A Markov Property for Covariance Structures

A. O'Hagan
University of Nottingham

September 29, 1998

Abstract

The problem of formulating covariance structures between two random variables or random vectors arises in many fields. One way to specify a covariance structure for a set of random variables defined at each point on a product space is by a Kronecker product form. Then the specification requires only a covariance structure on each of the component spaces. This note proves that a simple Markov-like assumption on the random variables implies that their covariance structure must, apart from an arbitrary scaling, follow a Kronecker product form. When a more restricted version of the Markov property is assumed, an autoregressive or Ornstein-Uhlenbeck form arises.

Keywords: AUTOREGRESSIVE MODELS; COVARIANCE STRUCTURES; KRONECKER PRODUCT; ORNSTEIN-UHLENBECK PROCESS; SPACE-TIME MODELLING

1. Covariance structures

Consider random variables $f(x, y)$ defined for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Each of the sets $\mathcal{X}$ and $\mathcal{Y}$ may be finite or infinite. In either case, but particularly when they are infinite, one might refer to $f(\cdot, \cdot)$ as a stochastic process on the product space $\mathcal{X} \times \mathcal{Y}$. We are interested in the structure of covariances between this set of random variables. Denote the covariance between $f(x, y)$ and $f(x', y')$ by $c\,[(x, y), (x', y')]$. We also introduce the notation

$$r\,[(x, y), (x', y')] = \frac{c\,[(x, y), (x', y')]}{\sqrt{c\,[(x, y), (x, y)]\, c\,[(x', y'), (x', y')]}}$$

for the corresponding correlations. Note that x or y may be random vectors, so that $\mathcal{X}$ and $\mathcal{Y}$ are themselves product spaces.

Modelling covariance structures is important in various fields. In geostatistics, the method of kriging is based on a spatial covariance structure, usually defined by a variogram (Cressie, 1991). Then x and y are the coordinates of a two-dimensional geographical space. It is often necessary to add a time dimension, and to model the covariance structure of a space-time process. In the application motivating this work, Kennedy and O'Hagan (1998), $\mathcal{X}$ is an arbitrary vector space and the y coordinate is a scalar complexity index for a computer code.

The complexity of possible covariance structures over $\mathcal{X} \times \mathcal{Y}$ can be a problem in modelling. In geostatistics, there is an arbitrariness in the choice of x and y. They may be defined according to the usual compass points, but could equally well be any two orthogonal directions. When x and y are naturally thought of in this way as *arbitrary* coordinates of a point in a higher dimensional space, then a covariance structure based on a distance measure invariant to rotations of the coordinates is appropriate. Thus it is normal in geostatistics to define the covariance function to have this property of isotropy. In other problems there is a natural coordinate system, the coordinates have quite different meanings, isotropy is clearly not appropriate, and it is generally more difficult to think of sensible structures. Sampson and Guttorp (1992) offer a process for estimating arbitrary covariance structures in a nonparametric way.

A special structure which can be useful is the Kronecker product form, where

$$c\left[(x,y),(x',y')\right] = c_x\left\{x,x'\right\} \times c_y\left\{y,y'\right\} , \tag{1.1}$$

and c_x, c_y define covariances on $\mathcal{X}$ and $\mathcal{Y}$ separately. Assuming the Kronecker form reduces the problem to the much simpler one of formulating c_x and c_y. This assumption is common in context of space-time models, for instance, where $\mathcal{X}$ is the two-dimensional space and $\mathcal{Y}$ is the one-dimensional time axis. It is made explicitly by Oehlert (1993). It is less obvious in Haslett and Raftery (1989) but questioned by several discussants. One purpose of the present paper is to provide a characterisation which may be useful in justifying the Kronecker product form.

The Kronecker product form is also assumed by Brown, Le and Zidek (1994), where x indexes a pollution monitoring site and y indexes a pollutant potentially measured at that site. A quite different example of the convenience offered by the Kronecker product covariance structure is that it is the form of the covariance matrix of a matrix normal distribution (Dawid, 1981).

It is also interesting to note a link between the Kronecker product and isotropy for a popular form of c_x and c_y. If these have the Gaussian forms

$$c_x\{x, x'\} = \exp\left\{-b(x-x')^2\right\}, \qquad c_y\{y, y'\} = \exp\left\{-b(y-y')^2\right\}$$

with the same roughness parameter b, then the Kronecker product (1.1) is isotropic, being a function of the Euclidean distance $(x-x')^2 + (y-y')^2$.

