INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH

ICER

WORKING PAPER SERIES

R. H. Mena, S. G. Walker

ON A CONSTRUCTION OF MARKOV MODELS IN CONTINUOUS TIME

Working Paper No. 25/2009

APPLIED MATHEMATICS
WORKING PAPER SERIES

ON A CONSTRUCTION OF MARKOV MODELS IN CONTINUOUS TIME

Ramsés H. Mena and Stephen G. Walker

Universidad Nacional Autónoma de México and University of Kent

Abstract

This paper studies a novel idea for constructing continuous-time stationary Markov models. The approach undertaken is based on a latent representation of the corresponding transition probabilities that conveys to appealing ways to study and simulate the dynamics of the constructed processes. Some well-known models are shown to fall within this construction shedding some light on both theoretical and applied properties. As an illustration of the capabilities of our proposal a simple estimation problem is posed.

Keywords: Gibbs sampler; Markov process; Stationary process

1. Introduction. Pitt, Chatfield and Walker (2002) introduced an approach to construct strictly stationary time series models with arbitrary but given marginal distributions. The idea goes as follows: Suppose that we wish to build up a Markovian model $\{X_n\}$ with the requirement that its marginal distribution belongs to a given parametric family, say it takes the form $\pi_X(x)$. Their approach consists of defining such a process by constructing the transition probabilities that govern it in such a way that the desired marginal remains invariant through time. Once the marginal form has been chosen, the construction of the transition probabilities is performed by imposing certain dependence through a latent variable with conditional density given by $f_{Y|X}(y \mid x)$. This conditional density is then used to construct the transition distribution, driving the process $\{X_n\}$, with transition density given in the following form

$$p(x_{n-1}, x_n) = \int f_{X|Y}(x_n \mid y)\, f_{Y|X}(y \mid x_{n-1})\, \eta_1(dy) \tag{1}$$

where

$$f_{X|Y}(x \mid y) \propto f_{Y|X}(y \mid x)\, \pi_X(x),$$

that is the posterior distribution under a likelihood based on a single observation, $f_{Y|X}$, and prior π_X. It is easy to show that $\pi_X(\cdot)$ constitutes an invariant density for the transition density (1), that is

$$\pi_X(x_n) = \int p(x_{n-1}, x_n)\pi_X(x_{n-1})\,\eta_2(dx_{n-1}). \tag{2}$$

Here, η_1 and η_2 denote certain reference measures, in practice the Lebesgue or counting measure.

The choice of the conditional density $f_{Y|X}(y \mid x)$ is quite open, and represents the main contribution to the Makovian dependence driving the constructed model. Pitt et al. (2002) used this idea to construct AR(1)-type models, in particular, they limited their choice of $f_{Y|X}(y \mid x)$ so the linear property $\mathbf{E}[X_t \mid X_{t-1} = x] = a\,x + b$ is attained. Further analysis outside this linearity property has been studied in Pitt and Walker (2005) and Mena and Walker (2007).

The main objective of this paper is to further explore this idea in the continuous time setting. In general, this leads us to consider a conditional distribution with density $f_{Y|X}$ such that the transition density resulting from (1), say $p(x_0, x_t)$, satisfies the well-known Chapman-Kolmogorov equations

$$p(x_0, x_{t+s}) = \int p(x_s, x_{t+s})\,p(x_0, x_s)\,\eta_2(dx_s). \tag{3}$$

Although it does not seem to be a general form for $f_{Y|X}$ under which the above is satisfied, we can establish some interesting results when $f_{Y|X}$ falls in some parametric families. In particular, this leads to the appealing representation (1) of transition densities corresponding to some well-known families of Markov models. Our approach consist of assigning to one of the underlying parameters, a time dimension and to examine the conditions under which the above equations are satisfied.

A Markov process constructed through the transition probability with density given by (1) clearly inherited some characteristic features of a Markov chain generated through a Gibbs sampler algorithm. In particular, all processes generated through this mechanism are *reversible*. The "latent" representation of the transition density, as given in (1), provides with an instrumental way of dealing with the law of the process which could be useful for many purposes such as the implementation of efficient estimation procedures. Clearly the nature of the state space of $\{X_t\}$, i.e. the nature of the support of π_X, and the kind of dependence induced through $f_{Y|X}$ might lead us to particular classifications of Markov processes, e.g. continuous time Markov chains and diffusion processes among others.

Describing the layout of the paper; in Section 2 we concentrate on the Gamma-Poisson model which leads, in particular, to the Cox-Ingersoll-Ross family of diffusion processes; in Section 3 we consider the normal-normal model which results in the Ornstein-Uhlenbeck class of diffusions; in Section 4 we look at the Poisson-Binomial model which leads to a type of birth-death process; in Section 5 we examine the Beta-Binomial model, where we utilise the construction of the Gamma-Poisson model to give a representation of the transition of the Wright-Fisher model; Section 6 contains a simple application derived from the proposed construction and finally, in Section 7, some discussion is provided.

2. Gamma-Poisson model. Let us start with a model contained in Pitt, Chatfield and Walker (2002) so we assume that $X_0 \sim \mathrm{Ga}(a,b)$, where Ga denotes the gamma distribution with mean a/b and $a, b > 0$. Hereafter, we will denote $\mathrm{D}(x;\theta)$ as the density/mass function corresponding to a random variable $X \sim \mathrm{D}(\theta)$. For $f_{Y|X}$, a natural choice, due to conjugacy properties, is the Poisson distribution. If $Y_1 \mid X_0 \sim \mathrm{Po}(\phi\, X_0)$ where $\phi > 0$, and consequently $X_1 \mid Y_1 \sim \mathrm{Ga}(a+Y_1, b+\phi)$ then the marginal density of X_1 also has a $\mathrm{Ga}(a,b)$ density. It is clear that X_1 is a Bayesian update of X_0 given Y_1. To proceed we take $Y_2 \mid X_1 \sim \mathrm{Po}(\phi\, X_1)$ and $X_2 \mid Y_2 \sim \mathrm{Ga}(a+Y_2, b+\phi)$, and so on. It is also clear that effectively a Gibbs sampler is being constructed based on the joint density

$$f(x,y) = \mathrm{Po}(y;\phi\, x)\, \mathrm{Ga}(x;a,b).$$

In this example the parameter ϕ controls the correlation of the process $\{X_n\}$. If ϕ is close to zero (equal to zero) then Y_1 is likely to be small (equal to zero) and so X_1 is close to (equal to) the $\mathrm{Ga}(a,b)$ density. On the other hand, if ϕ is large then so is Y_1 with high probability and so X_1 will be close to X_0 with high probability. As we mentioned in the introduction the resulting discrete time Markov process, $\{X_n\}$, enjoys all the properties of a chain generated by a Gibbs sampler with the distinctive feature that it is always on stationarity.

