MODELLING SOURCES OF VARIATION IN TRANSPORTATION SYSTEMS: THEORETICAL FOUNDATIONS OF DAY-TO-DAY DYNAMIC MODELS

David P. Watling, Institute for Transport Studies, University of Leeds, U.K.
d.p.watling@its.leeds.ac.uk
Giulio E. Cantarella, Department of Civil Engineering, University of Salerno, Italy
g.cantarella@unisa.it

Abstract – The last twenty years has seen a growing interest in models of transportation networks which explicitly represent the epoch-to-epoch adaptive behaviour of travellers, such as the day-to-day dynamics of drivers' route choices. These models may represent the system as either a stochastic or deterministic process. A body of theoretical literature now exists on this topic, and the purpose of the present paper is to both synthesise and advance this theory. To provide a focus to the work we analyse such models in terms of their ability to capture various contributory sources of variance in transportation systems. Dealing separately with the cases of uncongested and congested networks, we examine how moment-based deterministic dynamical systems may be exactly or approximately derived from some underlying stochastic process. This opens up such problems to the tools of both deterministic dynamical systems (e.g. stability analysis) and stochastic processes (e.g. Monte Carlo methods, statistical inference). In analysing these sources of variation, we also make several new advances to the existing body of theory, in terms of: extending the model assumptions (e.g. randomly-varying choice probabilities and stochastic demand); deriving exact, explicit connections between stochastic and deterministic processes in uncongested networks; applying stabililty analysis in novel ways to moment characterisations; and last, but not least, providing new limit theorems for asymptotic (large demand) analysis of the *dynamics* of stochastic process models in congested networks.

1. Introduction

For many decades transport modellers and planners have relied on the steady-state equilibrium paradigm as the basis for predicting the impacts of transport measures, be this through the Deterministic User Equilibrium (DUE) or Stochastic User Equilibrium (SUE) approach. More recently the focus of much development has been on within-day dynamic extensions of the equilibrium paradigm, with Dynamic Traffic Assignment (DTA) models. Common to all these approaches is, however, the implicit assumption of an essentially unchanging world over days (or, at least, the assumption that such variation is unimportant, or that the general pattern of traffic is not sensitive to such variation).

In contrast with equilibrium-based methods, dynamic process traffic assignment models emphasise the adaptive nature of traveller choices and the transport system over successive similar time periods, e.g. over days, weeks and months. We refer to such adaptations as 'day-to-day dynamics', even though they may actually occur over longer periods than a day; this distinguishes them from the 'within-day dynamic' issues that have been the focus of much DTA research. Such dynamic process traffic assignment approaches have a history dating back to the works of Horowitz (1984), Smith (1984) and Cascetta (1987, 1989), and so we can make some claim to this being a mature research area. Two recent companion papers to the present one review the main sub-classes of this research field, notably discrete-time Stochastic Process (SP) models (Watling & Cantarella,

2012) and discrete- and continuous-time Deterministic Process (DP) models (Cantarella & Watling, 2012), developing further the unified treatment of DP and SP models presented by Cantarella & Cascetta (1995). It is clear that there is now a substantial body of theory on these methods, and so with the first work almost thirty years ago, it is timely to consider how far this framework can take us in analysing particular kinds of problem.

While there is not the space here to describe all the contributions to this body of theory—rather the reader is referred to the reviews above and to the source references—some key themes that have emerged in the study of day-to-day dynamic process traffic assignment models are:

- **Coincidence with equilibrium models:** While day-to-day dynamic models avoid the premise of *equilibrium*, the notion of a fixed point of the dynamical system is still extremely important for understanding the modelled system. For example, if we know that all the point equilibria of a deterministic dynamical system are, say, DUE or SUE solutions, then we may exploit the body of literature on the multiplicity/uniqueness of DUE/SUE to infer properties of our day-to-day dynamical system. In this respect, we would highlight a recent strand of research exploring the coincidence between the point equilibria of certain classes of 'adjustment process' and DUE or boundedly-rational equilibrium solutions (Zhang et al, 2001; Yang & Zhang, 2009; Guo & Liu, 2010). For example, if a system is analysed shortly after an unplanned event (He & Liu, 2012), a planned change (Watling et al, 2012), or after some responsive control/information system is implemented (Hu & Mahmassani, 1997), then it may not be reasonable to assume that drivers have adapted sufficiently to be "near to a new equilibrium" (whatever that may mean precisely). However, even in such cases it is useful to know that a day-to-day model of such a system possesses point equilibria that it may be *headed toward*—the question of whether it really is headed towards such equilibria is then the task of stability analysis (see below).

- **Stability analysis of point equilibria:** The classical analyses of deterministic dynamical systems focus on examining the point equilibria, and asking: i) is a point equilibrium *locally stable* (essentially, if we start sufficiently close to it, is there a finite 'ball' that bounds its future distance from the equilibrium?); ii) is it *convergent* (essentially, given some initial conditions, will the system approach the equilibrium as time tends to infinity?); iii) is it *locally asymptotically stable* (is it both stable and convergent for some domain of attraction within a neighbourhood of the point equilibrium?); and iv) is it *globally asymptotically stable* (is the domain of attraction in iii) the entire state-space?). A range of stability results of this kind now exist for various kinds of system in both continuous-time (Smith, 1984; Friesz et al, 1994; Zhang & Nagurney, 1996; Han & Du, 2012) and discrete-time (Cantarella & Cascetta, 1995; Watling, 1999; Bie and Lo, 2010; Cantarella et al, 2012; Cantarella, 2013). Thus by characterising stability we are able to understand 'attractive' states towards which the day-to-day dynamic model will be ultimately drawn over time.

- **Existence and convergence of stationary probability distributions:** The analysis of stability, as described above, fundamentally lies within the domain of deterministic dynamical systems (i.e. DP models), yet corresponding issues arise in the study of SP models. Although relatively less attention has been paid to SP models, there still exist rather general frameworks for analysing SP traffic assignment models, that allow inferences to be made concerning i) *existence* of stationary probability distributions over ergodic subsets of the state-space (an ergodic subset being effectively a minimal region of the state-space for which there is zero-probability of the process leaving once it has entered entered), ii) *uniqueness* of a stationary distribution for any given process (if i) and ii) together are satisfied, then we say the process itself is *ergodic*); and iii) (strong) convergence of the process to a unique stationary distribution. Results relating to these issues may be found in Cascetta (1989), Cantarella & Cascetta (1995) and Watling & Cantarella (2012). Even if we use Monte Carlo simulation as a means of computing this process, knowledge as to whether it is headed towards a unique stationary distribution clearly has important implications for the way we interpret such simulations. Moreover, for ergodic processes it follows that an *unbiased* estimate of stationary means, variances, etc. of the process may be gained from a *single* pseudo-random realisation of the process, without the need multiple, replicated runs; clearly this has major implications for the computational load involved in implementing the approach.

- **Relationship between stationary probability distributions and equilibrium models:** As discussed above, there is a correspondence between point II in the world of DPs and point III in the world of SPs; it is natural to ask, then, does there exist a body of work corresponding to point I (which is clearly associated with DPs) but in the world of SPs? Indeed there does, though with the difference that the correspondence is not an exact one, but an asymptotic one that may be established as the OD demands and link capacities tend to infinity in tandem. Davis & Nihan (1993) provide the key source for this result, showing that in the asymptotic case the stationary distribution converges (in distribution) to a multivariate normal distribution, with mean given by an SUE model. Hazelton & Watling (2004) extended the applicability of this result by devising a practical method for computing the variance-covariance matrix of the stationary distribution, using only the SUE solution (i.e. without having to simulate the dynamic evolution of the process). Balijepalli & Watling (2005) subsequently extended this approach so as to be applicable to day-to-day dynamic models incorporating a within-day dynamic network loading. In the case of problems that possess multiple SUE, the behaviour of the corresponding SP may be more complex than this result suggests, yet still there are systematic procedures that can be followed for using SUE solutions to reveal the emergent phenomena of SP models in such cases (Watling, 1996). Even though we may apply our SP model in cases for which asymptotic (large demand) behaviour does not seem a plausible assumption, the asymptotic properties still reveal relatively easy-to-compute features that can aid in the interpretation of the results from, say, a simulation experiment, so have important practical consequences.

The purpose of the present paper is both to synthesise and extend these results in various ways. Our original contributions include extending the model assumptions, such as to randomly varying choice probabilities (section 3.3) and stochastic demand (section 4); explicitly deriving connections to deterministic processes in uncongested networks (sections 3 and 4); applying stabililty analysis in novel ways, to moment characterisations of stochastic process models (sections 3 and 4); and providing new limit theorems for asymptotic (large demand) analysis of the *dynamics* of stochastic process models (section 5.3.2).

Our purpose in doing so is both to extend the state-of-the-art, but also to explore (and encourage others) to use, extend and apply in novel ways the existing results in the literature. We do this, in part, by examining how to construct processes which satisfy the minimal necessary conditions needed to even consider applying the results above; we set out these conditions in the following section. Our purpose is to show that it is possible to be creative and to use these existing results in ways that may not have been envisaged by the authors, and we do this in order hopefully encourage others to be creative too in this way. Thus, one intention is that our paper acts as a bridge between modellers wishing to create day-to-day dynamic models, and the theory which currently exists.

In order to provide a focus to our approach, we examine such processes with a particular goal in mind—namely the representation, in a single consistent framework, of various kinds of time-dependent and stochastic variation. The 'consistency' we refer to arises from the requirement explicitly to represent how all elements of the transport system view, learn, assess and react to the variation; we cannot simply add the variation in an *ad hoc* way. This may be contrasted with the conventional equilibrium paradigm, which faces (and often avoids) the philosophical dilemma of *how* (and indeed, *whether*) equilibrium might be reached in a variable system. In fact, rather than rejecting equilibrium methods, our approach also helps to answer these questions.

In order to achieve this goal, we introduce the techniques in an original way, which we believe helps provide new and better insights into the foundations of dynamic process traffic assignment models, especially for the purpose we have in mind. In all cases our starting point is in terms of statistical models, and so fundamentally we adopt a SP approach. However, we shall see how DPs emerge for moments of such processes; firstly in the case of uncongested networks, where they emerge exactly to provide a full characterisation of the process, and secondly in the case of congested networks, where they arise from large demand approximations. This opens up the possibility to apply results on stability of DPs to the moment-based models, providing insights into both the deterministic and stochastic domains. Moreover, it allows an examination of how different model specifications, including different contributory sources of variation, influence the properties of such systems.

The paper is structured as follows. In section 2 we introduce the notation and some foundation concepts which guide the development of the models in subsequent sections. Section 3 considers the case of uncongested networks with fixed demand, and section 4 extends this to the case of stochastic demand, in each case using a series of models for a simple illustrative network. In the case of uncongested networks we show that it is possible to derive some exact analytic results. Section 5, on the other hand, considers congested networks, and develops asymptotic results for such a case (since exact results cannot be derived). Finally, in section 6 we draw together our conclusions and indicate promising avenues for future research.

2. Basic Concepts & Notation

A fundamental element of the models we consider will be the adaptive behaviour of travellers over time, as they repeat the requirement to make certain decisions related to travel or the location of activities. In the context of the present paper, the decision considered will be that of route choice, and we imagine that such decisions are potentially reviewed over *discrete* time periods which we refer to as 'days'. Days will be denoted by the letter t (for $t = 0, 1, 2,...$).

As stated in section 1, our objective in the present paper will be to show how various kinds of process may be cast in a form that allows the existing body of theory on day-to-day dynamic traffic assignment models to be applied; that is to say, our goal is to arrive at a situation in which subsequently it may be asked, are the sufficient conditions of a given theoretical piece of work satisfied? In order to do this, we shall require that the models are formulated in a way so as to satisfy four basic conditions:

a) There is a fixed and well-defined state-space $\mathcal{S}$ which describes the 'universe' of possible states corresponding to any single day t, with the state vector that describes day t as $\mathbf{x}^{(t)} \in \mathcal{S}$ (for $t = 0, 1, 2,...$).

b) If the model is a DP then $\mathcal{S}$ is part of m-dimensional Euclidean space ($\mathcal{S} \subseteq \mathbb{R}^m$) for some finite and constant m. If the model is a SP then *either* $\mathcal{S}$ is finite and formed from integer n-tuples (i.e. $\mathcal{S} \subseteq \mathbb{Z}^n$), or $\mathcal{S}$ is part of m-dimensional Euclidean space ($\mathcal{S} \subseteq \mathbb{R}^m$), or $\mathcal{S}$ is a combination of these two kinds of variable (i.e. $\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \subseteq \mathbb{Z}^n \times \mathbb{R}^m$).

c) The process satisfies the one-step Markov property. In the case of DPs this means that in order to determine the state at time t it is sufficient to know the state at time $t - 1$, and no additional information is imparted by knowing earlier states. In the case of SPs, we suppose that in order to determine the state *probability distribution* at time t it is sufficient to know the state probability distribution at time $t - 1$.

d) The process is time-homogeneous. In the case of DPs this means that the functional dependence of $\mathbf{x}^{(t)}$ on $\mathbf{x}^{(t-1)}$ is time-independent, all that matters is what the state $\mathbf{x}^{(t-1)}$ was, not the time $t - 1$ at which it was experienced (i.e. the same functional relationship would hold for the dependence of $\mathbf{x}^{(s+t)}$ on $\mathbf{x}^{(s+t-1)}$ for any positive integer s). In the case of SPs, time homogeneity relates to the fact that the conditional joint probability distribution/density of $\mathbf{x}^{(t)}$ given $\mathbf{x}^{(t-1)}$ is independent of t.

