Some stationary Markov processes in discrete time for unit vectors

L. Accardi, J. Cabrera, G.S. Watson[1]

[1]L. Accardi, University of Rome II, J. Cabrera, Rutgers University, and G.S. Watson, Princeton University.

Contents

1 Introduction

The statistics of directional data in which the observations are unit vectors, independently and identically distributed with various probability distributions (von Mises, Fisher, Scheidegger–Watson, etc.) has been extensively discussed see e.g. Mardia (1972), Watson (1983). There is also a literature on joint distributions of two unit vectors (see e.g. Jupp & Mardia (1980). Saw (1983), Rivest (1982, 1983)) and discussions of definitions of correlation between two vectors. In particular, Jupp & Mardia give a general joint density between unit vectors x and y which is proportional to

$$\exp(\kappa_1 x^t \mu + x^t y + \kappa_2 y^t \nu) \tag{1}$$

They also point out that natural definitions of the correlation between unit vectors will flow, like canonical correlations, from the covariance matrix

$$\begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \tag{2}$$

But for unit vectors there is no analogue of the stationary time series literature except for a brief note by Wehrly & Johnson (1979) giving a Markov process for the circular case. We show here a variety of simple models which lead to stationary Markov processes for directions with marginal distributions of the familiar types or generalizations or variations of them. These processes are easy to simulate if one can simulate the marginal $g(x)$, and conditional $f(x_{j+1}|x_j)$, probability densities. To calculate serial correlations, i.e. correlations between X_1 and X_n we would need to calculate the covariance matrices of X_1 and X_n with themselves and each other as in (2). The likelihoods of data from these processes can be written down simply so statistical inference can be done straightforwardly. Only passing remarks will however be made here on statistical problems.

While the method of construction given below arose when considering interacting quantum systems by using quantum Markov Chains, (see Accardi and Watson (1986)), the device as needed here is formally trivial. Let Ω be some space with measure $\omega(\cdot)$ and $f : \Omega \times \Omega \to \mathbb{R}^+$. Suppose $f(x_1, x_2) = f(x_2, x_1)$ and that

$$c^{-1} = \int_{\Omega\times\Omega} f(x_1, x_2)\omega(dx_1)\omega(dx_2) \quad \text{exists.}$$

Let

$$g(x) = \int_\Omega cf(x_1, x_2)\omega(dx_2)$$

Then $g(x)$ is the invariant density of a stationary Markov Chain and the conditional density of X_j, given $X_{j-1} = x_{j-1}$, is given by

$$f(x_j|x_{j-1}) = cf(x_{j-1}, x_j)/g(x_{j-1})$$

In this paper Ω will be

$$\Omega_q = \{x : \|x\| = 1,\ x\varepsilon\mathbb{R}^q\} \text{ and } \omega = \omega_q$$

the invariant density on Ω_q whose integral is the usual total measure of Ω_q e.g. 4π where $q = 3$. Useful forms for $\omega_q(dx)$ can be found by using exterior differential forms. Here will only need the simplest result: for x, $\nu\varepsilon\Omega_q$, set $x = \nu(\nu \cdot x) + \{1 - (\nu \cdot x)^2\}^{1/2}y$,

$$\|y\| = 1,\ y \cdot \nu \text{ where } \nu \cdot x = \nu^t x = r\ , \quad \|y\|^2 = y \cdot y$$

Then

$$\omega_q(dx) = (1 - r^2)^{(q-3)/2} dr\omega_{q-1}(dy)$$

The various models were derived by choosing the exponential part of f (thinking of this as an interaction between neighbouring vectors) as a function of various scalar products, the remaining parts of f coming from normalization considerations. Only these latter parts vary with the dimension q. So rather than giving the models for arbitrary q, we will give the results for $q = 3$ (the sphere) in § 2 and in § 3 for $q = 2$ (the circle), the practical cases, to facilitate referencing them in a future statistical paper. In § 4, we show some realizations for $q = 3$ models and discuss some of their properties. In § 5, one illustration is given of a generalization beyond the nearest neighbour interactions used in the rest of the paper.

