Bernoulli **6**(1), 2000, 183–190

A sufficiency property arising from the characterization of extremes of Markov chains

PAOLA BORTOT[1] and STUART COLES[2]

[1]*Department of Statistical Sciences, University of Padova, Padova, Italy.*
E-mail: bortot@hal.stat.unipd.it
[2]*Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK.*
E-mail: s.coles@lancs.ac.uk

At extreme levels, it is known that for a particular choice of marginal distribution, transitions of a Markov chain behave like a random walk. For a broad class of Markov chains, we give a characterization for the step length density of the limiting random walk, which leads to an interesting sufficiency property. This representation also leads us to propose a new technique for kernel density estimation for this class of models.

Keywords: extreme value theory; kernel density estimation; Markov chain; random walk; sufficient statistics

1. Introduction

The characterization of the extremal behaviour of Markov chains represents a major advance in the applicability of extreme value techniques to genuine time series of data. Smith (1992) obtained the following representation that is particularly amenable for statistical modelling. Let $\{X_i\}$ denote a stationary first-order Markov chain with bivariate distribution of consecutive pairs $P(X_1 \leqslant x_1, X_2 \leqslant x_2) = F(x_1, x_2)$ and marginal distribution function $F_1(x) = \exp(-\mathrm{e}^{-x})$, the standard Gumbel distribution. Under mild regularity conditions, as $u \to \infty$ and for fixed p, given $X_1 > u$, the differences $X_2 - X_1, X_3 - X_2, \ldots, \mathrm{X}_p - X_{p-1}$ are approximately independent of each other and of u. The implications for Markov chains with a more general marginal specification follow by marginal transformation.

For practical purposes Smith's characterization suggests first transforming an observed series to have Gumbel margins, and then modelling extremal episodes of the transformed chain – in the sense of an arbitrary observation X_i having exceeded a high threshold u – as a random walk with step length distribution given by

$$H(z) = \lim_{u\to\infty} P(\mathrm{X}_{i+1} \leqslant u + z | X_i = u). \tag{1}$$

A brief simulation study in Section 2 of this paper suggests that provided the temporal dependence is not too weak, the limiting random walk is a reasonable approximation to the true Markov chain at low enough levels for inference to be feasible. Smith *et al.* (1997) adopt the model for simulating the stochastic behaviour of functionals of extremal episodes, such as the distribution of the number or aggregate of exceedances of a high threshold u. These techniques require the specification of valid models for the step length distribution H in equation (1). The first aim of this paper, developed in Section 2, is to give an explicit characterization of admissible models for H. In Section 3 we go one stage further and investigate the inferential properties of the limiting random walk model itself – again, the simulations suggest that in most cases such inferences will give good approximations to the true Markov chain model. This gives rise to a curious and surprising sufficiency result. Finally, in Section 4, we consider the impact of these results for nonparametric estimation of the step length density, leading to a proposal for kernel estimation that respects the functional constraints on the admissible class.

2. Limiting random walk

2.1. Characterization

As in the previous section, we assume $\{X_i\}$ is a Markov chain having transitions determined by F, which is a bivariate distribution with Gumbel margins. Temporal dependence in the chain is determined by the dependence characteristics of F. Thus, it is the joint tail of F that characterizes extremal dependence in the chain. This, in turn, is determined by standard multivariate extreme value theory. Under weak regularity assumptions (Resnick 1987, Chapter 5), F lies in the domain of attraction of a bivariate extreme value distribution, G. This means that the distribution of linearly rescaled componentwise maxima of n independent variables having distribution F converges as $n \to \infty$ to G. More precisely, as $n \to \infty$,

$$\{F(x + \log n, y + \log n)\}^n \to \exp[-V(e^x, e^y)], \tag{2}$$

where V is a homogeneous function of order -1 satisfying $V(x, \infty) = V(\infty, x) = x^{-1}$, thus characterizing the bivariate extreme value family. Hence, adopting this limit as an approximation for large x and y,

$$\begin{aligned} F(x, y) &\approx \exp[-n^{-1} V(e^{x-\log n}, e^{y-\log n})] \\ &= \exp[-V(e^x, e^y)], \end{aligned} \tag{3}$$

by the homogeneity of V. Thus, the tail of F can be approximated by a bivariate extreme value distribution. Smith *et al.* (1997) suggested incorporating this approximation into the Markov chain to model the transitions at high levels. Thus, when $x_i > u$ for large u, $F(x_i, x_{i+1})$ is modelled by equation (3). Within this model we now seek to characterize the admissible class of limiting random walk step length distributions in (1).

