CALL #: Knight HA 1 .M4
LOCATION: ORU :: Main Library :: orum

Ariel
MC 10-20-05

TYPE: Article
USER JOURNAL TITLE: Metron
OCLC JOURNAL TITLE: Metron
ORU CATALOG TITLE: Metron
ARTICLE TITLE: Geometrical Modelling of Markovian Dependence
ARTICLE AUTHOR: Rogantin M. P.
VOLUME: 54
ISSUE: 1-2
YEAR: 1996
PAGES: 45-65
ISSN: 0026-1424
OCLC #:
CROSS REFERENCE ID: 157424
VERIFIED:

BORROWER: ORE :: Main Library
PATRON: Bulatov,Yaroslav
PATRON ID:
PATRON ADDRESS:
PATRON PHONE:
PATRON FAX:
PATRON E-MAIL:
PATRON DEPT: Engineering
PATRON STATUS: Undergraduate
PATRON NOTES:

MARIA PIERA ROGANTIN (*)

Geometrical modelling of Markovian dependence

CONTENTS:

1. INTRODUCTION

This paper is based on a series of previous works related to geometrical theory of statistical models jointly done with G. Pistone, to which we refer for a complete overview: see Pistone and Rogantin (1990), (1994), and (1995); see also Pistone and Sempi (1995).

In this paper we apply geometrical methods to analyse Markovian dependence, to show that suitable parameters can be splitted into two mutually orthogonal blocks, the first one representing the marginal probabilities and the latter representing the dependence.

Orthogonal parametrizations are useful to model situations where suitable functions of the parameters – marginal or not – linearly depend on covariates, see for example Azzalini (1994) (with Markovian dependence as treated here), Fitzmaurice and Laird (1994) (with non-Markovian dependence) and Cox and Reid (1987, with discussion) (general case).

We first describe in details a simplified case, considering dependence between two variables as a Markov chain with one transition.

(*) Dipartimento di Matematica, Università di Genova.

Then we study several classes of Markov chains; some of these are esponential models, so the well-known geometrical methods applie; other Markov chains – perhaps the most interesting in the applications – are curved exponential models; in section 3.3 and 3.4 we suggest a geometrical method to generalize the results about the exponential models to the curved models.

Let us briefly recall how an orthogonal parametrization of two subsets of parameters in the exponential models can be obtained; for a full discussion of this subject we refer to Amari (1982), Barndorff-Nielsen and Cox (1994), and Pistone and Rogantin (1995).

Let us consider a parametric and regular statistical model, that is a structure $(\Omega, F, (p_\theta)_{\theta\in\Theta}, \mu)$ where (Ω, F) is the measurable space with sample set Ω, μ is a reference measure, and $(p_\theta)_{\theta\in\Theta}$ is a family of probability μ–a.s. strictly positive densities w.r.t. μ with parameter $\theta\in\Theta$, Θ open set of R^d.

The parametric family of probability densities of the model can be showed to be a d-dimensional differential manifold in the set of all the probability densities equivalent to μ. When regularity conditions are satisfied (see Pistone and Rogantin, 1995), the functions defining the charts for the construction of the differential manifold take values in the parameters set, that is the parametrizations are the coordinate systems.

To realize a chart the mapping from θ to p_θ must be one-to-one: this is the *identifiability* condition of the parameters. The choice of parametrization do not affect the geometric description of the surface points.

The metrics of the manifold of the probability densities is defined by the Fisher information matrix, $I(\theta)$, which varies continuously from point to point. The element (i, j) of $I(\theta)$ is $E\left(\frac{\partial l}{\partial\theta_i}\frac{\partial l}{\partial\theta_j}\right)$, where l is the model log-likelihood.

The Fisher information matrix is covariant to parametrizations; in fact, denoting: by θ and η two different parametrizations of a probability density p, by $I_1(\theta)$ and $I_2(\eta)$ the respective Fisher information matrices, by f the one-to-one mapping of the parameters, $f(\theta) = \eta$, and by $J_{f^{-1}}$ the Jacobian matrix of inverse transformation, then:

$$I_2(\eta) = (J_{f^{-1}}(\theta))^t \quad I_1(\theta) \; J_{f^{-1}}(\theta)$$

In the exponential models, the log-likelihood $l(x; \theta)$ is $-\psi(\theta) + < T(x), \theta >$, where θ is in a convex open set of R^d, the function $\psi(\theta)$ is analytical and strictly convex, and T is a sufficient statistic in R^d whose expectation equals the gradient of ψ and variance equals the Hessian matrix of ψ. Fisher information matrix is the Hessian matrix of ψ.

Let us consider a parametrization such that the maximum likelihood estimates of new parameters are in the convex open set that coincides with the set where the sufficient statistic takes values. Let $f: R^d \supseteq \Theta \rightarrow H \subseteq R^d$, denote this transformation and let η denote the "means parameters":

$$\eta = f(\theta) = \nabla\psi(\theta).$$

In this context the Fisher information matrix $I_2(\eta)$ is the inverse of $I_1(f^{-1}(\eta))$, and the mutual position of the two reference systems of the parameters θ and η in the manifold allows to find a parametrization that makes orthogonal the estimates fo two subsets of parameters. In the following "mixed parametrization"

$$(\eta_1, \ldots, \eta_a, \theta_{a+1}, \ldots, \theta_d)$$

the first a parameters are orthogonal to the last $d-a$, see Amari (1982) and Barndorff-Nielsen and Cox (1994).

A short proof of the orthogonality follows. Let θ_A and θ_B be respectively the vectors $(\theta_1, \ldots, \theta_a)$ and $(\theta_{a+1}, \ldots, \theta_d)$, so that $\theta = (\theta_A, \theta_B)$; let η_A and η_B be respectively the vectors $(\eta_1, \ldots, \eta_a)$ and $(\eta_{a+1}, \ldots, \eta_d)$, so that $\eta = (\eta_A, \eta_B)$; let $\xi = (\eta_A, \theta_B)$ be the "mixed parameterization"; and let g be the transformation of parameters θ into parameters ξ, $g: \theta \rightarrow \xi$.