The method used in this paper is not of course the only method for building models for dependent unit vectors on the integers but leads to a wide class of models from which we have given useful examples. It seems much more difficult to make models for dependent vectors on the integer lattice in the plane, a case of some interest.

2 Stationary Markov motions on the sphere

We use the notation of § 1.

Model (2.1)

$$g(x) = (4\pi)^{-1} \tag{1}$$

$$f(x_{j+1}|x_j) = \{\lambda/(4\pi \sinh \lambda\} \exp \lambda x_j \cdot x_{j+1} \quad \lambda > 0 \tag{2}$$

Thus the marginal distribution is uniform and the conditional density of X_{j+1}, given $X_j = x_j$, is Fisher with modal vector x_j and concentration parameter λ which is also the interaction parameter i.e. when $\lambda = 0$ the successive X_j are independent. The joint distribution of two successive vectors is a special case (1). The joint density function (j.d.f) of $X_i, \ldots, X_n$ for $n \geq 2$ is given by

$$(4\pi)^{-n}(\lambda/\sinh \lambda)^{n-1} \exp \lambda(x_1 \cdot x_2 + \cdots + x_{n-1} \cdot x_n) \tag{3}$$

so that statistical inference will be particularly simple. For example the test of $\lambda = 0$ will be based on $x_1 \cdot x_2 + \cdots + x_{n-1} \cdot x_n$ Beran & Watson (1967) studied the permutation distribution of this statistic.

Model (2.2)

$$g(x) = 4\pi c_3(\lambda)(\lambda x \cdot \nu)^{-1} \sinh \lambda x \cdot \nu, \quad \lambda \geq 0 \tag{4}$$

$$f(x_{j+1}|x_j) = (\lambda x_j \cdot \nu)(4\pi \sinh \lambda x_j \cdot \nu)^{-1} \exp\{\lambda(x_j \cdot \nu)(x_{j+1} \cdot \nu)\} \tag{5}$$

The marginal distribution is novel, with mode at $x = \nu$, and λ, the interaction parameter, controls the concentration of the distribution of X which is rotationally symmetric about ν. But this distribution is a marginal of (1). It is also antipodally symmetric i.e. $g(x) = g(-x)$. The conditional density of X_{j+1}, given $X_j = x_j$ is Fisher about ν, with concentration parameter $\lambda(x_j \cdot \nu)$. The normalization function $c_3(\lambda)$ is given by

$$(4\pi)^2 c_3(\lambda) = \lambda/ \text{ shi } \lambda, \text{ shi } z = \int_0^z y^{-1} \sinh y dy \tag{6}$$

The j.d. f of $X_1, \ldots, X_n$ is given for $n \geq 3$

$$c_3(\lambda)\Pi_2^{n-2}(\lambda x_j \cdot \nu)(4\pi \sinh \lambda x_j \cdot \nu)^{-1} \exp\{\lambda \Sigma_1^{n-1}(x_j \cdot \nu)(x_{j+1} \cdot \nu)\} \tag{7}$$

Model 2.3

For κ, $\lambda \geq 0$, and a normalizing function $d_3(\kappa, \lambda)$,

$$g(x) = d_3(\kappa, \lambda) \exp(\kappa x \cdot \nu)\{\kappa(\lambda x \cdot y + 1)\}^{-1} 4\pi \sinh \kappa(\lambda x \cdot \nu + 1) \tag{8}$$

$$f(x_{j+1}|x_j) = \tag{9}$$

$$\{\kappa(\lambda x_j \cdot \nu + 1)\}\{\sinh \kappa(\lambda x_j \cdot \nu + 1)\}^{-1} \exp \kappa(\lambda x_j \cdot \nu + 1)x_{j+1} \cdot \nu)$$