Substituting (3) for F in equation (1), and using the fact that nth partial derivatives of V are homogeneous of order $-(n+1)$, gives, for $z \in \mathbb{R}$:

$$\begin{aligned} H(z) &= \lim_{u\to\infty}\left\{ f_1(u)^{-1}\left.\frac{\partial F(x,\, y)}{\partial x}\right|_{(x,y)=(u,u+z)}\right\} \\ &= \lim_{u\to\infty}\left\{\frac{-\mathrm{e}^u V_1(\mathrm{e}^u,\, \mathrm{e}^{u+z})\exp[-V(\mathrm{e}^u,\, \mathrm{e}^{u+z})]}{\mathrm{e}^{-u}\exp(-\mathrm{e}^{-u})}\right\} \\ &= \lim_{u\to\infty}\left\{\frac{-V_1(1,\, \mathrm{e}^z)\exp[-\mathrm{e}^{-u}V(1,\, \mathrm{e}^z)]}{\exp(-\mathrm{e}^{-u})}\right\} \\ &= -V_1(1,\, \mathrm{e}^z), \end{aligned} \tag{4}$$

where f_1 denotes the Gumbel density function, and V_1 is the derivative of V with respect to the first component. In some cases, the form of V leads to H having positive mass at $-\infty$, i.e. $H(-\infty) > 0$. Restricting attention to transition models for which H is continuously differentiable and $H(-\infty) = 0$, we obtain from (4) that the density of the step length of the limiting random walk is

$$h(z) = -\mathrm{e}^z V_{1,2}(1,\, \mathrm{e}^z), \tag{5}$$

where $V_{1,2}$ denotes the mixed partial derivative.

Equation (5) gives one characterization of the step length density, h, for the limiting random walk. From Coles and Tawn (1991, Theorem 2), the parametric family $V_{1,2}$ can always be written in the form

$$V_{1,2}(x,\, y) = \frac{-2}{(x+y)^3}\, g\left(\frac{x}{x+y}\right), \tag{6}$$

where g is a probability density function on [0, 1], having mean equal to $\frac{1}{2}$. Hence,

$$h(z) = \frac{2\mathrm{e}^z}{(1+\mathrm{e}^z)^3}\, g\left(\frac{1}{1+\mathrm{e}^z}\right) \tag{7}$$

which gives an alternative, but equivalent, characterization for h in terms of transformed densities on [0, 1]. It is easily checked that a variable Z having density (7) must have negative mean.

When modelling extremes it is usual to adopt a parametric family for either the homogeneous function V or the density g (Coles and Tawn 1991, for example). To emphasize the dependence on a possibly vector-valued parameter α, we denote these functions by $V(x,\, y;\, \alpha)$ and $g(w;\, \alpha)$, respectively. For example, one standard bivariate extreme value model is the logistic model (Gumbel, 1960), for which

$$V(x,\, y;\, \alpha) = (x^{-1/\alpha} + y^{-1/\alpha})^\alpha, \tag{8}$$

$x > 0$, $y > 0$, $0 < \alpha < 1$. Equivalently, from (6),

$$g(w;\, \alpha) = \tfrac{1}{2}(1/\alpha - 1)\{w(1-w)\}^{-1-1/\alpha}\{w^{-1/\alpha} + (1-w)^{-1/\alpha}\}^{\alpha-2}. \tag{9}$$

Hence, from either (5) or (7), the associated limiting random walk has density

$$h(z;\,\alpha) = (1/\alpha - 1)\mathrm{e}^{-z/\alpha}(1 + \mathrm{e}^{-z/\alpha})^{\alpha-2}, \qquad (10)$$

which is Example 2 of Smith (1992). The parameter α determines the degree of dependence in the Markov chain, with $\alpha \to 1$ corresponding to independence and $\alpha \to 0$ giving perfect dependence. The fact that g is symmetric about 0.5 for this model corresponds to time-reversibility of the Markov chain.