While these might seem, at first sight, rather trivial and weak conditions to satisfy, in order to satisfy them all it is often necessary to be creative, especially in the definition of what constitutes a 'state'. For example, on point a), it is not necessary that we describe a transportation system state only in terms of its traffic flows; see, for example, Watling (1999) for a description in terms of travel costs, which allows a model that is non-markovian in terms of flows to still satisfy the conditions above. On point b), our requirements are linked to the spaces which permit analysis with the theoretical tools available. For DPs, stability analysis is typically based on the analysis of the system Jacobian, which clearly only make sense for models expressed with a continuous state-space. On the other hand, the most well-known and directly-applicable results for SPs are in the case of discrete state-spaces; however corresponding tools do exist for a wider range of spaces, as discussed in Watling & Cantarella (2012). Regarding point c), this requirement might at first sight seem to preclude important classes of model in which drivers 'learn' from a sequence of past days in order to make their travel decisions for the forthcoming day; in fact, it does not preclude such approach, provided that we imbue our definition of the state variable with some notion of a finite history. Thus, we can see an important interdependence between a), b) and c). Finally, point d) is more of a pragmatic requirement. Although it is possible to analyse time-*in*homogeneous models (see Watling & Cantarella, 2012), little in the way of explicit theory has yet been developed for transportation systems, and in fact we are a long way from reaching the boundaries of application of existing theory of time-*homogenous* systems. For most problems of interest to transportation planners, the assumption of time-homogeneity will likely be acceptable.

The particular focus of the following sections 3 to 5 will be to analyse the extent to which various notions of variability may be included in day-to-day dynamics models, and how we might 'interrogate'/formulate these models in a way so as to satisfy the four requirements above. Given our interest in variability, then all of the models will be initially cast in the form of a SP; corresponding DPs will then emerge as a way of understanding the SP. Given the focus on SPs, we note here some basic notation that is used to describe them; this notation is not entirely conventional, but allows us to describe SPs under the various kinds of state-space permitted in point b) above. In particular, we suppose:

$\{q^{(t)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$ denotes the day t joint probability/probability-density function across the possible states $\mathbf{x} \in \mathcal{S}$ (for $t = 0, 1, 2...$); and
$\{\phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) : \mathbf{x}, \mathbf{y} \in \mathcal{S}\}$ denotes the *conditional* joint probability mass/density function across possible states $\mathbf{x} \in \mathcal{S}$ today, *given* that $\mathbf{y} \in \mathcal{S}$ was the state yesterday, assuming a model parameter vector $\boldsymbol{\theta} \in \Theta$ (we shall refer to this as the "transition function").

Property d) above for SPs then corresponds to the requirement that the transition function ϕ be time-independent, and the parameter vector $\boldsymbol{\theta}$ be fixed and time-independent, implying that the resulting process is *time-homogeneous*. We may then write our stochastic process as one of the following:

For any given initial distribution $\{q^{(0)}(\mathbf{x}) : \mathbf{x} \in \mathcal{S}\}$, then for $t = 1, 2, ...$:

i) Markov Chain:

$$q^{(t)}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})\, q^{(t-1)}(\mathbf{y}) \qquad (\mathbf{x} \in \mathcal{S} \subseteq \mathbb{Z}^n; \boldsymbol{\theta} \in \Theta)$$

ii) Markov Process:

$$q^{(t)}(\mathbf{x}) = \int_{\mathbf{y} \in \mathcal{S}} \phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta})\, q^{(t-1)}(\mathbf{y})\; \mathrm{d}\mathbf{y} \qquad (\mathbf{x} \in \mathcal{S} \subseteq \mathbb{R}^m; \boldsymbol{\theta} \in \Theta)$$

iii) Markov Process/Chain

$$q^{(t)}(\mathbf{x}) \equiv q^{(t)}((\mathbf{x}_{[1]},\mathbf{x}_{[2]})) =$$
$$\sum_{\mathbf{y}[1] \in \mathcal{S}[1]} \int_{\mathbf{y}[2] \in \mathcal{S}[2]} \phi((\mathbf{x}_{[1]},\mathbf{x}_{[2]}),(\mathbf{y}_{[1]},\mathbf{y}_{[2]}); \boldsymbol{\theta})\, q^{(t-1)}((\mathbf{y}_{[1]},\mathbf{y}_{[2]}))\; \mathrm{d}\mathbf{y}_{[2]}$$
$$(\mathbf{x} \equiv (\mathbf{x}_{[1]},\mathbf{x}_{[2]}), \mathbf{x}_{[1]} \in \mathcal{S}_{[1]} \subseteq \mathbb{Z}^n; \mathbf{x}_{[2]} \in \mathcal{S}_{[2]} \subseteq \mathbb{R}^m; \boldsymbol{\theta} \in \Theta).$$

The three specifications have the common feature that each includes the functions ϕ and q, and in each specification these functions perform the same role: in simple terms we might say that ϕ is "the model" (the thing that we calibrate and specify as modellers) and then (for time $t = 1, 2, ...$) q is "the unknown".

3. Uncongested Networks with Deterministic Demand

The examples in the present section reduce our problem to the simplest day-to-day process of all, and will aim to subsequently describe the development of techniques from this foundation. Here we aim to follow the analogy by which equilibrium methods originally stemmed from all-or-nothing and uncongested probabilistic assignment (and indeed which continue to use these foundation methods as building blocks). The examples given in sections 3.1 and 3.2 below are very clearly restricted examples of the wide family of models considered by Cantarella & Cascetta (1995). Example 3.3 does not so clearly fit within their framework, and so appears to extend it. In all examples, however, we aim to provide insights that are beyond this source reference, by making the restriction to uncongested cases. Although all of the examples consider only a network with two parallel routes, they are all extendable to the general case; we consider only the simplest case to aid description. For a general network description of the examples in section 3.1 and 3.2, for example, the reader is referred to Cantarella & Cascetta (1995), although the corresponding analysis was not performed there.

3.1 Modelling route choice behaviour: an introductory model – example 1

All the examples in this paper concern a network consisting of two parallel arcs/routes connecting a single OD pair.

In this sub-section we assume a fixed (non-random) integer OD demand $d \geq 1$, such that we may represent the state of the network by an integer *scalar* x denoting the flow on arc 1 (with the flow on arc 2 then clearly $d - x$). Thus, our state space $\mathcal{S} \subseteq \mathbb{Z}$ is given by $\mathcal{S} = \{0, 1, 2, \ldots, d\}$, and we are appealing to specification i) in section 2, that of a Markov Chain. Moreover, the demand d is assumed given, thus it is a parameter of the model.

We suppose that all the factors that affect route choice are combined into a single weighted measure which we call 'performance' (i.e. generalised cost), with the performance of arc i denoted by the fixed (non-random) value c_i (i = 1,2), with $c = c_1 - c_2$. Moreover the network is uncongested (or so lightly congested that it may considered to be uncongested), and so the flows do not affect the factors that motivate route choice such as travel time. Thus the cost difference c is a parameter of the model.

On any given day, each of the d drivers is supposed to choose between the routes independently and at random according to a logit choice random utility model, with scale parameter $\xi > 0$, based on arc performance values; the standard deviation of all sources of uncertainty, e.g. model aggregation or behavioural variation in terms of the user perception of performance, is proportional to ξ^{-1}. Thus the probability of choosing arc/route 1 is given by $\rho = 1/(1 + \exp(\xi c))$.

The parameters of the overall model may be collected together in the vector $\boldsymbol{\theta} = (d, c, \xi)$. The assumptions together imply that the *transition function*—i.e. the conditional probability distribution of the flow x on arc 1 on any one day, given that the flow on arc 1 was y yesterday—is given by the Binomial expression:

$$\phi(x, y; \boldsymbol{\theta}) = (d!/(x!\,(d-x)!)) \cdot \rho^x \cdot (1-\rho)^{d-x} = f(x; \boldsymbol{\theta}),$$

$$\text{say (for } x = 0,1,\ldots,d;\ y = 0,1,\ldots,d;\ \boldsymbol{\theta} = (d, c, \xi)).$$

That is to say, the transition function is independent of the previous day's flow y, and the evolution of the state probability distribution is given by:

$$q^{(t)}(x) = \Sigma_{y \in S}\ \phi(x, y; \theta)\ q^{(t-1)}(y) = f(x; \theta)\ \Sigma_{y \in S}\ q^{(t-1)}(y) = f(x; \theta) \qquad (x = 0,1,\ldots,d;\ t = 1,2,\ldots)$$

i.e. it is time-independent and coincides with the transition function. This occurs only because of the very restrictive assumption that the choice probabilities of the routes are constant, since the performance measures that driver these probabilities are flow-independent. In general, the state distribution and transition function, although both probability distributions, are two quite different entities (as we shall see in other examples below), but ones which seem to be often confused with each other by those unfamiliar with stochastic process methods. In this special case, where we obtain Binomial variation, it is simple to see that the state distribution has a mean and variance given by:

$$\mathrm{E}[X^{(t)}] = d \cdot \rho$$
$$\mathrm{var}[X^{(t)}] = d \cdot \rho \cdot (1 - \rho) \quad (t = 1,2,\ldots).$$

In general, stochastic process models also possess important *auto-correlations*, as measured by $\mathrm{cov}(X^{(t)}, X^{(s)})$ ($t \neq s$), but in the model given all such covariances are zero, the model assumes entirely independent draws over days, from an unchanging distribution.

The expected value $\mathrm{E}[X^{(t)}]$ thus coincides with the traffic flow that would be predicted by a standard logit-based stochastic assignment model. The variance $\mathrm{var}[X^{(t)}]$, being Binomial, increases with the actual driver population size d; this implies that, unlike with traditional equilibrium methods, it is no longer sufficient to think only of a flow *rate*, it makes a difference as to whether we are examining the behaviour of 2000 drivers over a 2 hour period or 1000 drivers over a 1 hour period. If, on the other hand, we fix d, then it is simple to show that the variance is at its maximum when $\rho = 0.5$, i.e. when $\xi c = 0$. Thus for fixed $c \neq 0$, it follows that $\mathrm{var}[X^{(t)}]$ increases as $\xi \to 0$. In other words, more behavioural variation in terms of the perception of performance (as measured by ξ^{-1}) implies more flow variation; this is rather intuitive, but is worth remarking on primarily as it does *not* necessarily follow in more complex versions of the model specification (as we shall see later). On the other hand, for fixed $\xi > 0$ then $\mathrm{var}[X^{(t)}]$ increases as $|c| \to 0$. As $|c|$ becomes large, then the alternatives are more clearly distinguished, whereas as $|c| \to 0$ the alternatives become more similar, such that ultimately it as if each driver tosses a coin to decide which route to take.

Obviously this model is highly simplistic and is yet to capture the real issue of day-to-day dynamics that we wish to highlight, and we can well see here the analogy we made earlier: while all-or-nothing assignment and stochastic network loading (Dial, 1971; Sheffi, 1985) have both proven to be critical building-blocks in the analysis of congested networks, the approaches themselves do not actually deal with the congestion issue. However, just as in the analogy, it is still valuable to analyse the features of the foundation method (in this case the modelling of uncongested networks through time-independent processes), since it allows us to understand to what extent the behavioural components of this model might be suitable for inclusion in a more general approach, such as that described later.

In particular we ask: What is this model actually assuming that drivers do, who make repeated trips? Firstly, the Binomial distribution implicitly assumes independence between the individual elements, in this case between the decisions of drivers. This could be defended, even in the case of a congested network, since the independence assumption resides in the transition function; i.e. in general, it is an assumption of *conditional* independence, essentially saying that *given* their past history (in which drivers will interact and so be interdependent), drivers make a decision today independently of one another. This seems a reasonable approximation, and seems broadly consistent with the notion of competitive, non-cooperative behaviour as assumed in DUE or SUE models. The distinction between independence and conditional independence is not clear from this simple example though (since this is a special case where they coincide), and so we will return to illustrate this issue later, with more complex examples.

Secondly, a quite different kind of independence emerges in this model, namely the *temporal* independence of decisions. That is to say, that while the decisions made on days $t-1$ and t, say, come from a common probability distribution, the actual realisation of any previous day does not affect a future day. This is a rather subtle issue to understand. Even though later we remove the assumption of a common probability distribution (through the use of learning process of the cost/performance measures), still we then have the issue of *temporal conditional independence*: i.e. the decisions on day t given the history up to day t are statistically independent of the decisions on day $t-1$ given the history up to day $t-1$. Thus, while the histories will typically be dependent over time, the decision-making given those histories is independent. This seems more difficult to defend.

Given the same group of drivers, it seems likely that there would be some consistency over time in their decisions, given an unchanging history on which to base those decisions – so what is the statistical variation representing? One interpretation is that if the logit probabilities gave choice probabilities of the two routes of 0.3:0.7, say, then each driver on each day they travel randomly chooses between these two routes with these probabilities. Perhaps a more plausible interpretation is that what we observe on different days is not the same d drivers, but a sample of d from a large population, and so we might try to match a sequence of drivers with similar characteristics over time, they are not actually the same d drivers, and this is some reason for the variation we see (the actual individual might not be so variable in their decisions). This is a similar logic used in the application of the paired t-test in cases where the entities being observed are not exactly the same in the two cases, but are paired together to be similar.

Having said this, the assumption of temporal conditional independence is a strong one, and it is natural to seek ways in which we might relax or control this in the model. The next example considers a simple device for achieving this.

3.2 Modelling route choice behaviour: a general model – example 2

In the example above we noted the difficulties in justifying the assumption of temporal conditional independence, given that travellers are likely to be somewhat repetitive in their behaviours over time if the underlying stimuli do not change so much. At least, it seems we should have a means of controlling/calibrating the level of repetition. This also provides us with the simplest way of generating a truly dynamic, stochastic process, and at the same time will allow us to explore "determinised" dynamic approximations to the stochastic process.

We thus adapt the very simple model described in Example 1, simply by assuming now:

- with probability α travellers reconsider their previous day's choice, and those that do decide to reconsider then make choices according to the logit model as before (possibly then repeating the previous day's choice); and
- with probability $1 - \alpha$ travellers choose between the available routes with probabilities equal to the fraction of travellers that actually chose those routes on the previous day.

Therefore we have one more parameter, assumed strictly positive, thus $\alpha \in]0,1]$, with $\alpha = 1$ leading to example 1 discussed in the previous sub-section.