The marginal distribution is novel but reduces to the Fisher distribution (x, κ) when the interaction parameter $\lambda = 0$. Its mode is at ν for all λ. The conditional density is however Fisher about ν with concentration $\kappa(\lambda x_j \cdot \nu + 1)$. For $n \geq 3$, the j.d.f. of $X_1, \ldots, X_n$ is given by

$$d_3(\kappa, \lambda)\Pi_2^{n-1}\{\kappa(\lambda x_j \cdot \nu + 1)/ \sinh \kappa(\lambda x_j \cdot \nu + 1)\}\times$$

$$\exp\{\kappa\Sigma_2^{n-1}(x_j \cdot \nu + \lambda x_j \cdot \nu x_{j+1} \cdot \nu + x_{j+1} \cdot \nu)\} \tag{10}$$

From (10) we see that model (2) is a special case of model (3) and further that the joint density of two successive obervations is a special form (1) — $\mu = \nu = \nu$, $A = \lambda\kappa\nu\nu^t$, $\kappa_1 = \kappa_2 = \kappa$. The joint density in Model 2.1 can be obtained by setting $\kappa_1 = \kappa_2 = 0$ and $A = \lambda$ I. Indeed, we could have made a Markov process from (1) by taking $\mu = \nu$, $\kappa_1 = \kappa_2$ and by choosing A to be any symmetric matrix but for general A the normalization is difficult so the simulation of the process will be more difficult again. We will see in the next section that the special cases behave differently so it is best to list them separately.

Model (2.1)

$$g(x) = (4\pi)^{-1} \tag{11}$$

$$f(x_{j+1}|x_j) = \{M(\lambda)\}^{-1} \exp \lambda(x_j \cdot x_{j+1})^2 \tag{12}$$

where

$$M(\lambda) = (2\pi) \int_{-1}^{1} \exp \lambda y^2 dy \tag{13}$$

The conditional density if X_{j+1}, given $X_j = x_j$, is Scheidegger–Watson ($S-W$) about x_j with concentration λ. It is antipodally symmetric. The j.d.f. of $X_1 \ldots X_n$ is for $n \geq 3$

$$(4\pi)^{-1}(M(\lambda))^{-(n-1)} \exp \lambda\Sigma(x_j \cdot x_{j+1})^2 \tag{14}$$

Model (2.5)

With the normalization function $b_3(\kappa, \lambda)$

$$g(x) = b_3(\kappa, \lambda) \exp \kappa (x \cdot \nu)^2 M[\kappa\{\lambda(x \cdot \nu)^2 + 1\}] \tag{15}$$

$$\begin{aligned} f(x_{j+1}|x_j) = {} & [M\{\kappa(\lambda(x_j \cdot \nu)^2 + 1)\}] \\ & \exp\{\kappa(\lambda(x_j \cdot \nu)^2 + 1)(x_{j+1} \cdot \nu)^2\} \end{aligned} \tag{16}$$

The marginal distribution is novel and reduces to the $S - W$ distribution when the interaction parameter $\lambda = 0$. The j.d.f. of $X_1, \ldots, X_n$ is for $n \geq 3$

$$b_3(\kappa, \lambda) \exp \kappa (x, \cdot \nu)^2 \Pi[M\{\kappa(\lambda(x_j \cdot \nu)^2 + 1)\}]^{-1} \times$$

$$\exp \kappa \Sigma\{(x_j \cdot \nu)^2 + \lambda(x_j \cdot \nu)^2 (x_j \cdot \nu)^2 + (x_{j+1} \cdot \nu)^2\} \tag{17}$$

In conclusion, models 2.2, 2.4, 2.5 will be applicable to axial data (undirected lines), while 2.1 and 2.3 are suitable only when the marginal distribution is uniform the remaining models are obtained. Model 2.3 has a marginal density that approximates the Fisher distribution — nome have this marginal exactly.

3 Markov motions on a circle

Here we use unit vectors in $\mathbb{R}^2$ and need the integral

$$\int_0^{2\pi} \exp \lambda \cos \theta d\theta = 2\pi I_0(\lambda) \tag{1}$$

Model (3.1)

$$g(x) = (2\pi)^{-1}$$

$$f(x_{j+1}|x_j) = \{2\pi I_0(\lambda)\}^{-1} \exp \lambda x_{j+1} \cdot x_j \tag{2}$$

so the conditional density is a von Mises density about x_j with concentration parameter λ. This process was constructed differently by Wehrly & Johnson (1979).

