Universidad Nacional Autónoma de México
Programa de Maestría y Doctorado en Ciencias de las Matemáticas
y de la Especialización en Estadística Aplicada

A dual construction of reversible Markov processes for filtering problems

Tesis
que para optar por el grado de
Doctor en Ciencias

presenta:
Freddy Palma Mancilla

tutor principal:
Dr. Ramsés H. Mena Chávez, IIMAS-UNAM

comité tutor:
Dr. Gerónimo Uribe Bravo, IMATE-UNAM
Dra. Eliane R. Rodrigues, IMATE-UNAM

Ciudad Universitaria CD. MX. Septiembre de 2017

Universidad Nacional Autónoma de México
UNAM
Biblioteca Central
Dirección General de Bibliotecas de la UNAM

UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Preface

The objective of this thesis is to build transition probabilities that drive a class of reversible Markov processes. Then, given these, we use them as signal processes within the stochastic filtering framework. In particular, appealing expressions for the optimal and prediction filters, associated to such signals, are provided. Furthermore, the conditional probability structure used to build the afore transitions guarantees the existence of a dual process. This allows to obtain an alternative way to compute the mentioned filters. Furthermore, the structure of our proposal is extendible to nonparametric processes.

The first chapter of this thesis reviews some relevant background material. For instance, we present concepts of the theory of Markov processes, which are borrowed from Ethier and Kurtz [8]. Our purpose here is to identify conditions under which a Markov process can be constructed from a valid set of transition probabilities. Also, we present the concepts of duality and stochastic filtering, dragged from Jansen and Kurt [15] and Cappé [3]. The final part of this chapter will introduce the reader to the nonparametric version of our proposal.

The discrete-time case of the construction is developed in Chapter 2. Here, besides explaining the methodology, we found the form of the corresponding dual and filters. Additionally, we present two remarkable models fall within two frameworks: the construction of AR(1) models, introduced by Pitt, Chatfield and Walker [37], and a novel set of processes derived from the Lancaster probabilities (*cf.* Lancaster [22] and Koudou [19]). This later class is related to the work of Diaconis, Khare and Saloff-Coste [6]. These models have an interest by their own, however their purpose is to derive a mechanism that could potentially be generalized to the continuous-time case.

In Chapter 3 we look for the continuous-time framework. Thus, based on the preliminar idea by Mena and Walker [30], we derive new properties associated to such of models, in particular when we use in stochastic filtering. Unlike the discrete-time case, here the Markovianity of the dual process can not always guaranteed. However, the whole structure of the model helps to build a Markovian dual. With these results, we are then able to find an alternative recursion for the filters computation. On other hand, in order to specify the model, we need two main components, namely two probability measures. Here is when the models described in Chapter 2 become helpful. In particular, we find that the relation between orthogonal polynomials and stochastic processes is a useful tool in this context.

The last chapter of this work is devoted to the nonparametric version of our proposal. In this context, the construction of random probability measures will play an important rol in the construction. Indeed, such measures will be the invariant measures associated to measure-valued Markov processes. Thus, using similar projective properties, as those used by Papaspiliopoulos, Ruggiero and Spanó [34], we will generalize the whole model to the nonparametric case.

Notation and abbreviations

Symbol	
$\mathbb{P}$	Probability measure
$\mathbb{E}$	Expectation operator
Var/Cov/Corr	Variance/Covariance/Correlation operator.
$\mathbb{R}$	Real numbers
$\mathbb{Z}$	Integer numbers
$\sigma(\mathbb{X})$	The σ-algebra generated by the measurable space $\mathbb{X}$.
$\mathcal{B}(\mathbb{X})$	The set of all real-valued, bounded, Borel measurable functions on $\mathbb{X}$.
$X \sim F$	The random variable X is distributed according to F.
$\mathscr{L}(X)$	The law of the random variable X.
$\mathrm{N}(\gamma, \tau)$	Normal distribution with mean γ and variance τ.
$\mathrm{Ga}(a, b)$	Gamma distribution with parameters (a, b).
$\mathrm{Be}(a_1, a_2)$	Beta distribution with parameters (a_1, a_2).
$\mathrm{Po}(\lambda)$	Poisson distribution with mean λ.
Bin	binomial
NB	negative-binomial
Hypgeo	hypergeometric
ED	exponential dispersion
$\propto$	proportional to
$\mathbf{I}_{\{\cdot\}}$	the indicator function
a.s.	almost sure
$\Gamma(\cdot)$	gamma function
$\mathsf{B}(\cdot)$	beta function
i.i.d	independent and identically distributed
i.e.	that is
e.g.	for example
$\stackrel{d}{=}$	equally in distribution

Contents

Chapter 1

Introduction

This thesis uses a conditional probability structure to build transition probabilities that drive reversible Markov processes, with arbitrary but given invariant distributions. Then, we use them as signal processes within the stochastic filtering framework. In particular, appealing expressions for the optimal and the prediction filters, associated to such signals, are presented. As a byproduct, we are able to obtain expressions for statistics corresponding to the filters. Moreover, our construction guarantees the existence of a dual process resulting in an alternative way to compute the filters. Furthermore, the flexibility of the model allows us to generalize the construction to a Bayesian nonparametric framework.

The rest of the chapter is organized as follows. First, we present some widely known distributional symmetries that are used throughout this work namely: exchangeability, stationarity and reversibility. Then, we provide some relevant background concerning to stochastic processes, which will help us, among other things, to define random measures in Chapter 4. Later, we summarize some basic concepts related to the theory of Markov process, to be precise, we focus on how to characterize this kind of processes via their transition probabilities. Thus, we will be ready to introduce the reader the concepts of stochastic filtering and duality between two Markov processes. In the final part we describe the Bayesian nonparametric model.