Fisher information matrix $I_1(\theta)$, exhibiting the blocks of dimension a and $d-a$, is $I_1(\theta) = \begin{pmatrix} I_1^{11}(\theta) & I_1^{12}(\theta) \\ I_1^{21}(\theta) & I_1^{22}(\theta) \end{pmatrix}$; the Jacobian matrix associated to the transformation g is: $J_g = \begin{pmatrix} I_1^{11}(\theta) & I_1^{12}(\theta) \\ 0 & Id \end{pmatrix}$. The Fisher information matrix of the "mixed parametrization" $I_3(g(\theta))$, that is $I_3(\xi)$, written in blocks showing the orthogonality between the two groups of parameters is:

$$I_3(\xi) = \begin{pmatrix} (I_1^{11}(\xi))^{-1} & 0 \\ 0 & I_1^{22}(\xi) - I_1^{21}(\xi)\,(I_1^{11}(\xi))^{-1}\, I_1^{12}(\xi) \end{pmatrix}$$

where $I_1^{ij}(\xi)$ is the matrix I_1^{ij} evaluated in the parameters ξ.

As a consequence of the orthogonality, the maximum likelihood estimates of the two blocks of parameters are asymptotically independent, see for example Barndorff-Nielsen and Cox (1994), p. 98 and Murray and Rice (1993), p. 216.

In the exponential models the dimension of parameter is the same as that of sufficient statistic. More general models, where parameter dimension is lower than sufficient statistic dimension, can be embedded in an ambient exponential super-model (where the dimension of parameter equals that of the sufficient statistic). That is the family of probability densities of this general model results a sub-manifold of that of the exponential model. Such models are called "curved exponential models", see for example Barndorff-Nielsen and Cox (1994), p. 65 and Pistone and Rogantin (1995), p. 33.

Denoting by θ the parameters of the exponential models and by α those of the curved exponential model, then $\theta = \theta(\alpha)$, and a version of the log-likelihood of the curved model is: $l(x; \theta(\alpha)) = -\psi(\alpha) + < T(x), \theta(\alpha) >$.

A curved sub-model inherits the geometry from the exponential model, that is the Fisher information matrix of curved sub-model can be expressed as $J_\theta' I_1(\theta(\alpha)) J_\theta$ (where J_θ is the Jacobian matrix associated to the transformation $\theta(\alpha)$).

This immersion is only an analytical tool that allows to treat more easily the curved model, and does not lose the power of inferential method.

2. Statistical dependence between two binary random variables

Let us first consider in details the investigation method of orthogonal parameters in case of only one transition.

2.1. *Orthogonal parametrization of statistical dependence*

Although the following results are largely known, we analyze them in details in order to explain the extension reported in successive sections.

Let (X, Y) be a couple of random variables with Bernoulli marginal distribution with respective parameters p_1 and p_2; dependence of Y from X can be expressed in Markovian form by the transition matrix P:

$$\begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Then: $p_2 = (1-p)\,\alpha + p_1\,(1-\beta)$.

Let $(X_i, Y_i)_{i=1,\dots,n}$ be an n-dimensional sample of (X, Y).

The free parameters in the model are three. We want to find a parametrization such that the third parameter is orthogonal to (p_1, p_2).

Let N_X and N_Y denote the random variables representing the sum respectively of X_i and Y_i, $i = 1, \dots, n$, which have binomial distribution with parameters respectively (n, p_1) and (n, p_2).

Let N^{00}, N^{01}, N^{10} and N^{11} denote the random variables representing the number of transition from X_i to Y_i, $i = 1, \dots, n$, respectively from 0 to 0, from 0 to 1, from 1 to 0 and from 1 to 1. Then: $N^{00} = \Sigma_{i=1}^{n} (1-X_i)(1-Y_i)$, $N^{01} = \Sigma_{i=1}^{n} (1-X_i)\,Y_i$, and so on. Such random variables have binomial distribution with parameters respectively $(n, (1-p_1)(1-\alpha))$, $(n, (1-p_1)\,\alpha)$, $(n, p_1\,\beta)$ and $(n, p_1\,(1-\beta))$.

Let δ denote the vector of parameters $(\delta_1, \delta_2, \delta_3) = (p_1, \alpha, \beta)$; the space of these parameters is $\Delta =]0, 1[^3$.

A version of the likelihood of the model, written in the parameters δ, is:

$$L\,((x_i, y_i)_{i=1,\dots,n}; \delta) = p_1^{n_X}\,(1-p_1)^{n-n_X}\,\alpha^{n^{01}}\,(1-\alpha)^{n^{00}}\,\beta^{n^{10}}\,(1-\beta)^{n^{11}}. \tag{1}$$

As described, the probability densities family $(p_\delta)_{\delta \in \Delta}$ can be viewed as a 3-dimensional differential manifold in the convex set of all the equivalent probability densities, and the metrics associated to parametrization δ is defined by the Fisher information matrix of the model:

$$I(\delta) = \begin{pmatrix} \dfrac{1}{p_1\,(1-p_1)} & 0 & 0 \\ 0 & \dfrac{1-p_1}{\alpha(1-\alpha)} & \\ & & \dfrac{p_1}{\beta(1-\beta)} \end{pmatrix}.$$

Notice that the three parameters are orthogonal, and that the area element is

$$\det (I(\delta))^{1/2}\, dp_1\, d\alpha\, d\beta = \alpha^{-1/2}(1-\alpha)^{-1/2}\beta^{-1/2}(1-\beta)^{-1/2}\, dp_1\, d\alpha\, d\beta;$$

then the "geometrically speaking uniform distribution" (or non informative prior) on the parameter set is the product of a uniform distribution on p_1, and of two Beta distributions with parameters (1/2, 1/2) on α and β.

Proposition 1. In the model (1) the parameter expressing the dipendence

$$\log\left(\frac{\alpha}{1-\alpha}\,\frac{\beta}{1-\beta}\right)$$

is orthogonal to the parameters of marginal probabilities (p_1, p_2).

Proof. It is possible to write the log-likelihood such that two elements of sufficient statistic are $\frac{N_X}{n}$ e $\frac{N_Y}{n}$, whose expected values are p_1 and p_2; then the third canonical parameter is orthogonal to (p_1, p_2). Choosing $\frac{N^{01}}{n}$ as the third element of sufficient statistic (where three other alternative choices lead to an analogous result), the log-likelihood is:

$$l((x_i, y_i)_{i=1,\dots,n};\delta) = \log((1-p_1)(1-\alpha)) +$$

$$\frac{n_X}{n}\log\left(\frac{p_1}{1-p_1}\,\frac{\beta}{1-\alpha}\right) + \frac{n_Y}{n}\log\left(\frac{1-\beta}{\beta}\right) + \frac{n^{01}}{n}\log\left(\frac{\alpha}{1-\alpha}\,\frac{\beta}{1-\beta}\right). \tag{2}$$

The vector θ of canonical parameters is:

$$\left(\log\left(\frac{p_1}{1-p_1}\,\frac{\beta}{1-\alpha}\right), \log\left(\frac{1-\beta}{\beta}\right), \log\left(\frac{\alpha}{1-\alpha}\,\frac{\beta}{1-\beta}\right)\right)$$

and $\psi(\theta)$ is $n \log(1 + e^{\theta_1} + e^{\theta_1+\theta_2} + e^{\theta_2+\theta_3})$.