Model 3.2

With a normalizing factor $c_2(\lambda)$,

$$g(x) = 2\pi c_2(\lambda) I_0(\lambda x \cdot \nu) \tag{3}$$

$$f(x_{j+1}|x_j) = \{2\pi I_0(\lambda x_j \cdot \nu)\}^{-1} \exp\{\lambda(x_j \cdot \nu)(x_{j+1} \cdot \nu)\} \tag{4}$$

The marginal density has not to our knowledge been exploited but the conditional density is von Mises about ν with concentration $\lambda x_j \cdot \nu$.

Model 3.3
With a normalizing function $d_2(\kappa, \lambda)$

$$g(x) = d_2(\kappa, \lambda) \exp(\kappa x \cdot \nu) 2\pi I_0\{\kappa(\lambda x \cdot \nu + 1)\} \tag{5}$$

$$f(x_{j+1}|x_j) = [2\pi I_0\{\kappa(\lambda x_j \cdot \nu + 1)\}]^{+1} \exp\{\kappa(\lambda x_j \cdot \nu + 1)(x_{j+1} \cdot \nu)\} \tag{6}$$

The marginal density is again and reduces to the von Mises when $\lambda = 0$. The conditional density is von Mises.

Model 3.4

$$g(x) = (2\pi)^{-1} \tag{7}$$

$$f(x_{j+1}|x_j) = \{N(\lambda)\}^{-1} \exp \lambda (x_j \cdot x_{j+1})^2 \tag{8}$$

where

$$N(\lambda) = \int_0^{2\pi} \exp \lambda \cos^2 \theta d\theta$$

Model 3.5

$$g(x) = b_2(\kappa, \lambda) \exp \kappa (x \cdot \nu)^2 N[\kappa\{\lambda(x \cdot \nu)^2 + 1\}] \tag{9}$$

$$f(x_{j+1}|x_j) = [N\{\kappa(\lambda(x_j \cdot \nu)^2 + 1)\}]^{-1} \exp \kappa(\lambda(x_j \cdot \nu)^2 + 1)(x_{j+1} \cdot \nu)^2 \tag{10}$$

The marginal density is novel but reduces to a known form when $\lambda = 0$.

Since these models parallel those in the last section, we only have to say what antipodal symmetry here means. Since

$$(x \cdot \nu)^2 = \cos^2 \theta = (1 + \cos 2\theta)/2$$

we may sometimes simplify the formulae and analysis.

4 Properties of the models

For new models of stationary processes of real random variables one would naturally work out serial correlations and spectra. It is not so clear what should be worked out here. There is no theory for spectra and no universally agreed definition of correlation. One could try to work out the correlation definitions given in Jupp & Mardia but we have only worked out the covariance matrix of X_1 and X_n for models 2.1 and 2.4, the easy cases. Such formal calculations may not show features of interest (see e.g. the discussion of model 2.2) so we have examined sample paths.

The simulation programs for the models of Section 3 were written in MACPASCAL and are available from GSW & or JC. In each case the program generates a vector from the invariant distribution to start the sequence. One may plot the vectors as points on equal area projections of the north and south hemispheres and/or join the successive points with straight lines — the figures show the rules when neighbouring points are in different hemispheres. For the models with a unit vector parameter it has been set at the north pole, that is at the center of te northern hemsphere.

Model 2.1 is a symmetric random walk on the sphere because (2) shows that the distribution of X_{j+1} is rotationally symmetric about x_j. As λ increases, the step size decreases and so the paths look more and more like a the paths of a continuous Brownian motion on the sphere. Roberts & Ursell (1960) have studied these motions. Figure 1 shows a path with 3597 steps which began at a random point in the northern hemisphere and was generated by $\lambda = 77$.