Under such a behavioural model, travellers now are assumed to have two reasons for choosing route 1 (over route 2): either they choose it out of habit based on a probability of y/d (where y is the number of drivers that actually chose route 1 yesterday), or they choose it according to a logit model as in Example 1. The probability of choosing for the first reason is $1-\alpha$ and for the second reason is α. Assuming they choose conditionally independently *between individuals* at any given time, this combination of assumptions implies:

$$X \mid Y = y \sim \text{Binomial}(d\,,\,(1-\alpha)(y/d) + \alpha \cdot \rho$$

and thus the transition function is:

$$\phi(x, y; \boldsymbol{\theta}) = (d!/(x!\,(d-x)!))\,.\,((1-\alpha)(y/d) + \alpha\cdot\rho)^{x} \cdot (1 - ((1-\alpha)(y/d) + \alpha\cdot\rho))^{d-x}$$
$$(\text{for } x = 0,1,...,d;\ y = 0,1,...,d;\ \boldsymbol{\theta} = (d, c, \xi)).$$

By standard properties of the Binomial we thus have for $t = 1,2,\ldots$:

$$\text{E}[X^{(t)}|X^{(t-1)}] = d\,((1-\alpha)(\,\text{X}^{(t-1)}/d) + \alpha\cdot\rho)$$
$$\text{var}[X^{(t)}|X^{(t-1)}] = d\,((1-\alpha)(\text{X}^{(t-1)}/d) + \alpha\cdot\rho)\,(1-((1-\alpha)(\text{X}^{(t-1)}/d) + \alpha\cdot\rho)).$$

These expressions show that the trajectory of the process, both in terms of its mean and variance, is dependent on actual choices made in the past (as reflected in $X^{(t-1)}$), unlike the model considered in Example 1. We can rearrange these expressions slightly to become:

$$\text{E}[X^{(t)}|X^{(t-1)}] = d\alpha\rho + (1-\alpha)X^{(t-1)}$$
$$\text{var}[X^{(t)}|X^{(t-1)}] = d\alpha\rho\,(1-\alpha\rho) + (1-\alpha)(1 - 2\alpha\rho)\,X^{(t-1)} - ((1-\alpha)^2/d)\,(X^{(t-1)})^2$$

and we may then use standard statistical identities to obtain expressions for the *unconditional* moments:

$$\text{E}[X^{(t)}] = \text{E}[\text{E}[X^{(t)}|X^{(t-1)}]] = d\alpha\rho + (1-\alpha)\text{E}[X^{(t-1)}]$$
$$\text{var}[X^{(t)}] = \text{E}[\text{var}[X^{(t)}|X^{(t-1)}]] + \text{var}[\text{E}[X^{(t)}|X^{(t-1)}]]$$
$$= d\alpha\rho\,(1-\alpha\rho) + (1-\alpha)(1 - 2\alpha\rho)\,\text{E}[X^{(t-1)}]$$
$$- ((1-\alpha)^2/d)\,(\text{var}[X^{(t-1)}] + (\text{E}[X^{(t-1)}])^2) + (1-\alpha)^2\,\text{var}[X^{(t-1)}]\,.$$

Thus, in the special case we are considering, we are able to understand the future evolution of the mean and variance from knowing only the mean and variance of the distribution in the previous time step, we do not need to know any higher order moments. Also, we can see in this particular case that if our interest were only in how the mean changes, then it is sufficient to know only the mean in the past. If we write $\mu^{(t)} = \text{E}[X^{(t)}]$, $\omega^{(t)} = \text{var}[X^{(t)}]$, then we can write the above simply as:

$$\mu^{(t)} = d\alpha\rho + (1-\alpha)\mu^{(t-1)}$$
$$\omega^{(t)} = d\alpha\rho(1-\alpha\rho) + (1-\alpha)(1-2\alpha\rho)\,\mu^{(t-1)} - (1-\alpha)^2\,d^{-1}(\omega^{(t-1)} + (\mu^{(t-1)})^2) + (1-\alpha)^2\,\omega^{(t-1)}\,.$$

An equilibrium/stationary probability distribution of such a system, would thus be characterised by a (mean, variance) combination (μ^*,ω^*) satisfying $\mu^{(t)} = \mu^{(t-1)} = \mu^*$ and $\omega^{(t)} = \omega^{(t-1)} = \omega^*$, which from the above is given by:

$$\mu^* = d\rho$$
$$\omega^* = d\rho\,(1-\alpha\rho + \alpha\,(1-\alpha)\,(1-2\rho))\,/(1-(1-d^{-1})(1-\alpha)^2)\,.$$

Figure 1 shows variance ω^* against parameter ξ (such that ρ = 1/(1 + exp(ξc) as introduced above) for α = 0.25, 0.50, 0.75, 1.00 (from top to bottom). As expected the variance decreases as parameter ξ increases (choice behaviour becomes less at random); moreover, the higher the reconsidering probability α, the smaller the variance ω^* is.

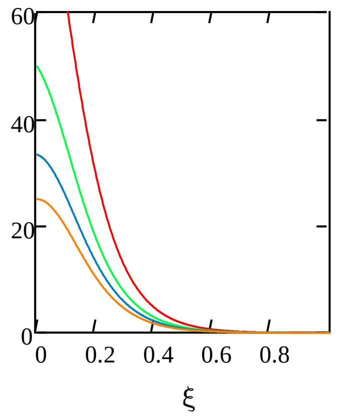


Figure 1: Variance against Logit dispersion parameter ξ for different values of α.

How might we use these results to understand something about our underlying system? It is not the purpose of the present paper to pursue this (rather our purpose has been to show how to open up the models for analysis), but it would seem that this formulation opens up several subsequent lines of enquiry, part of which is specific to this model and part of which draws on the theory described in section 1:

- From the expressions above we may examine how the derivative of the stationary variance ω^* depends on d, ρ and α, and thereby understand both the direction and relative importance of these components of flow variation. This may be done partly to ask whether the model is reasonable, but also as a potential means of suggesting a candidate parameterisation of the variation which might be estimated from observed data. It is notable in this model that μ^* is independent of α, so α does not influence the stationary mean of process, only the stationary variance.
- Although, as noted above, α does not affect the stationary distribution, it is clear from the expressions that it does affect how fast stationarity is approached from some given initial conditions. This may be seen by re-writing the process for the mean in the form of a process relative to the stationary mean, where it is evident that the greater is α then the faster is the decrease in distance to stationarity:

$$(\mu^{(t)} - \mu^*) = (1-\alpha)(\mu^{(t-1)} - \mu^*)\,.$$

Finally, we may also examine—through the stability theorems for Deterministic Processes (DPs) cited in section 1 (see Cantarella, 2013, for an updated presentation) — the *coupled system* above in terms of the binary state variable $(\mu^{(t)}, \omega^{(t)})$ considered as a DP over the mean and variance space, $\mu^{(t)}$ and $\omega^{(t)}$; in this case the pair of stationary moments (μ^*, ω^*)

plays the role of the (unique) fixed-point of the DP. To study the stability properties of this fixed-point it is useful to look at the Jacobian matrix of the DP. The Jacobian is clearly triangular, with entries on the main diagonal independent of the state variables, $(\mu^{(t)}, \omega^{(t)})$. Thus it is easy to compute its determinant δ, moreover the two eigenvalues η_1, η_2 are given by the entries on the main diagonal:

$$\delta = (1-\alpha)^3(1 - d^{-1})$$
$$\eta_1 = (1-\alpha)$$
$$\eta_2 = (1-\alpha)^2(1 - d^{-1})$$

As far as $\alpha \in\ ]0, 1]$ and $d \geq 1$, the determinant of the Jacobian is non negative and less than one, that is the system is dissipative, that is converges to some attractor; moreover both the eigenvalues are non negative and less than one, thus the (unique) fixed-point (μ^*, ω^*) is always (asymptotically) stable.

3.3 Modelling route choice behaviour: an extension – example 3

We have thus seen one simple way to relax the assumption of temporal conditional independence of decisions, and to control the level of dependence through the parameter α. This parameter has been seen to have important impacts on the dynamic trajectory of the process as well as on the levels of variability observed, even when the process is stationary. Now we explore how we might parameterise some additional elements of choice variation.

We generalise Example 2 by now supposing that $X\,|Y$, rather than following a Binomial distribution with a fixed choice probability, follows a Beta-Binomial distribution, derived from compounding a beta-distributed choice probability with a Binomial distribution across routes conditional on the choice probability. The Beta-Binomial model has been widely used in the statistical analysis of many social systems, such as consumer purchasing including car fuel purchasing (Chatfield and Goodhart, 1970), as an error distribution for regression where we wish to identify components of variance (Crowder, 1978), in Bayesian inference for brand choice of consumers (Lee & Sabavala, 1987), and for representing the variations in choice probabilities between trials of subjects given sensory or preference tests to perform (Ennis & Bi, 1998). Wilcox (1981) reviews the development and use of this model for representing individuals performing various psychological functions, noting the problem with the simpler binomial assumption applied to individuals (rather than situation in nature) is that it 'ignores an individual's wisdom, determination, pessimism and experience'; and, besides, the beta-binomial has been shown to fit well to *populations* of individuals performing such psychological functions. It seems that it is not a great step, then, to argue that the Beta-Binomial may be a suitable model for representing population of individuals repeatedly making path choice decisions; if nothing else, it provides a means of representing over-dispersion in choices relative to the Binomial model, and from real data we may then test whether this over-dispersion is significant.

In our particular case, for the mean of the beta distribution we shall use the expression for choice probability derived in Example 2, namely $(1-\alpha)(y/d) + \alpha\rho$, remembering the meaning of ρ. Since we wish to specify the mean of the beta distribution, we use a commonly-used re-parameterisation of the distribution in terms of its mean and a parameter $\upsilon > 0$ such that we suppose the choice probability P is distributed as:

$P \sim \text{Beta}(a,b)$ where $a = \upsilon\,\text{E}[P]$ and $b = \upsilon\,(1 - \text{E}[P])$

where in our case we set $\text{E}[P] = (1-\alpha)(y/d) + \alpha\rho$, and the density function of the beta r.v. is $f(p) = p^{a-1}\,(1-p)^{b-1} / B(a,b)$, with $B(a,b) = \Gamma(a+b) / (\Gamma(a)\,\Gamma(b))$, $\Gamma(.)$ being the gamma function. Therefore we have one more parameter, assumed strictly positive, $\upsilon > 0$, with $\upsilon \to \infty$ leading to example 2 discussed in the previous sub-section. Figure 2 shows the pdf of Beta for different values of of *E[P]* and υ.

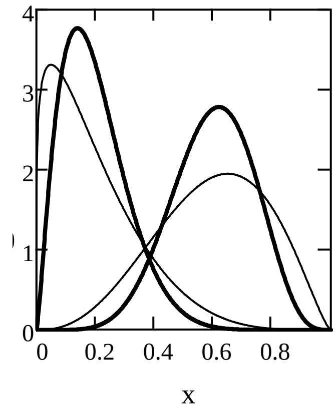


Figure 2: Beta density function for different values of E[P] = 0.2 (left) or 0.6 (right) and υ = 6 (thin line) or 12 (thick line).

Supposing, then, that $X \mid (Y, P)$ is Binomial(d, P), then the composition of these assumptions leads to the Beta-Binomial density for the transition function (i.e for $X \mid Y$) of:

$$\phi(x, y; \boldsymbol{\theta}) = (d!/(x!(d-x)!))\,.\,B(x + \upsilon((1-\alpha)(y/d)+\alpha\rho),\, d - x + \upsilon(1-(1-\alpha)(y/d)-\alpha\rho)) \,/\, B(\upsilon((1-\alpha)(y/d)+\alpha\rho),\, \upsilon(1-(1-\alpha)(y/d)-\alpha\rho))$$

(for $x = 0,1,\ldots,d$; $y = 0,1,\ldots,d$; $\boldsymbol{\theta} = (d, \rho(=1/(1+\exp(\xi c))), \upsilon)$)

Supposing, then, that $X \mid (Y, P)$ is Binomial(d, P), then the composition of these assumptions leads to the Beta-Binomial density for the transition function (i.e for $X \mid Y$) of:

$$\phi(x, y; \boldsymbol{\theta}) = (d!/(x!(d-x)!))\,.\,B(x + \upsilon((1-\alpha)(y/d)+\alpha\rho),\, d - x + \upsilon(1-(1-\alpha)(y/d)-\alpha\rho)) \,/\, B(\upsilon((1-\alpha)(y/d)+\alpha\rho),\, \upsilon(1-(1-\alpha)(y/d)-\alpha\rho))$$

(for $x = 0,1,\ldots,d$; $y = 0,1,\ldots,d$; $\boldsymbol{\theta} = (d, \rho(=1/(1+\exp(\xi c))), \upsilon)$)

where the function $B(.,.)$ is defined above.

By standard properties of the Beta-Binomial distribution, we thus have for $t = 1,2,\ldots$:

$$\text{E}[X^{(t)}|X^{(t-1)}] = d((1-\alpha)(X^{(t-1)}/d)+\alpha\rho)$$
$$\text{var}[X^{(t)}|X^{(t-1)}] = d((1-\alpha)(X^{(t-1)}/d)+\alpha\rho)(1-(1-\alpha)(X^{(t-1)}/d)-\alpha\rho)(1+(d-1)/(\upsilon+1)).$$

That is to say, we obtain the same conditional mean as in Example 2, and a similar conditional variance but just with the additional multiplier $(1+(d-1)/(\upsilon+1))$ above. Thus as $\upsilon \to \infty$ we approach the Binomial (fixed choice probabilities) case; the variance in the choice probability, like $\text{var}[X^{(t)}|X^{(t-1)}]$, is inversely proportional to υ.

Since the multiplier $(1+(d-1)/(\upsilon+1))$ is independent of $X^{(t-1)}$, then we can use the results derived for Example 2 but just with the additional multiplier inserted. Thus the *unconditional* mean and variance now behave as:

$$\mu^{(t)} = d\alpha\rho + (1-\alpha)\mu^{(t-1)}$$
$$\omega^{(t)}/\ (1+(d-1)/(\upsilon+1)) = d\alpha\rho\,(1-\alpha\rho) + (1-\alpha)(1-2\alpha\rho)\ \mu^{(t-1)}$$
$$-\ (1-\alpha)^2\ d^{-1}(\omega^{(t-1)} + (\mu^{(t-1)})^2) + (1-\alpha)^2\ \omega^{(t-1)}$$

and thus as υ is reduced (indicating increasing choice probability variance), so the factor $(1+(d-1)/(\upsilon+1))$ is increased and so the variance $\omega^{(t)}$ is relatively more sensitive to changes in the right-hand side variables $(\mu^{(t-1)}, \omega^{(t-1)})$.
The stationary distribution is again fully characterised by (μ^*, ω^*), now satisfying:

$$\mu^* = d\rho$$
$$\omega^* = d\rho\,(1 - \alpha\rho\ + \alpha\,(1 - \alpha)\,(1 - 2\rho))\ /((1/(1+(d-1)/(\upsilon+1))) - (1 - d^{-1})(1-\alpha)^2)\ .$$

What might we conclude from such an approach? Again, there are many suggestive features for further investigation:

- From the process for $(\mu^{(t)}, \omega^{(t)})$, we can see that the mean process is not affected at all by the new parameter υ, but the variance process is.