The vector η, computed in (p_1, p_2, α), is: $(p_1, p_2, \alpha(1-p_1))$.

Then a "mixed parametrization" is

$$\xi = (\eta_1, \eta_2, \theta_3) = \left(p_1, p_2, \log\left(\frac{\alpha}{1-\alpha}\,\frac{\beta}{1-\beta}\right)\right)$$

and the parameters (p_1, p_2) are orthogonal to the parameter $\log\left(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}\right)$.□

The inverse transformation from the parameters ξ to the parameters δ, and then to parameters θ, is computable but does not have easy closed form because the parameter α is solution of the following quadratic equation:

$$\alpha^2 (1 - p_1) (e^{\theta_3} - 1) + \alpha ((e^{\theta_3} - 1) (p_1 - p_2) - e^{\theta_3}) + e^{\theta_3} p_2 = 0.$$

In order to compute $I_3 (\xi)$, let us write the Fisher information matrix $I_1 (\theta)$ with distinct blocks, for convenience in the parameters (p_1, p_2, α), as follows:

$$I_1 (p_1, p_2 \alpha) =$$

$$\left(\begin{array}{cc|c} p_1 (1-p_1) & (1-p_1)(p_2-\alpha) & (1-p_1)(p_2-\alpha-p_1(1-\alpha)) \\ (1-p_1)(p_2-\alpha) & p_2(1-p_2) & (1-p_1)(1-p_2)\alpha \\ \hline (1-p_1)(p_2-\alpha-p_1(1-\alpha)) & (1-p_1)(1-p_2)\alpha & (1-p_1)\alpha(1-\alpha(1-p_1)) \end{array}\right)$$

To compute a prior distribution with the parameters (p_1, p_2, θ_3) we must have an explicit expression of the determinant of I_3 – for the same reason computed as before in (p_1, p_2, α) – is:

$$p_1^3 (1 - p_1) (1 - \alpha)^2 + (1 - p_2)^3 p_2 \alpha^2 +$$

$$+ (1 - p_1) (p_2 - \alpha) ((p_2 - \alpha) (p_1 - 2 (1 - p_2) \alpha) - p_1 (1 - \alpha) (p_1 - (1 - p_2) \alpha))$$

Notice that the parameter $\log (\frac{\alpha}{1-\alpha} \frac{\beta}{1-\beta})$ equals 0 if $\alpha = 1 - \beta$, that is when Y and X are independent.

When $\alpha = \beta$ (that is the probability of changing state does not depend from the initial state) the matrix is bi-stochastic and the process is also a backward Markov chain. In this case the log-likelihood is:

$$\frac{n_X}{n} \log\left(\frac{p_1}{1-p_1}\right) + \left(\frac{n^{01}}{n} + \frac{n^{10}}{n}\right) \log\left(\frac{\alpha}{1-\alpha}\right) + \log((1-p_1)(1-\alpha));$$

and the parameter p_1 is orthogonal to the parameter $\log (\frac{\alpha}{1-\alpha})$ expressing the transition probabilities.

2.2. *Analogy with contingency tables*

The statistical dependence between two random variables can be expressed by two-way contingency tables. Then the previous results are also useful to find parameters that express interactions between marginal variables and that are orthogonal to the marginal probabilities of the table.

Let N_X and N_Y be the marginal random variables, with binomial distributions of parameters respectively (n, p_1) e (n, p_2). Let $(N^{ij})_{i,j=0,1}$ be the random variables corresponding to the cells of the table. As before, we introduce two parameters α e β so that the row profiles of the table are independent and have binomial distributions with parameters respectively $(n - n_X, 1 - \alpha)$ and (n_X, β).

Then the parameter $\log\left(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}\right)$, describing the interactions between the marginal variables, is orthogonal to marginal parameters p_1 e p_2.

This parameter, denoted by θ, is a natural parameter in the conditional likelihood of the contingency table:

$$L\left((n^{ij})_{i,j=0,1}; n_X, n_Y, p_1, \alpha, \beta\right) = \frac{\binom{n-n_Y}{n^{10}}\binom{n_Y}{n^{01}}\theta^{n^{01}}}{\sum_{i=0}^{n_Y}\binom{n-n_Y}{n-n_X-i}\binom{n_Y}{i}\theta^{i}}.$$

The study and the use of parameters that describe the interactions between random variables and that are orthogonal to marginal probabilities is widely discussed in literature, see for example, Plackett (1977), Lang and Agresti (1994) and Molenberghs and Lasaffre (1994).

3. Binary Markov chains

Let us consider a Markov chain $(X_t)_{t=1,\ldots,T}$, where each random variable X_t has Bernoulli distribution with parameter p_t, and let us consider an its n-dimensional sample, denoted by $((X_{tk})_{k=1,\ldots,n})_{t=1,\ldots,T}$. Let the transition matrix from X_t to X_{t+1} be:

$$P_t = \begin{pmatrix} 1-\alpha_t & \alpha_t \\ \beta_t & 1-\beta_t \end{pmatrix} \qquad \text{with } t = 1, \ldots, T-1$$

with $\alpha_t \neq \beta_t$; then: $p_{t+1} = (1 - p_t)\alpha_t + p_t(1 - \beta_t)$.

For $t = 1, \ldots, T,$
sampling number of
dom variables have
For $t = 1, \ldots, T,$
senting the sampling
t; these random var
$(n, (1 - p_t)(1 - \alpha_t)),$
We shall consid
1. non-homogen
matrices are differen
2. Markov chai
matrices are differen
stant and the initial
3. homogeneous
trices are equal at e
4. stationary M
time and the initial
The first chain i
one transition; but in
parametrized. More i
kov chains, but these
we propose a genera
tion". The "log-odds
one and the third one
case, but it remains i

3.1. *Non-homogene*

Let us consider
ary Markov chain.