Following (1), we can obtain the transition density for the target process $\{X_n\}$ given by

$$\begin{aligned} p(x_{n-1},x_n) &= \sum_{y=0}^{\infty} \mathrm{Ga}(x_n; y+a, \phi+b)\, \mathrm{Po}(y; x_{n-1}\phi) \\ &= \frac{\exp\{-[\phi(x_n+x_{n-1})+bx_n]\}}{(\phi+b)^{-(a+1)/2}\,\phi^{(a-1)/2}} \\ &\times \left(\frac{x_n}{x_{n-1}}\right)^{\frac{a-1}{2}} \mathrm{I}_{a-1}\left(2\sqrt{x_n x_{n-1}\phi(\phi+b)}\right), \end{aligned} \tag{4}$$

where $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind with index ν. See Abramowitz and Stegun (1992).

The following interpretation follows; if we take n samples from the Poisson distribution then the sum is a sufficient statistic for the mean parameter and this would give us a $\mathrm{Po}(S_n X_0)$ distribution, where S_n is the sum of the n Poisson values. We can clearly generalize this to $\mathrm{Po}(\phi X_0)$ and remove the need for ϕ to be an integer. It turns out that a requirement for the process $\{X_t\}$ to be Markov and continuous in time can be attained by considering specifications of ϕ_t, namely as a function of t, that result in the Chapman-Kolmogorov equations being satisfied.

In other words, to introduce time continuous dependence in the above model is by allowing the parameter ϕ, that controls the correlation, to vary with time, so we write ϕ_t, and find the form of this function such that the resulting process, $\{X_t\}$, is still Markov and exists.

Here, the state spaces corresponding to the processes we will present, are complete and separable, hence the existence of a Markov process with the prescribed laws, can be ensured by the accomplishment of the Chapman-Kolmogorov equations. Further assumptions would be required for more general spaces. See, for example, Pollard (1984).

2.1. Choice of ϕ_t satisfying Chapman-Kolmogorov equations. If we assume that ϕ_t is a strictly positive deterministic function, the process resulting when generalizing the (4) results in a time homogeneous transition density given by

$$\begin{aligned} p(x_0, x_t) =& \sum_{y=0}^{\infty} \mathrm{Ga}(x_t; y+a, \phi_t+b)\, \mathrm{Po}(y; x_0\phi_t) \\ =& \frac{e^{-[\phi_t(x_t+x_0)+bx_t]}}{(\phi_t+b)^{-(a+1)/2}\, \phi_t^{(a-1)/2}} \left(\frac{x_t}{x_0}\right)^{\frac{a-1}{2}} \mathrm{I}_{a-1}\left(2\sqrt{x_t\, x_0 \phi_t(\phi_t+b)}\right), \end{aligned} \tag{5}$$

where $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind with index ν.

In order to see for which values of ϕ_t expression (5) satisfies the Chapman-Kolmogorov equations, it is easier to deal with the Laplace transform than with the transition density. Denote the Laplace transform of the random variable Z as $\mathcal{L}_Z(\lambda) := \mathbf{E}[e^{\lambda Z}]$, where in general $\lambda \in \mathbb{C}$, obvious restrictions for the domain of λ will follow depending of the distribution at issue. Hence, if $Z \sim \mathrm{Ga}(a, b)$, then $\mathcal{L}_Z(\lambda) = (1 - b^{-1}\lambda)^{-a}$ and if $Z \sim \mathrm{Po}(\eta)$, then $\mathcal{L}_Z(\lambda) = \exp\{\eta(e^\lambda - 1)\}$.

The Laplace transform for the transition (5) can be easily found by using the latent decomposition in the variable Y as follows

$$\begin{aligned}
\mathcal{L}_{X_t|X_0=x_0}(\lambda) &= \mathbf{E}[\mathcal{L}_{X_t|Y_t}(\lambda) \mid X_0 = x_0] \\
&= \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a} \mathcal{L}_{Y_t|X_0}\left(-\ln\left(1-(\phi_t+b)^{-1}\lambda\right)\right) \\
&= \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a} \exp\left\{\frac{x_0\,\phi_t\lambda}{\phi_t+b-\lambda}\right\}. \qquad (6)
\end{aligned}$$

Proposition 1. A stationary gamma Markov process $\{X_t\}$ defined through transition densities given by equation (5) satisfies the Chapman-Kolmogorov equations if

$$\phi_t := \frac{b}{e^{c\,t}-1}, \qquad c > 0. \tag{7}$$

Proof. In terms of Laplace transforms the Chapman-Kolmogorov equations are satisfied if the following equality holds

$$\mathbf{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0\right] = \mathcal{L}_{X_{t+s}|X_0}(\lambda). \tag{8}$$

Therefore, in this case

$$\begin{aligned}
& \mathbf{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0\right] \\
= \;& \left\{1-(\phi_t+b)^{-1}\lambda\right\}^{-a} \mathcal{L}_{X_s|X_0}\left(\frac{\phi_t\lambda}{\phi_t+b-\lambda}\right) \\
= \;& \left\{1-\frac{\lambda(\phi_t+\phi_s+b)}{(\phi_t+b)(\phi_s+b)}\right\}^{-a} \exp\left\{\frac{x_0\lambda\phi_t\phi_s}{(\phi_t+b)(\phi_s+b)-\lambda(\phi_t+\phi_s+b)}\right\}
\end{aligned}$$

which equals to $\mathcal{L}_{X_{t+s}|X_0}(\lambda)$ if and only if ϕ satisfies

$$\phi_{t+s} = \frac{\phi_t\phi_s}{\phi_t+\phi_s+b}. \tag{9}$$

Multiplying equation (9) by b and adding one in each side of the equality we obtain

$$\frac{\phi_{t+s}+b}{\phi_{t+s}} = \frac{(\phi_t+b)(\phi_s+b)}{\phi_t\phi_s}. \tag{10}$$

Now, if we define $\varphi_t := (b+\phi_t)/\phi_t$ then we get

$$\varphi_{t+s} = \varphi_t\,\varphi_s, \tag{11}$$

known as the exponential Cauchy equation, and for which positive solution is given by $\varphi_t = e^{ct}$. Applying the corresponding substitutions, we obtain the desired result. □

Although equations (5) and (6) fully characterize the law that regulates the dynamics of the constructed Markov process $\{X_t\}$, it is of interest to see whether the resulting process can be identified within a particular class of Markov processes, such as Markov chains, diffusion processes, Lévy processes, etc. In order to endeavor this classification task, the first thing we look at is the nature of the state space of the process at issue, which also match the support of the chosen stationary distribution. In the particular case of this section, the choice of a gamma distribution conveys to a Markov process with the positive real line as a state space which in turn suggests that the model might fall in the class of diffusion processes.