1.1 Distributional symmetries

The symmetry known as exchangeability implies that the probability law of a sequence of random variables is unaltered under finite permutation of any subsequences. The following definition gives us the mathematical implication of such symmetry.

Definition 1. *Consider a complete and separable metric space* $\mathbb{Y}$ *endowed with its Borel* σ*-algebra* $\mathscr{Y}$*. A sequence* $\{Y_n\}_{n\geq 1}$ *of* $\mathbb{Y}$*-valued random variables, defined in some probability space* $(\Omega, \mathscr{F}, \mathbb{P})$*, is said to be exchangeable, if the probability distribution of the random vector* $(Y_1, ..., Y_n)$ *coincides with the probability distribution of* $(Y_{\tau(1)}, ..., Y_{\tau(n)})$*, for any* $n \geq 1$ *and any permutation* τ *of the indices* $\{1, ..., n\}$*.*

In this context, de Finetti's representation theorem allows to express the joint probability distribution of any subsequence of an exchangeable sequence as the integral of a product of conditionally independent probabilities. In fact, some cases conceive such probabilities as the empirical limit attained to a set of observations. Our proposal uses a conditional probability structure generated from exchangeability to build stationary and reversible Markov models.

Let us recall that, stationarity states that the probability law of a sequence of random variables does not change under shift translations. That is to say, a sequence $Y = \{Y_n\}_{n\geq 1}$ is called stationary if for any finite subsequence $(Y_1, ..., Y_n)$ of Y, its joint probability satisfies the following,

$$\mathbb{P}\big[\, Y_1 \in A_1, ..., Y_n \in A_n \,\big] = \mathbb{P}\big[\, Y_{1+s} \in A_1, ..., Y_{n+s} \in A_n \,\big],$$

for any $n, s \geq 1$ and any measurable collection of sets $\{A_i\}_{i\geq 1}$. On other hand, reversibility implies that the probability law of a sequence of observations does not change under a time reversibility, *i.e.*

$$\mathbb{P}\big[\, Y_1 \in A_1, ..., Y_n \in A_n \,\big] = \mathbb{P}\big[\, Y_n \in A_1, ..., Y_1 \in A_n \,\big],$$

for any $n \geq 1$. Later, these distributional symmetries will play an important rol in our construction. For instance, the mechanism that we will use to build a set of transition probabilities allows us to find a duality between two Markov processes. Before to introduce the reader the concept of such duality, let us give a brief review of the theory of stochastic processes.

1.2 Stochastic processes

Letting $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space, $(\mathbb{X}, \mathscr{X})$ be a measurable space and $\mathcal{T}$ be a completely ordered set. A standard definition of a stochastic process is given below.

Definition 2. *A stochastic process with common state-space* $(\mathbb{X}, \mathscr{X})$ *is a collection of random variables* $\{X(t, \omega); t \in \mathcal{T}, \omega \in \Omega\}$ *such that* $X(t) : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{X}, \mathscr{X})$, *for any fixed* $t \in \mathcal{T}$. *Moreover, the function* $X : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{X}^T, \mathscr{X}^T)$ *is referenced to the realized path of the stochastic process.*

In practice, it is common for the set $\mathcal{T}$ to be a subset of $\mathbb{R}_+ \cup \{0\}$. Hence, hereafter, if $\mathcal{T}$ is countable we will denote the process by $\{X_n\}_{n\in\mathcal{T}}$, letting clear that is a discrete-time process. Whereas, $\{X_t\}_{t\in\mathcal{T}}$ will denote a continuous-time process. Thus, we associate the letter n to the discrete case and the t for the continuous case. Now, recall that, a process X is measurable if $X : \mathcal{T} \times (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{X}, \mathscr{X})$ is $\mathscr{T} \times \mathscr{F}$-measurable, where $\mathscr{T} = \mathscr{B}(\mathcal{T})$. Additionally, we say that X is (almost sure) continuous (right continuous, left continuous) if for (almost) every $\omega \in \Omega$, the sample path $X(\cdot, \omega)$ is continuous (right continuous, left continuous).

Stochastic filtering problems, as we will see below, deal with the information available of a random phenomenon in order to assign it a probability measure. Hence, clearly, this is related with the filtration associated to a stochastic process. Then, a collection $\{\mathscr{F}_t\}_{t\in\mathcal{T}}$ of σ-algebras of sets in $\mathscr{F}$ is a filtration if $\mathscr{F}_t \subset \mathscr{F}_{t+s}$, for $t, s \in \mathcal{T}$. Thus $\mathscr{F}_t$ correspond to the information known at time t. In particular, for a process X we define $\{\mathscr{F}_t^X\}_{t\in\mathcal{T}}$ by $\mathscr{F}_t^X = \sigma(X_s; s \leq t)$; that is, $\mathscr{F}_t^X$ is the information obtained by observing X up to time t. Then, a process X is adapted to the filtration $\{\mathscr{F}_t\}_{t\in\mathcal{T}}$ if X is $\mathscr{F}_t$-measurable for each $t \geq 0$. Hence, since $\mathscr{F}_t$ is increasing in t, X is $\mathscr{F}_t$-adapted if and only if $\mathscr{F}_t^X \subset \mathscr{F}_t$, for each $t \geq 0$. Additionally, a process X is $\mathscr{F}_t$-progressive if for each $t \geq 0$ the restriction of X to $\mathcal{T} \times \Omega$ is $\mathscr{T} \times \mathscr{F}_t$-measurable. Then, every right (left) continuous $\mathscr{F}_t$-adapted process is $\mathscr{F}_t$-progressive.