Proposition 2. T
chain is an exponent

$\theta_t = \log\Big($ …

are orthogonal to th

For $t = 1, \ldots, T$, let N_t denote the random variable representing the sampling number of values equal to 1 achieved at the time t; these random variables have binomial distributions with parameters $(n, p_t)_{t=1,\ldots,T}$.

For $t = 1, \ldots, T$, let $(N_t^{01})_{t=1,\ldots,T}$ denote the random variable representing the sampling number of transition from 0 to 1 happening at time t; these random variables have binomial distributions with parameters $(n, (1-p_t)(1-\alpha_t))_{t=1,\ldots,T}$.

We shall consider four classes of Markov chains:

1. non-homogeneous and non-stationary Markov chain: the transition matrices are different at each time and the initial probability is general;
2. Markov chain with a constraint on the parameters: the transition matrices are different but the odds ratio of the transitions matrices is constant and the initial probability is general;
3. homogeneous and non-stationary Markov chain: the transition matrices are equal at each time and the initial probability is general;
4. stationary Markov chain: the transition matrices are equal at each time and the initial probability is a stationary probability.

The first chain is an extension of that described before in the case with one transition; but in the applications is not realistic to have models so over-parametrized. More interesting are the homogeneous and the stationary Markov chains, but these ones are "curved" exponential models; for these chains we propose a generalization of the exponential models "mixed parametrization". The "log-odds stable" chain is an intermediate case between the first one and the third one, and we study it because is a restriction of the general case, but it remains in the family of the exponential models.

3.1. *Non-homogeneous and non-stationary Markov chain*

Let us consider an n-sample of a non-homogeneous and non-stationary Markov chain.

Proposition 2. The set of non-homogeneous and non-stationary Markov chain is an exponential model. The parameters expressing the dependence

$$\theta_t = \log\left(\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}\right) \qquad \text{with } t = T+1, \ldots, 2T-1 .$$

are orthogonal to the parameters of marginal probabilities $(p_t)_{t=1,\ldots,T}$.

Proof. The model parameters are T for the marginal probabilities and $T-1$ for the transition probabilities, then in total there are $2T-1$ parameters.

In order to find $T-1$ parameters depending by the transition probabilties that are orthogonal to the marginal probabilities, we follow the same method as in the one transition case. In fact it is possible to write the exponential model so that the first T elements of the sufficient statistic are the random variables $(\frac{N_t}{n})_{t=1,\dots,T}$, whose expected values are $(p_t)_{t=1,\dots,T}$.

A version of the model likelihood is:

$$p_1^{n_1}(1-p_1)^{n-n_1}\prod_{t=1}^{T-1}\alpha_t^{n_t^{01}}(1-\alpha_t)^{n-n_t-n_t^{01}}(1-\beta_t)^{n_{t+1}-n_t^{01}}\beta_t^{n_t-n_{t+1}+n_t^{01}}$$

and its logarithm is:

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}\frac{\beta_1}{1-\alpha_1}\right)+\sum_{t=2}^{T-1}\frac{n_t}{n}\log\frac{\beta_t}{1-\alpha_t}\frac{1-\beta_{t-1}}{\beta_{t-1}}+\frac{n_T}{n}\log\left(\frac{1-\beta_{T-1}}{\beta_{T-1}}\right)+$$

$$\sum_{t=1}^{T-1}\frac{n_t^{01}}{n}\log\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}+\left(\log(1-p_1)+\sum_{t=1}^{T-1}\log(1-\alpha_t)\right) \qquad (3)$$

With the "mixed parametrization" the last $T-1$ elements of the canonical parameter $(\theta_{T+1},\dots,\theta_{2T-1})$, where

$$\theta_t=\log\left(\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}\right) \qquad t=T+1,\dots,2T-1,$$

are orthogonal to the parameters $(p_t)_{t=1,\dots,T}$, that express the marginal probabilities because $(p_t)_{t=1,\dots,T}$ are the expected values of the first T elements of the sufficient statistic. □

3.2. *"Log-odds stable" and non stationary Markov chain*

DEFINITION 3. A non-homogeneous and non-stationary Markov chain with the following constraint on the parameters:

$$\log\left(\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}\right)=\theta \qquad \text{with } \theta \text{ constant } \neq 0$$

is called "log-odds stable" Markov chain.

Let us consider an n-dimensional sample of such a Markov chain.

This model has $T+1$ parameters, and it is a sub-model of the previous model (3) and it includes as particular case the homogeneous Markov chain (with transition matrix independent from the time).

Proposition 4. The set of the "log-odds stable" and non-stationary Markov chain is an exponential model. The parameter expressing the dependence

$$\theta = \log\left(\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}\right)$$

is orthogonal to the parameter of the marginal probabilities $(p_t)_{t=1,\dots,T}$.

Proof. Denoting by N^{01} the random variable representing the whole number of transitions from 0 to 1, a version of the model log-likelihood is:

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}\frac{\beta_1}{1-\alpha_1}\right)+\sum_{t=2}^{T-1}\frac{n_t}{n}\log\left(\frac{\beta_t}{1-\alpha_t}\frac{1-\beta_{t-1}}{\beta_{t-1}}\right)+\frac{n_T}{n}\log\left(\frac{1-\beta_{T-1}}{\beta_{T-1}}\right)+$$

$$+\frac{n^{01}}{n}\log\left(\frac{\alpha_t}{1-\alpha_t}\frac{\beta_t}{1-\beta_t}\right)+\left(\log(1-p_1)+\sum_{t=1}^{T-1}\log(1-\alpha_t)\right) \qquad (4)$$

With the "mixed parametrization" the last element of the canonical parameter θ is orthogonal to the parameters of the marginal probabilities, $(p_t)_{t=1,\dots,T}$, being these the expected values of the first T elements of the sufficient statistic. □

3.3. *Homogeneous and non-stationary Markov chain*

Let us consider an n-dimensional sample of an homogeneous Markov chain where the transition is: $\begin{pmatrix}1-\alpha & \alpha\\ \beta & 1-\beta\end{pmatrix}$.