2.2. The Gamma-Poisson diffusion process. Given a time-homogeneous Markov process with transition density, $p_t(x_0, x_t)$, we could test whether it is a diffusion process. This can be done by verifying

$$\lim_{t \downarrow 0} \frac{1}{t} \int_{|x_t - x_0| > \varepsilon} p_t(x_0, x_t)\, \mathrm{d}x_t = 0, \tag{12}$$

for $\varepsilon > 0$. Condition (12) essentially prevents a process to have instantaneous jumps. An application of Chebyshev inequality ensures that (12) is satisfied if

$$\lim_{t \downarrow 0} \frac{1}{t}\, \mathbf{E}\{|X_t - X_0|^h \mid X_0 = x_0\} = 0, \quad \text{for } h > 2. \tag{13}$$

With this condition being satisfied the well know connection with stochastic differential equations, with drift coefficient $\mu(x)$ and diffusion coefficient $\sigma(x)$, can be established through

$$\mu(x) \;:=\; \lim_{t \downarrow 0} \frac{1}{t}\, \mathbf{E}\{X_t - X_0 \mid X_0 = x\} \tag{14}$$

$$\sigma(x) \;:=\; \lim_{t \downarrow 0} \frac{1}{t}\, \mathbf{E}\{|X_t - X_0|^2 \mid X_0 = x\} \tag{15}$$

In order to check these limits for the gamma-Poisson process, let us first define $\mathbf{E}_{x_0}(\cdot) := \mathbf{E}(\cdot \mid X_0 = x_0)$. Now note that if $Z \sim \mathrm{Ga}(a, b)$ then $\mathbf{E}[Z^j] = (a)_j / b^j$, where $(a)_j := a(a+1) \cdots (a+j-1)$ denotes the ascending factorial, also known as the Pochhammer symbol. Therefore, for the Gamma-Poisson process we have

$$\mathbf{E}_{x_0}\left[X_t^j\right] = \frac{\mathbf{E}_{x_0}\left[(y + a)_j\right]}{(b + \phi_t)^j}, \tag{16}$$

where the expectation in the right hand side is taken with respect to a $\mathrm{Po}(x_0 \phi_t)$ distribution.

Hence, in order to check condition (13), it is enough to consider $h = 4$, in which case we get

$$
\begin{aligned}
&\mathbf{E}_{x_0}\ \left(|X_t - X_0|^4\right) \qquad\qquad (17)\\
&= \quad \mathbf{E}_{x_0}(X_t^4) - 4\,x_0\,\mathbf{E}_{x_0}(X_t^3) + 6\,x_0^2\,\mathbf{E}_{x_0}(X_t^2) - 4\,x_0^3\,\mathbf{E}_{x_0}(X_t) + x_0^4\\
&= \quad \frac{\{12\,x_0^2\phi_t^2 + \phi_t[12x_0^3 b^2 - x_0^2 b(24+24a) + x_0(24+36a+12a^2)]}{(b+\phi_t)^4}\\
&+ \quad x_0^4\,b^4 - 4x_0^3 b^3 a + x_0^2 b^2(6a+6a^2) - x_0 b(8a+12a^2-4a^3) + (a)_4\}\,.
\end{aligned}
$$

Furthermore, it is easily seen that for $\phi_t = b\,(e^{\,ct}-1)^{-1}$

$$\lim_{t\downarrow 0}\frac{1}{t}(b+\phi_t)^{-4} = \lim_{t\downarrow 0}\frac{1}{t}\phi_t^2(b+\phi_t)^{-4} = \lim_{t\downarrow 0}\frac{1}{t}\phi_t(b+\phi_t)^{-4} = 0. \qquad (18)$$

Therefore, condition (13) follows. Analogously, applying (14) and (15), it can be seen that

$$\mu(x) = c\,(a/b - x) \quad \text{and} \quad \sigma(x) = \sqrt{\frac{2\,c}{b}x} \qquad (19)$$

Hence, the Gamma-Poisson process can be seen as the law of the solution to a stochastic differential equation given by

$$\mathrm{d}X_t = c\,(a/b - X_t)\,\mathrm{d}t + \sqrt{\frac{2\,c}{b}X_t}\,dW_t \qquad (20)$$

where $\{W_t\}$ denotes a Brownian motion. This process constitutes a simple reparameterization of the Cox-Ingersoll-Ross (CIR) model widely used as a model for interest rates. See Cox, Ingersoll and Ross (1985). The form of the drift function clearly identifies the reverting mean effect towards its equilibrium value a/b, which is precisely the mean of the invariant density. By making $b = 1$ and $\gamma := \sqrt{2c}$ we get the typical reparameterization of the CIR model. Note that the diffusion resulting from (20) can hit zero when $a < 1$, however this in not incompatible with the stationarity of the process. Furthermore, it is also worth noting that the proposed construction establishes directly the reversibility of the diffusion process.

2.3. A non-stationary Gamma-Poisson process. One interesting question is whether a construction of the type (5) is also available for non-stationary models. Here we present a possibility using the same construction as in the stationary Gamma-Poisson model.

Assume that instead of constructing a process with marginal distribution being $\mathrm{Ga}(a, b)$, we want a process with "marginal" measure with Lebesgue density given by $q(x) = x^{a-1}$ with $a > 0$, which is clearly not integrable on $\mathbb{R}_+$. As before, let us introduce the dependence by assuming

$Y_t \mid X_0 \sim \mathrm{Po}(x_0\, \phi_t)$. After a simple application of Bayes' theorem, it turns out that the posterior distribution is integrable on $\mathbb{R}_+$ and given by $X_t \mid Y_t \sim \mathrm{Ga}(y_t + a, \phi_t)$, in Bayesian terms, a proper posterior under an improper prior.

Following (9), the condition on ϕ that leads to Chapman-Kolmogorov equations to be satisfied is given by

$$\phi_{t+s} = \frac{\phi_t \phi_s}{\phi_t + \phi_s}, \tag{21}$$

which have positive solution when $\phi_t = 1/ct$, for $c > 0$. Hence, the process $\{X_t\}$ is defined through the conditional distributions

$$Y_t \mid X_0 \sim \mathrm{Po}(X_0/ct), \quad X_t \mid Y_t \sim \mathrm{Ga}(Y_t + a, 1/ct). \tag{22}$$

In the same way as in the gamma-Poisson model, we can verify that the constructed law corresponds to diffusion process. As before, using the moments (16), condition (13) can be verified for $h = 4$. Computing the corresponding limits in (14) and (15), it is easily seen that

$$\mu(x) = a\, c \quad \text{and} \quad \sigma(x) = \sqrt{2\, c\, x}. \tag{23}$$

Therefore the associated diffusion corresponds to the solution of a SDE given by

$$\mathrm{d}X_t = a\, c\, \mathrm{d}t + \sqrt{2\, c\, X_t}\, dW_t. \tag{24}$$

Again, a simple reparameterization leads to a well-known diffusion process. Take $a := \delta/2$ and $c := 2$, then the resulting diffusion is known as the δ-dimensional squared Bessel process, typically denoted by $\mathrm{BESQ}^\delta(x)$. The square root of this process measures the Euclidean distance of a δ-dimensional Brownian motion from the origin and plays an important role in mathematical finance; see Yor (2001). The transition density for this model has the same expression as in (5) with $b = 0$ and $\phi_t = 1/(ct)$.