2. The symmetric Markov property

2.1. A Markov assumption

We now consider an assumption on the covariance function $c(\cdot, \cdot)$ based on the following heuristic. Suppose that I wish to estimate $f(x, y)$ and I observe $f(x, y')$ for some $y' \neq y$. Then it may be reasonable to assume that I will gain no further useful information about $f(x, y)$ by observing $f(x', y')$ for any other $x' \neq x$. This is a kind of Markov property, that $f(x, y)$ depends on $\{f(x', y') : x' \in \mathcal{X}\}$ for given x, y and y' only through the nearest point $f(x, y')$.

Formally, we define the property $\mathcal{M}(\mathcal{X}; y, y')$ as asserting that for all

$$c[(x, y), (x', y')] = \frac{c[(x, y), (x, y')]\, c[(x, y'), (x', y')]}{c[(x, y'), (x, y')]}. \tag{2.1}$$

If the joint distribution of the $f(x, y)$s is normal (or more accurately, to allow for $\mathcal{X}$ and/or $\mathcal{Y}$ being infinite, if $f(\cdot, \cdot)$ is a Gaussian process on $\mathcal{X} \times \mathcal{Y}$), (2.1) is equivalent to the heuristic argument because it says that $f(x, y)$ and $f(x', y')$ are independent given $f(x, y')$. Otherwise it says that the corresponding partial correlation is zero, equivalent to a linear independence.

If $\mathcal{M}(\mathcal{X}; y, y')$ is assumed for all $y, y' \in \mathcal{Y}$, we denote this by $\mathcal{M}(\mathcal{X}; \mathcal{Y})$. Note that now, although both the heuristic and (2.1) are expressed asymmetrically in x and y, $\mathcal{M}(\mathcal{X}; \mathcal{Y})$ is effectively symmetric because covariances are. Simply applying (2.1) with $c[(x', y'), (x, y)]$ on the left hand side gives

$$c[(x, y), (x', y')] = \frac{c[(x, y), (x', y)]\, c[(x', y), (x', y')]}{c[(x', y), (x', y)]}. \tag{2.2}$$

Thus, $f(x, y)$ and $f(x', y')$ are also (linearly) independent given $f(x', y)$. Notice that switching the roles of y and y' in this way is not legitimate in the case of the more restrictive Markov property $\mathcal{M}(\mathcal{X}; y, y')$. In fact, it is precisely the fact that $\mathcal{M}(\mathcal{X}; \mathcal{Y})$ implies both $\mathcal{M}(\mathcal{X}; y, y')$ and $\mathcal{M}(\mathcal{X}; y', y)$ that gives it the extra strength and leads to the stronger theorem below.

2.2. Relationship to Kronecker product

It is easy to show that (2.1) holds for all $y, y' \in \mathcal{Y}$, and therefore (2.2) holds, if the covariances follow the Kronecker product form (1.1). The fact that assuming the Kronecker product structure implies an assumption of the symmetric Markov property $\mathcal{M}(\mathcal{X}; \mathcal{Y})$ is interesting in itself. We now explore the extent to which $\mathcal{M}(\mathcal{X}; \mathcal{Y})$ conversely implies the Kronecker product form (1.1). The following theorem shows that the implication is not complete, but that if the Markov property holds then there is some scaling of $f(\cdot, \cdot)$ whose covariances have the Kronecker product structure.

Theorem 2.1. *The symmetric Markov property $\mathcal{M}(\mathcal{X}; \mathcal{Y})$ holds if and only if there exists a function a on $\mathcal{X} \times \mathcal{Y}$ such that the covariance structure of the random variables $g(x, y) = f(x, y)/a(x, y)$ has a Kronecker product form.*

The proof is given in section 3.

