based on the so-called global Markov property: if two subsets of variables A and C are separated by a subset B in the sense that all paths in G that connect a variable in A with a variable in C goes through B, then the joint distribution of the variables in A is conditionally independent from the joint conditional distribution of variables in C, given the variables in B.

Graphical log-linear models are useful in modeling complex association structures among the variables, but many of the important substantive research problems require the analysis of effects (i.e. directed associations) and these are, intuitively, best modeled by using directed acyclic graphs (DAGs). A DAG is a simple directed graph (an arrow always goes between two different nodes and there is at most one arrow between any two nodes) without a directed loop, that is, without a path following the direction of the arrows starting in a node and ending in the same node. A node is called a parent of another one if there is an arrow pointing from the former one to the latter one. A node is called a descendant of another node if there is a directed path leading from the latter one to the former one.

A DAG model is specified by a list of conditional independence restriction, requiring that

$$V_i \perp\!\!\!\perp \mathrm{nd}(V_i) | \mathrm{pa}(V_i), \tag{2}$$

for every i, where $\mathrm{nd}(V_i)$ is the set of nodes that are not descendants of V_i and $\mathrm{pa}(V_i)$ is the set of parents of V_i.

3 DAG models as marginal models

Because for every variable, $\mathrm{pa}(V_i) \subseteq \mathrm{nd}(V_i)$, the conditional independencies in (2) that characterize a DAG model, apply to subsets of the variables, namely $\{V_i\} \bigcup \mathrm{nd}(V_i)$. In this sense, DAG models belong to the class of marginal models introduced in [BR02], see also [RB04] for several possible applications of these models.

Marginal models in [BR02] are defined by imposing linear or affine restrictions on marginal log-linear parameters. Marginal log-linear parameters are ordinary log-linear parameters computed from a marginal of the contingency table, rather than from the entire table. Therefore, every marginal log-linear parameter is characterized by two subsets of the variables: the marginal in which it is computed and the the effect to which it applies. For example, in an $ABCD$ table, the AB effect in the ABC marginal, when all the variables are binary, is

$$\lambda^{ABC}_{1\,1\,*} = \frac{1}{2} \sum_k \log \left(\frac{p_{11k+} p_{22k+}}{p_{12k+} p_{21k+}} \right)^{1/4}, \tag{3}$$

that is, the marginal log-linear parameter $\lambda^{ABC}_{i\,j\,*}$ of the $ABCD$ table is related to the average conditional log odds ratio between A and B, conditioned on and averaged over C, after marginalization over D. When the variables are not binary, the marginal

log-linear parameter is matrix-valued. For example, the parameter of the AB effect has $(I-1)(J-1)$ linearly independent values, if A has I and B had J categories. The notation λ^{ABC}_{AB*} refers to all these values, where the upper index is the marginal and the lower index is the effect. The effect is always a subset of the marginal.

Marginal log-linear parameters provide the analyst with a flexible tool to parameterize the joint distribution of the variables on Ω. The exact rules and several properties of these parameterizations are described in [BR02]. To obtain a parameterization, one needs to select certain marginals $M_1, M_2, \ldots, M_t$ of the table, including the entire table and order them in a way that if $M_i \subseteq M_j$ then M_i precedes M_j. Then, every subset of the variables should appear as an effect, within the first marginal where it is possible (i.e. within the first marginal that contains it). Such a parameterization is called a hierarchical marginal log-linear parameterization.

4 Parameterization of DAG models

To obtain a hierarchical marginal log-linear parameterization of all the joint distributions on Ω, in which path DAG models and path models may be conveniently parameterized, consider first a so-called well-numbering of the variables. A well-numbering is an order, in which $i < j$ implies that $V_i \in \mathrm{nd}(V_j)$. If the variables are well-numbered, then (2) is equivalent to

$$V_i \perp\!\!\!\perp \left(\mathrm{nd}(V_i) \bigcap \mathcal{V}_{<i}\right) \setminus \mathrm{pa}(V_i) \mid \mathrm{pa}(V_i), \tag{4}$$

where $\mathcal{V}_{<i}$ is the set of variables preceding V_i, see [LDLL90].

The hierarchical marginal log-linear parameterization will be based on the series of subsets $\mathcal{V}_{<i} \bigcup \{V_i\}$, $i = 1, \ldots, q$. All effects will appear within the marginal that comes first from among those that contain it.