The marginal probabilities of success p_t, with $t \geq 2$, can be expressed by the parameters (p_1, α, β) as follows:

$$p_t = \frac{\alpha}{\alpha+\beta} - \left(\frac{\alpha}{\alpha+\beta} - p_1\right)(1-\alpha-\beta)^{t-1}.$$

Denoting by N_A the random variable $\Sigma_{t=2}^{T-1} N_t$, a version of the model log-likelihood is:

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}\frac{\beta_1}{1-\alpha}\right) + \frac{n_A}{n}\log\left(\frac{1-\beta}{1-\alpha}\right) + \frac{n_T}{n}\log\left(\frac{1-\beta}{\beta}\right) +$$

$$+\frac{n^{01}}{n}\log\left(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}\right) + \log\left((1-p_1)(1-\alpha)^{T-1}\right). \qquad (5)$$

Proposition 5. The parametric family of probability densities of the homogeneous and non-stationary Markov chain (5) is 3-dimensional.

The model (5) is curved: it is not an exponential model.

The smallest exponential model that contains the model (5) is 4-dimensional, and a version of its log-likelihood, written in the canonical form, is:

$$\frac{n_1}{n}\theta_1 + \frac{n_A}{n}\theta_2 + \frac{n_T}{n}\theta_3 + \frac{n^{01}}{n}\theta_4 - \psi_1(\theta_1, \theta_2 \theta_3, \theta_4). \qquad (6)$$

Let $\mathcal{M}_4$ denote the manifold corresponding to this exponential model.

The model (5) corresponds to a sub-manifold – denoted by $\mathcal{S}_3$ – embedded in the manifold $\mathcal{M}_4$.

Proof. The model (5) is 3-dimensional; in fact the Jacobian matrix associated to the transformation of the parameters (p_1, α, β) into the parameters $(p_1, p_2, \ldots, p_T, \log(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}))$ has rank three (for example the minor whose rows correspond to the gradient of (p_1, p_2, p_3) has strictly negative determinant).

Parametrizing the sub-model (5) with (p_1, β, θ), a version of the log-likelihood is:

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}(\beta + \right.$$

$$+\frac{n_T}{n}\log\left(\frac{1-\beta}{\beta}\right) + $$

The model (5)
tistic is greater than
lation between the p
linear).
We refer to Ba
of the smallest expo
A parametric r

$\mathcal{S}_3$

□

Notice that the
exponential model
itial time, at the fi
ities of success; in
sponding to $\mathcal{M}_4$ is
$\nabla\psi_1$ are $(p_1, p_A, p_T$
The method us
an orthogonal para
plicable here.

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}(\beta+e^{\theta}(1-\beta))\right)+\frac{n_A}{n}\log\left(\frac{1-\beta}{\beta}(\beta+e^{\theta}(1-\beta))\right)+$$

$$+\frac{n_T}{n}\log\left(\frac{1-\beta}{\beta}\right)+\frac{n^{01}}{n}\theta+\log\left((1-p_1)\left(\frac{\beta}{\beta+e^{\theta}(1-\beta)}\right)^{T-1}\right). \qquad (7)$$

The model (5) is curved; in fact the dimension of the sufficient statistic is greater than the dimension fo the parameters (for example the relation between the parameter $\theta = \log\left(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}\right)$ and the other ones is not linear).

We refer to Barndorff-Nielsen and Cox (1994), p. 65, for the form of the smallest exponential model that contains it.

A parametric representation of the sub-manifold $\mathcal{S}_3$ of $\mathcal{M}_4$ is:

$$\mathcal{S}_3:\begin{cases}\theta_1 &= \log\left(\frac{p_1}{1-p_1}(\beta+e^{\theta}(1-\beta)\right)\\ \theta_2 &= \log\left(\frac{1-\beta}{\beta}(\beta+e^{\theta}(1-\beta)\right)\\ \theta_3 &= \log\left(\frac{1-\beta}{\beta}\right)\\ \theta_4 &= \theta\end{cases}$$

□

Notice that the parametrization of the marginal probabilities in the exponential model (6) emphasizes the probability of success at the initial time, at the final time, and the sum of the intermediate probabilities of success; in fact the sufficient statistic of the model (6) corresponding to $\mathcal{M}_4$ is $\frac{1}{n}(N_1, N_A, N_T, N^{01})$ and the first three elements of $\nabla\psi_1$ are (p_1, p_A, p_T).

The method used for the models (3) and (4), in order to investigate an orthogonal parameter to the marginal probabilities, is not directly applicable here.

3.3.1. *"Mixed parametrization" for curved exponential models*

The topics showed in this section are illustrated in Figure 1.

Let us consider a point $\hat{P} = (\hat{p}_1, \hat{\beta}, \hat{\theta})$, $\hat{P} \in \mathcal{S}_3$, and the coordinate line $\gamma(\theta)$ through $\hat{P}$ lying on $\mathcal{S}_3$; a parametric representation is:

$$\gamma(\theta): \begin{cases} \theta_1 & = \log\left(\dfrac{\hat{p}_1}{1-\hat{p}_1}(\hat{\beta} + e^{\theta}(1-\hat{\beta})\right) \\ \theta_2 & = \log\left(\dfrac{1-\hat{\beta}}{\hat{\beta}}(\hat{\beta} + e^{\theta}(1-\hat{\beta})\right) \\ \theta_3 & = \log\left(\dfrac{1-\hat{\beta}}{\hat{\beta}}\right) \\ \theta_4 & = \theta \end{cases} \tag{8}$$

and the direction of the tangent vector to $\gamma(\theta)$ evaluated in $\hat{P}$ is furnished by the derivative of (8):

$$\gamma'(\hat{\theta}) = (\hat{\alpha}, \hat{\alpha}, 0, 1)$$

where $\hat{\alpha}$ is the value of the parameter α computed in $\hat{P}$, that is $\dfrac{1-\hat{\beta}}{\hat{\beta} + e^{\hat{\theta}}(1-\hat{\beta})}$.