2.3. Discussion

As will be seen in the proof below, the function a is certainly not unique but can in particular be set to $a(x, y) = \pm\sqrt{c\left[(x, y), (x, y)\right]}$, i.e. to plus or minus the standard deviation of $f(x, y)$. So apart from the sign this is just a standardising transformation making the variance of $g(x, y)$ one for all x and y. It is therefore effectively the correlation structure $r\left[(x, y), (x', y')\right]$ which has the Kronecker product form (again, apart from the sign changes). From the point of view of modelling, it is interesting to separate the exercise into two parts, modelling the variances and modelling the correlations. The theorem shows that the Markov property refers essentially to correlations, and it may be realistic to assume a Kronecker product form for the correlations. (Although the Markov property only implies this for the squared correlations, it is hard to see how one would allow sign changes in a practical model.) Then the variances may be modelled separately. There is an interesting connection for Bayesian modelling with the work of Barnard, McCulloch and Meng (1997), who propose a prior distribution for a covariance matrix in which arbitrary distributions are assumed for the variances, but the conditional distribution of the correlations given the variances is that of an inverse Wishart distribution.

3. Proof of Theorem 2.1

3.1. Necessity

We first prove that the Markov property holds if the random variables $g(x, y)$ have the Kronecker product covariance structure, for arbitrary scaling $a(x, y)$. Denote the covariance function of $g(x, y)$ by $c^*\left[(x, y), (x', y')\right]$, so that

$$c\left[(x, y), (x', y')\right] = a(x, y)\, a(x', y')\, c^*\left[(x, y), (x', y')\right] \ .$$

Now if c^* has the Kronecker product form (1.1), it follows immediately that

$$\begin{aligned} c\left[(x, y), (x', y')\right] &= a(x, y)\, a(x', y')\, c^*_x\left\{x, x'\right\}\, c^*_y\left\{y, y'\right\} , \\ c\left[(x, y), (x, y')\right] &= a(x, y)\, a(x, y')\, c^*_x\left\{x, x\right\}\, c^*_y\left\{y, y'\right\} , \end{aligned}$$

with similar expressions for $c\left[(x, y'), (x', y')\right]$ and $c\left[(x, y'), (x, y')\right]$. (2.1) follows immediately.

3.2. Sufficiency

3.2.1. Squared correlations

To prove the converse is more complex. First, note that (2.1) and (2.2) give

$$\frac{c\left[(x, y), (x, y')\right] c\left[(x, y'), (x', y')\right]}{c\left[(x, y'), (x, y')\right]} = \frac{c\left[(x, y), (x', y)\right] c\left[(x', y), (x', y')\right]}{c\left[(x', y), (x', y)\right]} , \tag{3.1}$$

and by considering the two forms for $c\left[(x, y'), (x', y)\right]$ we also have

$$\frac{c\left[(x, y), (x, y')\right] c\left[(x, y), (x', y)\right]}{c\left[(x, y), (x, y)\right]} = \frac{c\left[(x, y'), (x', y')\right] c\left[(x', y), (x', y')\right]}{c\left[(x', y'), (x', y')\right]} . \tag{3.2}$$

Multiplying (3.1) and (3.2), we have

$$r\left[(x, y), (x, y')\right]^2 = r\left[(x', y), (x', y')\right]^2 \ . \tag{3.3}$$

Therefore $r\left[(x,y),(x,y')\right]^2$ does not depend on x, and we can therefore denote it by $r_y\left\{y,y'\right\}^2$.

Similarly (by cross-multiplying (3.1) and (3.2)), we find that $r\left[(x,y),(x',y)\right]^2$ does not depend on y, and so will be denoted by $r_x\left\{x,x'\right\}^2$. Now squaring (2.1) and dividing by $c\left[(x,y),(x,y)\right]c\left[(x',y'),(x',y')\right]$ gives

$$\begin{aligned} r\left[(x,y),(x',y')\right]^2 &= r\left[(x,y),(x,y')\right]^2 r\left[(x,y'),(x',y')\right]^2 \\ &= r_x\left\{x,x'\right\}^2 r_y\left\{y,y'\right\}^2 . \end{aligned} \tag{3.4}$$

We have now shown that the squared correlations of the $f(x,y)$s have a Kronecker product form. It follows that their correlations also have a Kronecker product form except that some correlations may have the opposite signs. Thus, if we take $r_x\left\{x,x'\right\}$ and $r_y\left\{y,y'\right\}$ to be (either positive or negative) square roots of $r_x\left\{x,x'\right\}^2$ and $r_y\left\{y,y'\right\}^2$ then

$$r\left[(x,y),(x',y')\right] = \pm\, r_x\left\{x,x'\right\} r_y\left\{y,y'\right\} .$$