As a very simple example, consider three variables, A, B and C, with arrows pointing from A to C and from B to C. A well-numbering is A, B, C and the relevant marginals are A, AB, ABC. A hierarchical marginal log-linear parameterization consists of the following parameters:

$$\lambda^{A}_{\emptyset},\ \lambda^{A}_{A},\ \lambda^{AB}_{*\,B},\ \lambda^{AB}_{AB},\ \lambda^{ABC}_{*\,*\,C},\ \lambda^{ABC}_{A*\,C},\ \lambda^{ABC}_{*\,BC},\ \lambda^{ABC}_{ABC}. \tag{5}$$

When all the variables are binary, all the above parameters have essentially one value, and one has eight parameters. The conditional independencies in (4), in the present case, reduce to:

$$A \perp\!\!\!\perp B$$

and this is true if and only of

$$\lambda^{AB}_{AB} = 0.$$

Consequently, the remaining parameters in (5) parameterize all the distributions in the DAG model.

Contrary to intuitive expectation, in addition to an overall effect $\lambda^{A}_{\emptyset}$, main effects λ^{A}_{A}, $\lambda^{AB}_{*\,B}$ and $\lambda^{ABC}_{*\,*\,C}$, and effects related to the arrows $\lambda^{ABC}_{A*\,C}$ and $\lambda^{ABC}_{*\,BC}$, there also remains an additional non-zero effect λ^{ABC}_{ABC}. This last effect represents the joint effect of A and B on C, in spite of the fact that A and B are marginally independent.

Further, this effect cannot be associated with the arrows present in the DAG. For a precise interpretation of how these parameters measure effects the size of the related effect, see [BR03] and [RB04].

Path models are supposed to have only effects related to arrows and will be defined here by assuming the parameters in the hierarchical marginal log-linear parameters not associated with arrows are zero.

5 Path models

Consider a DAG with the variables forming Ω being the nodes and all strictly positive probability distributions on Ω parameterized by hierarchical marginal log-linear paramneters based on the marginals $\mathcal{V}_{<i} \bigcup \{V_i\}$, $i = 1, \ldots, q$ in a well-numbering of the variables. Then, the assumption, inspired by the modified path models in [Goo73], that

$$\lambda_E^M = 0 \text{ unless } |E| \leq 1 \text{ or } E \text{ is an arrow in the DAG} \tag{6}$$

is the path model associated with the DAG. Models defined by the restriction (6) are marginal log-linear models and, in addition to having a straightforward interpretation, such models have a number of desirable statistical properties, as it was demonstrated in [BR02].

First, the parameters not set to zero in (6) are variation independent from each other, making individual interpretations of the parameters meaningful. Second, maximum likelihood estimates under Poisson or multinomial sampling are stationary points of the likelihood with probability converging to one as the sample size increases. Third, the maximum likelihood estimates have an asymptotic normal distribution. Fourth, the likelihood ratio statistic has an asymptotic normal distribution. Fifth, the maximum likelihood estimates of the parameters not set to zero in (6) are equal to their observed values.

6 Maximum likelihood estimation

Several algorithms for fitting models with restrictions on the marginal distributions of contingency tables have been proposed in the literature. Lagrange multiplier techniques were discussed by [Hab85], [LA94] and [Ber97]. These methods are not guaranteed to converge, however, since they seek a saddle point (rather than a local extreme) of the Lagrangian likelihood, which may be difficult to find. An alternative method, based on maximizing a reparameterized likelihood, was discussed by [MN89] and [GM95]. This approach has the advantage over the aforementioned Lagrange multiplier methods that a maximum is sought, which tends to be easier to find using gradient methods than a saddle point. However, because of the reparameterization, 'iteration within iteration' is needed, making the algorithm computationally cumbersome.

An alternative approach which does not suffer from these drawbacks was introduced by [BR05]. It is a quasi-Newton algorithm applied to a certain exact penalty function which was first introduced by [Chr95]. This penalty function has the constrained maximum likelihood estimate as its unconstrained maximum. For marginal

models, this method has the advantages, compared to the previously proposed methods, that (i) each iterative step can be performed fast and (ii) convergence is easy to achieve since a maximum rather than a saddle point is sought. We shall now describe this method.