- From the expressions for the stationary moments, it is clear that reducing υ (i.e. increasing choice probability variance) therefore increases the stationary flow variance ω^*, as one may expect from intuition.

Figure 3 shows variance ω^* against parameter ξ (such that $\rho = 1/(1 + \exp(\xi c)$ as introduced above) for $\alpha = 0.75$, and υ = 10, 100, 1000, 1000 (from top to bottom). As expected the smaller the parameter υ, the greater the variance ω^* is.

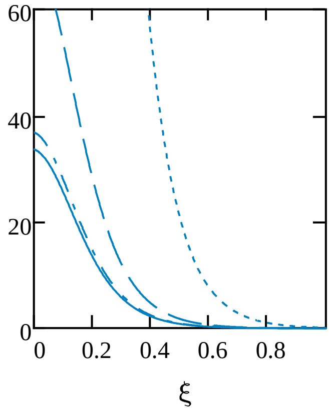


Figure 3: Variance a*gainst Logit dispersion parameter ξ for different values of υ ($\alpha = 0.75$).*

As in example 2, we can consider the DP over the mean and variance space, $\mu^{(t)}$ and $\omega^{(t)}$, with a triangular Jacobian with entries on the main diagonal independent of the state variables. Thus, the determinant and the two eigenvalues are given by:

$$\delta = (1-\alpha)^3(1 - d^{-1})\ (1+(d-1)/(\upsilon+1))$$
$$\eta_1 = (1-\alpha)$$
$$\eta_2 = (1-\alpha)^2(1 - d^{-1})\ (1+(d-1)/(\upsilon+1))$$

As far as $\alpha \in\ ]0, 1]$, $d \geq 1$, and $\upsilon > 0$ the determinant and both the eigenvalues are non negative. Thus the system is dissipative if the determinant is less than one, that is if the following condition holds:

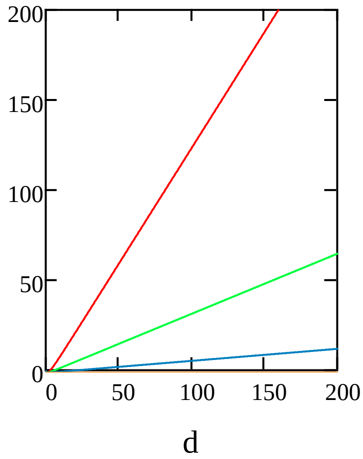
$$\upsilon > \upsilon_1(d, \alpha) = d\,((1-\alpha)^3\,(d-1)-1)\,/\,(d-(1-\alpha)^3\,(d-1))$$

Moreover, the first eigenvalue is always less than one but, the second eigenvalue is less than one for great enough values of parameter υ only, that is if the following condition holds:

$$\upsilon > \upsilon_0(d, \alpha) = d\,((1-\alpha)^2\,(d-1)-1)\,/\,(d-(1-\alpha)^2\,(d-1))$$

If the above condition holds, the fixed-point (μ^*, ω^*) is (asymptotically) stable, the stability region depending on d, α, *and* υ, It is also worth noting that for $\upsilon \to \infty$ both the above conditions always holds, consistently with results in the previous sub-section.

Function $\upsilon_0(d, \alpha)$ increases against d, whilst decreases against α, with $\upsilon_0(d, \alpha = 1) = -1$ (Fig. 4). This function takes negative values for $\alpha > 1 - 1\,/\,(d-1)^{1/2}$. Thus increasing the habit parameter α and/or decreasing the demand flow d enlarge the stability region. Similar considerations hold for function $\upsilon_1(d, \alpha)$.



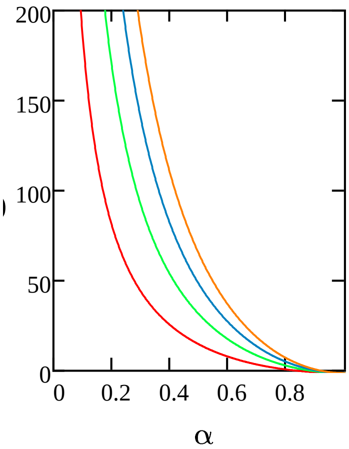


Figure 4: Boundary stability function $\upsilon_0(d, \alpha)$ for parameter υ,
left side: function $\upsilon_0(d, \alpha)$ against d for α = 0.25, 0.50, 0.75 (top to bottom)
right side: function $\upsilon_0(d, \alpha)$ against α for d = 50, 100, 150, 200 (left to right)

The investigation of the stability properties of the process $(\mu^{(t)}, \omega^{(t)})$ suggests that reducing υ, while increasing the choice variation, will tend to reduce the stability domain, which may seem counter to intuition in which adding variation tends to make stability better. This can be better understood by the interpretation often given to υ in such a Beta-Binomial specification as acting as a kind of sample size, such that a smaller υ is like saying the system is based on a smaller number of individuals and so behaves in a more 'lumpy' way—reducing υ is in a sense working in the opposite direction to the asymptotic, law of large numbers results exploited later in section 5, and this respect an interesting feature is the contrasting effects of d and υ on issues such as stability.

4. Uncongested Networks with Stochastic Demand

In the previous section we considered how to model variation in the behavioural elements of route choice, and the manner in which such elements impact on the variation in flows we might observe in the real-world. In the present section we consider how we might model variation in elements or processes that are exogenous to our modelled world, especially through variation in the parameters supplied to the model. Examples of such parameters we might consider include:

- variation in the origin-destination demand levels, reflecting daily changes in activity patterns;
- variation in any of the behavioural parameters represented in section 3, perhaps reflecting daily variation in the composition of the driving population;
- variation in the elements of the congestion relationships, as considered in section 5, e.g. variability in capacities.

However, we shall focus in the present section on one example of such variation, namely that in the origin-destination demand levels, and will propose several alternative formulations of this phenomenon. Two different approaches will be presented below.

4.1 Binomially-distributed stochastic demand – example 4

Our first approach aims to generalise Example 2 of section 3.2. assuming Binomially-distributed stochastic demand. We suppose that there is a larger pool of potential travellers, given by some integer $d_{\mathrm{pot}} > 0$, and that this potential pool of travellers each chooses to travel on any one day independently (of each other and days) with constant probability ε ($0 < \varepsilon \le 1$). Thus, if the random variable X_i denotes the flow on route i (i = 1,2), then the allocation to the three options (route 1, route 2, no-travel) is given by the random variables (X_1, X_2, $d_{\mathrm{pot}} - X_1 - X_2$), which are conditionally Multinomially distributed given the flows on the previous day, with probabilities ($\varepsilon \cdot p$, $\varepsilon \cdot (1-p)$, $1-\varepsilon$) if p is the conditional probability of choosing route 1. The probability p will then be defined using the techniques described in section 3.2.

It is convenient to slightly change the parameterisation introduced earlier to use:

$$\rho_1 = \rho \qquad\qquad \rho_2 = 1 - \rho$$

and with our state variable now two dimensional, $\mathbf{x} = (x_1, x_2)$, the above combination of assumptions implies:

$$(X_1, X_2) \mid (Y_1, Y_2) = (y_1, y_2) \sim \text{Multinomial}(d_{\mathrm{pot}}, \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}y_1+\alpha\rho_1), \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}y_2)+\alpha\rho_2))$$

and thus the transition function is:

$$\phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = (d_{\mathrm{pot}}!/(x_1!\, x_2!\, (d_{\mathrm{pot}} - x_1 - x_2)!)) \,.\, (\varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}y_1+\alpha\rho_1))^{x_1} \,.\, (\varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}y_2+\alpha\rho_2))^{x_2} \,.\, (1 - \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}(y_1+y_2) - \varepsilon\alpha)^{d_{\mathrm{pot}}-x_1-x_2}$$

$$(\text{for } \mathbf{x}, \mathbf{y} \in (0,1,\ldots,d_{\mathrm{pot}})^2;\ \boldsymbol{\theta} = (d_{\mathrm{pot}}, \varepsilon, \alpha, \rho_1, \rho_2)).$$

By standard properties of the Multinomial we thus have for t = 1,2,… and for i = 1, 2:

$$\mathrm{E}[X_i^{(t)} \mid (X_1^{(t-1)}, X_2^{(t-1)})] = d_{\mathrm{pot}}\, \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}X_i^{(t-1)} + \alpha\rho_i)$$

$$\mathrm{var}[X_i^{(t)} \mid (X_1^{(t-1)}, X_2^{(t-1)})] = d_{\mathrm{pot}}\, \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}X_i^{(t-1)} + \alpha\rho_i) \,.(1- \varepsilon((1-\alpha)d_{\mathrm{pot}}^{-1}X_i^{(t-1)} + \alpha\rho_i))$$

and

$$\mathrm{cov}[X_1^{(t)},X_2^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})] = -d_{\mathrm{pot}}\,\varepsilon^2\,\Pi_{i=1,2}\left((1-\alpha)d_{\mathrm{pot}}^{-1}X_i^{(t-1)} + \alpha\rho_i\right).$$

In order to make a connection to Example 2, we shall change our parameterisation by introducing the mean demand level $d = d_{\mathrm{pot}}\,\varepsilon$, thus writing our expressions in (d, ε) rather than $(d_{\mathrm{pot}}, \varepsilon)$. This has the advantage that we can examine the effect of changing our assumption about the level of demand variability (i.e. through the choice of value for ε), *while keeping the mean demand d constant*. Thus, below, we will examine the impact of demand variability on flow variability, under a given mean demand level. This re-parameterisation gives, after simplification:

$$\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})] = d\alpha\rho_i + \varepsilon(1-\alpha)X_i^{(t-1)}$$
$$\mathrm{var}[X_i^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})] = d\alpha\rho_i(1-\varepsilon\alpha\rho_i) + (1-\alpha)\varepsilon(1-2\varepsilon\alpha\rho_i)X_i^{(t-1)} - \varepsilon^3 d^{-1}(1-\alpha)^2(X_i^{(t-1)})^2$$
$$\mathrm{cov}[X_1^{(t)},X_2^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})] = -d\varepsilon\alpha^2\rho_1\rho_2 - \varepsilon^2\alpha(1-\alpha)\rho_2 X_1^{(t-1)} - \varepsilon^2\alpha(1-\alpha)\rho_1 X_2^{(t-1)} - d^{-1}\varepsilon^3(1-\alpha)^2 X_1^{(t-1)}X_2^{(t-1)}.$$

Note that in the special case $\varepsilon = 1$ (deterministic demand, i.e. each traveller chooses to travel each day with probability 1), the expressions for $\mathrm{E}[X_1^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})]$ and $\mathrm{var}[X_1^{(t)}|(X_1^{(t-1)}, X_2^{(t-1)})]$ above collapse to those given in section 3.2, as we would expect.

Using the statistical identities as previously:

$$\mathrm{E}[X_i^{(t)}] = \mathrm{E}[\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] = d\alpha\rho_i + \varepsilon(1-\alpha)\mathrm{E}[X_i^{(t-1)}]$$

$$\begin{aligned}\mathrm{var}[X_i^{(t)}] &= \mathrm{E}[\mathrm{var}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] + \mathrm{var}[\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]]\\ &= d\alpha\rho_i(1-\varepsilon\alpha\rho_i) + (1-\alpha)\varepsilon(1-2\varepsilon\alpha\rho_i)\,\mathrm{E}[X_i^{(t-1)}]\\ &\quad - \varepsilon^3 d^{-1}(1-\alpha)^2(\mathrm{var}[X_i^{(t-1)}] - (\mathrm{E}[(X_i^{(t-1)})])^2) + \varepsilon^2(1-\alpha)^2\,\mathrm{var}[X_i^{(t-1)}]\end{aligned}$$

$$\begin{aligned}\mathrm{cov}[X_1^{(t)}, X_2^{(t)}] &= \mathrm{E}[\mathrm{cov}[X_1^{(t)}, X_2^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]]\\ &\quad + \mathrm{cov}[\mathrm{E}[X_1^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})],\mathrm{E}[X_2^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]]\\ &= -d\varepsilon\alpha^2\rho_1\rho_2 - \varepsilon^2\alpha(1-\alpha)\rho_2\,\mathrm{E}[X_1^{(t-1)}] - \varepsilon^2\alpha(1-\alpha)\rho_1\mathrm{E}[X_2^{(t-1)}]\\ &\quad - d^{-1}\varepsilon^3(1-\alpha)^2\,(\mathrm{cov}[X_1^{(t-1)}, X_2^{(t-1)}] - \mathrm{E}[X_1^{(t-1)}]\mathrm{E}[X_2^{(t-1)}])\\ &\quad + \varepsilon^2(1-\alpha)^2\,\mathrm{cov}[X_1^{(t-1)}, X_2^{(t-1)}].\end{aligned}$$

and so we may characterise the process fully in terms of the means $\mu_i^{(t)}$ $(i = 1,2)$, variances $\omega_i^{(t)}$ $(i = 1,2)$, and covariance $\kappa^{(t)}$ as:

$$\mu_i^{(t)} = d\alpha\rho_i + \varepsilon(1-\alpha)\mu_i^{(t-1)} \qquad (i = 1,2)$$
$$\omega_i^{(t)} = d\alpha\rho_i(1-\varepsilon\alpha\rho_i) + \varepsilon(1-\alpha)(1-2\varepsilon\alpha\rho_i)\,\mu_i^{(t-1)} + \varepsilon^3 d^{-1}(1-\alpha)^2(\mu_i^{(t-1)})^2 + \varepsilon^2(1-\alpha)^2(1-\varepsilon d^{-1})\,\omega_i^{(t-1)} \quad (i = 1,2)$$
$$\kappa^{(t)} = -d\varepsilon\alpha^2\rho_1\rho_2 - \varepsilon^2\alpha(1-\alpha)\rho_2\mu_1^{(t-1)} - \varepsilon^2\alpha(1-\alpha)\rho_1\mu_2^{(t-1)} + d^{-1}\varepsilon^3(1-\alpha)^2\,\mu_1^{(t-1)}\mu_2^{(t-1)} + \varepsilon^2(1-\alpha)^2(1-\varepsilon d^{-1})\kappa^{(t-1)}.$$

Again we note that in the special case $\varepsilon = 1$, the expressions above (for $\mu_1^{(t)}$ and $\omega_1^{(t)}$) collapse to those given in section 3.2, as we would expect.