To compute the covariance matrix of X_n and X_1, we first observe that $EX_n = 0$ because of the uniform marginal. As the conditional density of X_n is Fisher x_{n-1}, λ, it is easy to show that

$$E(X_n|X_{n-1} = x_{n-1}) = x_{n-1}L(\lambda)$$

where L is the Langevin function $\coth\lambda - \lambda^{-1}$ which is known to lie in $[0, 1]$ attaining the ends of this interval only when λ is zero or infinity. It is also known that

$$EX_nX_n^t = I_3/3$$

Hence

$$EX_nX_0 = L(\lambda)^n I_3/3$$

and so any definition of correlation will decrease geometrically as n tends to infinity.

Model 2.2 exhibits a very interesting behaviour as is seen in Figure 2. This path, using $\lambda = 10$, started at a random point in the southern hemisphere and at $N = 1395$ moved to th northern hemisphere — the transition may be seen in the figure — and remained there for the rest of the run. The antipodal symmetry of the invariant density guarantees that the point will endlessly switch hemispheres — let's call these switches "reversals". Now this behaviour reminds us of the earth's magnetic polarity. It is known that magnetic north wanders around in the region of its current position but occasionally switches to the region of the current south pole — these are called reversals in geo–magnetism.

Of course the dispersion about the usual positions is much smaller than is seen in this run. This dispersion is reduced by increasing λ which also increases the times between reversals e.g. none were seen in a 3 hour run with $\lambda = 30$.

Thus an open problem is to calculate the distribution of the reversal time for this model. One might guess that an approximation to the mean value would be the inverse of the probability the $X_{j+1} \cdot \nu > 0$ given that $X_j \cdot \nu < 0$, a formula for which is easy to find.

The marginal distribution (4) is novel. Its density falls off slower, as one moves away from either of the two modes, than the Scheidegger–Watson density (proportional to $\exp \lambda(x \cdot \nu)^2$ appearing in Models 2.4 and 2.5. From (7) we see that, for large n, inference will be simple. The maximum likelihood estimator of ν is the eigen–vector of $(\Sigma x_j x_{j+1}^t + \Sigma x_{j+1} x_j^t)/2$ associated with its largest eigenvalue.

Model 2.3 gives paths centered on the vector ν, the north pole in Figure 3, in a rotationally symmetric way since the density depends on scalar products. On comparing (7) and (10), we see that when λ is very large, this model should behave somewhat like Model 2.2, as is seen in Figure 3. That is, when a vector appears in one hemisphere the path should stay there for a while, although the hemispheres are not here quite symmetric (that containing ν is favoured). The combined effect of κ & λ is to tighten up the distribution, κ having the larger effect.

Model 2.4 has no orientation and must produce paths the cover the sphere uniformly. It is a random walk for axes just as model 2.1 is for vectors λ controls the step length, as may be seen from Figures 4a,b. The sufficient statistics is $\Sigma(x_j \cdot x_{j+1})^2$. To show how the serial correlations must behave for this model, we observe that $EX_n = 0$ and that

$$E(X_n | X_{n-1} = x_{n-1}) = x_{n-1}\{t \exp \lambda t^2 dt / M(\lambda)\}$$

where the factor in braces lies in $[0, 1]$. Thus as in Model 2.1 all definitions of correlations of X_n and X_0 must tend geometrically to zero as n increases.

Model 2.5 is the axial version of model 2.3. The step length is mainly controlled by κ. Figure 5 shows a path with $\kappa = 0.5$, $\lambda = 10$. A path with $\kappa = 2.5$, $\lambda = 10$ stayed within 40 degrees of ν for 1000 steps.

The models in section 3 behave similarly but one cannot illustrate their behaviour in any neat way.

5 Extension

Returning to the general construction of Section 1, suppose that $f\ (x_1, x_2, x_3)$ is a density on $\Omega \times \Omega \times \Omega$ such that the marginal density of X_1 and X_2 has the same functional form as that of X_2 and X_3. Call it $h(\dots)$. If $g(x_j, x_{j+1})$ is the invariant density of two successive random variables, the joint distribution of $X_2, \dots, X_n$ would be

$$g(x_1, x_2)\{f(x_1, x_2, x_3)/h(x_1, x_2)\} \dots \{f(x_{n-2}, x_{n-1}, x_n)\}$$

Thus it is clear that $g = h$ and that we have a way of generating stationary sequences with 2 step dependence.