2.2. Accuracy of the random walk model

Exact inference for the Markov model of Section 1 has been discussed in detail by Smith *et al.* (1997). A simpler approach is based directly on the limiting random walk characterization. Let $Z_1, \ldots, Z_n$ correspond to differences between consecutive variables above a high threshold u of a Markov chain; that is, $Z_i = X_{(i)+1} - X_{(i)}$, where the $X_{(i)}$ are the realizations of the chain exceeding u. By the characterizations of Sections 1 and 2, the Z_i can be modelled as independent variables with density function h as in (5) or (7). Simulation studies suggest this approximation is accurate provided dependence in the chain is not too weak. We will demonstrate this for Markov chains whose transition distribution F is given by the right-hand side of (2) with V given by (8). For values of α in the range 0.3 to 0.9 a series of 10 000 points from the associated Markov chain was simulated. Accordingly, the correct limiting model for h is given by equation (10). In each case the quality of the limiting model for steps above $u = u_{0.95}$, the 95% marginal quantile, was examined in two ways: first, to assess the independence of successive steps; and second, to assess the validity of density (10). Judged in terms of the sample correlation of successive transitions above $u_{0.95}$, the assumption of independent transitions was found to be reasonable for each value for α: in the worst case, the correlation was found to be about -0.1. For reasonably small values of α, the assumption of the limiting density (10) was also found to be a good approximation, though problems arise for larger values of α, corresponding to weakly dependent chains. Figure 1, giving probability plots of the transition densities above $u_{0.95}$ with respect to the true limit model (10), illustrates the problem. For $\alpha = 0.9$ and to a lesser extent for $\alpha = 0.7$, the limit model is poor in the lower tail, while for smaller values of α the model is almost perfect. In the case of $\alpha = 0.9$ a probability plot based on a threshold corresponding to $u_{0.999}$, the 99.9% quantile, is also shown: there is some improvement, but the fit in the lower tail is still very poor. This suggests that for large values of α much higher thresholds are necessary before the limiting model gives a workable approximation. Such a limitation is not surprising, since the limit model of (10) as $\alpha \to 1$ corresponds to unit mass at $-\infty$.

3. A sufficiency property of the random walk model

We now restrict our attention to Markov chains having the same time-reversibility property as the logistic model (8), so that $g(w;\,\alpha) = g(1 - w;\,\alpha)$, or equivalently, $V(x,\,y;\,\alpha) = V(y,\,x;\,\alpha)$. In this section we show that the corresponding model for the limiting random walk h in equation (7) possesses a special sufficiency property. Surprisingly, despite the fact that h is asymmetric with negative mean generally depending on the parameter value α,

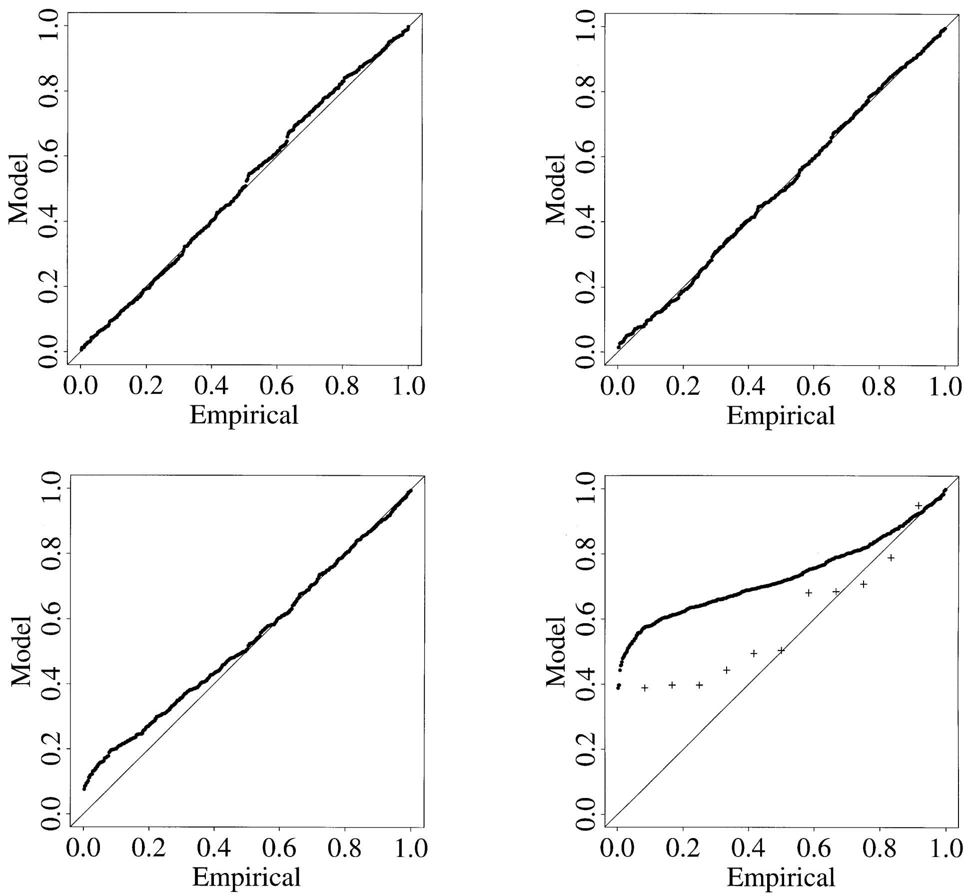

Figure 1. Probability plots of the limiting random walk model for a Markov chain with transitions given by the logistic bivariate extreme value distribution. Distributions are based on steps above a threshold equal to the 95% marginal quantile for each value of the dependence parameter α, and also, plotted as '+', for the 99.9% marginal quantile in the case $\alpha = 0.9$.