The models in (6) can be described by constraints on the vector of cell probabilities π of the following form:

$$h(\pi) = B \log A\pi = 0$$

where B is a contrast matrix, i.e., has rows summing to zero (see [LA94, Ber97]). With p the vector of observed proportions which are assumed to be strictly positive, we seek to maximize the likelihood kernel

$$L(\pi) = p' \log(\pi) - 1'\pi \ ,$$

subject to the model constraint $h(\pi) = 0$. We denote this maximum likelihood estimate (MLE) as $\hat{\pi}$. As shown by [La96] the MLE of the cell probability vector π is obtained by determining the saddle point of the kernel of the Lagrangian log likelihood function

$$L(\pi, \lambda) = p' \log(\pi) - 1'\pi - \lambda' h(\pi) \ ,$$

where λ is a vector of unknown Lagrange multipliers. The Hessian of the constraint function $h(\pi)$ is

$$H(\pi) = \frac{\partial h(\pi)'}{\partial \log \pi} = D_\pi A' D_{A\pi}^{-1} B'$$

where D_π is the diagonal matrix with π on the main diagonal. Now differentiating $L(\pi, \lambda)$ with respect to π, equating to zero, and solving for λ yields

$$\lambda(\pi) = \left(H(\pi)'H(\pi)\right)^{-1} H(\pi)'(p - \pi),$$

see [BR05]. We may now consider the function $L(\pi, \lambda(\pi))$, which depends only on the unknown probability vector π, not on the Lagrange multipliers. However, $L(\pi, \lambda(\pi))$ does not in general have a maximum at $\hat{\pi}$. Next, we define a function which does have this property.

With

$$V(\pi) = \left(H(\pi)'H(\pi)\right)^{-1}$$

and $G(\pi)$ the matrix with (i, j)th coordinate

$$g_{ij} = -\delta_{ij}\pi_{ij} - \sum_k \lambda_k(\pi) \frac{dh_k(\pi)}{d \log \pi_i \, d \log \pi_j}$$

let

$$W(\pi) = V(\pi)H(\pi)'G(\pi)H(\pi)V(\pi)$$

(see [BR05] for a closed form matrix expression for $G(\pi)$). We can now define the exact penalty function

$$P(\pi) = L(\pi, \lambda(\pi)) + \frac{1}{2}h(\pi)'W(\pi)h(\pi) + \frac{1}{2}h(\pi)'V(\pi)h(\pi).$$

The function $P(\pi)$ is called 'exact' since it does not depend on the Lagrange multiplier vector λ, and 'penalty function' because of the third term which penalizes for deviations of $h(\pi)$ from zero. The important property of $P(\pi)$ is that it has $\hat{\pi}$ as an unconstrained maximum [Chr95, BR05]. There are several possibilities for maximizing $P(\pi)$. Direct application of Newton's method involves up to fourth order derivatives of the constraint function and is therefore impractical. Newton's method using so-called automatic differentiation was proposed by [Chr95]. Alternatively, [BR05] proposed a quasi-Newton method based on first and second derivatives with respect to π of the function

$$P(\pi;\pi_0) = L(\pi,\lambda(\pi)) + \frac{1}{2}h(\pi)'W(\pi_0)h(\pi) + \frac{1}{2}h(\pi)'V(\pi_0)h(\pi),$$

where the occurrences of $\lambda(\pi)$ and $h(\pi)$ and the Hessian are replaced by their MLEs, i.e., by 0. This yields for the gradient of P

$$\begin{aligned}\nabla P(\pi;\pi_0) = p - \pi - H(\pi)\lambda(\pi) - G(\pi)H(\pi)V(\pi)h(\pi) + \\ H(\pi)W(\pi_0)h(\pi) + H(\pi)V(\pi_0)h(\pi)\end{aligned}$$

and for its Hessian $\nabla^2 P(\pi;\pi_0) = D_\pi$. The algorithm is as follows:

$$\begin{aligned}&\log \pi^{(0)} = \log p \\ &\log \pi^{(k+1)} = \log \pi^{(k)} - t^{(k)} \left[\nabla^2 P\left(\pi^{(k)};\pi^{(k)}\right)\right]^{-1} \nabla P\left(\pi^{(k)};\pi^{(k)}\right)\end{aligned}$$

where $k = 0, 1, \ldots$ and $0 < t^{(k)} \leq 1$ is a step size which must be chosen small enough to ensure that

$$P(\pi^{(k+1)};\pi^{(k)}) > P(\pi^{(k)};\pi^{(k)})$$

The algorithm appears to have good numerical properties for a wide variety of models. It has been applied to problems with many thousands of cells, the main limitation apparently being the amount of computer memory available. Further numerical experience is given in [BR05].