Proposition 6. There exist an exponential model corresponding to manifold $\mathcal{M}_4$ having in $\hat{P}$ the fourth coordinate parallel to $\gamma'(\hat{\theta})$; a version of its log-likelihood is:

$$\frac{n_1}{n}\zeta_1 + \frac{n_A}{n}\zeta_2 + \frac{n_T}{n}\zeta_3 + \left(\hat{\alpha}\left(\frac{n_1}{n} + \frac{n_A}{n}\right) + \frac{n^{01}}{n}\right)\zeta_4 - \psi_1(\zeta_1, \zeta_2, \zeta_3, \zeta_4) \tag{9}$$

Moreover:

$$\left(\frac{\partial \psi_1}{\partial \zeta_1}, \frac{\partial \psi_1}{\partial \zeta_2}, \frac{\partial \psi_1}{\partial \zeta_3}\right) = (p_1, p_A, p_T) = \left(p_1, \sum_{2}^{T-1} p_t, p_T\right).$$

Proof. The lo
as:

$\frac{n_1}{n}(\zeta_1 + \hat{\alpha}\zeta_4) + \frac{n_A}{n}$

This model act
$\theta_2, \theta_3, \theta_4$) of $\mathcal{M}_4$ are
($\zeta_1, \zeta_2, \zeta_3, \zeta_4$) throug

$h: (\zeta_1, \zeta_2, \zeta_3, \zeta_4$

Denoting by J
$\psi_1(\zeta_1, \zeta_2, \zeta_3, \zeta_4)$ of
$(\zeta_1, \zeta_2, \zeta_3, \zeta_4) J_h$. In

$\left(\frac{\partial \psi_1}{\partial \zeta_1}, \frac{\partial \psi_1}{\partial \zeta_2}, \frac{\partial \psi_1}{\partial \zeta_3}\right)$

$\frac{1}{n}(E_\zeta(N_1), E_\zeta(N_A$

□

Proposition 7.
rameters (p_1, p_A, p_T
tio of the odds of
Then the coor
noted by $\mathcal{H}(p_1, p_A,$

Proof. The p
ity of the paramet
model correspondi

Proof. The log-likelihood of the model (9) can be also written as:

$$\frac{n_1}{n}(\zeta_1 + \hat{\alpha}\zeta_4) + \frac{n_A}{n}(\zeta_2 + \hat{\alpha}\zeta_4) + \frac{n_T}{n}\zeta_3 + \frac{n^{01}}{n}\zeta_4 - \psi_1(\zeta_1, \zeta_2, \zeta_3, \zeta_4).$$

This model actually corresponds to $\mathcal{M}_4$ because the parameters (θ_1, θ_2, θ_3, θ_4) of $\mathcal{M}_4$ are expressible as linear combinations of the parameters (ζ_1, ζ_2, ζ_3, ζ_4) through the mapping:

$$h : (\zeta_1, \zeta_2, \zeta_3, \zeta_4) \mapsto (\theta_1, \theta_2, \theta_3, \theta_4) \text{ such that } \begin{cases} \theta_1 = \zeta_1 + \hat{\alpha}\zeta_4 \\ \theta_1 = \zeta_2 + \hat{\alpha}\zeta_4 \\ \theta_1 = \zeta_3 \\ \theta_1 = \zeta_4 \end{cases}$$

Denoting by J_h the Jacobian matrix of the mapping h, the function ψ_1 (ζ_1, ζ_2, ζ_3, ζ_4) of the model (9) is such that $\nabla_\theta \psi_1 (\theta_1, \theta_2, \theta_3, \theta_4) = \nabla_\zeta \psi_1 (\zeta_1, \zeta_2, \zeta_3, \zeta_4) J_h$. In particular:

$$\left(\frac{\partial\psi_1}{\partial\zeta_1}, \frac{\partial\psi_1}{\partial\zeta_2}, \frac{\partial\psi_1}{\partial\zeta_3}\right) = \left(\frac{\partial\psi_1}{\partial\theta_1}, \frac{\partial\psi_1}{\partial\theta_2}, \frac{\partial\psi_1}{\partial\theta_3}\right) =$$

$$\frac{1}{n}(E_\zeta(N_1), E_\zeta(N_A), E_\zeta(N_T)) = (p_1, p_A, p_T) = \left(p_1, \sum_2^{T-1} p_t, p_T\right). \quad (10)$$

□

Proposition 7. In the exponential model corresponding to $\mathcal{M}_4$ the parameters (p_1, p_A, p_T) are orthogonal to the parameter θ, that is the log-ratio of the odds of the transition matrix.

Then the coordinate surface of $\mathcal{M}_4$ corresponding to (p_1, p_A, p_T) – denoted by $\mathcal{H}(p_1, p_A, p_T)$ – is orthogonal to the direction of $\gamma'(\hat{\theta})$.

Proof. The parameter ζ_4 of the model (9) is θ and the orthogonality of the parameters results from the "mixed parametrization" of the model corresponding to $\mathcal{M}_4$. □

THEOREM 8. There exist two parameters of the curve exponential model of the homogeneous and non-stationary Markov chain depending by the marginal probabilities (p_1, p_A, p_T) that are orthogonal to the parameter θ, that express the transition probabilities.

Proof. Let $\mathscr{C}$ denote the surface obtained intersecting the surface $\mathscr{H}(p_1, p_A, p_T)$, defined in the Prop. 7, with the sub-manifold $\mathscr{S}_3$ corresponding to the curved model (5).

The dimension of $\mathscr{C}$ is 2. In fact, for each $\hat{P}$ of $\mathscr{S}_3$, the tangent space to $\mathscr{H}$ at the point $\hat{P}$ is orthogonal to the tangent space of $\mathscr{S}_3$ where $\gamma'(\hat{\theta})$ lies; then the whole dimension of the two tangent spaces is 4 (dimension of $\mathscr{M}_4$). Moreover, as already noticed, the dimension of each of these tangent spaces is 3. Because of the Gassmann formula

$$\dim(\mathscr{C}) = \dim(\mathscr{H} \cap \mathscr{S}_3) = \dim(\mathscr{H}) + \dim(\mathscr{S}_3) - \dim(\mathscr{H} + \mathscr{S}_3) = 3 + 3 - 4.$$

Let consider another point $\tilde{P} = (\tilde{\zeta}, \tilde{\theta})$ that belongs to $\mathscr{C}$; let $\gamma'_{\tilde{P}}(\theta)$ be the coordinate surface through $\tilde{P}$. The orthogonal space to the tangent space $\gamma'_{\tilde{P}}(\tilde{\theta})$ is still $\mathscr{H}$, as it results from (10); then the previous conclusion does not depend on the selected point.

The surface $\mathscr{C}$ can be represented by two parameters function of (p_1, p_A, p_T):

$$\mathscr{C}(u, v) = (\varphi_1(p_1, p_A, p_T), \varphi_2(p_1, p_A, p_T)).$$

□

Notice that the degenerate case $\alpha = \beta$, meaning that the transition matrix is bi-stochastic, the model log-likelihood is:

$$\frac{n_1}{n}\log\left(\frac{p_1}{1-p_1}\right) + \left(\frac{n_1}{n} - \frac{n_T}{n} + 2\frac{n^{01}}{n}\right)\log\left(\frac{\alpha}{1-\alpha}\right) + \log\left((1-p_1)(1-\alpha)^{T-1}\right).$$

This model, that is a sub-model of the previous, is an exponential model. In this case the parameter p_1 is orthogonal to the parameter $\log\left(\frac{\alpha}{1-\alpha}\right)$ expressing the dependence.