The stationary distribution is again fully characterised by $(\mu^*, \omega^*, \kappa^*)$, whose rather cumbersome expressions, given below, are not completely developed for brevity's sake:

$\mu_i^* = d\alpha\rho_i \;/\; (1 - \varepsilon\,(1-\alpha))$ $(i = 1,2)$

$(1 - \varepsilon^2(1-\alpha)^2\,(1 - \varepsilon\, d^{-1}))\ \omega_i^* = d\alpha\rho_i\,(1 - \varepsilon\alpha\rho_i)$
$+\ \varepsilon\,(1-\alpha)(1 - 2\varepsilon\alpha\rho_i)\ \mu_i^* + \varepsilon^3 d^{-1}(1-\alpha)^2\,(\mu_i^*)^2$ $(i = 1,2)$

$(1 - \varepsilon^2(1-\alpha)^2\,(1 - \varepsilon\, d^{-1}))\ \kappa^* = -d\varepsilon\alpha^2\rho_1\rho_2$
$-\varepsilon^2\alpha(1-\alpha)\rho_2\mu_1^* - \varepsilon^2\alpha(1-\alpha)\rho_1\mu_2^* + d^{-1}\varepsilon^3(1-\alpha)^2\ \mu_1^*\mu_2^*$

As in examples above, a stability analysis of the fixed-point $(\mu_1^*, \mu_2^*, \omega_1^*, \omega_2^*, \kappa^*)$ can be carried out by looking at the Jacobian matrix of the DP over the means, variances and covariance space, $\mu_i^{(t)}$, $\omega_i^{(t)}$ and $\kappa^{(t)}$. The Jacobian is clearly triangular with entries on the main diagonal independent of the state variables, thus it is easy to compute its determinant δ, and its five eigenvalues are given by the entries on the main diagonal:

$\eta_1 = \varepsilon\,(1-\alpha)$ $\eta_2 = \varepsilon\,(1-\alpha)$
$\eta_3 = \varepsilon^2(1-\alpha)^2\,(1 - \varepsilon\, d^{-1})$ $\eta_4 = \varepsilon^2(1-\alpha)^2\,(1 - \varepsilon\, d^{-1})$
$\eta_5 = \varepsilon^2(1-\alpha)^2\,(1 - \varepsilon\, d^{-1})$

As far as $\alpha \in\,]0, 1]$, $d \geq 1$, and $\varepsilon \in\,]0, 1]$ the determinant of the Jacobian is non negative and less than one, that is the system is dissipative; moreover all the eigenvalues are non negative and less than one, thus the (unique) fixed-point $(\mu_1^*, \mu_2^*, \omega_1^*, \omega_2^*, \kappa^*)$ is always (asymptotically) stable.

Same results as above are obtained if we generalise Example 2 of section 3.2 by introducing D, which denotes the random OD demand, and X, whichn as before denotes the flow on route 1, with now (D, X) our state variable, we thus have (extending the fixed demand results of section 3.2):

$D^{(t)} \mid (D^{(t-1)}, X^{(t-1)}) \sim \text{Binomial}(d_{\text{pot}}\,, \varepsilon)$
$X^{(t-1)} \mid (D^{(t-1)}, X^{(t-1)}, D^{(t)}) \sim \text{Binomial}(D^{(t)}, (1-\alpha)(X^{(t-1)}/d_{\text{pot}}) + \alpha\rho)$.

It is worth remarking that the fact that the distribution of $D^{(t)} \mid (D^{(t-1)}, X^{(t-1)})$ is independent of the previous state $(D^{(t-1)}, X^{(t-1)})$ is a special case, and is an expression of inelastic demand; thus our model is one of stochastic but inelastic demand. A quite simple generalisation would be to include a demand function in this specification, so that then demand could be both elastic and stochastic. However, in keeping with the focus of the present paper, we shall focus only on the case of inelastic, stochastic demand.

4.2 Poisson distributed stochastic demand – example 5

What if instead we had started with a Poisson demand distribution for the demand? Approaching this using the kind of formulation used for Example 4, say that now the OD demand is distributed as:

$X_1 + X_2 \sim \text{Poisson}(d)$

and that given flows of (Y_1, Y_2) yesterday:

$X_i \mid (Y_1, Y_2, X_1 + X_2) \sim \text{Binomial}(X_1 + X_2, (1-\alpha)(Y_i/d) + \alpha\rho_i)$ $(i = 1,2)$.

Then $X_1 \mid (Y_1, Y_2)$ and $X_2 \mid (Y_1, Y_2)$ are independently distributed as:

$X_i \mid (Y_1, Y_2) \sim \text{Poisson}(d((1-\alpha)(Y_i/d) + \alpha\rho_i))$ $(i = 1,2)$

i.e.:

$X_i \mid (Y_1, Y_2) \sim \text{Poisson}((1-\alpha)Y_i + d\alpha\rho_i)$ $(i = 1,2)$

with the transition function thus:

$$\phi(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) = ((1-\alpha)y_1+d\alpha\rho_1)^{x_1} ((1-\alpha)y_2 +d\alpha\rho_2)^{x_2}.\ \exp(-(1-\alpha)(y_1+y_2)-d\alpha)/(x_1!x_2!)$$
$$(\text{for } \mathbf{x}, \mathbf{y} \in \mathbb{N}^2; \boldsymbol{\theta} = (d, \alpha, \rho_1, \rho_2)).$$

Standard properties of the Poisson thus yield:

$$\mathrm{E}[X_i\,|(Y_1,Y_2)] = \mathrm{var}[X_i\,|(Y_1,Y_2)] = (1-\alpha)Y_i + d\alpha\rho_i \qquad (i = 1,2)$$

or:

$$\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})] = \mathrm{var}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})] = (1-\alpha)X_i^{(t-1)} + d\alpha\rho_i \qquad (i = 1,2)\ .$$

Applying the statistical identities used previously yields, for $i = 1,2$:

$$\begin{aligned}
\mathrm{E}[X_i^{(t)}] &= \mathrm{E}[\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] = (1-\alpha)\mathrm{E}[X_i^{(t-1)}] + d\alpha\rho_i \\
\mathrm{var}[X_i^{(t)}] &= \mathrm{E}[\mathrm{var}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] + \mathrm{var}[\mathrm{E}[X_i^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] \\
&= \mathrm{E}[(1-\alpha)X_i^{(t-1)} + d\alpha\rho_i] + \mathrm{var}[(1-\alpha)X_i^{(t-1)} + d\alpha\rho_i] \\
&= (1-\alpha)\ \mathrm{E}[X_i^{(t-1)}] + d\alpha\rho_i + (1-\alpha)^2\mathrm{var}[X_i^{(t-1)}]
\end{aligned}$$

and:

$$\begin{aligned}
\mathrm{cov}[X_1^{(t)}, X_2^{(t)}] &= \mathrm{E}[\mathrm{cov}[X_1^{(t)}, X_2^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] \\
&\quad + \mathrm{cov}[\mathrm{E}[X_1^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})],\mathrm{E}[X_2^{(t)}|(X_1^{(t-1)},X_2^{(t-1)})]] \\
&= \mathrm{E}[0] + \mathrm{cov}[(1-\alpha)X_1^{(t-1)} + d\alpha\rho_1, (1-\alpha)X_2^{(t-1)} + d\alpha\rho_2] \\
&= (1-\alpha)^2\ \mathrm{cov}[X_1^{(t-1)}, X_2^{(t-1)}]
\end{aligned}$$

and so we may characterise the process fully in terms of the means $\mu_i^{(t)}$ ($i = 1,2$), variances $\omega_i^{(t)}$ ($i = 1,2$), and covariance $\kappa^{(t)}$ as:

$$\begin{aligned}
\mu_i^{(t)} &= d\alpha\rho_i + (1-\alpha)\mu_i^{(t-1)} + && (i = 1,2) \\
\omega_i^{(t)} &= d\alpha\rho_i + (1-\alpha)\ \mu_i^{(t-1)} + (1-\alpha)^2\omega_i^{(t-1)} && (i = 1,2) \\
\kappa^{(t)} &= (1-\alpha)^2\ \kappa^{(t-1)}\ .
\end{aligned}$$

Thus, the mean flow on each route evolves exactly as in Example 2, again being separable from the evolution of the other moments. The evolution of the variance follows a different rule to that in Example 2. It is interesting to note that the model also will include a non-zero covariance term (if initialised at time $t = 0$ with a non-zero value), even though the flows are *conditionally* independent.

The stationary means, variances and covariance are given by:

$$\begin{aligned}
\mu_i^* &= d\rho_i && (i = 1,2) \\
\omega_i^* &= d\rho_i\,/\,(1 - (1-\alpha)^2) && (i = 1,2) \\
\kappa^* &= 0\ .
\end{aligned}$$

From the expressions for the stationary moments, it is clear that increasing any parameter, d, ρ_i, α, increases the stationary flow variance ω^*, as one may expect from intuition.

As in example 4 above, a stability analysis of the fixed-point (μ^*, ω^*, κ^*) can be easily carried out by looking at the Jacobian matrix of the DP, which is triangular with entries on the main diagonal independent of the state variables. It is easy to compute its determinant δ and its five eigenvalues, given by the entries on the main diagonal:

$$\begin{aligned}
\eta_1 &= (1-\alpha) & \eta_2 &= (1-\alpha) \\
\eta_3 &= (1-\alpha)^2 & \eta_4 &= (1-\alpha)^2 \\
\eta_5 &= (1-\alpha)^2
\end{aligned}$$

Thus, as far as $\alpha \in\ ]0, 1]$, $d \geq 1$, and $\varepsilon\ \in\ ]0, 1]$ the system is dissipative, and the (unique) fixed-point (μ_1^*, μ_2^*, ω_1^*, ω_2^*, κ^*) is always (asymptotically) stable.

5. Congested Networks

In sections 3 and 4 our focus was on the case of uncongested networks, as a foundation to examining dynamic process problems. In the examples given, a 'positive' feedback occurs from the decisions made on one day to the decisions made on subsequent days; it is positive in the sense that increased use of a route on one day will, in the model, lead to increased use of that route on the following day. On the other hand, in all but the last (micro-simulation) example there are 'mass effects', in that the behaviour of the group influences the behaviour of the individual (through the habitual tendency). Congestion also is a kind of mass effect, but a negative one in that increased use of a link will tend to increase its travel time, and thereby reduce its perceived attractiveness for future journeys. Further than this, it implies that some of the attributes that typically motivate route choice—such as those related to travel time—cannot be known in advance by travellers; moreover, this seems to be beyond something that can be captured by stationary probability distribution of "mis-perceptions" as in a random utility model, there is likely to be something *systematic* about the 'learning process' of information acquisition

In practice, we may acquire information from many sources, such as personal experience of travelling some route, or talking with others about their experiences, or by accessing fixed or real-time information systems. Capturing such details is beyond the scope of the present paper, and instead we utilise the simple aggregate learning processes described in the literature which make no explicit reference to how the learning was done, effectively representing some combination of all the sources mentioned to capture how travellers as a group may acquire information (e.g following some systematic change to the network). This is not intended to suggest that the techniques are limited to such approaches since they are not, the real limitation is that there is still relatively little evidence of how drivers actually acquire information, at least for developing a suitable model of such acquisition.

Unlike the case of uncongested networks, it is difficult to obtain analytic results to describe the system evolution of congested networks, even for simple two-route examples. However, it is possible *approximately* to characterise the relevant distributions, using the asymptotic results for general network structures described in Hazelton & Watling (2004) (an application of results first derived by Davis & Nihan, 1993), on which the example below draws. As in sections 3 and 4 we shall focus on the simplest possible networks, consisting of a single OD pair joined by a pair of parallel routes, and our example aims to generalise Example 1 in section 3.1 (this simplest example already turns out to be sufficiently complex to convey our main points). All of the analysis presented easily applies to the case of *n* parallel routes joining a single OD pair, if all mentions of Binomial are changed to Multinomial, and all mentions of 2 routes changed to *n*. The extension to several OD pairs is relatively straightforward, and will be discussed in a future paper.

5.1 Modelling congestion: cost functions

In order to make some progress in analytically capturing the evolution of this process, the analysis is based on an asymptotic analysis whereby we examine the behaviour of the process as the OD demand, denoted by ς, becomes large, but in a special sense. Since simply scaling the demand would not give any meaningful results, what we analyse is what happens when the demand is 'scaled' for the purposes of modelling route choice, but the scaling is reversed when it is substituted in the congestion relationships. We might think of this process, intuitively, as one in which OD demands and link capacities are scaled in tandem, if we are adopting travel cost functions whose actual argument is the ratio flow – capacity, as almost always the case in transportation network analysis. Thus $c_{i\varsigma}(x_i)$ denotes the travel cost on route 1 when the flow on route 1 is x_i under an OD demand of ς, for i =1,2 . Noting that the "real" flow (reversing the scaling) would be $\varsigma^{-1}x_i$ we are thus motivated to consider functions of the form:

$$c_{i\varsigma}(x_i) = c_i(\varsigma^{-1}x_i)$$

where $c_i(.)$ is a function independent of ς. We use $\mathbf{c}_\varsigma(\mathbf{x}) = (c_{1\varsigma}(x_1), c_{2\varsigma}(x_2))^{\mathrm{T}}$ and $\mathbf{c}(\varsigma^{-1}\mathbf{x}) = (c_1(\varsigma^{-1}x_1), c_2(\varsigma^{-1}x_2))^{\mathrm{T}}$ to denote the corresponding vector mappings.