7 An example

The data in Table 1 are from [MN89], originally published in [Sch70] and were also analyzed by [Fie70] and by [BFH75]. The data concern the daytime habits of two species of lizard, grahami and opalinus and were collected by observing occupied sites or perches and recording species involved, time of day, height and diameter of perch and whether the site was sunny or shady.

Suppose we wish to identify how the factors that possibly determine whether a perch is occupied by a grahami or an opalinus lizard are related to each other. A plausible model for the determinants is shown in Figure 1 (Model 1), which is a DAG model.

The model assumptions are as follows: (1) the physical characteristics of an occupied perch (height and diameter) are independent from the time of observation, (2) the diameter of an occupied perch does not affect directly whether the perch is sunny or shady, that is, these variables are conditionally independent given the other

Table 1. Site preferences of two species of lizard. H: perch height, D: perch diameter, S: sunny/shady, T: time of day, S: G - grahami, O - opalinus

			T					
	Perch		Early		Mid-day		Late	
S	D (in)	H (ft)	G	O	G	O	G	O
Sun	≤ 2	< 5	20	2	8	1	4	4
		≥ 5	13	0	8	0	12	0
	> 2	< 5	8	3	4	1	5	3
		≥ 5	6	0	0	0	1	1
Shade	≤ 2	< 5	34	11	69	20	18	10
		≥ 5	31	5	55	4	13	3
	> 2	< 5	17	15	60	32	8	8
		≥ 5	12	1	21	5	4	4

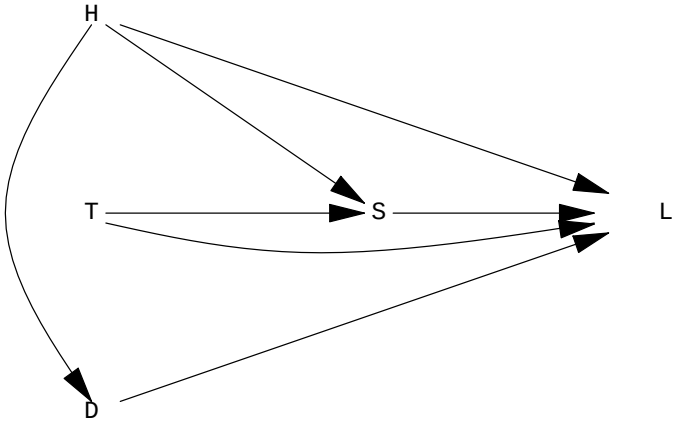

Fig. 1. Model 1, DAG Model

explanatory variables. Formally these assumptions can be defined by the following conditional independencies:

$$HD \perp\!\!\!\perp T \text{ and } D \perp\!\!\!\perp S \mid TH. \tag{7}$$

A well-numbering is H, D, T, S, L. The marginals involved in the parameterization are H, HD, HDT, $HDTS$, $HDTSL$. The corresponding zero effects in the DAG model are

$$\lambda^{HDT}_{H*\,T},\ \lambda^{HDT}_{*\,DT},\ \lambda^{HDT}_{HDT},\ \lambda^{HDTS}_{*\,D*\,S},\ \lambda^{HDTS}_{HD*\,S},\ \lambda^{HDTS}_{*\,DTS},\ \lambda^{HDTS}_{HDTS},$$

and the parameters of the distributions in the model are

$$\lambda^{H}_{\emptyset},\ \lambda^{H}_{H},\ \lambda^{HD}_{*\,D},\ \lambda^{HD}_{HD},\ \lambda^{HDT}_{*\,*\,T},\ \lambda^{HDTS}_{*\,*\,*\,S},\ \lambda^{HDTS}_{H*\,*\,S},$$

$$\lambda^{HDTS}_{*\,T*S},\ \lambda^{HDTS}_{HT*S},\ \lambda^{HDTSL}_{****L},\ \lambda^{HDTSL}_{H***L},\ \lambda^{HDTSL}_{*\,D**L},\ \lambda^{HDTSL}_{**T*L},$$
$$\lambda^{HDTSL}_{***SL},\ \lambda^{HDTSL}_{HD**L},\ \lambda^{HDTSL}_{H*\,T*L},\ \lambda^{HDTSL}_{H**SL},\ \lambda^{HDTSL}_{*\,DT*L},\ \lambda^{HDTSL}_{*\,D*SL},$$
$$\lambda^{HDTSL}_{**\,TSL},\ \lambda^{HDTSL}_{HDT*L},\ \lambda^{HDTSL}_{HD*SL},\ \lambda^{HDTSL}_{H*\,TSL},\ \lambda^{HDTSL}_{*\,DTSL},\ \lambda^{HDTSL}_{HDTSL},$$