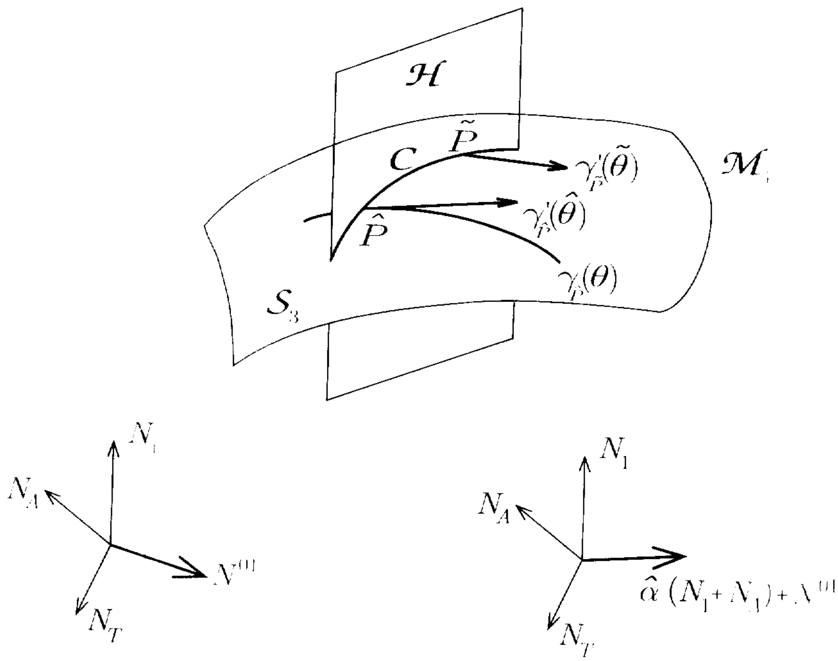

Fig. 1

3.4. *Stationary Markov chain*

Let us consider an n-dimensional sample of a stationary Markov chain.

The initial probability of success equals the marginal probability and, written depending on the transition probabilities, is $\log\frac{\alpha}{\alpha+\beta}$.

Using the previous notation a version of model log-likelihood is:

$$\frac{n_1}{n}\log\left(\frac{\alpha}{1-\alpha}\right)+\frac{n_A}{n}\log\left(\frac{1-\beta}{1-\alpha}\right)+\frac{n_T}{n}\log\left(\frac{1-\beta}{\beta}\right)+$$

$$+\frac{n^{01}}{n}\log\left(\frac{\alpha}{1-\alpha}\frac{\beta}{1-\beta}\right)+\log\left(\frac{\beta}{\alpha+\beta}(1-\alpha)^{T-1}\right). \qquad (11)$$

Proposition 9. The parametric family of probability densities associated to this model is a 2-dimensional sub-manifold embedded in $\mathcal{M}_4$, denoted by $\mathcal{S}_2$.

Proof. It is possible to verify that the Jacobian matrix of the mapping from $\mathcal{M}_4$ to $\mathcal{S}_2$ has rank 2.

A parametric equation of $\mathcal{S}_2$ in $\mathcal{M}_4$ is:

$$\mathcal{S}_3 : \begin{cases} \theta_1 & = \theta + \log\left(\dfrac{\beta}{1-\beta}\right) \\ \theta_2 & = \log\left(\dfrac{1-\beta}{\beta}\,(\beta + e^{\theta}\,(1-\beta))\right) \\ \theta_3 & = \log\left(\dfrac{1-\beta}{\beta}\right) \\ \theta_4 & = \theta \end{cases}$$

□

Then it is possible to follow the proofs related to the sub-manifold $\mathcal{S}_3$ of the homogeneous and non-stationary Markov chain.

Let us consider a point $\hat{P}_1 = (\hat{\beta}, \hat{\theta})$, $\hat{P}_1 \in \mathcal{S}_2$ and the coordinate curve $\gamma_1(\theta)$ through $\hat{P}_1$ lying on $\mathcal{S}_2$; the direction of tangent vector to $\gamma_1(\theta)$ at the point $\hat{\theta}$ is:

$$\gamma'(\hat{\theta}) = (1, \hat{\alpha}, 0, 1)$$

being $\hat{\alpha}$ as the previous Markov chain.

Proposition 10. There exists an exponential model corresponding to manifold $\mathcal{M}_4$ that in $\hat{P}_1$ has the fourth coordinate parallel to $\gamma_1'(\hat{\theta})$; a version of its log-likelihood is:

$$n_1\zeta_1 + n_A\zeta_2 + n_T\zeta_3 + (n_1 + n_A\hat{\alpha} + n^{01})\,\zeta_4 - \psi_2(\zeta_1, \zeta_2, \zeta_3, \zeta_4). \qquad (12)$$

Moreover

$$\left(\frac{\partial\psi_2}{\partial\zeta_1}, \frac{\partial\psi_2}{\partial\zeta_2}, \frac{\partial\psi_2}{\partial\zeta_3}\right) = (p_1, p_A, p_T).$$

Proof. The model (12) corresponds actually to $\mathcal{M}_4$ because the parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$ of $\mathcal{M}_4$ are expressible as linear combinations of the parameters $(\zeta_1, \zeta_2, \zeta_3, \zeta_4)$. It follows:

$$\left(\frac{\partial\psi_2}{\partial\zeta_1}, \frac{\partial\psi_2}{\partial\zeta_2}, \frac{\partial\psi_2}{\partial\zeta_3}\right) = \frac{1}{n}(E_\zeta(N_1), E_\zeta(N_A), E_\zeta(N_T)) = (p_1, p_A, p_T).$$

Proposition 11. In the model corresponding to $\mathcal{M}_4$ the parameters (p_1, p_A, p_T) are orthogonal to the parameter θ. Then the surface of $\mathcal{M}_4$ corresponding to (p_1, p_A, p_T) – denoted by $\mathcal{H}_1(p_1, p_A, p_T)$ – is orthogonal to the direction of $\gamma_1'(\theta)$ and this does not depend on the selected point $\hat{P}$.