Consider the example for unit vectors in 3 dimensions. Let

$$f = c \exp\{\kappa(x_1 \cdot \nu + x_2 \cdot \nu + x_3 \cdot \nu) + \lambda(x_1 \cdot x_2 + x_2 \cdot x_3)\}$$

Then

$$h = c \exp\{\kappa(x_1 \cdot \nu + x_2 \cdot \nu) + \lambda x_1 \cdot x_2\} \exp\{\kappa x_3 \cdot \nu + \lambda x_2 \cdot x_3\}\omega(dx_3)$$

$$= c \exp\{\kappa(x_1 \cdot \nu + x_2 \cdot \nu) + \lambda x_1 \cdot x_2\}[4\pi \sin|\kappa\nu + \lambda x_2|/|\kappa\nu + \lambda x_2|]$$

One gets the same function on integrating out x_3 so $h = g$. The marginal density of the process is found by integrating out x_1 from h and its

$$c\{4\pi \sinh|\kappa\nu + \lambda x_2|/|\kappa x_2|\}^2 \exp \kappa\nu \cdot x_2$$

An xample of a process for axes may be found by starting out with

$$f = c \exp\{\kappa\Sigma(x_1 \cdot \nu)^2 + \lambda(x_1 \cdot x_2 + x_2 \cdot x_3)\}$$

but integrating out is here a problem. The construction could in principe, provide examples with any range dependence.

A very different way of constructing processes on the sphere is as follows. Let $\{t_n\}$ be an arbitrary stationary process where $-1 < t_n < 1$. Let $\{\zeta\}$ be a sequence of independent unit vectors, orthogonal to the unit vector ν and uniformly distributed. Define

$$X_n = \nu t_n + (1 - t_n^2)^{1/2} \zeta_n$$

Then the sequence $\{X_n\}$ is a stationary sequence of unit vectors — the dimension is arbitrary. Models 2.2, 2.3, 2.5 all have this form.

ACKNOWLEDGEMENTS

The work of Accardi and Watson was partially supported by Contract N00014–84–K–0421 from the Office of Naval Research and Grant MCS821381 from the National Science Foundation.

References

[1] ACCARDI, L., WATSON, G.S. (1986). *Markov states of the quantum electromagnetic field*. Phys. Rev. A **35** (1987) 1275–1283.

[2] BERAN, N., WATSON, G.S. *Testing a sequence of unit vectors for serial correlation*. J. Geophys. Rev., **72** (22) 5655–5659.

[3] JUPP, P.E., MARDIA, K.Y. (1980). *A general correlation coefficient of directional data and related regression questions*. Biometrika, **67** (1) (1980) 163–173.

[4] MARDIA, K.Y. *Statistics of directional data*. Academic Press, New York, pp. 355 (1972).

[5] RIVEST, J.P. *Bivariate circular distributions and correlation*. J. Roy. Stat. Soc., **11** (1982) 81–90.

[6] RIVEST, J.P. *Notions of symmetry for bivariate distributions*. Ann. Statist., **12** (3) (1984) 1085–1089.

[7] SAW, J.G. *Bivariate models with rotational correlation and arbitrary marginals*. Biometrika, **70** (1983) 1085–1089.

[8] WATSON, G.S. *Statistics on Spheres*. Wiley, New York, pp. 238 (1983).

[9] Vehrly, T.E., Johnson, R.A. *Bivariate models for dependence of angular observations and a related Markov process*. Biometrika, **66** (1) (1979) 255–256.

Summary

Some simple parametric models are given for stationary sequences of unit vectors. Their behaviour is mainly shown by simulations. One process is reminescent of the changing polarity of the earth's magnetic field. Some extensions are suggested.

Riassunto

Dei semplici modelli parametrici vengono dati per serie stazionarie di vettori unitari. Il loro comportamento viene illustrato attraverso delle simulazioni. Uno dei processi può rappresentare le variazioni nella polarità del campo magnetico terrestre. Vengono suggerite alcune estensioni.