The second model we consider is a path model derived from Model 1 by eliminating the following parameters according to (6):

$$\lambda^{HDTS}_{HT*S},\ \lambda^{HDTSL}_{HD**L},\ \lambda^{HDTSL}_{H*\,T*L},\ \lambda^{HDTSL}_{H**SL},\ \lambda^{HDTSL}_{*\,DT*L},\ \lambda^{HDTSL}_{**TSL},$$
$$\lambda^{HDTSL}_{HDT*L},\ \lambda^{HDTSL}_{HD*SL},\ \lambda^{HDTSL}_{H*\,TSL},\ \lambda^{HDTSL}_{*\,DTSL},\ \lambda^{HDTSL}_{HDTSL},$$

For example, setting the λ^{HDTS}_{HT*S}, parameter to zero can be interpreted, if S is considered a response to H and T, as assuming that these have separate effects only. Or, setting the λ^{HDTSL}_{HD**L}, parameter to zero can be interpreted as assuming that H and D have only separate effects on L. Notice that there is an effect from H to D and there is no arrow between H and T, still, in a path model neither pair is assumed to have a joint effect on its respective response.

The goodness-of-fit of the models is characterized here by the likelihood-ratio statistics LR, although it is well-known that asymptotic p values may be unreliable when analyzing sparse contingency tables like the one we have here (see, e.g. [Rud86]), however LR is useful to compare nested models. Table 2 displays goodness-of-fit tests for the two models (note that one half was added to the empty cells before model fitting)

Table 2. Goodness-of-Fit Tests for Models for Table 1

Model	df	LR	p
Model 1	12	13.3	.35
Model 2	32	24.8	.81
Model 2 / Model 1	20	11.5	.93

The increase in LR, when moving to the more restrictive path model from the DAG model is far from being significant. Thus, removing all effects from Model 1, except for those consisting of a child and one of its parents, seems sensible. Further, the observed explanatory variables appear to give a reasonably good prediction of the response variable.

One may want to go further, looking for an even simpler model. Backward elimination can be used, sequentially eliminating terms from the parameters defining Model 2. In the first step, the smallest increase in LR, compared to Model 2, results from eliminating the λ^{HDTS}_{H**S} parameter. After removing it, each additional parameter elimination causes significant increase in LR. Figure 2 presents parameter estimates for this final path model. Only non-redundant parameters are displayed and, naturally, no parameter estimates for the HS effect that is removed. For every hierarchical marginal log-linear parameter displayed, there is an additional last (redundant) value, the sum of which with the given value(s) is zero.

For each effect, the strength and direction of the dependencies can be read off from the parameter estimates. As it can be seen, small-diameter occupied sites are more likely to be in high positions. Occupied sites observed at mid-day are more likely to be shady than sites observed at another time.

A perch in a high position is more likely to be occupied by a grahami lizard, as opposed to an opalinus lizard than a perch in a low position. The same can be said about small versus large perches and about sunny versus shady perches and about perches observed early or at mid-day, versus late. These effects may be separated and their strengths are measured here by the the appropriate value of the relevant marginal log-linear parameter (the value given is one quarter of the logarithm of the conditional odds ratio for binary explanatory variables, see (3)).

Regarding the latter finding: previous analyses (see [MN89]) using GLM arrived at the same conclusion. One main advantage of the approach presented here is the possibility of tracing direct and indirect paths, e.g. T has a direct effect on Lizard (at mid-day it is more likely to observe a grahami than an opalinus) and an indirect effect through S as well: a lizard is more likely to occupy a shady site at mid-day, than at another time, and an occupied shady perch is more likely to be occupied by an opalinus.