Proof. Analogous to that of the homogeneous and non-stationary Markov chain. □

THEOREM 12. There exists one parameter of the homogeneous and non-stationary Markov chain depending on the marginal probabilities (p_1, p_A, p_T) that is orthogonal to the dependence parameter θ.

Proof. Intersecting $\mathcal{H}_1(p_1, p_A, p_T)$ with the sub-manifold $\mathcal{S}_2$, corresponding to the curved model (11), we obtain a curve that can be represented with one parameter depending on (p_1, p_A, p_T), then this parameter is orthogonal to the parameter θ. □

4. Conclusions

We have shown the existence of a "mixed parametrization" for two *curved* exponential models corresponding to homogeneous Markov chain and to stationary Markov chain.

The explicit computation of the parameters depending on (p_1, p_A, p_T) presents some difficulties in order to invert the formulas of p_A and p_T as function of (p_1, β, θ) (and T). this computation is easy only for specific values of parameters.

The results presented in this paper for binary Markov chain are more general. If a block of parameters is one-dimensional it is well-known that one can find a second block of parameters orthogonal to the first one, see for example Cox and Reid (1987) and Barndorff-Nielsen and Cox (1994). But the geometrical methods used in this paper are independent of the parameters dimension. A paper extending the previous results to Markov chains with a number of states greater than 2 is in progress.

ACKNOWLEDGEMENTS

The Author thanks G. Pistone for the constant encouragement and the constructive comments and A. Azzalini, A. Forcina and G. Marchetti for the discussions about the state of research on this topic.

REFERENCES

AMARI, S.I. (1982) Differential geometry of curved exponential families-curvature and information loss, *The Annals of Statistics*, 18 n. 2, 357-385.

AZZALINI, A. (1994) Logistic regression for autocorrelated data with application to repeated measures, *Biometrika*, 81, 767-75.

BARNDORFF-NIELSEN, O.E. and COX, D.R. (1994) *Inference and Asymptotics*, Monographs in statistics and applied probability, 52, London: Chapman & Hall.

COX, D.R. and REID, N. (1987) Parameter orthogonality and approximate conditional inference (with discussion), *J. R. Statist. Soc.*, B 49, 1-39.

FITZMAURICE, G.M. and LAIRD, N.M. (1994) A likelihood-based methods for analyzing longitudinal binary responses, *Biometrika*, 81, 141-51.

LANG, J.B. and AGRESTI, A. (1994) Simultaneously modelling joint and marginal distributions of multivariate categorical responses, *J. Amer. Statist. Ass.*, 89, 625-632.

MOLENBERGHS, G. and LESAFFRE, R. (1994) Marginal modelling of correlated ordinal data using a multivariate Plackett distribution, *J. Amer. Statist. Ass.*, 89, 633-644.

MURRAY, M.K. and RICE, J.W. (1993) *Differential geometry and statistics*. Monographs in statistics and applied probability, 48 London: Chapman & Hall.

PISTONE, G. and ROGANTIN, M.P. (1990) Gli strumenti della geometria differenziale nell'inferenza statistica. In *Memorial Beniamino Gulotta, Giornate di lavoro di probabilità e statistica*, 85-99, Istituto Gramsci, Palermo.

PISTONE, G. and ROGANTIN, M.P. (1994) *The transformation of the non-parametric statistical manifold under conditioning and sampling*, 57th IMS and 3rd WC BS, Chapel Hill June 20-25 1994.

PISTONE, G. and ROGANTIN, M.P. (1995) *Geometria differenziale: strumenti dell'inferenza statistica*, Corso di Formazione, XXXVII Riunione Scientifica Società Italiana di Statistica, Sanremo 9 aprile 1994, Rapporto interno n. 1/1995; Dipartimento di Matematica del Politecnico di Torino.

PISTONE, G. and SEMPI, C
of all the probabil

PLACKETT, R.L. (1977) Th

Geom

We search for a para
splits into two mutually o
ities and the latter one re
ods of differential geomet
In this paper we first
dence between two rando
amine some classes of M
well-known geometrical m
interesting in the applicat
pose a geometrical metho
tial models.

Modellizza

Si cerca una paramet
parametri sia suddiviso i
probabilità marginali e il
strumenti della geometria
Dapprima è illustrato
pendenza di due variabili
gono poi studiate alcune
tipo esponenziale e quin
– forse le più significati
delli di modelli esponenz
metrico – che permette d
esponenziali.

KEY WORDS

Exponential models;
chains.

PISTONE, G. and SEMPI, C. (1995) An infinite dimensional geometric structure on the space of all the probability measures equivalent to a given one, *Annals of Statistics*, 23, 5.

PLACKETT, R.L. (1977) The marginal totals of a 2×2 table, *Biometrika*, 64, 37-42.

Geometrical modelling of Markovian dependence

SUMMARY

We search for a parametrization of a Markov process such that the set of parameters splits into two mutually orthogonal blocks, the first one representing the marginal probabilities and the latter one representing the dependence. This aim is achieved by using methods of differential geometry.

In this paper we first explain in detail a simplified case where we consider the dependence between two random variables as a Markov chain with one transition. Then we examine some classes of Markov chains. Some of those are exponential models, so that the well-known geometrical methods applie. Instead other Markov processes – perhaps the most interesting in the applications – are curved exponential models; for those models we propose a geometrical method which generalize the "mixed parametrization" of the exponential models.

Modellizzazione geometrica della dipendenza markoviana

RIASSUNTO

Si cerca una parametrizzazione di un processo markoviano in modo che l'insieme dei parametri sia suddiviso in due blocchi fra loro ortogonali il primo dei quali esprima le probabilità marginali e il secondo esprima la dipendenza. A tal scopo si utilizzano alcuni strumenti della geometria differenziale.

Dapprima è illustrato nei dettagli un caso semplificato in cui si considera la sola dipendenza di due variabili (vista come catena di Markov con una sola transizione). Vengono poi studiate alcune classi di catena di Markov. Alcune di queste sono modelli di tipo esponenziale e quindi i metodi geometrici noti possono essere applicati; altre classi – forse le più significative dal punto di vista delle applicazioni – invece sono sottomodelli di modelli esponenziali e per questi viene proposto un metodo – sempre di tipo geometrico – che permette di generalizzare ai cosiddetti modelli curvi i risultati sui modelli esponenziali.

KEY WORDS

Exponential models; mixed parametrization; orthogonal parameters; binary Markov chains.