3. Normal-Normal model. Here we start with

$$X_0 \sim \mathrm{N}(\mu, \tau)$$

and impose the dependence in the model through

$$Y_t \mid X_0 \sim \mathrm{N}\left(X_0, \phi_t\, \tau\right), \tag{25}$$

with $\mu \in \mathbb{R}$, $\tau, \alpha > 0$ and $\phi_t > 0$ for all $t > 0$. Once more, Bayes theorem implies that

$$X_t \mid Y_t \sim \mathrm{N}\left(\frac{Y_t + \phi_t\,\mu}{1+\phi_t}, \frac{\tau\,\phi_t}{1+\phi_t}\right). \tag{26}$$

With the conditional distributions (25) and (26) and using (1), the transition density driving the stationary process $\{X_t\}$ with Normal marginals is given by

$$p(x_0, x_t) = \mathrm{N}\left(x_t; \frac{x_0 + \phi_t\,\mu}{1+\phi_t}, \tau\,[1-(1+\phi_t)^{-2}]\right). \tag{27}$$

Hence, if $Z \sim \mathrm{N}(\mu, s)$, we obtain $\mathcal{L}_Z(\lambda) = \exp\{\lambda\mu - \lambda^2\, s/2\}$. Therefore, the Laplace transform corresponding to the transition density (27) is given by

$$\begin{aligned}\mathcal{L}_{X_t|X_0}(\lambda) &= \exp\left\{\lambda x_0(1+\phi_t)^{-1} + \lambda\mu\phi_t\,(1+\phi_t)^{-1} - \right. \\ &\times \left. \frac{\lambda^2\tau}{2}[1-(1+\phi_t)^{-2}]\right\}.\end{aligned} \tag{28}$$

The Chapman-Kolmogorov equations are satisfied if $\mathbf{E}[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0] = \mathcal{L}_{X_{t+s}|X_0}(\lambda)$. Hence we get

$$\begin{aligned}\mathbf{E}\left[\mathcal{L}_{X_{t+s}|X_s}(\lambda) \mid X_0 = x_0\right] &= \exp\left\{\frac{\lambda\,\phi_t\,\mu}{1+\phi_t} - \frac{\lambda^2\tau}{2}[1-(1+\phi_t)^{-2}]\right\} \\ &\times \ \mathcal{L}_{X_s|X_0}\left(\frac{\lambda}{1+\phi_t}\right) \\ &= \ \exp\left\{\lambda\mu\,\left[1-(1+\phi_t)^{-1}(1+\phi_s)^{-1}\right] + \lambda\,x_0(1+\phi_t)^{-1}(1+\phi_s)^{-1}\right\} \\ &\times \ \exp\left\{-\frac{\lambda^2\,\tau}{2}\left[1-(1+\phi_t)^{-2}(1+\phi_s)^{-2}\right]\right\}\end{aligned} \tag{29}$$

which equals $\mathcal{L}_{X_{t+s}|X_0}(\lambda)$ if

$$\phi_{t+s} = \phi_t\,\phi_s + \phi_t + \phi_s. \tag{30}$$

The functional equation (30) arises frequently in probability theory and its positive solution is given by $\phi_t = e^{\alpha\,t} - 1$, $\alpha > 0$. See Aczél and Dhombres (1989). Hence, the transition density (27) can be rewritten as

$$p_t(x_0, x_t) = \mathrm{N}\left(x_t; x_0\,e^{-\alpha t} + \mu\,(1-e^{-\alpha t}), \tau\,[1-e^{-2\alpha t}]\right). \tag{31}$$

In the same way as in Section 2.2, we can verify condition (13) to see whether a diffusion process can be associated with the transition density given by (31). It turns out that in this case we have

$$\begin{aligned}\mathbf{E}_{x_0}\ \left(|X_t - X_0|^4\right) &= (x_0-\mu)^4\,(e^{-\alpha t}-1)^4 + 3\tau^2(e^{-2\alpha t}-1)^2 \\ + &\quad (2e^{-3\alpha t} - 2e^{-\alpha t} - e^{-4\alpha t} + 1)(6\tau x^2 - 12x\mu\tau + 6\mu^2\tau),\end{aligned}$$

from which is seen that

$$\lim_{t\downarrow 0}\frac{1}{t}\mathbf{E}_{x_0}\left(|X_t - X_0|^4\right) = 0 \tag{32}$$

In the same manner, limits (14) and (15) can be obtained to get the drift and diffusion coefficients, yielding to a diffusion process solution of a SDE given by

$$\mathrm{d}X_t = -\alpha(x-\mu)\,\mathrm{d}t + \sqrt{2\tau\alpha}\,\mathrm{d}W_t. \tag{33}$$

If we put $\tau = \sigma^2/2\alpha$ we obtain the SDE

$$\mathrm{d}X_t = -\alpha(x-\mu)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \tag{34}$$

known as the mean reverting Ornstein-Uhlenbeck model.

4. Poisson-Binomial model. Following the approach in the previous sections, let us assume that we want to construct a Markov chain in continuous case with a Poisson stationary distribution. With this in mind let us choose $\pi_X(x,\lambda) = \mathrm{Po}(x;\lambda e^\theta)$, serving as the stationary distribution of the process to be constructed. Having set the stationary behavior, the corresponding conditional density is chosen by

$$f_{Y|X}(y \mid x;\xi) = \mathrm{bin}(y;x,\xi) = \binom{x}{y}(1-\xi)^{x-y}\;(\xi)^y\;I_{\{0,\ldots,x\}}(y),$$

where $0<\xi<1$. After an application of Bayes theorem, we get

$$f_{X|Y}(x \mid y;\xi) = \frac{[(1-\xi)\lambda]^{x-y}}{(x-y)!}\,\exp\left\{(x-y)\theta - \lambda e^\theta(1-\xi)\right\} I_{\{y,\ldots,\infty\}}(x).$$

Hence, as before, the idea is to construct a continuous time stationary process $X = \{X_t; t\in\mathbb{R}_+\}$ by introducing a latent process $Y = \{Y_t; t\in\mathbb{R}_+\}$ via the updating mechanism

$$\begin{aligned}\{Y_t \mid X_0 = x_0\} &\sim f_{Y|X}(\cdot \mid x_0;\xi_t)\\ \{X_t \mid Y_t = y_t\} &\sim f_{X|Y}(\cdot \mid y_t;\xi_t),\end{aligned}$$

where ξ_t is a function in $(0,1)$ which has to satisfy Chapman-Kolmogorov equations (3).

In this case the transition probability (1) has mass probability function given by

$$
\begin{aligned}
& p\left(x_0, x_t\right)=\sum_{y=0}^{\infty} f_{X \mid Y}\left(x_t \mid y ; \xi_t\right) f_{Y \mid X}\left(y \mid x_0 ; \xi_t\right) \\
= & \frac{\left[\lambda e^\theta\left(1-\xi_t\right)\right]^{x_t} \exp \left\{-\lambda e^\theta\left(1-\xi_t\right)\right\}\left(1-\xi_t\right)^{x_0}}{x_t!} \\
\times & \sum_{y=0}^{x_0 \wedge x_t} \frac{\left[\frac{\xi_t}{\lambda e^\theta\left(1-\xi_t\right)^2}\right]^y x_0!x_t!}{y!\left(x_t-y\right)!\left(x_0-y\right)!} \\
= & \frac{\left[\lambda e^\theta\left(1-\xi_t\right)\right]^{x_t} \exp \left\{-\lambda e^\theta\left(1-\xi_t\right)\right\}\left(1-\xi_t\right)^{x_0}}{x_t!} \\
\times & { }_2 F_0\left(-x_0,-x_t, \frac{\xi_t}{\lambda e^\theta\left(1-\xi_t\right)^2}\right) \\
= & \operatorname{Po}\left(x_t ; \lambda e^\theta\left(1-\xi_t\right)\right)\left(1-\xi_t\right)^{x_0}{ }_2 F_0\left(-x_0,-x_t, \frac{\xi_t}{\lambda e^\theta\left(1-\xi_t\right)^2}\right)
\end{aligned} \tag{35}
$$

where $a \wedge b$ stands for $\min\{a, b\}$ and ${}_2F_0()$ is a generalized hypergeometric function, see Abramowitz and Stegun (1992), formulas 15.4.1 and 15.4.2. For this expression, we have used the relation $1/(x-y)! = (-1)^y(-x)_y/x!$.