Note that our assumption will be that the only source of randomness in the actual travel costs will be due to the randomness in flows. This is an extreme and unnecessarily restrictive assumption, and in practice there are likely to be many other unobserved sources of variation on the actual travel costs, e.g. due to weather, incidents, vehicle-mix. The model defined could be extended to represent such variations, either through postulating a probability distribution of elements of the parameters of the cost functions, and/or by assuming additional additive variation on the distribution of travel costs generated by variables flows and/or variable parameters (i.e. this would be in addition to the flow-based variation captured in the postulated model). These are important factors to consider, yet in line with the rest of the paper we neglect them here in order to obtain a simple illustrative model.

5.2 Modelling route choice behaviour: learning and choice processes

A central assumption to this Hazelton & Watling approximation method is the premise of a learning process for travellers for the measured disutility $U_i^{(t-1)}$ of each route i perceived at the end of travelling on day t – 1, that is used when making decisions for the following day t. This measured disutility is assumed to be the accumulated knowledge based on a weighted average of a finite number m of previous actual experiences, i.e. on days $t - 1$, $t - 2$, …, $t - m$, with exponentially-decreasing weights depending on the lag between the current day and the time at which the experience was had. Since we shall examine an aggregate model, the 'experience' to which we refer is the experience of the whole driver population, and so—as mentioned earlier—this process is intended to represent some combination of information sources, both direct and indirect personal experience.

The actual experiences of cost derive directly from applying the travel cost functions (as defined in section 5.1) to the flows on the relevant days. Thus, if the random variable $X_i^{(t)}$ denotes the flow on route i on day t (under the demand-scaling defined in section 5.1), then the vector random variable $\mathbf{U}^{(t-1)} = (U_1^{(t-1)}, U_2^{(t-1)})^{\mathrm{T}}$ of perceived "learnt" measured disutilities is assumed to be related to the vectors of flow random variables $\mathbf{X}^{(t-k)} = (X_1^{(t-k)}, X_2^{(t-k)})^{\mathrm{T}}$ for k =1 ,2,…,m through:

$$\mathbf{U}^{(t-1)} = (s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{c}_\varsigma(\mathbf{X}^{(t-k)})$$

where m is a given positive integer parameter, where the weighting parameter satisfies $0 < \lambda < 1$, and where $s(\lambda) = \sum_{k=1,2,\ldots,m} \lambda^{k-1} = (1-\lambda^m)/(1-\lambda)$, such that the implied weights $(s(\lambda))^{-1}\lambda^{k-1}$ (k = 1,2,…,m) are positive, decreasing (in k for any λ) and sum to unity.

We assume, generalising Example 1 of section 3.1, that conditionally on the vector of learnt costs $\mathbf{U}^{(t-1)}$ at the end of day t – 1, the fixed demand of ς travellers on an day t each choose a route independently of one another, with choice probabilities given by a random utility model $\mathbf{p}(\mathbf{U}^{(t-1)}) = (p_1(\mathbf{U}^{(t-1)}), p_2(\mathbf{U}^{(t-1)}))^{\mathrm{T}}$.

5.3 The overall model

It follows that the combination of assumptions in sections 5.1–5.2 describe an m-dependent Markov process, whereby the probability distribution of the state on any day t, as represented through the vector random variable $\mathbf{X}^{(t)}$, is fully determined by the previously-realised values of the states $\{\mathbf{X}^{(t-k)} : k = 1, 2, \ldots, m\}$. The assumptions may be summarised as:

$$\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)} \sim \text{Binomial}(\varsigma, \mathbf{p}(\mathbf{U}^{(t-1)}))$$

where

$$\mathbf{U}^{(t-1)} = (s(\lambda))^{-1} \sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{c}_{\varsigma}(\mathbf{X}^{(t-k)})$$

and where

$$s(\lambda) = (1 - \lambda^{m})/(1 - \lambda)$$

$$\mathbf{c}_{\varsigma}(\mathbf{x}) = \mathbf{c}(\varsigma^{-1}\mathbf{x})$$

for some vector of cost functions $\mathbf{c}(.)$, choice model $\mathbf{p}(.)$, OD demand ς, learning weight $0 < \lambda < 1$, and where asymptotic analysis will mean examining $\varsigma \to \infty$.

Before proceeding it is worth clarifying that $\mathbf{p}(\mathbf{U}^{(t-1)})$ as used above is describing quite a complex entity. On the one hand, $\mathbf{U}^{(t-1)}$ is a random variable that evolves based on a weighted average of a finite number of past (also random) experiences. So it is a "mean" in some sense, through the expression relating $\mathbf{U}^{(t-1)}$ to the experienced travel costs, but a "mean" that itself follows an unfolding, day-to-day varying probability distribution. When used in a random utility model, when we write $\mathbf{p}(\mathbf{U}^{(t-1)})$, then we are considering $\mathbf{U}^{(t-1)}$ as a mean in a quite different sense—there is not a unique interpretation, but a useful one to have in mind is that the distribution of $\mathbf{U}^{(t-1)}$ contains information on between-day variation, whereas the distribution we consider in random utility theory could be said to represent between-individual (inter-personal) variation (as well as other sources of randomness such as aggregation modelling errors). Thus we are presuming that the coefficient of the random utility model defined through the function $\mathbf{p}(.)$ are parameterising this inter-personal variation, relative to the inter-day variation contained in the distribution of $\mathbf{U}^{(t-1)}$. However, at the time a (conditional) choice is made all the variation in $\mathbf{U}^{(t-1)}$ is conditioned out, so that as far as the random utility model is concerned $\mathbf{p}(\mathbf{U}^{(t-1)})$ is a fixed, not random, entity. In fact, as this discussion may suggest, the theoretical connection between random utility models and stochastic process models is far from simple. This is still more the case when one considers that stochastic process models are generally aiming to capture real observable variation (in flows, travel times etc.), whereas part of the variation captured by the distribution in a random utility model will represent the *modeller's uncertainty* in understanding choice behaviour, so in this sense is not an observable phenomenon of the transportation system. This is a philosophical and technical issue that we shall leave for future research to resolve.

5.3.1 An approach to the asymptotic analysis of the stationary distribution

The specification of the model presented is especially useful, by virtue of two results. The first, as established by Cascetta (1989), noted that if the random utility model $\mathbf{p}(.)$ is such that a non-zero probability is assigned to all feasible alternatives (as satisfied by regular random utility models defined on an infinite support), then the process above has a unique stationary probability distribution to which it converges, regardless of the initial conditions. Davis & Nihan (1993) subsequently established that for the process described, this unique stationary distribution as represented by the random variable $\mathbf{X}^*$ satisfies the following limit result:

$$\varsigma^{-0.5}(\mathbf{X}^* - \mathbf{x}_{\text{SUE}}) \to_d \text{MVN}(\mathbf{0}, \mathbf{V}^*) \qquad \text{as } \varsigma \to \infty$$

for some covariance matrix $\mathbf{V}^*$, where $\to_d$ denotes convergence in distribution, where MVN denotes the multivariate normal distribution, and where $\mathbf{x}_{\text{SUE}}$ is the (assumed) unique SUE solution satisfying:

$$\mathbf{x}_{\text{SUE}} = \varsigma\, \mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}}))\ .$$

Why is this result significant? As noted by Hazelton & Watling (2004), a direct implication of this result is that if $\boldsymbol{\mu}^*$ denotes the mean of $\mathbf{X}^*$ then:

$$\varsigma^{-1}\boldsymbol{\mu}^* = \varsigma^{-1}\mathbf{x}_{\text{SUE}} + O(\varsigma^{-0.5})$$

where $f(\varsigma)$ is $O(\varsigma^n)$ if $\lim_{\varsigma\to\infty} f(\varsigma)/\varsigma^n = k < \infty$ for some finite constant k. For large ς we thus have a justification to approximate $\varsigma^{-1}\boldsymbol{\mu}^*$ by $\varsigma^{-1}\mathbf{x}_{\text{SUE}}$, since $\boldsymbol{\mu}^*$ and $\mathbf{x}_{\text{SUE}}$ both grow with ς. Expressed a different way, which might make this clearer, if $\boldsymbol{\rho}^*$ and $\boldsymbol{\rho}_{\text{SUE}}$ are respectively vectors denoting the stationary and SUE *proportions* of demand on the two routes, then the result above states that:

$$\boldsymbol{\rho}^* = \boldsymbol{\rho}_{\text{SUE}} + O(\varsigma^{-0.5})\ .$$

Hazelton & Watling went on to produce analogous results for the stationary covariance matrix $\mathbf{V}^*$, with the logic that the mean and covariance matrix were sufficient to characterise the full stationary distribution, based on Davis & Nihan's asymptotic theorem of a multivariate normal limiting distribution. The results for the covariance matrix were based on two distributional approximations, derived from Davis & Nihan's result, namely assuming $\mathbf{c}_\varsigma(.)$ and $\mathbf{p}(.)$ to be continuously differentiable:

$$\mathbf{c}_\varsigma(\mathbf{X}) = \mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}}) + \varsigma^{-1}\,\mathbf{B}\,(\mathbf{X} - \mathbf{x}_{\text{SUE}}) + O_p(\varsigma^{-0.5})$$
$$\mathbf{p}(\mathbf{U}) = \mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}})) + \mathbf{D}\,(\mathbf{U} - \mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}})) + O_p(\varsigma^{-0.5})$$

where $\mathbf{B}$ and $\mathbf{D}$ are respectively the Jacobian matrix of $\mathbf{c}_\varsigma(.)$ evaluated at $\mathbf{x}_{\text{SUE}}$ and the Jacobian matrix of $\mathbf{p}(.)$ evaluated at $\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}})$. Note that since these are statements about relationships between random variables, then so must the order notation logically be a statement about distributions. In particular we say a random variable $Y = O_p(\varsigma^n)$ if there exists an a such that $\lim_{\varsigma\to\infty} \Pr(|Y/\varsigma^n| > a) = 0$. In simple terms, this indicates that as $\varsigma \to \infty$ then we can regard the transformation $\mathbf{c}_\varsigma(\mathbf{X})$ of the random variable $\mathbf{X}$ as a *linear* transformation, given by the first order Taylor series approximation about the SUE solution.

It is worth pausing at this stage with the development, to ensure that it is clear what the expressions for $\mathbf{c}_\varsigma(\mathbf{X})$ and $\mathbf{p}(\mathbf{U})$ are aiming to convey. As an example, consider the expression for $\mathbf{p}(\mathbf{U})$. The left hand-side, namely $\mathbf{p}(\mathbf{U})$, is describing a non-linear transformation of the random variable $\mathbf{U}$, and so the result of this transformation will itself be a *random variable*. On the right-hand side, however, $\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}}))$ and $\mathbf{D}$ describe *functions* evaluated at a particular value $\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}})$, and so themselves simply return a single value. Thus, all terms in $\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}})) + \mathbf{D}\,(\mathbf{U} - \mathbf{c}_\varsigma(\mathbf{x}_{\text{SUE}}))$ except for $\mathbf{U}$ are single values, and so this expression is simply describing a linear transformation of the random variable $\mathbf{U}$, the result of which is also another random variable. The expression is, then, fundamentally an expression about the equality of random variables, i.e. that they follow the same probability distribution. The correspondence between the non-linear and linear transformations is not exact, however, and thus the $O_p(\varsigma^{-0.5})$ term captures the error in approximating the probability distribution of $\mathbf{p}(\mathbf{U})$ by the probability distribution of $\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{SUE})) + \mathbf{D}\,(\mathbf{U} - \mathbf{c}_\varsigma(\mathbf{x}_{SUE}))$.

Hazelton & Watling subsequently show, through a series of steps, that the stationary covariance matrix of the process may be related to properties of the SUE solution. It is helpful for this purpose to define a function that returns the covariance matrix of a Binomial(ς, $\mathbf{p}$) random variable:

$$\boldsymbol{\Theta}(\varsigma, \mathbf{p}) = \varsigma\,(\mathrm{diag}(\mathbf{p}) - \mathbf{p}\,\mathbf{p}^{\mathrm{T}})$$

and in particular let:

$$\boldsymbol{\Theta}_{SUE} = \boldsymbol{\Theta}(\varsigma, \mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{SUE}))\,.$$

They then show that the stationary covariance matrix $\mathbf{V}^*$ is related to $\boldsymbol{\Theta}_{SUE}$ through:

$$\varsigma^{-1}\mathbf{V}^* = \varsigma^{-1}\boldsymbol{\Theta}_{SUE} + (s(\lambda))^{-2}\,\varsigma^{-1}\,\mathbf{D}\,\{\Sigma_{i=0,1,2,\ldots\infty}\,\mathbf{M}^i\,(\mathbf{B}\boldsymbol{\Theta}_{SUE}\mathbf{B}^{\mathrm{T}})\,(\mathbf{M}^i)^{\mathrm{T}}\}\,\mathbf{D}^{\mathrm{T}} + O(\varsigma^{-0.5} + \lambda^{m-1})$$

where

$$\mathbf{M} = (s(\lambda))^{-1}\,\mathbf{BD} + \lambda\,\mathbf{I}\,.$$

This expression for $\mathbf{V}^*$, like the expression earlier for the mean ($\varsigma^{-1}\boldsymbol{\mu}^* = \varsigma^{-1}\mathbf{x}_{SUE} + O(\varsigma^{-0.5})$), indicates that in principle we can construct *relatively accurate* estimates for these moments (and hence the whole limiting distribution) based on knowledge of the SUE solution. Here 'relatively accurate' means accurate relative to the OD demand ς. Note that we cannot expect to get estimates with high absolute accuracy, since the mean and covariance matrix grow with ς.

The expression above, involving an infinite sum, does not give a practical method of constructing estimates of $\mathbf{V}^*$ however; instead they propose truncating the sum at the first two terms to yield a practical estimator of:

$$\widehat{\mathbf{V}} = \boldsymbol{\Theta}_{SUE} + (s(\lambda))^{-2}\,(\mathbf{DB}\boldsymbol{\Theta}_{SUE}(\mathbf{DB})^{\mathrm{T}} + \mathbf{DMB}\boldsymbol{\Theta}_{SUE}(\mathbf{DMB})^{\mathrm{T}})\,.$$

5.3.2 Asymptotic analysis of the dynamics of the process

The objective of the source paper for the material in sub-section 5.3.1 was, therefore, to relate the stationary distribution of the stochastic process to properties of the process (as contained in $\mathbf{D}$, $\mathbf{M}$, $\mathbf{B}$ and λ) and the SUE solution. Departing from this objective in the present section, we may also consider to what extent these theoretical results shed light on the *dynamics* of the process, not just the stationary distribution.

