A CONSTRUCTION FOR PROCESSES WITH HISTORY-DEPENDENT TRANSITION INTENSITIES

P. Whittle
University of Cambridge

Summary

At the Sheffield meeting in honour of Joe Gani I presented the material of Whittle (1990); this was essentially Theorem 1 below and applications of it in network contexts. In the ensuing discussion John Bather raised an interesting point, which I now pursue.

1. Introduction. Suppose that a continuous-time Markov process has state variable x and infinitesmal generator T. Its equilibrium density $\pi(x)$ relative to a suitable measure μ then satisfies $T'\pi = 0$, where T' is the operator adjoint to T in that $\int(\pi Tf)\mu(dx) = \int(fT'\pi)\mu(dx)$ for functions $\pi(x)$, $f(x)$. Consider now a family of such processes parametrised by a vector $\nu = \{\nu_j\}$ for which the generator has the form

$$T(\nu) = \sum_{n} \nu_j T_j \tag{1}$$

Here the T_j are a set of fixed generators and ν_j a set of variable parameters, scalar and non-negative. One can regard T_j as the generator of transitions by a particular mode (the j^{th} mode); these different modes being weighted differently as we vary the parameters ν_j. We shall refer to the vector $\nu = \{\nu_j\}$ as the <u>rate vector</u>; its essential property is that it enters linearly into the generator.

Let the family of processes generated as ν varies in some given set N be denoted F. Let us also suppose that the equilibrium density $\pi(x|\nu)$ of the process with intensity (1) is unique and known, for all ν in N. Note that prescription (1) includes also the non-homgeneous case

$$T(\nu) = T_0 + \sum_{j \neq 0} \nu_j T_j \tag{2}$$

This corresponds to the case when $\nu_0 \equiv 1$ for ν in N.

Consider now the mixed density

$$\tilde{\pi}(x) = \int \pi(x|\nu)\phi(d\nu) \tag{3}$$

where ϕ is a probability measure on N. Consider the process with generator

$$\tilde{T} = \sum_j \tilde{\nu}_j(x) T_j \tag{4}$$

where

$$\tilde{\nu}_j(x) = \frac{\int \nu_j \pi(x|\nu)\phi(d\nu)}{\int \pi(x|\nu)\phi(d\nu)} \tag{5}$$

Then it is a two-line matter to verify

Theorem 1. *The mixed density $\tilde{\pi}(x)$ is the equilibrium density of the process with infinitesmal generator $\tilde{T}$.*

Despite the formal immediacy of the result, there are substantial and interesting implications. First, note that we have extended the family of processes F with generator (1) for varying ν to a family $\tilde{F}$ with generator (4) for varying mixing measure ϕ. The theorem states that the equilibrium density for a member of this extended class of processes is just the mixed density $\tilde{\pi}$.

One can view ϕ as a prior distribution for the parameter ν, and $\tilde{\nu}$ as a corresponding Bayesian estimate of ν based upon the assumption that the fixed-ν process has reached equilibrium and that one has observed the current state value x. (The estimate might more appropriately be termed a 'forgetful Bayesian' estimate in that one utilises only the current state value and not previously observed values. This is the very point we take up in later sections.) The effect of modifying ν to $\tilde{\nu}$ is to introduce a feedback into the process, by which a transition rule of form (1) is adopted, but the rate vector is adapted to that value which seems most concordant with the current value of x. One might conjecture that the effect of this feedback is to perpetuate the status quo; to adapt the transition rule so that it will tend to yield as outcome the very state value which is current.

This effect, if it exists, should be particularly marked in the case when ϕ gives mass only to a discrete set of values $\nu^{(r)}$, so that

$$\tilde{\pi}(x) = \sum_r \phi_r \pi(x|\nu^{(r)}) \tag{6}$$

say. State space will then decompose itself into sets D_r, where D_r is the set in which $\tilde{v}(x) \sim v^{(r)}$. (More correctly, D_r is the set in which the maximum likelihood estimate $\hat{v}(x)$ equals $v^{(r)}$. The forgetful Bayesian estimate $\tilde{v}(x)$ supplies a smoothed version of $\hat{v}(x)$.) The 'perpetuation of the status quo' conjecture would then amount more specifically to the conjecture that the D_r constitute basins of attraction in state space, escape from any given D_r being rendered more difficult by the presence of feedback. The r^{th} term in (6) could then presumably be identified as the component of distribution corresponding to sojourn in D_r.

One may now ask whether the 'forgetful Bayesian' estimate $\tilde{v}$ might not be converted into a genuine Bayesian estimate based upon a history rather than a moment of observation (supposedly, for a fixed-v process in equilibrium). This was John Bather's query. One imagines that the effect of such an increased dependence upon past history would be to increase attachment to the status quo. In view of this natural conjecture, the results below are interesting.