As in Section 2.1, we can proceed to find the form of ξ_t such that Chapman-Kolmogorov equations are satisfied. It is easily seen that

$$
\begin{aligned}
\mathcal{L}_{Y \mid X=x}(\psi) & =\left\{1-\xi_t+\xi_t e^\psi\right\}^x \quad \text{and} \\
\mathcal{L}_{X \mid Y=y}(\psi) & =e^{y \psi} \exp \left\{\lambda e^\theta\left(1-\xi_t\right)\left(e^\psi-1\right)\right\} .
\end{aligned}
$$

Therefore the Laplace transform corresponding to the transition mass function (35) can be computed as

$$
\begin{aligned}
\mathcal{L}_{X_t \mid X_0=x}(\psi) & =\mathbf{E}\left[\mathcal{L}_{X_t \mid Y_t}(\psi) \mid X_0=x\right] \\
& =\exp \left\{\lambda e^\theta\left(1-\xi_t\right)\left(e^\psi-1\right)\right\} \mathcal{L}_{Y_t \mid X_0=x}(\psi) \\
& =\exp \left\{\lambda e^\theta\left(1-\xi_t\right)\left(e^\psi-1\right)\right\}\left[1-\xi_t+\xi_t e^\psi\right]^{x_0} .
\end{aligned} \tag{36}
$$

As before, using the Laplace transform (36), to satisfy the Chapman-Kolmogorov equations is equivalent to satisfy

$$
\mathbf{E}\left[\mathcal{L}_{X_{t+s} \mid X_s}(\psi) \mid X_0=x\right]=\mathcal{L}_{X_{t+s} \mid X_0=x}(\psi). \tag{37}
$$

Hence, we have

$$
\begin{aligned}
& \mathbf{E}\ \left[\mathcal{L}_{X_{t+s}|X_s}(\psi) \mid X_0 = x\right] \\
= \ & \exp\{\lambda e^{\theta}(1-\xi_t)(e^{\psi}-1)\}\, \mathcal{L}_{X_s|X_0=x}(\widehat{\psi}) \\
= \ & \exp\{\lambda e^{\theta}(1-\xi_t\xi_s)(e^{\psi}-1)\}\left[1-\xi_t\xi_s+\xi_t\xi_s e^{\psi}\right]^{x_0}, \qquad (38)
\end{aligned}
$$

where $\widehat{\psi} = \log\left(1-\xi_t+\xi_t\, e^{\psi}\right)$ The above quantity equals $\mathcal{L}_{X_{t+s}|X_0}(\psi)$ if and only if $\xi_{t+s} = \xi_t\xi_s$ which, once more, leads to $\xi_t = e^{C\,t}$. Restricting such a solution to the constrain $0 < \xi_t < 1$ implies $\xi_t = e^{-\alpha t}$ with $\alpha > 0$.

4.1. The Poisson-Binomial continuous time Markov process. Just as in the Gamma-Poisson example, the Poisson-Binomial model presented in the previous section can be identified within a well-known class of Markov chain model.

In fact, for this model the infinitesimal generator $Q = \{q_{ij}\}$ of the semigroup $P_t = \{p_t(i,j)\}$ is given by

$$
q_{ij} \;=\; \begin{cases} -\lim_{t\downarrow 0} \frac{1-p_t(i,i)}{t}, & j = i \\ \lim_{t\downarrow 0} \frac{p_t(i,j)}{t}, & i \neq j \end{cases} = \begin{cases} -\alpha(i+\lambda e^{\theta}), & j = i \\ \alpha\lambda e^{\theta}, & j = i+1 \\ i\alpha, & j = i-1 \\ 0, & \text{otherwise}, \end{cases}
$$

where $p_t(i,j)$ is given as in (35) with $\xi_t = e^{-\alpha t}$. It can be seen that in this case the above process is strong Markov, since all the states are stable, that is $0 \le -q_{ij} < \infty$. This infinitesimal generator is immediately recognized as that corresponding to a conservative birth and death process with birth rate $\alpha\lambda e^{\theta}$ and death rate $i\alpha$. If we set $\theta = 0$ we obtain a stationary model with $\mathrm{Po}(\lambda)$ marginal distributions.

5. Beta-Binomial model. A model particularly appealing in genetical applications arises from the following construction. First, let us notice that if $X_i \stackrel{\text{ind}}{\sim} \mathrm{Ga}(a_i, 1)$, $i = 1, 2$ then $\mathsf{P} = X_1/(X_1 + X_2) \sim \mathrm{Be}(a_1, a_2)$ and is independent of $X = X_1 + X_2 \sim \mathrm{Ga}(a_1 + a_2, 1)$. Second, it is well known that if $Y_i \stackrel{\text{ind}}{\sim} \mathrm{Po}(\phi X_i)$, $i = 1, 2$, then $\{Y_1 \mid Y\} \sim \mathrm{Bin}(Y, X_1/(X_1 + X_2))$, where $Y := Y_1 + Y_2$.

Hence, following the construction of Section 2, we now construct two Gamma–Poisson processes by means of the following conditional representations:

$$\{X_{jt} \mid Y_{jt}\} \sim \mathrm{Ga}(a_j + Y_{jt}, 1 + \phi_t) \quad \text{and} \quad \{Y_{jt} \mid X_{j0}\} \sim \mathrm{Po}(\phi_t \, X_{j0})$$

for $j = 1, 2$ and define the process $\{\mathsf{P}_t\}$ through the following transformation

$$\mathsf{P}_t := \frac{X_{1t}}{X_{1t} + X_{2t}}. \tag{39}$$

Therefore, $\{\mathsf{P}_t \mid Y_{1t}, Y_t\} \sim \mathrm{Be}(a_1 + Y_{1t}, a - a_1 + Y_t - Y_{1t})$, where $Y_t = Y_{1t} + Y_{2t}$. In the notation of Section 2 we have set $b = 1$, and therefore $\phi_t = (e^{\,ct} - 1)^{-1}$. Conditioning on the event $\{Y_t = m\}$ we can construct the transition density driving this process as follows:

$$\begin{aligned}
& p\,(\mathsf{p}_t \mid \mathsf{p}_0, Y_t = m) = \sum_{k=0}^{m} \mathrm{Be}(\mathsf{p}_t; a_1 + k, a - a_1 + m - k)\, \mathrm{Bin}(k; m, \mathsf{p}_0) \\
= \quad & \Gamma(a + m)[(1 - \mathsf{p}_t)(1 - \mathsf{p}_0)]^m \, \mathsf{p}_t^{a_1 - 1} \, (1 - \mathsf{p}_t)^{a_2 - 1} \\
\times \quad & \sum_{k=0}^{m} \frac{\varrho^k \binom{m}{k}}{\Gamma(a_1 + k)\Gamma(a_2 + m - k)} \\
= \quad & \frac{\Gamma(a + m)[(1 - \mathsf{p}_t)(1 - \mathsf{p}_0)]^m \, \mathsf{p}_t^{a_1 - 1} \, (1 - \mathsf{p}_t)^{a_2 - 1}}{\Gamma(a_1)\Gamma(a_2 + m)} \\
\times \quad & {}_2F_1(-m, -a_2 - m + 1; a_1; \varrho) \, d\mathsf{p}_t,
\end{aligned} \tag{40}$$