3.2.2. Patterns of signs

Now it is clear that if we let $a(x,y) = \pm\sqrt{c\left[(x,y),(x,y)\right]}$, i.e. we scale by plus or minus the standard deviation of $f(x,y)$, then the covariances $c^*\left[(x,y),(x',y')\right]$ of $g(x,y)$ are just $\pm\, r\left[(x,y),(x',y')\right]$. To complete the proof, we need to prove that the sign of $a(x,y)$ can be chosen for each (x,y) in such a way as to make

$$c^*\left[(x,y),(x',y')\right] = r_x\left\{x,x'\right\} r_y\left\{y,y'\right\} , \tag{3.5}$$

which is the Kronecker product form as required, with r_x and r_y playing the roles of c_x^* and c_y^*.

Consider the patterns of sign changes that are possible. By examining the derivations of (3.3) and (3.4) we see that we **either** have

$$\begin{aligned} r\left[(x,y),(x,y')\right] &= r\left[(x',y),(x',y')\right] , \\ r\left[(x,y),(x',y)\right] &= r\left[(x,y'),(x',y')\right] , \\ r\left[(x,y),(x',y')\right] &= r\left[(x',y),(x,y')\right] , \end{aligned}$$

or else

$$\begin{aligned} r\left[(x,y),(x,y')\right] &= -\, r\left[(x',y),(x',y')\right] , \\ r\left[(x,y),(x',y)\right] &= -\, r\left[(x,y'),(x',y')\right] , \\ r\left[(x,y),(x',y')\right] &= -\, r\left[(x',y),(x,y')\right] . \end{aligned}$$

Now if we change the sign of **one** of the random variables $f(x,y)$, $f(x',y)$, $f(x,y')$ or $f(x',y')$, we flip between these two different cases because we change the sign of the correlation on one side of each equation but not on the other. It holds for all rectangles (x,y), $(x'.y)$, (x,y'), (x',y') in $\mathcal{X}\times\mathcal{Y}$, but for degenerate rectangles (i.e. $x=x'$ or $y=y'$) only the first case is possible. This observation is the key to proving that we can choose the signs of the $a(x,y)$s so as to make all the correlations agree, since it is clear that the proof is complete if we can choose the sign of $a(x,y)$ so that the first case above holds for every rectangle. This is not straightforward, because changing the sign of any $f(x,y)$ flips the case in *every* rectangle with (x,y) at one vertex.

3.2.3. Induction

First consider arbitrary orderings on $\mathcal{X}$ and $\mathcal{Y}$, such that the smallest elements in each are x_0 and y_0. Define $r_x\{x,x'\}$ to be $r\left[(x,y_0),(x',y_0)\right]$ and $r_y\{y,y'\} = r\left[(x_0,y),(x_0,y')\right]$, thereby fixing the signs of these functions. Then we can set $a(x_0,y)$ and $a(x,y_0)$ to have positive signs. We now proceed by induction. For given $(x_1,y_1)\in\mathcal{X}\times\mathcal{Y}$, define $A(x_1,y_1)$ to be that region of $\mathcal{X}\times\mathcal{Y}$ for which $x\le x_1$ and $y\le y_1$ but excluding the point (x_1,y_1) itself. We assume that the sign of $a(x,y)$ has been fixed for all $(x,y)\in A(x_1,y_1)$, such that (3.5) holds for all rectangles in $A(x_1,y_1)$, and we prove that the sign of $a(x_1,y_1)$ can then be chosen to make it hold also for all rectangles comprising (x_1,y_1) and three points in $A(x_1,y_1)$.

Denote by P the point (x_1,y_1). Consider any two rectangles $PQQ'P'$ and $PRR''P''$, where Q,Q',P',R,R'' and P'' are in $A(x_1,y_1)$. We will show that either the signs all agree on both rectangles, i.e. the first case above holds for both rectangles, or else the second case holds for both rectangles. First suppose that the two rectangles have a side in common, by letting $R=Q$. Then $P'Q'R''P''$ is a rectangle, and it is also entirely in $A(x_1,y_1)$, so the first case applies to this rectangle by the inductive assumption. It is then easy to see that for the two rectangles $PQQ'P'$ and $PRR''P''$ the same case must apply, either both the first case or both the second. In general, when $R\ne Q$, we can prove the same result by examining the two pairs of rectangles $(PQQ'P', PQQ''P'')$ and $(PQQ''P'', PRR''P'')$. It follows that the same case applies to *all* rectangles $PQQ'P'$, either all are the first case or all are the second case. The sign of $a(x_1,y_1)$ can therefore be positive or negative accordingly, whereupon the inductive step is proven.