2. Incorporation of State-History in the Construction. Rather than assuming a complete history of observation we shall assume the state variable augmented by variables which store statistical information from past observations. Suppose that the state variable x is supplemented by a variable y such that (i) y(t) is a deterministic function of current x-history $\{x(s); s \le t\}$ and (ii) the process $\{x.y\}$ is Markov. This implies that y must obey a deterministic Markov up-dating rule, so that $y(t+\Delta)$ is a function of y(t) and $\{x(s); t < s \le t+\Delta\}$ for any $\Delta > 0$. For example, y(t) might be a partial history: $y(t) = \{x(s); t\text{-}h < s \le t\}$. A more manageable choice would be that in which y was a finite-dimensional vector satisfying a differential equation

$$\dot{y} = a(x, y) \tag{7}$$

One could imagine a scheme of this kind for which y(t) supplied quite a good summary of those aspects of the history of x relevant for estimation of v; we give an example in the next section.

We shall now apply the ideas of the last section to the Markov process with state variable (x, y). The infinitesmal generator defined previously by (1) will now be modified so that it has the action

$$T^* (v) = \Sigma v_j T_j^* + T^o. \tag{8}$$

Here T_j^* induces the transitions in (x, y) corresponding to the transitions in x induced by T_j, and T^0 reflects the changes in y in the absence of any change in x: the purely deterministic change implied by its deterministic Markov evolution rule.

The (x, y) process will have an equilibrium density $\pi(x, y|\nu)$ relative to an appropriate measure. This will have the equilibrium density $\pi(x|\nu)$ of the simple x-process as marginal x-density, because x-dynamics are not affected by the values of y.

We can now introduce the mixed density

$$\tilde{\pi}(x, y) = \int \pi(x, y|\nu)\,\phi(d\nu) \tag{9}$$

the 'forgetful Bayesian' estimate

$$\tilde{\nu}(x, y) = \frac{\int \nu\pi(x, y|\nu)\,\phi(d\nu)}{\int \pi(x, y|\nu)\,\phi(d\nu)} \tag{10}$$

and the process in which $\tilde{\nu}$ replaces ν in the transition rule (8).

Theorem 2. *The process for which* $\tilde{\nu}(x, y)$ *replaces* ν *in the definition* (8) *of the infinitesmal generator has equilibrium distribution* $\tilde{\pi}(x, y)$. *Furthermore, the marginal* x*-density of this distribution is the density* $\tilde{\pi}(x)$ *defined previously by* (3).

Proof. Since the infinitesmal generator is linear in ν, Theorem 1 is applicable to the compound process $\{x, y\}$ and the first assertion follows. The second assertion follows from (9) and the fact that $\pi(x|\nu)$ is the marginal of $\pi(x, y|\nu)$. □

Thus, we can indeed modify the mixing construction so that $\tilde{\nu}$ is an estimate of ν based upon aspects of x-history as well as upon current x. Surprisingly, this does not affect the equilibrium distribution of the x-process. However, it must affect the transient behaviour. In particular, one would expect it to increase the 'persistence' of the process, and to deepen the basins of attraction D_r. We shall investigate this conjecture by example.

3. An Example. We shall take a linear Gaussian process as example, because this permits explicit analysis. There would be more interest in a case for which the mixing distribution ϕ was discrete; the fact that ν will follow a continuous distribution will mean that there will not be discrete basins of attraction, and 'perpetuation of the status quo' will at best correspond to an increased sluggishness of movement rather than to temporary capture in some subset of state space. However, the case we consider is substantial enough to throw up some surprises.

Suppose that the x-process has a scalar state variable obeying the stochastic differential equation

$$\dot{x} = \alpha(\nu - x) + \varepsilon \tag{11}$$

where $\alpha > 0$ and ε is white noise of power c. The infinitesmal generator has the effect

$$T\psi(x) = \alpha(\nu - x)\psi'(x) + (c/2)\psi''(x)$$

and so is indeed linear in ν. In equilibrium the process has a $N(\nu, (c/2\alpha))$ density; this defines $\pi(x|\nu)$. The parameter ν is then interpretable as the equilibrium mean of x. If we choose the mixing measure ϕ so that ν is $N(0, \tau)$ then we find that

$$\tilde{\nu} = \frac{2\alpha x/c}{(2\alpha/c) + (1/\tau)} = x - x(1 + 2\tau\alpha\ c)^{-1}$$

For the mixed process x is $N(0, (c/2\alpha) + \tau)$ in equilibrium; this defines $\tilde{\pi}(x)$. The effect of feedback is to replace a drift towards ν by a more sluggish drift, ultimately to the origin.