where

$${}_2F_1(a, b; c; z) = \sum_{n=0}^{\infty} \frac{(a)_n \, (b)_n}{(c)_n} \frac{z^n}{n!}$$

denotes the Gauss hypergeometric function, $\varrho := \mathsf{p}_t \, \mathsf{p}_0 \, [(1 - \mathsf{p}_t)(1 - \mathsf{p}_0)]^{-1}$ and $a := a_1 + a_2$. On the other hand, we have

$$\begin{aligned}
\Pr(Y_t = m) \quad &= \quad \mathbf{E}_{X_0} \{\mathrm{Po}(y_t; \phi_t X_0)\} = \mathbf{E}_{X_0} \left\{ e^{-\phi_t X_0} \frac{(\phi_t X_0)^m}{m!} \right\} \\
&= \quad \frac{\phi_t^m}{m!} \frac{(a)_m}{(1 + \phi_t)^{a+m}},
\end{aligned} \tag{41}$$

since $X_0 \sim \mathrm{Ga}(a, 1)$. Therefore the transition density for the process $\{\mathsf{p}_t\}$ is given by

$$\begin{aligned}
& p\,(\mathsf{p}_t \mid \mathsf{p}_0) = \mathbf{E}_{Y_t} \left\{ \sum_{k=0}^{Y_t} \mathrm{Be}(\mathsf{p}_t; a_1 + k, a - a_1 + Y_t - k)\, \mathrm{Bin}(k; Y_t, \mathsf{p}_0) \right\} \\
= \; & \sum_{m=0}^{\infty} \left\{ \sum_{k=0}^{Y_t} \mathrm{Be}(\mathsf{p}_t; a_1 + k, a - a_1 + Y_t - k)\, \mathrm{Bin}(k; Y_t, \mathsf{p}_0) \right\} \Pr(Y_t = m),
\end{aligned}$$

which using (40) and substituting $\phi_t = (e^{c\,t} - 1)^{-1}$ in (41) leads to the transition density

$$p\,(\mathsf{p}_t \mid \mathsf{p}_0) = \frac{\mathsf{p}_t^{a_1-1}(1-\mathsf{p}_t)^{a_2-1}(1-e^{-ct})^a}{\Gamma(a)\Gamma(a_1)}$$
$$\times \sum_{m=0}^{\infty} \frac{\Gamma(a+m)^2}{\Gamma(a_2+m)m!}[e^{-ct}(1-\mathsf{p}_t)(1-\mathsf{p}_0)]^m {}_2F_1(-m, -a_2-m+1; a_1; \varrho). \tag{42}$$

Hence, we have constructed a stationary process $\{\mathsf{p}_t\}$ with $\mathrm{Be}(a_1, a_2)$ marginal distributions.

5.1. The Beta-Binomial diffusion process. As before, we can associate a diffusion process to the model described in the previous section. If we notice that for $\mathsf{p} \sim \mathrm{Be}(a, b)$ we have the moments $\mathbf{E}[\mathsf{p}^j] = (a)_j/(a+b)_j$ and therefore

$$\begin{aligned} \mathbf{E}\left[\mathsf{P}_t^j \mid \mathsf{P}_0 = \mathsf{p}_0, Y_t = m\right] &= \frac{1}{(a+m)_j}\mathbf{E}\left[(a_1 + Y_{1t})_j \mid \mathsf{P}_0 = \mathsf{p}_0, Y_t = m\right] \\ &= \frac{1}{(a+m)_j}\sum_{k=0}^{m} \mathrm{Bin}(k, m, \mathsf{p}_0)(a_1+k)_j. \end{aligned}$$

Now, substituting $\phi_t = (e^{c\,t} - 1)^{-1}$ in (41) we get

$$\Pr(Y_t = m) = \frac{(a)_m}{m!}\,(1-e^{-ct})^a\,e^{-m\,c\,t}.$$

Therefore the unconditioned moments are given by

$$\sum_{m=0}^{\infty} \mathbf{E}\left[\mathsf{P}_t^j \mid \mathsf{P}_0 = \mathsf{p}_0, Y_t = m\right]\Pr(Y_t = m).$$

In particular, we can see that

$$\mathbf{E}[(\mathsf{P}_t - \mathsf{P}_0) \mid \mathsf{P}_0 = \mathsf{p}_0] = \frac{(1-e^{-ct})^a\,(a_1 - a\mathsf{p}_0)\,{}_2F_1(a, a; a+1; e^{-ct})}{a}.$$

Hence, following (14) and assuming that $a > 1$ we get

$$\mu(\mathsf{p}) = \frac{c}{a-1}(a_1 - a\mathsf{p}).$$

Analogously, we could get

$$\sigma^2(\mathsf{p}) = \frac{2c}{a-1}\,\mathsf{p}(1-\mathsf{p}).$$

and verify that condition (13) is satisfied. Therefore, the associated diffusion process can be seen as the solution to the SDE given by

$$\mathrm{d}\mathsf{P}_t = \frac{c}{a-1}(a_1 - a\mathsf{P}_t)\mathrm{d}t + \sqrt{\frac{2c}{a-1}\,\mathsf{P}_t(1-\mathsf{P}_t)}\,\mathrm{d}W_t.$$

If we put $c = (a-1)/2$ then we get

$$\mathrm{d}\mathsf{P}_t = \frac{1}{2}(a_1 - a\mathsf{P}_t)\mathrm{d}t + \sqrt{\mathsf{P}_t(1-\mathsf{P}_t)}\,\mathrm{d}W_t,$$

which is known as the reversible mutation Wright-Fisher diffusion model or a particular case of the Jacobi SDE.

6. Estimation example. In this section we simply exemplify a potential application of the latent representation of the transition, in the context of estimation. If a tractable analytic expression for the transition density $p_t(x_0, x_t)$ is available, then for a given data set $\mathbf{x} = (x_{t_1}, \ldots, x_{t_T})$ for $t_1 \leq t_2 \leq \cdots \leq t_T$ we could compute

$$L_{\mathbf{x}}(\theta) = q_X^\theta(x_{t_1}) \prod_{i=1}^{T-1} p_{(t_{i+1}-t_i)}(x_{t_i}, x_{t_{i+1}}), \tag{43}$$

that could be used for instance in maximum likelihood or Bayesian estimation methods.