$\{|Z_1|, |Z_2|, \ldots, |Z_n|\}$ constitutes a sufficient statistic for this family. Thus, all likelihood inferences, such as the maximum likelihood estimate and likelihood-based confidence intervals, are invariant to the sign of the Z_i. The result is proved as follows. From the symmetry and homogeneity properties of V, equation (5) can be rewritten as:

$$\begin{aligned} h(z;\, \alpha) &= -\mathrm{e}^{z} V_{1,2}(1,\, \mathrm{e}^{z};\, \alpha) \\ &= -\mathrm{e}^{-2z} V_{1,2}(\mathrm{e}^{-z},\, 1;\, \alpha) \\ &= -\mathrm{e}^{-2z} V_{1,2}(1,\, \mathrm{e}^{-z};\, \alpha). \end{aligned}$$

Hence,

$$h(z;\,\alpha) = -m(z)V_{1,2}(1,\, \mathrm{e}^{|z|};\, \alpha),$$

where

$$m(z) = \begin{cases} \mathrm{e}^{z} & z \geqslant 0, \\ \mathrm{e}^{-2z} & z < 0. \end{cases}$$

This provides the factorization required for the sequence of absolute values of the step lengths, $\{|Z_i|\}$, $i = 1, \ldots, n$, to be sufficient for α.

The sufficiency property can also be reasoned from an intuitive viewpoint. By the Markov property, the log-likelihood for α is a sum of contributions associated with the individual Z_i. But by the time-reversibility of the process, positive and negative steps in the limiting random walk contain the same information about the level of dependence in the Markov chain. Thus, inferences on α should depend only on the magnitude of each individual step.

Representation (7) also suggests a way of deriving other models sharing the same sufficiency property. Let $g(w;\,\alpha)$ be a family of probability density functions on [0,1], symmetric about $\frac{1}{2}$. Applying the logistic transformation

$$W = (1 + \mathrm{e}^{Z})^{-1} \tag{11}$$

leads to Z having density

$$h^{*}(z;\,\alpha) = \frac{\mathrm{e}^{z}}{(1+\mathrm{e}^{z})^{2}}\, g\left(\frac{1}{1+\mathrm{e}^{z}};\, \alpha\right). \tag{12}$$

Up to a multiplication factor not involving α, this is equal to h in equation (7) and, hence, has the same sufficiency property.

4. Implications for inference

Consider now the problem of inference for h, the limiting random walk step length density. In common with other inferential problems involving multivariate extremes, the entire admissible class for h has no finite parametrization, being limited in this case only by the representation (7). In analogous situations, both parametric and nonparametric procedures have been considered; see Coles and Tawn (1991) for a discussion of parametric procedures, de Haan and Resnick (1993) and Capéraà *et al.* (1997) for a discussion of nonparametric procedures, and Smith *et al.* (1990) for a comparison of the two approaches. In the present context, as discussed in Section 2, the parametric approach would be to adopt a sub-family for h satisfying constraint (7), such as model (10), and to estimate parameters by, say, maximum likelihood. The restriction of parametric procedures to sub-families is avoided by the use of nonparametric models, but at the expense of difficulty in constraining function estimates to lie within their functional domain, representation (7) in this specific case. Moreover, because of the sufficiency argument in Section 3, invoking the likelihood principle, positive and negative values of Z should contain the same inferential information. Implicitly, any likelihood-based technique such as maximum likelihood or Bayesian inference is bound to respect this criterion, but this is not necessarily the case for nonparametric estimators. However, the discussion of Section 3 suggests a way to construct a kernel density estimate of h which respects both the representation (7) and the property of invariance to the sign of the Z_i.