Now, from standard properties of the Binomial distribution we know that:

$$\mathrm{E}[\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)}] = \varsigma\,\mathbf{p}(\mathbf{U}^{(t-1)})$$

Applying the statistical identity as used previously:

$$\begin{aligned}\mathrm{E}[\mathbf{X}^{(t)}] &= \mathrm{E}[\mathrm{E}[\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)}]] \\ &= \varsigma\,\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})]\end{aligned}$$

and using the distributional approximation for $\mathbf{p}(\mathbf{U})$ in the neighbourhood of stationarity:

$$\varsigma^{-1}\,\mathrm{E}[\mathbf{X}^{(t)}] = \mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})] = \mathbf{p}(\mathbf{c}_{\varsigma}(\mathbf{x}_{\mathrm{SUE}})) + \mathbf{D}\,(\mathrm{E}[\mathbf{U}^{(t-1)}] - \mathbf{c}_{\varsigma}(\mathbf{x}_{\mathrm{SUE}})) + O(\varsigma^{-0.5})$$
$$= \varsigma^{-1}\mathbf{x}_{\mathrm{SUE}} + \mathbf{D}\,(\mathrm{E}[\mathbf{U}^{(t-1)}] - \mathbf{c}_{\varsigma}(\mathbf{x}_{\mathrm{SUE}})) + O(\varsigma^{-0.5})\ .$$

Now, also we have that:

$$\mathrm{E}[\mathbf{U}^{(t-1)}] = (s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathrm{E}[\mathbf{c}_{\varsigma}(\mathbf{X}^{(t-k)})]$$
$$= (s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, (\mathbf{c}_{\varsigma}(\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\,\mathbf{B}\,(\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5}))$$
$$= \mathbf{c}_{\varsigma}(\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\,(s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{B}\,(\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})\ .$$

Combining these expressions and denoting $\boldsymbol{\mu}^{(t)} = \mathrm{E}[\mathbf{X}^{(t)}]$ yields:

$$\varsigma^{-1}\,(\boldsymbol{\mu}^{(t)} - \mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}\,(s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{DB}\,(\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})\ .$$

Thus, asymptotically with small error relative to ς, we can relate the mean $\boldsymbol{\mu}^{(t)}$ of the process to the means $\{\boldsymbol{\mu}^{(t-k)} : k = 1,2,\ldots,m\}$ on the preceding m days, at least approximately in a neighbourhood of stationarity where Davis & Nihan's result may be assumed to approximately hold.

Note that in stationarity, $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)} = \boldsymbol{\mu}^{(t-2)} = \ldots = \boldsymbol{\mu}^{(t-m)} = \boldsymbol{\mu}^{*}$ (say), and the dynamic equations above give:

$$\varsigma^{-1}\,(\boldsymbol{\mu}^{*} - \mathbf{x}_{\mathrm{SUE}}) = \varsigma^{-1}\,(s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{DB}\,(\boldsymbol{\mu}^{*} - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})$$
$$= \varsigma^{-1}\,\mathbf{DB}\,(\boldsymbol{\mu}^{*} - \mathbf{x}_{\mathrm{SUE}})\,(s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1} + O(\varsigma^{-0.5})$$
$$= \varsigma^{-1}\,\mathbf{DB}\,(\boldsymbol{\mu}^{*} - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})$$

implying that:

$$\varsigma^{-1}\,(\mathbf{I} - \mathbf{DB})\,\boldsymbol{\mu}^{*} = \varsigma^{-1}\,(\mathbf{I} - \mathbf{DB})\,\mathbf{x}_{\mathrm{SUE}} + O(\varsigma^{-0.5})\ .$$

Now, as discussed in Cantarella et al (2010), the invertibility of $\mathbf{I} - \mathbf{DB}$ is a condition that may be adopted for assuming uniqueness of the SUE solution. In particular, it is weaker than (is implied by) assuming that the Jacobian $\mathbf{B}$ of the travel cost functions is positive definite (possibly non-symmetric) and that the choice model is a regular random utility model (meaning that $\mathbf{D}$ is a negative semi-definite symmetric matrix). Thus, under the assumption that $(\mathbf{I} - \mathbf{DB})^{-1}$ exists, we obtain:

$$\varsigma^{-1}\,\boldsymbol{\mu}^{*} = \varsigma^{-1}\,\mathbf{x}_{\mathrm{SUE}} + (\mathbf{I} - \mathbf{DB})^{-1}\,O(\varsigma^{-0.5}) = \varsigma^{-1}\,\mathbf{x}_{\mathrm{SUE}} + O(\varsigma^{-0.5})$$

i.e. the result stated earlier for stationarity, as derived in Hazelton & Watling (2004).

Let us now derive the analogous result that describes the dynamics of the covariance matrix of the process. Again, from standard properties of the Binomial distribution we know that:

$$\mathrm{var}[\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)}] = \boldsymbol{\Theta}(\varsigma, \mathbf{p}(\mathbf{U}^{(t-1)}))$$

where $\boldsymbol{\Theta}(.\,,.)$ was defined above. Applying the statistical identity as used previously:

$$\mathrm{var}[\mathbf{X}^{(t)}] = \mathrm{E}[\mathrm{var}[\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)}]] + \mathrm{var}[\mathrm{E}[\mathbf{X}^{(t)} \mid \mathbf{U}^{(t-1)}]]$$
$$= \varsigma\,(\mathrm{diag}(\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})]) - \mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})(\mathbf{p}(\mathbf{U}^{(t-1)}))^{\mathrm{T}}]) + \varsigma^{2}\,\mathrm{var}[\mathbf{p}(\mathbf{U}^{(t-1)})].$$

and therefore:

$$\varsigma^{-1}\,\mathrm{var}[\mathbf{X}^{(t)}] = \mathrm{diag}(\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})]) - \mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})(\mathbf{p}(\mathbf{U}^{(t-1)}))^{\mathrm{T}}] + \varsigma\,\mathrm{var}[\mathbf{p}(\mathbf{U}^{(t-1)})]\ .$$

Thus we require three elements, $\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})]$, $\mathrm{var}[\mathbf{p}(\mathbf{U}^{(t-1)})]$ and $\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})(\mathbf{p}(\mathbf{U}^{(t-1)}))^{\mathrm{T}}]$.

Firstly, from results we deduced earlier we have immediately that:

$$\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})] = \varsigma^{-1}\, \mathbf{x}_{\mathrm{SUE}} + \varsigma^{-1}\, (s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{DB}\, (\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})\,.$$

Secondly, for the variance we have:

$$\mathrm{var}[\mathbf{p}(\mathbf{U}^{(t-1)})] = \mathrm{var}[\mathbf{p}(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})) + \mathbf{D}(\mathbf{U}^{(t-1)} - \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))] + O(\varsigma^{-1})$$

which after some steps can be shown to be expressible as:

$$\begin{aligned}\mathrm{var}[\mathbf{p}(\mathbf{U}^{(t-1)})] &= (s(\lambda))^{-2}\, \varsigma^{-2}\, \mathbf{DB}\, (\textstyle\sum_k \lambda^{2k-2}\, \mathrm{var}[\mathbf{X}^{(t-k)}])\, (\mathbf{DB})^{\mathrm{T}} \\ &\quad + 2(s(\lambda))^{-2}\, \varsigma^{-2}\, \mathbf{DB}\, (\textstyle\sum_{j,k\,(j<k)} \lambda^{j+k-2}\, \mathrm{cov}[\mathbf{X}^{(t-j)}, \mathbf{X}^{(t-k)}])\, (\mathbf{DB})^{\mathrm{T}} + O(\varsigma^{-1})\,.\end{aligned}$$

With some similar steps of substitution and simplification, the third and final term is:

$$\begin{aligned}&\mathrm{E}[\mathbf{p}(\mathbf{U}^{(t-1)})(\mathbf{p}(\mathbf{U}^{(t-1)}))^{\mathrm{T}}] \\ &\quad = (\varsigma^{-1}\mathbf{x}_{\mathrm{SUE}} - \mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))(\varsigma^{-1}\mathbf{x}_{\mathrm{SUE}} - \mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}} + 2(\varsigma^{-1}\mathbf{x}_{\mathrm{SUE}} - \mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))\, \mathbf{D}^{\mathrm{T}}\, (\mathrm{E}[\mathbf{U}^{(t-1)}])^{\mathrm{T}} \\ &\qquad + \mathbf{D}\, \mathrm{E}[\mathbf{U}^{(t-1)}(\mathbf{U}^{(t-1)})^{\mathrm{T}}]\, \mathbf{D}^{\mathrm{T}} + O(\varsigma^{-1})\,.\end{aligned}$$

As already noted above:

$$\mathrm{E}[\mathbf{U}^{(t-1)}] = \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}(s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{B}\, (\mathrm{E}[\mathbf{X}^{(t-k)}] - \mathbf{x}_{\mathrm{SUE}}) + O(\varsigma^{-0.5})\,.$$

Furthermore,

$$\begin{aligned}\mathrm{E}[\mathbf{U}^{(t-1)}(\mathbf{U}^{(t-1)})^{\mathrm{T}}] &= \mathrm{E}[((s(\lambda))^{-1} \textstyle\sum_j \lambda^{k-1}\, \mathbf{c}_\varsigma(\mathbf{X}^{(t-j)}))((s(\lambda))^{-1} \textstyle\sum_k \lambda^{k-1}\, \mathbf{c}_\varsigma(\mathbf{X}^{(t-k)}))^{\mathrm{T}}] \\ &= (s(\lambda))^{-2} \textstyle\sum_{j,k} \lambda^{j+k-2}\, \mathrm{E}[\mathbf{c}_\varsigma(\mathbf{X}^{(t-j)})(\mathbf{c}_\varsigma(\mathbf{X}^{(t-k)}))^{\mathrm{T}})\end{aligned}$$

Then it may be shown, after some steps, that we may express:

$$\begin{aligned}&\mathrm{E}[\mathbf{c}_\varsigma(\mathbf{X}^{(t-j)})(\mathbf{c}_\varsigma(\mathbf{X}^{(t-k)}))^{\mathrm{T}})] \\ &\quad = (\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) - \varsigma^{-1}\mathbf{B}\mathbf{x}_{\mathrm{SUE}})(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) - \varsigma^{-1}\mathbf{B}\mathbf{x}_{\mathrm{SUE}})^{\mathrm{T}} + 2(\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) - \varsigma^{-1}\mathbf{B}\mathbf{x}_{\mathrm{SUE}})\, \varsigma^{-1}\, (\mathrm{E}[\mathbf{X}^{(t-k)}])^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} \\ &\qquad + \varsigma^{-2}\, \mathbf{B}\, \mathrm{cov}[\mathbf{X}^{(t-j)}, \mathbf{X}^{(t-k)}]\, \mathbf{B}^{\mathrm{T}} + \varsigma^{-2}\, \mathbf{B}\mathrm{E}[\mathbf{X}^{(t-j)}](\mathrm{E}[\mathbf{X}^{(t-k)}])^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + O(\varsigma^{-1})\,.\end{aligned}$$

If we now denote $\mathbf{V}^{(t)} = \mathrm{var}[\mathbf{X}^{(t)}]$ and $\mathbf{W}^{(s,t)} = \mathrm{cov}[\mathbf{X}^{(s)}, \mathbf{X}^{(t)}]$ (for $s < t$), then bringing together all these results then after simplification, we arrive at our final result:

$$\begin{aligned}&\varsigma^{-1}\, \mathbf{V}^{(t)} \\ &= \varsigma^{-1}\, \mathrm{diag}\, (\mathbf{x}_{\mathrm{SUE}} + \mathrm{diag}((s(\lambda))^{-1} \textstyle\sum_{k=1,2,\ldots,m} \lambda^{k-1}\, \mathbf{DB}\, (\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}})) \\ &+ \varsigma^{-1}(s(\lambda))^{-2}\mathbf{DB}(\textstyle\sum_k \lambda^{2k-2}\, \mathbf{V}^{(t-k)})(\mathbf{DB})^{\mathrm{T}} + \varsigma^{-1}2(s(\lambda))^{-2}\, \mathbf{DB}(\textstyle\sum_{j,k\,(j<k)} \lambda^{j+k-2}\, \mathbf{W}^{(t-j,\,t-k)})(\mathbf{DB})^{\mathrm{T}} \\ &- \varsigma^{-2}\, \mathbf{x}_{\mathrm{SUE}}\, (\mathbf{x}_{\mathrm{SUE}})^{\mathrm{T}} + \varsigma^{-1}\, \mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})\, (\mathbf{x}_{\mathrm{SUE}})^{\mathrm{T}} + \varsigma^{-1}\, \mathbf{x}_{\mathrm{SUE}}\, (\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}} - (\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))(\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}} \\ &- 2\varsigma^{-1}\, \mathbf{x}_{\mathrm{SUE}}\, \mathbf{D}^{\mathrm{T}}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) + 2\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})\mathbf{D}^{\mathrm{T}}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) - 2\varsigma^{-2}\mathbf{x}_{\mathrm{SUE}}\mathbf{D}^{\mathrm{T}}(s(\lambda))^{-1} \textstyle\sum_k \lambda^{k-1}\mathbf{B}(\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) \\ &+ 2\varsigma^{-1}\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})\mathbf{D}^{\mathrm{T}}\, (s(\lambda))^{-1} \textstyle\sum_k \lambda^{k-1}\, \mathbf{B}(\boldsymbol{\mu}^{(t-k)} - \mathbf{x}_{\mathrm{SUE}}) \\ &- (s(\lambda))^{-2}\, \mathbf{D}\, \Big\{ \textstyle\sum_{j,k} \lambda^{j+k-2} \Big\{ \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})\, (\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}} - \varsigma^{-1}(\mathbf{B}\, \mathbf{x}_{\mathrm{SUE}}\, (\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}} + \mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})(\mathbf{B}\mathbf{x}_{\mathrm{SUE}})^{\mathrm{T}}) \\ &\qquad + \varsigma^{-2}\, \mathbf{B}\mathbf{x}_{\mathrm{SUE}}(\mathbf{B}\mathbf{x}_{\mathrm{SUE}})^{\mathrm{T}} + \varsigma^{-1}\mathbf{B}\boldsymbol{\mu}^{(t-j)}\, (\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}) - \varsigma^{-1}\mathbf{B}\mathbf{x}_{\mathrm{SUE}}) + \varsigma^{-1}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})(\boldsymbol{\mu}^{(t-k)})^{\mathrm{T}}\, \mathbf{B}^{\mathrm{T}} \\ &\qquad - 2\varsigma^{-2}\, \mathbf{B}\, \mathbf{x}_{\mathrm{SUE}}\, (\boldsymbol{\mu}^{(t-k)})^{\mathrm{T}}\, \mathbf{B}^{\mathrm{T}} + \varsigma^{-2}\, \mathbf{B}\boldsymbol{\mu}^{(t-j)}(\boldsymbol{\mu}^{(t-k)})^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} \Big\}\Big\}\, \mathbf{D}^{\mathrm{T}} \\ &+ O(\varsigma^{-0.5})\end{aligned}$$