Furthermore, let us present the formal definition of the set of finite dimensional distributions associated to any stochastic process, which allows to give the necessary conditions to prove their existence.

Definition 3. *Let* $\tilde{\mathcal{T}} = \{(t_1, ..., t_n); 0 \leq t_1 \leq \cdots \leq t_n, n = 1, 2, ...\}$ *be the collection of all finite increasing sequences in* $\mathcal{T}$. *The set of finite dimensional distributions* *associated to the stochastic process* $X = \{X_t\}_{t\in\mathcal{T}}$ *is the collection of functions* $\mathcal{P} = \{\mathbb{P}_{(t_1,\dots,t_n)}(\cdot); (t_1,\dots,t_n) \in \tilde{\mathcal{T}}\}$ *such that*

$$\mathbb{P}_{(t_1,\dots,t_n)}(A) = \mathbb{P}\big[\,(X_{t_1},\dots,X_{t_n}) \in A\,\big],$$

for any measurable set $A \in \mathscr{X}^n$.

Recall that, when working with discrete-time stochastic processes, the above definition is commonly written in terms of consecutive and finite sequences of $\mathcal{T}$.

Letting $\pi_E^F : (\mathbb{X}^F, \mathscr{X}^F) \to (\mathbb{X}^E, \mathscr{X}^E)$ be the projections defined by $\pi_E^F((f_t)_{t\in F}) = (f_t)_{t\in E}$, where $E \subset F \subset \tilde{\mathscr{T}}$. It is known that, the product σ-algebra $\mathscr{X}^F$ coincides with the minimum σ-algebra over $\mathbb{X}^F$ with respect to which the projections $\{\pi_t^F\}_{t\in F}$ are measurable. Hence, one can easily prove that the family $\mathbb{P}_{(t_1,\dots,t_n)}(A)$ satisfies the following consistency conditions,

$$\mathbb{P}_E = \mathbb{P}_F \circ \big[\,\pi_E^F\,\big]^{-1}, \qquad \forall E \subset F. \tag{1.1}$$

Conversely, the so-called Kolmogorov extension theorem tells us that, given a family of probabilities, there exists a stochastic process with finite dimensional distributions that coincides with that family if and only if these probabilities satisfy the consistency condition (1.1).

The above finite dimensional distributions help to derive some notions of equivalence between two stochastic processes. In particular, if X and Y are stochastic processes with the same finite dimensional distributions, we say that Y is is a version of X. In this case, these processes do not need to be defined in the same probability space. Also, if X and Y are defined in the same probability space and for each $t \geq 0$, $\mathbb{P}[X_t = Y_t] = 1$, then we say that Y is a modification of X. Finally, if there exists $\mathscr{N} \in \mathscr{F}$ such that $\mathbb{P}[\mathscr{N}] = 0$ and $X(\cdot;\omega) = Y(\cdot;\omega)$ for all $\omega \notin \mathscr{N}$, then we say that X and Y are indistinguishable.

1.3 Markov processes

Markov processes are one of the most important class of stochastic processes that are useful for modeling phenomena with limited historical dependency. Indeed, the evolution of a Markov process at a future time, conditioned on its present and past values, depends only on its present value. In order to present a formal definition of these class of processes, consider a stochastic process $X = \{X_t\}_{t\in\mathcal{T}}$, where $\mathcal{T}$ is a completely ordered set, defined over the filtered probability space $(\Omega, \mathscr{F}, \mathbb{P}, \{\mathscr{F}_t\}_{t\in\mathcal{T}})$ and with state space $(\mathbb{X}, \mathscr{X})$.

Definition 4. *A stochastic process X is said to be Markov, with respect to the filtration $\{\mathscr{F}\}_{t\in\mathcal{T}}$, if X_t is $\mathscr{F}_t$-adapted and,*

$$\mathbb{E}\big[\, f(X_t)\big|\mathscr{F}_s\,\big] = \mathbb{E}\big[\, f(X_t)\big|X_s\,\big],$$

for any $\mathscr{X}$-measurable and bounded function f and, $0 \le s < t$. Equivalently, if

$$\mathbb{P}\big[\, X_t \in A\big|\mathscr{F}_s\,\big] = \mathbb{P}\big[\, X_t \in A\big|X_s\,\big],$$

for $A \in \mathscr{X}$. The above equality is commonly called the Markov property.

Hereafter, we consider Markov processes with respect to its canonical filtration. Then, the probability law of $\{X_t\}_{t\in\mathcal{T}}$ is characterized by a system of probabilities $\mathbb{P}[X_t \in A|X_s]$, here denoted by $\mathtt{k}_{s,t}(X_s, A)$, termed as a system of transition probability functions. Moreover, the kernel $\mathtt{k}_{s,t} : \mathbb{X}\times\mathscr{X} \to [0, 1]$ is a transition probability function if it satisfies the following conditions: $\mathtt{k}_{s,t}(x, \cdot)$ is a probability measure on $(\mathbb{X}, \mathscr{X})$, for any $x \in \mathbb{X}$; the function $\mathtt{k}_{s,t}(\cdot, A)$ is $\mathscr{X}$-measurable, for any $A \in \mathscr{X}$; and

$$\mathtt{k}_{s,t}(x, A) = \int \mathtt{k}_{s,r}(x, \mathtt{d}v)\mathtt{k}_{r,t}(v, A),$$