4. The asymmetric Markov property

4.1. A representation

It is clear that the asymmetric form $\mathcal{M}(\mathcal{X}; y, y')$ of the Markov property does not imply the symmetric form, and so will not have an implication like the Kronecker product covariance structure. Instead, we have the following result.

Theorem 4.1. *The asymmetric Markov property $\mathcal{M}(\mathcal{X}; y, y')$ holds for given y and y' if and only if there exist constants $r(x)$ for all $x \in \mathcal{X}$ such that the two sets of random variables $\{f(x, y) - r(x)f(x, y') : x \in \mathcal{X}\}$ and $\{f(x, y') : x \in \mathcal{X}\}$are uncorrelated.*

Proof: For given y and y', let $e(x) = f(x, y) - r(x)f(x, y')$ and $g(x) = f(x, y')$.

Suppose first that $e(x)$ and $g(x')$ are uncorrelated for all x and x'. Let $cov(e(x), e(x')) = c_e(x, x')$ and $cov(g(x), g(x')) = c_g(x, x')$. Then for $y \neq y'$ we have $c\,[(x, y),(x', y')] = cov(r(x)g(x) + e(x), g(x')) = r(x)c_g(x, x')$. Similarly, $c\,[(x, y),(x, y')] = r(x)c_g(x, x)$. Since $c\,[(x, y'),(x', y')] = c_g(x, x')$ and $c\,[(x, y'),(x', y')] = c_g(x, x)$, the Markov property (2.1) holds (for $y \neq y'$; it holds trivially for $y = y'$).

Now suppose that the Markov property (2.1) holds and let $r(x) = c\,[(x, y),(x, y')]\,/\,c\,[(x, y'),(x, y')]$. Then $cov(e(x), g(x')) = c\,[(x, y),(x', y')] - r(x)c\,[(x, y'),(x', y')] = 0$ from (2.1).

4.2. Discussion

The theorem shows that when the Markov property holds we can represent $f(x, y) = r(x)f(x, y') + e(x)$, a kind of autoregressive model. This is for any given y and y'. If we assume it for all ordered pairs $y > y' \in \mathcal{Y}$, then the theorem induces a kind of independent increments model for $f(x, y)$. In principle we have a different $r(x)$ for all x and for all pairs y, y', but arguments of stationarity will often suggest that there should be a single constant $r(x) = r$ for all $x \in \mathcal{X}$, and that this should depend only on a measure of distance between y and y'. When $\mathcal{Y} = \mathbb{R}$, we thereby obtain Ornstein-Uhlenbeck processes $f(x, \cdot)$ for each x, correlated across $x \in \mathcal{X}$.

Kennedy and O'Hagan (1998) employ a simple autoregressive model where $f(x, y)$ and $f(x, y')$ are the outputs of a computer code run at two different levels of accuracy, for input x.

References

Barnard, J., McCulloch, R. E. and Meng, X-L. (1997). Modeling variance matrices in terms of standard deviations and correlations with application to shrinkage. Technical Report 438, Department of Statistics, University of Chicago.

Brown, P. J., Le, N. D. and Zidek, J. V. (1994). Multivariate spatial interpolation and exposure to air pollutants. *Canad. J. Statist.* 22, 489–509.

Cressie, N. A. C. (1991). *Statistics for Spatial Data.* Wiley, N. Y.

Dawid, A. P. (1981). Some matrix-variate distribution theory. *Biometrika* 68, 265–274.

Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *Applied Statistics* 38, 1–50.

Kennedy, M. C. and O'Hagan, A. (1998). Predicting the output from a complex computer code when fast approximations are available. Statistics Research Report 98–09, University of Nottingham.

Oehlert, G. W. (1993). Regional trends in sulfate wet deposition. *J. Amer. Statist. Soc.* 88, 390–399.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric estimation of nonstationary covariance structure. *J. Amer. Statist. Soc.* 87, 108–119.