Consider now the supplementation of x by a scalar y obeying the equation

$$\dot{y} = \beta(x - y)\,. \tag{12}$$

where $\beta > 0$. Thus y is an exponentially-damped average of past x, and, for a fixed-ν process, would provide good information on ν. As β becomes infinite the damping becomes infinite and y reduces to x itself. As β tends to zero then y converges to a long-term average of x. It is for small β that one would expect the mixed version of the process to show considerable persistence.

Consider the process {x, y} specified by (11) and (12). The variables x and y are jointly normal when the process has reached equilibrium, both having ν as mean. Calculation shows the spectral density matrix to be

$$f(\omega) = \frac{c}{\omega^2 + \alpha^2}\begin{bmatrix} 1 & \beta/(-i\omega + \beta) \\ \beta/(i\omega + \beta) & \beta^2/(\omega^2 + \beta^2) \end{bmatrix},$$

this implies an equilibrium covariance matrix

$$\operatorname{cov}(x, y) = \frac{c}{2\alpha}\begin{bmatrix} 1 & \beta/(\alpha+\beta) \\ \beta/(\alpha+\beta) & \beta/(\alpha+\beta) \end{bmatrix}.$$

The joint density can then be verified to have the form

$$\pi(x, y|\,\nu) = A(x, y)\exp\left[-\frac{\alpha(\alpha+\beta)}{\beta c}(y-\nu)^2\right]$$

where A(x, y) is independent of ν. Thus, on the basis of observations x and y the variable y is sufficient for inference on the parameter ν. Taking again the N(0, τ) mixing distribution for ν we have then

$$\tilde{\nu}(x, y) = \frac{(2\alpha(\alpha+\beta)\,y/\beta c)}{(2\alpha(\alpha+\beta)/\beta c) + (1/\tau)} = \gamma y \tag{13}$$

say. This converges to y itself as β tends to zero.

For the mixed process relation (11) then becomes modified to

$$\dot{x} = \alpha(\gamma y - x) + \varepsilon \tag{14}$$

this again to be coupled with relation (12). One finds from these relations that the x-process has spectral density function

$$f(\omega) = \frac{c(\omega^2+\beta^2)}{|(i\omega+\alpha)(i\omega+\beta) - \alpha\beta\gamma|^2} = \frac{c(\omega^2+\beta^2)}{(\omega^2+\chi^2)(\omega^2+\eta^2)} \tag{15}$$

say, where

$$\chi + \eta = \alpha + \beta$$

$$\chi\eta = \alpha\beta(1-\gamma) = \alpha\beta\{1 + 2\alpha(\alpha+\beta)\tau/(\beta c)\}^{-1}. \tag{16}$$

The autocovariance of the x-process is then

$$\tau(s) = \frac{c}{2\chi\eta(\chi^2-\eta^2)}[(\chi^2-\beta^2)\eta e^{-\chi s} - (\eta^2-\beta^2)\chi e^{-\eta s}] \tag{17}$$

In particular,

$$\tau(0) = \frac{c(\beta^2+\chi\eta)}{2\chi\eta(\chi+\eta)} = (c/2\alpha) + \tau.$$

This is independent of β, in agreement with the second assertion of Theorem 2.

A measure of persistence might be $-\tau'(0)/\tau(0)$; the rate of decay of the autocorrelation coefficient at zero lag. We find from expression (17) that $\tau'(0) = -c/2$ so that

$$\tau'(0)/\tau(0) = \alpha c(c+2\alpha\tau)^{-1}.$$

This expression decreases with increasing τ, but is quite independent of β. So, by this measure, incorporation of process history into the estimate of ν, and so into the dynamics of the process, has no effect at all upon the degree of persistence of the process.

We can see what effect this incorporation has by examining the autocorrelation function (17). It follows from (16) that one of χ or η become small as β becomes small. Specifically, we can take it that $\chi = \alpha + O(\beta)$ and $\eta = O(\beta)$. Thus, for small β the term $\exp(-\eta s)$ in (17) is slowly decaying. On the other hand, it has a coefficient of order β. One could thus say that, for small β, the variable x has a persistent component (i.e. a component of slowly decaying autocovariance) but that this component is of small amplitude.

Whether this conclusion is valid more generally has yet to be established. What it seems to imply is that the incorporation of a long-term state-average into the dynamics of the process by the construction of section 2 indeed produces a correspondingly persistent component in the process, but one of negligible magnitude, possibly because of the very stochastic stability of this average.

This work was carried out during the author's tenure of a Science and Engineering Research Council Senior Fellowship.

Reference

Whittle, P. (1990). A construction for multi-modal processes, and a potential memory device. To appear *J. Appl. Prob.*, **27** , 146-155.