for $0 \le s < r < t$. The last equality is known as Chapman-Kolmogorov's property. In particular, we will say that X is a standard Markov process if $\mathtt{k}_{s,t}$ is a transition probability function and $\lim_{t\ge 0} \mathtt{k}_{s,s+t}(x_0, \mathtt{d}x) = \delta_x(x_0)$, for $x_0 \in \mathbb{X}$. Also, if the transition probability function $\mathtt{k}_{s,t}$ only depends on $t - s$, *i.e.* when $\mathtt{k}_{s,t} = \mathtt{k}_{0,t-s}$, then we say that $\mathtt{k}_{s,t}$, and so $\{X_t\}_{t\in\mathcal{T}}$, is time-homogeneous. Hence, we denote it by $\mathtt{k}_{0,t} = \mathtt{k}_t$. In this case, Chapman-Kolmogorov's equations takes the form

$$\mathtt{k}_{t+s}(x, A) = \int_{\mathbb{X}} \mathtt{k}_s(x, \mathtt{d}v)\mathtt{k}_t(v, A), \tag{1.2}$$

for $x \in \mathbb{X}$, $A \in \mathscr{A}$ and $t, s > 0$. It is worth noticing that, if $\{X_t\}_{t\in\mathcal{T}}$ is not time-homogeneous, then the bivariate process $\{(t, X_t)\}_{t\in\mathcal{T}}$ is time-homogeneous. Hence, in what follows we only consider time-homogeneous Markov processes.

The collection of finite dimensional distributions associated to a time-homogeneous Markov process is composed by elements of the form,

$$\mathbb{P}[(X_{t_1}, ..., X_{t_n}) \in B] = \int \cdots \int_B \mathtt{k}_{t_{n-1}}(x_{t_{n-1}}, \mathtt{d}x_{t_n}) \cdots \mathtt{k}_{t_1}(x_{t_1}, \mathtt{d}x_{t_2})\pi(\mathtt{d}x_{t_1}),$$

for $0 \leq t_1 < \cdots < t_n$ and any set $B \in \mathscr{X}^n$, where π is the initial distribution of the process, *i.e.* the law of X_0. Clearly, the law of a time-homogeneous Markov process is completely characterized by specifying the family of transition probability functions and some initial distribution. For the discrete case, the family of transition probability functions $\{\mathtt{k}_t\}_{t\in\mathcal{T}}$ is determined by the one-step transition probability function via $\mathtt{k}_n = \mathtt{k}^n$. Thus, the existence of the probability law of the whole process, induced by the invariant measure and transition distribution functions, follows form the Kolmogorov existence theorem. As a matter of fact, if $(\mathbb{X}, \mathscr{X})$ is complete and separable, then there exists a Markov process whose finite dimensional distribution are given by (1.3). This is precisely the result on which our construction is based, borrowed from the book of Ethier and Kurtz [8].

Now, let us link the distributional symmetries of the previous section with the context of Markov processes. First, we say that π is an invariant measure of $\{X_t\}_{t\in\mathcal{T}}$ if its transition probability functions satisfy the following equation

$$\pi(A) = \int_{\mathbb{X}} \pi(\mathtt{d}x)\mathtt{k}_t(x, A), \tag{1.3}$$

for $A \in \mathscr{X}$. As a consequence, if the initial distribution of a Markov process X is an invariant measure, then X is a stationary process. Also, a time-homogeneous Markov process is reversible with invariant measure π if $\pi(\mathbb{X}) < \infty$ and

$$\int_{B'} \pi(\mathtt{d}x)\mathtt{k}_t(x, B) = \int_B \pi(\mathtt{d}x)\mathtt{k}_t(x, B'), \tag{1.4}$$

for $B, B' \in \mathscr{X}$. Then, if π is finite, then any reversible Markov process is also stationary. Moreover, the invariant distribution function is unique, and all the transition distribution functions will eventually converge to the invariant distribution of the process.

Furthermore, the transition probabilities associated to Markov processes allow us to define an operator, known as infinitesimal generator, which also serves to characterize such processes. Hence, let us present its formal definition and some of its properties. Let T_t be a bounded linear operator defined by

$$T_t f(x) \equiv \int_{\mathbb{X}} f(x)\mathtt{k}_t(x, \mathtt{d}v)$$

for $x \in \mathbb{X}$ and $f \in \mathcal{B}(\mathbb{X})$, where $\mathcal{B}(\mathbb{X})$ denotes the set of all real-valued, bounded, Borel measurable functions on $\mathbb{X}$. The family of operators $\{T_t\}_{t\in\mathcal{T}}$ defines a contraction semigroup on $\mathcal{B}(\mathbb{X})$, *i.e.* $||T_t f|| \leq ||f||$, $T_0 = 1$ and $T_{s+t} = T_s T_t$. Now, the infinitesimal generator $\mathscr{A}$ associated to the contraction semigroup $\{T_t\}_{t\in\mathcal{T}}$, or to the Markov process $\{X_t\}_{t\in\mathcal{T}}$, is the linear operator defined by

$$\mathscr{A} f(x) = \lim_{t\downarrow 0} \frac{T_t f(x) - f(x)}{t} \tag{1.5}$$

for all $f \in \mathcal{B}(\mathbb{X})$, such that the right side converges to some function uniformly in x. The class of all function f such that the limit (1.5) exists determines the domain of $\mathscr{A}$, denoted by $\mathcal{D}_{\mathscr{A}}$. A widely known property that satisfy this operator is the well-known *Backward* equation, *i.e.*

$$\mathscr{A} T_t f = \frac{\partial}{\partial t} T_t f$$

for all $f \in \mathcal{B}(\mathbb{X})$. The above equation is used in this work within the dual framework in Chapter 3.