However, closed expressions for (1), are not always available or they might be hard to compute, therefore as an alternative approach one could resort to the augmented likelihood

$$L_{\mathbf{x},\mathbf{y}}^{aug}(\theta) = q_X^\theta(x_{t_1}) \prod_{i=1}^{T-1} f_{X|Y}^\theta(x_{t_{i+1}} \mid y_{t_{i+1}})\, f_{Y|X}^\theta(y_{t_{i+1}} \mid x_{t_i}), \tag{44}$$

and, for instance, proceed through an expectation-maximization algorithm, that in order to maximise (43) computes iteratively a sequence $\theta_1, \ldots, \theta_j, \ldots$ converging to the maximum value of (43), with the following two steps

- E-step. For given data set $\mathbf{x}$ and current parameter value θ_j, compute the following expectation

$$Q(\theta \mid \theta_{(j)}, \mathbf{x}) = \mathbf{E}_{\theta_{(j)}}\left[\log L_{\mathbf{x},\mathbf{y}}^{aug}(\theta)\right], \tag{45}$$

  where the expectation $\mathbf{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $\mathbf{F}_{\mathbf{Y}|\mathbf{X}}^{\theta_{(j)}}$.

- M-step. Maximise $Q(\theta \mid \theta_{(j)}, \mathbf{x})$ in θ and define

$$\theta_{(j+1)} = \arg\max_\theta Q(\theta \mid \theta_{(j)}, \mathbf{x}). \tag{46}$$

The EM iterations satisfy

$$Q(\theta_{(j+1)} \mid \theta_{(j)}, \mathbf{x}) \geq Q(\theta_{(j)} \mid \theta_{(j)}, \mathbf{x}), \tag{47}$$

which implies that the sequence θ_j is always moving towards the maximum.

We can write

$$\mathbf{F}^{\theta}_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} \mid \mathbf{x}) \propto \prod_{t=1}^{T-1} f^{\theta}_{X|Y}(x_{t_{i+1}} \mid y_{t_{i+1}})\, f^{\theta}_{Y|X}(y_{t_{i+1}} \mid x_{t_i}).$$

Furthermore, if the E-step is not easy to evaluate, then, under the assumption that a set $\mathbf{y}$ of latent random numbers is easily simulated from $\mathcal{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$, we can proceed from a Monte Carlo point of view and approximate Q as follows

$$\hat{Q}(\theta \mid \theta_{(j)}, \mathbf{x}) = \frac{1}{m}\sum_{k=1}^{m} \log(L^{aug}_{\mathbf{x},\mathbf{y}^{(k)}}(\theta)), \tag{48}$$

where $\mathbf{y}^{(k)} \sim \mathcal{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$. See Tanner and Wong (1987) and Wei and Tanner (1990). Due to the independence structure underlying this construction, we can simulate each component of $\mathbf{y}^{(k)}$ individually. That is, for a given k we can simulate $y^{(k)}$ by sampling individually each $y_{t_{i+1}}$ from a distribution with density

$$f(y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}) \propto f^{\theta}_{X|Y}(x_{t_{i+1}} \mid y_{t_{i+1}})\, f^{\theta}_{Y|X}(y_{t_{i+1}} \mid x_{t_i}) \tag{49}$$

for $i = 1, \ldots, T-1$.

Let us consider the Gamma-Poisson diffusion model (20), in this case $\theta = (a, b, c)$, for which an analytical maximum likelihood estimator is not available. Hence we could alternatively use the estimator (48) where

$$f(y_{t_{i+1}} \mid x_{t_{i+1}}, x_{t_i}) \quad \propto \quad \mathrm{Ga}(x_{t_{i+1}}; y_{t_{i+1}} + a, \phi(\tau_{i+1}) + b)\, \mathrm{Po}(y_{t_{i+1}}; \phi(\tau_{i+1}) x_{t_i})$$

If we take $\rho = e^{-c}$ and $a = 1$ then the augmented likelihood is given by

$$L^{aug}_{\mathbf{x},\mathbf{y}}(\theta) = \mathrm{Ga}(x_{t_1}; 1, b) \prod_{i=1}^{T-1} \mathrm{Ga}\left(x_{t_{i+1}}; y_{t_{i+1}} + 1, \phi(\tau_{i+1}) + b\right)\, \mathrm{Po}(y_{t_{i+1}}; \phi(\tau_{i+1}) x_{t_i}),$$

with corresponding scores given by

$$\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial b} = \frac{1}{b}\left(T + 2\sum_{i=1}^{T-1} y_{t_{i+1}}\right) - x_{t_1} - \sum_{i=1}^{T-1} \frac{\rho^{\tau_{i+1}} x_{t_i} + x_{t_{i+1}}}{1 - \rho^{\tau_{i+1}}} \tag{50}$$

and

$$\frac{\partial l^{aug}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \rho} = \sum_{i=1}^{T-1} \; \frac{(y_{t_{i+1}}+1)\tau_{i+1}\rho^{\tau_{i+1}-1}}{1-\rho^{\tau_{i+1}}} - \frac{bx_{t_{i+1}}\tau_{i+1}\rho^{\tau_{i+1}-1}}{(1-\rho^{\tau_{i+1}})^2}$$
$$+ \quad y_{t_{i+1}}\left(\frac{\tau_{i+1}}{\rho} + \frac{\tau_{i+1}\rho^{\tau_{i+1}-1}}{1-\rho^{\tau_{i+1}}}\right) - \frac{bx_{t_i}\tau_{i+1}\rho^{\tau_{i+1}-1}}{(1-\rho^{\tau_{i+1}})^2}. \tag{51}$$

Equating expression (50) to zero and solving for b we get the following estimator

$$\hat{b}_{aug} = \frac{T + 2\sum\limits_{i=1}^{T-1} y_{t_{i+1}}}{x_{t_1} + \sum\limits_{i=1}^{T-1} \frac{x_{t_{i+1}}+x_{t_i}}{1-\rho^{\tau_{i+1}}}}.$$

In general, the estimator for ρ based on the augmented likelihood, is not directly available. However, if we assume that the observations $\mathbf{x}$ are uniformly spaced, $\tau_i = 1$ for all $i = 2, \ldots, T$, then an estimate for ρ is given by solving the following quadratic equation

$$\rho^2\left(1 - T - \sum_{i=1}^{T-1} y_{t_{i+1}}\right) - \rho\left[b\left\{\sum_{i=1}^{T-1}(x_{t_{i+1}} + x_{t_i})\right\} - T + 1\right] + \sum_{i=1}^{T-1} y_{t_{i+1}} = 0,$$

At this point we have, at least, two alternatives to MLE in order to estimate the parameters in the Poisson-gamma stationary model: a MCEM scheme, in which we need to simulate from the latent vector $\mathbf{Y} \mid \mathbf{X}$ or an EM method where the E-step is obtained analytically.

Let us consider two data sets, $\mathbf{x}$ and $\mathbf{x}^{\tau}$, of size T=1000, simulated from transition (5). This can be easily done by first simulating a $\{y_{t_i+1} \mid \; x_{\tau_i}\} \sim \mathrm{Po}(x_{\tau_i}\phi(\tau_{i+1}-\tau_i))$ and then $\{x_{t_i+1} \mid y_{t_i+1}\} \sim \mathrm{Ga}(y_{t_i+1}+a, \phi(\tau_{i+1}-\tau_i)) + b)$ with $a = 1, b = 3$ and $\rho = 0.7$. For this latter specification of the a parameter we do not have a strictly positive diffusion process. For the data set $\mathbf{x}$, we assumed equally spaced data, that is $\tau_i = 1$ for all $i = 2, \ldots, T$. For the data set $\mathbf{x}^{\tau}$, the data were generated at exponential times with intensity parameter $\lambda = 0.5$. The simulated data together with their corresponding ACF's are displayed in Figure 1.