The key in deriving this expression is that each of the individual components in the sum involves terms that grow at the same asymptotic rate with ς (namely $\boldsymbol{\mu}^{(t)}$, $\mathbf{V}^{(t)}$, $\mathbf{W}^{(s,t)}$ and $\mathbf{x}_{\mathrm{SUE}}$) as opposed to those that by construction do not grow with ς (namely $\mathbf{D}$, $\mathbf{B}$, λ and by the special way in which the cost functions incorporate demand, also $\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})$). Thus, for example, picking one component of the sum at random: $\varsigma^{-1}\mathbf{x}_{\mathrm{SUE}}\, (\mathbf{D}\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}}))^{\mathrm{T}}$ has a term $\mathbf{x}_{\mathrm{SUE}}$ that grows with ς and so $\varsigma^{-1}\mathbf{x}_{\mathrm{SUE}}$ does not grow with ς; by construction, neither $\mathbf{D}$ nor $\mathbf{c}_\varsigma(\mathbf{x}_{\mathrm{SUE}})$

vary with ς, and so the overall terms does not vary with ς. The same applies to all the terms, except the $O(\varsigma^{-0.5})$ terms which clearly decay to zero as $\varsigma \to \infty$, hence justifying an approximation that neglects the order terms for large ς. It is noted that a term also exists involving the autocorrelations $\mathbf{W}^{(s,t)}$. In order to complete our characterisation of the process, then, we would need to derive a third set of approximating expressions, relating the autocorrelations to the moments. In principle, it seems that this may be achieved by following the same strategy as was used to derive the $\mathbf{V}^{(t)}$ limit, but an explicit derivation is beyond the scope of the present paper (justified by our focus primarily on understanding the components of variance, rather than the dynamics *per se*).

To conclude, from these expressions, we may make the following observations:

- As noted also in the series of examples in sections 3 and 4, the process for the means decouples, in the sense that we can relate the mean $\boldsymbol{\mu}^{(t)}$ of the process to the means $\{\boldsymbol{\mu}^{(t-k)} : k = 1,2,\ldots,m\}$ on the preceding m days without knowledge of further moments. However, this is directly as a result of the asymptotic approximation effectively linearising the process, and so some greater care needs to be taken in interpreting this effect; it could as well be interpreted as a sign of the weakness in the approximation used. Primarily the purpose of the approximation is to provide insights into variances.

- Overall we have defined a deterministic dynamical system in which $(\boldsymbol{\mu}^{(t)}, \mathbf{V}^{(t)})$ may be computed from $\{\boldsymbol{\mu}^{(t-k)} : k = 1,2,\ldots,m\}$, $\{\mathbf{V}^{(t-k)} : k = 1,2,\ldots,m\}$ and $\{\mathbf{W}^{(t-j,\, t-k)} : j = 1,2,\ldots, k-1;\ k = 1,2,\ldots,m\}$. This process would, in principle at least, seem tobe amenable to stability analysis as a deterministic process with state variables given by concatenated sequences (over successive periods of m days) of the moments $(\boldsymbol{\mu},\mathbf{V})$ and autocorrelations $\mathbf{W}$. This is certainly not, however, straightforward.

- Using the moment equations to solve for stationarity, whereby $\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(t-1)} = \boldsymbol{\mu}^{(t-2)} = \ldots = \boldsymbol{\mu}^{(t-m)} = \boldsymbol{\mu}^*$ and $\mathbf{V}^{(t)} = \mathbf{V}^{(t-1)} = \mathbf{V}^{(t-2)} = \ldots = \mathbf{V}^{(t-m)} = \mathbf{V}^*$ is not such an attractive option, since a complex equation would arise which would also include the unknown autocorrelations (which persist even when the process is stationary). In this case, i.e. when the interest is in the stationary distribution, the expression derived in Hazelton & Watling (2004) is more attractive since it is both explicit (for general networks) and does not require knowledge of the autocorrelations. This is achieved by making further approximating arguments; for example, it is supposed that λ is small, so the Hazelton & Watling approximation is a limit both in terms of $\varsigma \to \infty$ and $\lambda \to 0$, whereas the dynamic analysis above makes no assumption about λ.

6. Conclusions & Future Research Directions

6.1 Major findings

In this paper we have aimed to show how the relatively mature body of theoretical results on day-to-day dynamic process models may be creatively applied to problems, including those for which the developers of the theory did not envisage them being applied. We have seen that provided that we can satisfy the minimal necessary requirements—namely, defining a state-space of appropriate form in order that the Deterministic Process (DP) or Stochastic Process (SP) is markovian and time-homogeneous—then we are in a position to apply a range of results and gain insights into the nature of the transportation system under study.

The main thrust of the paper, in terms of its application, has been to examine the way in which SP models may be used to represent *in an internally consistent manner* various sources of variation, and how these sources combine and follow through to the variation in flows that we might observe on-street. To this end, several SP models have been proposed and discussed aiming at both presenting a step-by-step approach to an SP model specification, and analysing (some of) the several sources of dispersion that can be captured by the variance explicitly introduced by an SP model (in comparison with a DP model). In this way our aim has been to highlight how dispersion in the real world can effectively be modelled through SP models. This issue is relevant for project appraisal, improving our understanding of the real world and at the same time supporting more robust estimates of the effects of a project implementation.

6.2 Extensions

The reported SP instances in our paper are deliberate simple, mainly since all of them refer to a two-link network. However, the extension to general networks may be relatively straightforward, and we have provided source references in which such extensions are defined (though not investigated in the same way as for our paper). Certainly it would be interesting to expand the investigation of the present paper to more general networks, and in such cases a number of new issues would deserve attention, such as:

- the distinction between link vs. route variables is no longer irrelevant, the former generally leading to easier-to-solve models, the latter being (arguably) somehow the natural way to describe users' behaviour; and
- covariances between OD pairs which will emerge when extending the section 4 results, even though ODs are assumed conditionally independent (see Hazelton & Watling, 2004, for some consideration of this issue already);

Other issues worth of further analysis with such models include:

- the development of SP models with a real (as opposed to discrete) state-space, which is especially relevant for 'infinite' learning models, such as exponential smoothing filters; the use of such models may need to include learnt disutility as a state variable to retain the Markov assumption (see Watling & Cantarella, 2012);
- the learnt disutility may be described by a stochastic equation, analogous to the one used for flow, involving another ϕ function;
- limitations of the asymptotic theory used for the analysis in section 4, e.g. for near-periodic systems;

- problems that arise with SP models when the corresponding SUE model possesses multiple solutions (see Watling, 1996), and the subsequent difficulty in applying the asymptotic theory of section 4 (can it be applied *locally*?)

While many of the above-enlisted issues have already been partially addressing in the literature, a general framework for their analysis is still needed (possibly founded on that described in Cantarella & Cascetta, 1995).

6.3 Research perspectives

Apart from the extensions in the above sub-section, several wider issues seem worthy of further research effort, such as those enlisted below:

- More research on moment characterisations of SP models, stability issues (with respect to the DP describing the evolution of moments), and the relationship to DP models (e.g. how much of the variation in the SP is captured through a DP model of the mean?).
- Alternative methods of solution for SP models to Monte Carlo, following on from the point above, and developing the results in Hazelton & Watling (2004).
- Further links to real-life data on variation (e.g. building on work such as that in Guo & Liu, 2010), and the corresponding need for statistical estimation methods, e.g. Bayesian techniques. Now it seems we have rich enough models to hope for such an approach.

Acknowledgments. This work was partially supported by UNISA local grant ORSA091208 (financial year 2009) and ORSA118135 (financial year 2011), and by UK EPSRC grant ref. EP/I00212X/1 (2011-15). We would like to thank all three anonymous reviewers for their helpful and constructive criticisms of an earlier version.

REFERENCES

Balijepalli, N.C. and Watling, D.P. (2005). Doubly Dynamic Equilibrium Approximation Model for Dynamic Traffic Assignment. In: Transportation and Traffic Theory: Flow, Dynamics and Human Interaction (ed.: H.Mahmassani), Elsevier, Oxford, UK, 741-760.

Bie J. & Lo H.K. (2010). Stability and attraction domains of traffic equilibria in a day-to-day dynamical system formulation. Transportation Research Part B 44, 90–107.

Cantarella G.E. (2013). Day-to-day dynamic models for intelligent Transportation Systems design and appraisal. In *Transportation Research part C*, in press.

Cantarella G.E & Cascetta E. (1995). Dynamic Process and Equilibrium in Transportation Networks: Towards a Unifying Theory. Transportation Science 29, 305-329.

Cantarella G.E & Watling D.P. (2012). Modelling road traffic assignment as a dynamic process: State-of-the-art and research perspectives. Submitted for publication.

Cantarella G.E. Gentile G. & Velonà P. (2010). Uniqueness of Stochastic User Equilibrium. In: Proceedings of 5th IMA conference on Mathematics in Transportation, London, UK, April 2010.

Cantarella G. E., Velonà P., Vitetta A. (2012). **S**ignal setting with demand assignment: Global optimization with day-to-day dynamic stability constraints In Journal of Advanced Transportation, vol. 46, p. 254-268.

Cascetta E. (1987) Static and dynamic models of stochastic assignment to transportation networks. In: Flow control of congested networks (eds: Szaego G., Bianco L. & Odoni A.), Springer Verlag, Berlin.
Cascetta E. (1989). A stochastic process approach to the analysis of temporal dynamics in transportation networks. Transportation Research Part B 23, 1-17.
Chatfield C. & Goodhart G.J. (1970). The Beta-Binomial Model for Consumer Purchasing Power. Journal of the Royal Statistical Society Series C, 19(3), 240-250.
Crowder M.J. (1978). Beta-binomial Anova for Proportions. Journal of the Royal Statistical Society Series C, 27(1), 34-37.
Davis G. & Nihan N. (1993). Large population approximations of a general stochastic traffic assignment model. Operations Research 41, 169-178.
Dial R. (1971). A probabilistic multipath traffic assignment model which obviates path enumeration. Transportation Research 5(2), 83-111.
Ennis D.M. & Bi J (1998). The beta-binomial model: Accounting for inter-trial variation in replicated difference and preference tests. Journal of Sensory Studies 13(4), 389-412.
Friesz, T.L., Bernstein, D., Mehta, N.J., Tobin R.L., Ganjalizadeh (1994). Day-to-day dynamic network disequilibria and idealized traveller information systems. Operations Research 42(6), 1120–1136.
Guo X. & Liu H. (2010). Bounded rationality and irreversible network change. Transportation Research B 45, 1606-1618.
Han L. & Du L. (2012). On a link-based day-to-day traffic assignment model. Transportation Research Part B 46, 72–84.
Hazelton M.L. & Watling D.P. (2004). Computation of equilibrium distributions of Markov traffic assignment models. Transportation Science 38, 331-342.
He X. & Liu H.X. (2012). Modeling the day-to-day traffic evolution process after an unexpected network disruption. Transportation Research Part B 46, 50–71.
Horowitz J.L. (1984). The stability of stochastic equilibrium in a two-link transportation network. Transportation Research Part B 18, 13-28.
Hu T.-Y. & Mahmassani H.S. (1997). Day-to-day evolution of network flows under real-time information and reactive signal control. Transportation Research C 5, 51-69.
Lee J.C. & Sabavala D.J. (1987). Bayesian estimation and prediction for the beta-binomial model. Journal of Business & Economic Statistics 5(3), 357-367.
Sheffi Y. (1985). Urban transportation networks. Prentice Hall, New Jersey.
Smith M. J. (1984). The stability of a dynamic model of traffic assignment: An application of a method of Lyapunov. Transportation Science 18, 245-252.
Watling D.P. (1996). Asymmetric problems and stochastic process models of traffic assignment. Transportation Research Part B 30, 339-357.
Watling D.P. (1999). Stability of the stochastic equilibrium assignment problem: a dynamical systems approach. Transportation Research Part B 33, 281-312.
Watling D.P. & Cantarella G.E. (2012). Model representation and decision-making in an ever-changing world: The role of stochastic process models of transportation systems. Paper presented at 4th International Symposium on Dynamic Traffic Assignment, Martha's Vineyard, June 4th-6th 2012, submitted for publication.
Watling D, Milne D & Clark S (2012). Network Impacts of a Road Capacity Reduction: Empirical Analysis & Model Predictions. Transportation Research 46A, 167-189.
Wilcox R.R. (1981). A review of the beta-binomial model and its extensions. Journal of Educational and Behavioral Statistics 6(1), 3-32.
Yang F. & Zhang D. (2009). Day-to-day stationary link flow pattern. Transportation Research Part B 43, 119-126.

Zhang D., Nagurney A., (1996). On the local and global stability of a travel route choice adjustment process. Transportation Research B 30(4), 245–262.

Zhang, D., Nagurney A., Wu, J. (2001). On the equivalence between stationary link flow patterns and traffic network equilibria. Transportation Research 35B, 731–748.