Some classical examples of Markov are commonly specified via its infinitesimal generator. Next, we will give a brief review of birth and death processes and, diffusion processes. This models are used throughout this work, in particular, those that fall into our construction. Also, we will give a review of processes with independent and stationary increments. These are commonly used in the Bayesian nonparametric framework. Hence, they will be helpful in Chapter 4.

1.3.1 Birth and death processes

Birth and death processes play an important role in the theory and applications in continuous-time Markov chains. These class of models are continuous-time Markov process with parameter set $\mathcal{T} = [0, \infty)$ and state space $\mathbb{X} = \{-1, 0, 1, 2, ...\}$ with stationary transition probabilities, denoted by $\mathsf{p}_{i,j}(t) = \mathtt{k}_t(i, j)$ for $i, j \in \mathbb{X}$,

a. $\mathsf{p}_{i,i+1}(h) = \lambda_i h + o(h)$ as $h \to 0$, $i \in \mathbb{X}$;

b. $\mathsf{p}_{i,i-1}(h) = \mu_i h + o(h)$ as $h \to 0$, $i \geq 0$;

c. $\mathsf{p}_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$ as $h \to 0$, $i \in \mathbb{X}$;

d. $\mathsf{p}_{i,j}(h) = \delta_{i,j}$;

e. $\mathsf{p}_{-1,-1}(t) = 1$, $\mathsf{p}_{-1,i}(t) = 0$, for $t \geq 0$, $i \neq -1$,

with $\mu_0 \geq 0$, $\lambda_0 > 0$, $\lambda_i, \mu_i > 0$, for $i \geq 1$. The parameters λ_i and μ_i are called, respectively, the birth and death rates. In postulates (a.) and (b.) is assumed that if the process starts in state i, then in a small interval of time the probabilities of going one state up or down are essentially proportional to the length of the interval.

If $\mu_0 > 0$ then we have an absorbing state -1; once we enter we can never leave it. If $\mu_0 = 0$ we have a reflecting state 0. After entering to 0 we will always go back to state 1 after some time. In this case the state -1 can never be reached and so we ignore it and take $\mathbb{X} = \{0, 1, 2, ...\}$. Thus, the infinitesimal generator $\mathscr{A}$ of a birth and death process is given by

$$\mathscr{A} f(i) = \lambda_i f(i+1) - (\lambda_i + \mu_i) f(i) + \mu_i f(i-1), \tag{1.6}$$

for $i \geq 0$ and all bounded real-valued function $f \in \mathcal{B}(\mathbb{X})$. Later, we will make some connection between the operator (1.6) with the theory of orthogonal polynomials.

1.3.2 Diffusion processes

A continuous-time stochastic process which possesses the strong Markov property and for which the sample paths are almost always continuous functions of the time is called a diffusion process. In fact, the theory of diffusion processes introduces a dependency between second order differential operators of the Markov processes. Thus, given $X_s = x$, for (infinitesimal) small times t, the displacement $X_{s+t} - X_s = X_{s+t} - x$ has mean and variance $t\mu(x)$ and $t\sigma^2(x)$, respectively. Here, $\mu(x)$ and $\sigma^2(x)$ are functions of the state x. The existence of infinitesimal mean and variance parameters, that does not require the existence of finite moments, can be guaranteed as following. Let $\epsilon > 0$ and consider the equalities

$$\mathbb{E}[(X_{t+s} - X_s)\mathcal{I}_{\{|X_{t+s}-X_s|\leq\epsilon\}}|X_s = x] = t\mu(x) + o(t) \tag{1.7}$$

$$\mathbb{E}[(X_{t+s} - X_s)^2\mathcal{I}_{\{|X_{t+s}-X_s|\leq\epsilon\}}|X_s = x] = t\sigma^2(x) + o(t) \tag{1.8}$$

$$\mathbb{P}[|X_{t+s} - X_s| > \epsilon|X_s = x] = o(t) \tag{1.9}$$

where $\mathcal{I}_A = 1$ if A holds and is zero otherwise. We say that a Markov process X_t on the state space $\mathbb{X} = (a, b)$ is a diffusion with drift coefficient $\mu(x)$ and diffusion coefficient $\sigma^2 > 0$, if it has continuous sample paths, and the relations (1.7)-(1.9) hold for all $x \in \mathbb{X}$.

Thus, all twice continuously differentiable functions f, vanishing outside a closed bounded subinterval of $\mathbb{X}$, belong to $\mathcal{D}_{\mathscr{A}}$, which implies that its infinitesimal generator takes the form

$$\mathscr{A} f(x) = \mu(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x).$$

for every f. In Chapter 3 we will use the above operator in order to derive an expression for the spectral representation of the transition probability density associated to it. This will help us to relate, as the birth and death processes, with the theory of orthogonal polynomials.

1.3.3 Processes with independent and stationary increments

Let $X = (X_t)_{t\geq0}$ be a stochastic process taking values in $\mathbb{R}^d$ and $0 \leq t_1 < t_2$. The random variable $X_{t_2} - X_{t_1}$ is called the increment of the process X over the interval $[t_1, t_2]$. Then, a stochastic process is said to be a process with independent increments if the increments over non overlapping intervals are stochastically independent, *i.e.* X_t has *càdlàg* sample paths and the random variables $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}}$, are independent, for $n > 0$. Hence, discontinuities of the process can only occur by jumps.