Table 1 shows the behavior of the above estimate as the sample size m increases. We observed that only a few simulations are required in order to get a relatively good estimation. With simulations larger than 30, the resulting estimates, obtained from the MCEM, did not show a significant improvement. Being the latter our main objective, we initially implemented the MCEM fixing $m = 30$.

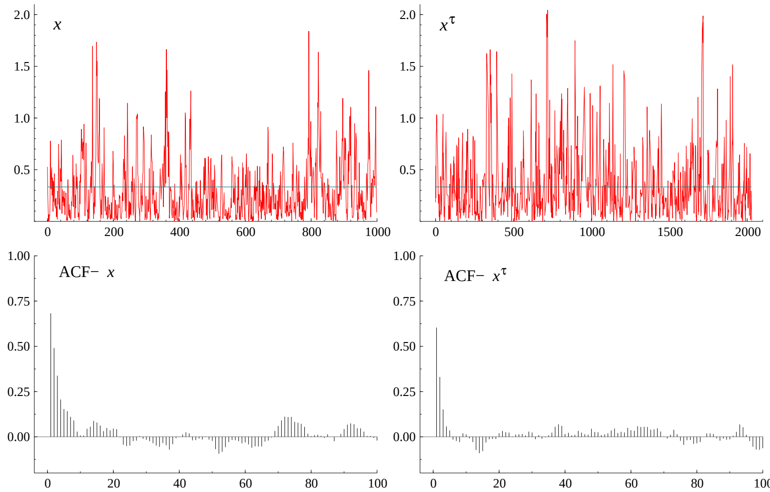

Figure 1: Simulated data and respective ACF's from the Poisson-gamma model with parameters $a = 1, b = 3$ and $\rho = 0.7$. The simulations were performed under two schemes: $\mathbf{x}$ denotes an equally-spaced sample and $\mathbf{x}^\tau$ denotes an exponentially-distributed sample (with intensity parameter $\lambda = 0.5$).

Tables 2 and 3 show the MCEM iterations for the uniformly spaced data set $\mathbf{x}$ and the exponentially spaced data set $\mathbf{x}^\tau$ respectively. It is worth noticing that the scale parameter, b, is not fully recovered, but this might be due to the inefficiency of having only one trajectory for inference purposes. For the randomly spaced data set the parameter ρ, which represents the correlation of the model, is not as close to the theoretical value as in the case of the uniformly spaced data set. This is mainly due to the fact that we have ignored the randomness of the time-gap between observations. Ways to correct this issue are studied in Yacine and Mykland (2003).

7. Discussion. The construction presented here allows for a nice statistical interpretation of a class of continuous time Markov processes. It allows for either a direct or latent availability of the transition density, which can be used in the understanding, study, estimation and construction of new models. Of particular interest is the representation of transition densities corresponding to diffusion processes, where estimation procedures are not always based on the likelihood due to the unavailability and/or intractability of the transitions.

Acknowledgements. R.H. Mena is grateful for the support from the International Centre for Economic Research (ICER), Turin, Italy. R.H. Mena and S.G. Walker also acknowledge partial

Simulations (m)	$\hat{Q}(\theta \mid \theta, \mathbf{x})$	$\hat{Q}(\theta \mid \theta, \mathbf{x}^{\tau})$
(Model)	-731.49	-787.71
1	-688.53	-921.40
5	-711.89	-590.12
10	-708.63	-723.58
30	-717.68	-782.63
100	-721.03	-781.31
1000	-731.94	-786.10

Table 1: Monte Carlo approximation for the E-step, evaluated at the true parameter-value.

support from CONACyT, grant J50160.

References.

Abramowitz, M. and Stegun, I. A., editors (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Dover Publications Inc., New York. Reprint of the 1972 edition.

Aczél, J. and Dhombres, J. (1989). *Functional Equations in Several Variables.* Cambridge Univ. Press.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.

Pitt, M. K. and Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association*, **100**, 554–564.

Pitt, M. K., Chatfield, C., and Walker, S. G. (2002). Constructing first order autoregressive models via latent processes. *Scandinavian Journal of Statistics*, **29**, 657–663.

Mena, R. H. and Walker, S. G. (2007). On the stationary version of the generalized hyperbolic ARCH model. *Annals of the Institute of Statistical Mathematics*, **59**, 325-348.

MCEM-Iter.	$\hat{Q}$	l_θ	a	b	ρ
Model		478.72	1.0	3.0	0.7
1	-35.918	281.74	0.99767	3.3630	0.12458
2	-62.689	294.19	1.0081	3.3982	0.14979
3	-122.25	309.94	1.0076	3.3966	0.18435
4	-148.64	322.86	1.0070	3.3946	0.21350
5	-190.99	339.73	1.0066	3.3934	0.25236
10	–345.00	398.47	1.0111	3.4088	0.39295
20	-525.91	454.29	0.99835	3.3666	0.54879
30	-634.71	474.07	1.0055	3.3913	0.62272
40	-706.49	479.18	0.99591	3.3593	0.66021
50	-695.56	479.21	1.0154	3.4251	0.65675
100	-774.23	480.23	0.99182	3.3458	0.68584

Table 2: MCEM iterations for the simulated Poisson-gamma model ($a = 1, b = 3$ and $\rho = 0.7$) based on the uniformly simulated data ($\mathbf{x}$). For the results in this table the MC simulations required within each iteration where reduced by half in each iteration with a minimum of 5 simulations.

MCEM-Iter.	$\hat{Q}$	l_θ	a	b	ρ
Model		374.48	1.0	3.0	0.7
1	-120.80	142.73	0.95280	2.6325	0.063404
2	-32.749	153.78	1.0729	3.1216	0.053279
3	-27.059	157.77	1.0895	3.1878	0.060540
4	-30.732	159.52	1.0926	3.1982	0.064691
5	-37.874	163.66	1.0793	3.1529	0.078280
10	-79.441	185.96	1.0781	3.1455	0.14994
20	-135.59	219.68	1.0837	3.1561	0.26650
30	-216.14	250.83	1.0487	3.0281	0.38240
40	-249.71	263.20	1.0465	3.0232	0.42071
50	-266.83	267.86	1.0321	2.9609	0.44180
100	-263.01	269.29	1.0551	3.0484	0.43593

Table 3: MCEM iterations for the simulated Poisson-gamma model ($a = 1, b = 3$ and $\rho = 0.7$) based on the exponentially spaced data ($\mathbf{x}^\tau$). For the results in this table the MC simulations required within each iteration where reduced by half in each iteration with a minimum of 5 simulations.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-Verlag. New York.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, **82**, 528–550.

Wei, G. C. and Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Yacine, A. S., and Mykland, P. A. (2003). The effect of random and discrete sampling when estimating continuous-time diffusions *Econometrica*, **71**, 483–549.

Yor, M. (2001). *On Exponential Functionals of Brownian Motion and Related Processes*. Springer.