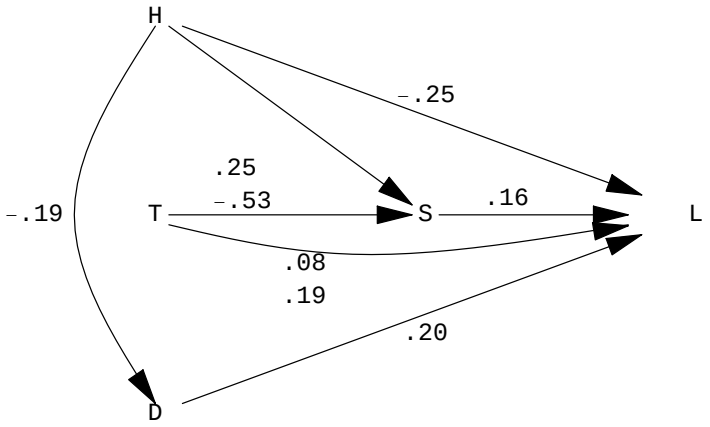

Fig. 2. Parameter Estimates, Path Model

References

[Agr02] Agresti,A.: Categorical Data Analysis. 2nd ed. Wiley, New York (2002)

[Ber97] Bergsma,W. P.: Marginal Models for Categorical Data. Tilburg University Press, Tilburg (1997)

[BR02] Bergsma, W. P. Rudas, T.: Marginal models for categorical data. Ann. Stat. **30**, 140–159 (2002)

[BR03] Bergsma, W.P., Rudas, T.: On conditional and marginal association. Annales de la Faculte des Sciences de Toulouse, **11**, 455-468 (2003)

[BR05] Bergsma, W. P., Rapcsak, T.: An exact penalty method for smooth equality constrained optimization with application to maximum likelihood estimation. Submitted. (2005)

[BFH75] Bishop, Y. V. V., Fienberg, S. E., Holland, P. W.: Discrete Multivariate Analysis. MIT Press, Cambridge, MA.(1975)

[Chr95] Christianson, B.: Geometric approach to Fletcher's ideal penalty function. Journal of Optimization Theory and Applications, **84**, 433-441 (1995)

[Fie70] Fienberg, S.E.: The analysis of multidimensional contingency tables. Ecology, **51**, 419-433 (1970)

[Goo73] Goodman, L. A. The analysis of multidimensional contingency table when some variables are posterior to others: a modified path analysis approach. Biometrika **60**, 179-192 (1973)

[GM95] Glonek, G. J. N., McCullagh, P.: Multivariate logistic models. J. Roy. Statist. Soc. Ser. B", **57**, 533-546(1995)

[Hab85] Haber, M.: Maximum likelihood methods for linear and loglinear models in categorical data. Comput. Statist. Data anal. **3**, 1-10 (1985)

[Lan96] Lang, J. B. Maximum likelihood methods for a generalized class of log-linear models. Ann. Stat.", **24**, 726-752 (1996)

[LA94] Lang, J. B., Agresti,A.: Simultaneously modelling the joint and marginal distributions of multivariate categorical responses. J. Am. Stat. Ass., **89**, 625-632 (1994)

[Lau96] Lauritzen, S. L.: Graphical Models. Clarendon Press, Oxford (1996)

[LDLL90] Lauritzen, S.L., Dawid, A. P., Larsen, B. N., Leimer, H.-G.: Independence Properties of Directed Markov Fields. Networks, **20**, 491-505 (1990)

[MN89] McCullagh, P., Nelder,J. A.: Generalized Linear Models. Chapman and Hall, London (1989)

[Rud86] Rudas, T.: A Monte Carlo comparison of the small sample behaviour of the Pearson, the likelihood ratio and the Cressie-Read statistics. Journal of Statistical Computation and Simulation, **24**, 107-120 (1986)

[Rud98] Rudas, T.: Odds Ratios in the Analysis of Contingency Tables. Sage, Thousand Oaks (1998)

[Rud02] Rudas, T. Canonical representation of log-linear models. Communications in Statistics (Theory and Methods), **31(12)**, 2311–2323 (2002)

[RB04] Rudas, T., Bergsma, P.: On Applications of Marginal Models to Categorical Data. Metron, **42**, 15-37 (2004)

[Sch70] Schoener, T. W.: Nonsynchronous spatial overlap of lizards in patchy habitats. Ecology, **51**, 408-418 (1970)