Additionally, if the distribution of the increments $X_{t+s} - X_s$ depends only on t, *i.e.* $X_{t+s} - X_s \stackrel{d}{=} X_t$ for all $s, t, > 0$, then we say that the process has stationary increments. As a consequence, stationary increments excludes the possibility of having fixed jumps. A stochastic process with stationary and independent increments is known as Lévy process. From these assumptions it is also clear that a Lévy process satisfies the Markov property and thus a Lévy process is a special type of Markov process.

It is known that, if X is a Lévy process then there exists a continuous function $\psi : \mathbb{R}^d \to \mathbb{R}$ called the exponent characteristic of X, such that

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\{t\,\psi(u)\},$$

where

$$\psi(u) = -\frac{1}{2}uAu + i\gamma u + \int_{\mathbb{R}^d}\big(e^{iux} - 1 - iux\mathtt{I}_{|x|\le 1}\big)\nu(\mathtt{d}x).$$

with $A \in \mathbb{R}^d \times \mathbb{R}^d$ is a positive defined matrix, $\gamma \in \mathbb{R}^d$ and ν is a positive Radon measure on $\mathbb{R}^d$ such that:

$$\int_{|x|\le 1} |x|^2\nu(\mathtt{d}x) < \infty \qquad \text{and} \qquad \int_{|x|\ge 1} \nu(\mathtt{d}x) < \infty.$$

The measure ν is known as the Lévy measure and the value $\nu(A)$ represents the expected number, per unit time, of jumps whose size belong to A. The vector (A, γ, ν) is called the characteristic triplet of X. The expression for the exponent ψ is known as the Lévy-Khinchin representation. Moreover, if $d = 1$ and $\nu(\mathbb{R}) = \infty$ (infinite activity case), the set of jump times of every trajectory of the Lévy process is countably infinite and dense in $[0, \infty)$.

Furthermore, one can prove that if X is a Lévy process then the distribution of X_t has an infinitely divisible distribution, for every $t \ge 0$. Thus, the Lévy-Khinchin formula also gives a general representation for the characteristic function of any infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution, then there exists a Lévy process X such that the distribution of X_1 has distribution F.

Increasing Lévy processes are also called subordinators because they can be used as time changes for other Lévy processes. Thus, a Lévy process is a subordinator if any of the following statements hold:

- $X_t \ge 0$ a.s. for some $t \ge 0$.
- Sample paths of X are almost surely nondecreasing: $t < s \Rightarrow X_t \ge X_s$.

- The characteristic triplet of X satisfies the following $A = 0$, $\nu((-\infty, 0]) = 0$, $\int_0^\infty (x \wedge 1)\nu(\mathrm{d}x) < \infty$ and $b > 0$.

The most common examples of infinitely divisible laws are: the Gaussian distribution, the gamma distribution and the α-stable distributions. The class of increasing Lévy processes turn out to be useful for defining random measures.

1.4 Duality

The conditional probability structure used to built reversible Markov processes guarantees the existence of a dual processes. In this sense, duality of Markov processes with respect to a duality function has been used to develop the connections of fundamental structures or properties of Markov processes, such as time reversal, stochastic monotonicity, intertwining, to name a few. However, the existence of the dual associated to a given Markov process has not been yet fully resolved. Such property and some new results in this field were presented by Jansen and Kurt [15], who also made a recent review of a theoretical background. It is worth to mention that, the existence of the dual for a given Markov processes implies that the associated martingale problem is well defined. Thus, besides guaranteeing the existence of the dual, the construction unveils the form of the duality function. A formal definition of this kind of duality goes as follows.

Definition 5. *Let X and Y be two Markov processes with state spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$, respectively, and $h : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}$ a bounded measurable function. Then X and Y are dual to each other with respect to the duality function h if and only if*

$$\mathbb{E}[h(X_t; y)|X_0 = x] = \mathbb{E}[h(x; Y_t)|Y_0 = y]. \tag{1.10}$$

for all $x \in \mathbb{X}$, $y \in \mathbb{Y}$ and $t \geq 0$.

Letting $\mathscr{A}$ and $\mathscr{G}$ be the infinitesimal generators associated to X and Y, with domains $\mathcal{D}(\mathscr{A})$ and $\mathcal{D}(\mathscr{G})$, respectively. It is straightforward to prove that if $\mathtt{h}(x; \cdot) \in \mathcal{D}(\mathscr{G})$ and $\mathtt{h}(\cdot; y) \in \mathcal{D}(\mathscr{A})$, then

$$\mathscr{A}\mathtt{h}(\cdot; y)(x) = \mathscr{G}\mathtt{h}(x; \cdot)(y),$$

for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$. The converse is true as well, under certain conditions. See Jansen and Kurt [15].

In order to illustrate the mentioned duality consider two Markov processes defined over the same Polish space X with a partial order $\leq$, and $\mathtt{h}(x, y) := \mathbf{I}_{\{x \leq y\}}$. These processes are dual with respect to the duality function $\mathtt{h}$ if and only if,

$$\mathbb{P}[X_t \leq y | X_0 = x] = \mathbb{P}[x \leq Y_t | Y_0 = y].$$

This kind of duality occurs in many contexts. For instance, the Brownian motion reflected at 0 and the Brownian motion absorbed at 0 are dual in this sense. This type of duality, sometimes called Siegmund duality, is related to time reversal in a sense that it reverses the role of entrance and exit laws.