Our model is that $Z_1, \ldots, Z_n$, representing step lengths at high levels of the chain, are independent variables with density of the form (7). Applying the logistic transformation (11) generates independent variables $W_1, \ldots, W_n$ with density

$$h_W(w) = 2wg(w), \qquad 0 \leq w \leq 1. \tag{13}$$

Now define, for $i = 1, \ldots, n$,

$$W_i^* = \begin{cases} W_i & \text{with probability } \frac{1}{2} \\ 1 - W_i & \text{with probability } \frac{1}{2}. \end{cases} \tag{14}$$

Then each W_i^* has density

$$\begin{aligned} h_{W^*}(w) &= \tfrac{1}{2}\{2wg(w) + 2(1-w)g(1-w)\} \\ &= g(w) \end{aligned}$$

by the symmetry of g. Thus, applying the logistic transform (11) not only to $Z_1, \ldots, Z_n$, but also to $Z_{n+1}, \ldots, Z_{2n}$, where $Z_{n+i} = -Z_i$ for $i = 1, \ldots, n$, generates a non-independent sample $W_1^*, \ldots, W_{2n}^*$ having the same marginal distribution as the variable (14), i.e., having marginal density g. Moreover, the sample is symmetric about 0.5. Standard, though bounded, kernel density estimation on the W_i^* therefore yields an estimate $\tilde{g}$ of g which is itself symmetric about 0.5. Substitution of $\tilde{g}$ for g in (7) then generates an estimate, $\tilde{h}$ of h, which by construction is an admissible step length density for a limiting random walk in the sense defined in Section 1.

A simulation study also suggests that, in circumstances where the limiting model gives a good approximation, this approach has better mean-square-error properties than a naive kernel density estimator applied to the Z_i. Again we focus on model (2) with V given by (8), but now restricted to the case $\alpha \leq 0.7$, for which the random walk limit with step length density (10) was found previously to give a good approximation.

For each α, series of 10 000 points, X_i, were again simulated from the Markov chain, and steps $Z_i = X_{i+1} - X_i$ for which X_i exceeded $u_{0.95}$, or $u_{0.99}$ in the case of $\alpha = 0.7$, were assumed to be independent with step length density (10). This procedure was repeated 100 times, to obtain estimates of mean integrated square error in the estimation of h using the two different kernel density estimators. The results are summarized in Table 1.

Similar empirical rules were used to control the degree of smoothing for each of the two estimators in order to obtain a fair comparison. However, because of the boundedness of the domain in the logistic space, reflected kernels were used for the transformed estimator.

For small values of α the mean integrated square error of the naive density estimator is reduced by about 50% when using the transformed estimator. For larger values of α the

Table 1. Mean integrated square errors in the estimation of step length density h, using naive kernel density estimation on step lengths (MISE ($\tilde{h}_1$)) and transformed kernel density estimation in logistic space (MISE ($\tilde{h}_2$)); in each case, estimates are based on steps $Z_i = X_{i+1} - X_i$ for which X_i exceeded $u_{0.95}$, or $u_{0.99}$ in the case of $\alpha = 0.7$

α	0.3	0.4	0.5	0.6	0.7
MISE ($\tilde{h}_1$)	0.0030	0.0023	0.0018	0.0020	0.020
MISE ($\tilde{h}_2$)	0.0015	0.0011	0.0008	0.0016	0.020

improvement is less marked to the extent that there is no improvement in the case $\alpha = 0.7$ despite the fact that a higher threshold was selected. Interpretation is difficult in this case, and for larger values of α, since the quality of the limiting model itself is poor. However, it is encouraging that when the model is appropriate the transformed estimator seems to have preferred inference properties, in addition to the benefit of falling within the true functional space of the model.

Acknowledgements

We are grateful to Richard Smith and Jonathan Tawn for comments on an earlier draft of this paper, and to a referee and editor whose comments enabled us to improve the clarity of our exposition.

References

Capéraà, P., Fougères, A.L. and Genest, C. (1997) A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika*, **84**, 567–577.

Coles, S.G. and Tawn, J.A. (1991) Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B*, **53**, 377–392.

De Haan, L. and Resnick, S.I. (1993) Estimating the limit distribution of multivariate extremes. *Comm. Statist. Stochastic Models*, **9**, 275–309.

Gumbel, E.J. (1960) Bivariate exponential distributions. *J. Amer. Statist. Assoc.*, **59**, 794–816.

Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* New York: Springer-Verlag.

Smith, R.L. (1992) The extremal index for a Markov chain. *J. Appl. Probab.*, **29**, 37–45.

Smith. R.L., Tawn, J.A. and Yuen, H.K. (1990) Statistics of multivariate extremes. *Internat. Statist. Rev.*, **58**, 47–58.

Smith, R.L., Tawn, J.A. and Coles, S.G. (1997) Markov models for threshold exceedances. *Biometrika*, **84**, 249–268.

Received October 1997 and revised June 1998