The Wright-Fisher model represents another example that possesses the above duality. In fact, such discrete-time model widely used for genetic evolution consists of a finite haploid population of size $2N$ and where each individual is of type A or a. Time start at $n = 0$ and at each time $n \in \mathbb{N}$, each individual randomly chooses an individual from the previous generation and adopt its type. This procedure follows until one of the types fixates in the population, *i.e.* until only one of the types remains due to extinction of the other. In this case, the transition probability function is given by a binomial kernel and, the probability that type a goes extinct equals the initial fraction of A's in the population. An interesting question is what we can say about the time until fixation, denoted by τ. In particular, the expectation of τ is computed using the generic variability H'n of the population at time n, *i.e.* the probability that two different individuals, randomly drawn from the population at time n, are of different type. Hence, it is hereby useful to look at the population's genealogical history. Indeed, the backward ancestral path, which generates the dual of the Wright-Fisher model, leads us to the equality $\mathbb{E}[\tau] = 2NH_0$. Thus, duality lets us to derive explicit formulas for the more complicated process from the simpler one.

It is worth noticing that, since we require (1.10) holds for any arbitrary initial condition, Markov process duality is in fact a property of the transition kernels of two Markov processes, rather than two concrete processes (*cf.* Sturm and Swart [41]). Nonetheless, this property turns out to be the one which allows to obtain an alternative way to compute the filters.

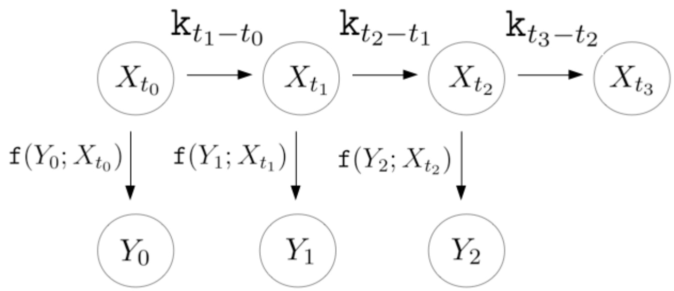

Figure 1.1: Dynamics of a hidden Markov model

1.5 Stochastic filtering

The aim of stochastic filtering is to estimate an evolving dynamical system, the signal, customarily modeled by a stochastic process. Such signal can not be measured directly, however a partial measurement can be obtained, via an observation process. See Cappé [3]. In this thesis, we build transition probabilities, $\mathtt{k}_t$, that drive reversible Markov signals, $X = (X_t)_{t\geq 0}$, with invariant measure π. We will denote by $Y = \{Y_t\}_{t\geq 0}$ the observation process whose conditional distribution $\{Y_t | X_t\}$ is denoted by $\mathtt{f}(Y_t; X_t)$, known as emission density. Clearly, the signal parametrizes the law of the observations. Then, considering a discrete-time sample $\{X_{t_n}\}_{n\geq 0}$ of X and, a sequence of conditionally independent observations $\{Y_n\}_{n\geq 0}$. The model $\{(X_{t_n}, Y_n)\}_{n\geq 0}$ is called a hidden Markov model and its dynamics can be seen in Figure 1.1.

Let $\mathscr{Y} = \{\mathscr{Y}_t\}_{t\geq 0}$ be the filtration generated by the observation process Y, *i.e.* $\mathscr{Y}_t = \sigma(Y_s, s \in [0,t])$, for $t \geq 0$. Thus, the σ-algebra $\mathscr{Y}_t$ is the information available from observations up to time t. Such information is used to make inferences about the signal. In particular, the computation or approximation of quantities of the form $\mathbb{E}[\varphi(X_t)|\mathscr{Y}_t]$, where φ is a real-valued function. Each of these statistics will provide fragments of information about X_t. Mathematically, this means computing the conditional distribution of X_t given $\mathscr{Y}_t$, which is progressively measurable with respect to $\mathscr{Y}_t$ so that

$$\mathbb{E}\left[\,\varphi(X_t)\,\middle|\,\mathscr{Y}_t\,\right] = \int_{\mathbb{X}} \varphi(x)\mathbb{P}\left[\,X_t \in \mathtt{d}x\,\middle|\,\mathscr{Y}_t\,\right], \tag{1.11}$$

for all statistics φ, for which both terms of the above identity make sense. Thus, knowing $\mathbb{P}[X_t \in \mathtt{d}x|\mathscr{Y}_t]$ allows us, at least theoretically, to compute any inference of X_t given $\mathscr{Y}_t$.

The filters are given by the laws of the signal at different times given past and present observations, *i.e.*

$$\mathscr{L}(X_{t_k}|Y_1, ..., Y_n) \tag{1.12}$$

for $n \geq 0$. When $t_k = n$, the law (1.12) is termed as the optimal or exact filters; if $t_k > n$ then the law (1.12) is known as prediction filter and otherwise the law (1.12) is called as marginal smoothing filter. These filters may be calculated recursively by explicit algorithms, however iterations become rapidly intractable and exact formulae are usually difficult to obtain.

Without loss of generality, we consider observations equally spaced in time, *i.e.* $t_{n+1} - t_n = \Delta$ for all $n \geq 0$ and $\Delta > 0$. Letting $\nu_n := \mathscr{L}(X_n|Y_{0:n})$, with $Y_{0:n} := (Y_0, ..., Y_n)$, the exact or optimal filters, starting with $\nu_{-1} = \pi$, are given by the recursion

$$\nu_n = \phi_{Z_n}(\psi(\nu_{n-1})),$$

for $n \geq 0$. Here, the update operator ϕ is obtained via Bayes' theorem and, the prediction operator is associated to the transition function $\mathtt{k} := \mathtt{k}_1$. Thus,

$$\phi_y(\nu)(\mathtt{d}x) = \frac{\mathtt{f}(y; x)\nu(\mathtt{d}x)}{\mathsf{m}_\nu(y)}, \qquad \text{with} \qquad \mathsf{m}_\nu(y) = \int_{\mathbb{X}} \mathtt{f}(\mathtt{d}y; x)\nu(\mathtt{d}x);$$

and

$$\psi(\nu)(\mathtt{d}x) = \int_{\mathbb{X}} \nu(v)\mathtt{k}_\Delta(v, \mathtt{d}x)\mathtt{d}v.$$

See Papaspiliopoulos and Ruggiero [33].

On the other hand, the predictor filters are obtained via the recursion $\psi(\nu_n)$, for $n \geq 0$. Since π is an invariant measure of the transition $\mathtt{k}_t$, we have that $\psi(\nu_\pi) = \nu_\pi$. One of the main problems of stochastic filtering is to derive computable or tractable expression for the filters. The signals proposed will allow to obtain appealing expression for the optimal and the prediction filters. As a result, we will be able to compute statistic associated to them.

1.6 Bayesian nonparametric paradigm

The last part of this thesis extends our construction to a Bayesian nonparametric approach. This will be done using the conditional probability structure inherit in exchangeable random variables, without the parametric simplification. That is to say, in the full extend of exchangeability, that leads us to infinite dimensional spaces. For this purpose, we will borrow the construction of random probability measures from the Bayesian nonparametric literature. These measures, commonly referred as nonparametric priors distributions, will be invariant measures of the signal processes.

Let $Y = (Y_n)_{n\geq 1}$ be an infinite sequence of $(\mathbb{Y}, \mathscr{Y})$-valued random variables over the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Also, let $\mathscr{P}_{\mathbb{X}}$ be the space of all probability measures on $(\mathbb{X}, \mathscr{X})$, the topology of the weak convergence makes such a space Polish. The sequence Y is exchangeable if and only if there exists a probability measure Π on $\mathscr{P}_{\mathbb{Y}}$ such that the joint distribution of any finite subsequence of Y has the following integral representation

$$\mathbb{P}\big[\,(Y_1, ..., Y_n) \in A\,\big] = \int_{\mathscr{P}_{\mathbb{Y}}} \prod_{i=1}^{n} \mu(A_i)\Pi(\mathsf{d}\mu), \tag{1.13}$$

for any $n \geq 1$ and $A = (A_1, ..., A_n) \in \mathbb{X}^{(n)}$. The probability measure Π is commonly named as de Finetti's measure of Y. Furthermore, it is very common to formulate the exchangeability assumption as following,

$$\begin{aligned} Y_i|\mu &\sim \mu, \qquad \text{for } i \geq 1, \\ \mu &\sim \Pi. \end{aligned}$$

Thus, the random probability measure μ, which is $\mathscr{P}_{\mathbb{X}}$-valued, has prior distribution Π. Note that, given μ, the conditional distribution of $(Y_1, ..., Y_n)$ is $\mu^{(u)} = \prod_{i=1}^{n} \mu$. Let us notice that, in order to carry out our proposal, we will require mathematical tractability of the posterior distributions associated to such prior distributions. Hence, part of the Chapter 4 reviews some literature concerning to random probability measure.

Then, the conditional probability structure obtained from (1.13) will be used to build transition probabilities that drive measure-valued Markov processes. Also, using similar projective properties, as those used in Papaspiliopoulos, Ruggiero and

Spanó [34], we will obtain that the sequence Y and μ are dual to each other with respect to a duality function. Mathematically, the projective properties consider measurable partitions $A = (A_1, ..., A_k)$ of $\mathbb{Y}$, for any $k \geq 1$, such that $(\mu_t(A_1), ..., \mu_t(A_k))$ is distributed according to $\Pi^{\alpha(A)}$, where α is the parameter of Π and $\alpha(A) = (\alpha(A_1), ..., \alpha(A_k))$. Thus, we will also be able to compute the optimal and the prediction filters.

Chapter 2

A discrete-time Markov construction for filtering problems

The construction of discrete-time stationary Markov models with known stationary distributions has been widely studied in the literature. In particular, stationary time series models with non-normal marginal have been proposed by Lawrence and Lewis [24] and Jacobs and Lewis [13] for the case of exponential margins; models with gamma, Poisson and negative binomial margins can be found in Lawrence [23], McKenzie ([27], [28]), Lewis, McKenzie and Hugus [25], Al-Osh and Aly [1], to name a few. Later, a unification and generalization of some of these models was proposed by Joe [16], his work was based on a thinning operation applied to some class of infinite divisible distributions closed under convolutions. This idea was also used by Jorgensen and Song [17] with the difference that they considered exponential dispersion margins. A generalization of this kind of models was proposed by Pitt, Chatfield and Walker [37]. Although, such work restricted their attention to purely autoregressive processes. Having this in mind, Pitt and Walker [38] extended the later methodology to state-space models and autoregressive conditional heteroscedasticity models (ARCH); also, they proved that these class of models perform well compared with competing methods for the applications considered. A more extensive review about how to built stationary Markov model via latent processes can be found in Mena and Walker [29], who also proposed an stationary version of the generalized hyperbolic ARCH model.

This chapter provides a mechanism to built transition probabilities that drive discrete-time reversible Markov processes. This methodology unifies some of the models mentioned above. Indeed, there exists a resemblance to the conditional distributional properties of the Gibbs sampler method. The unification of this kind of time-discrete models allows us to derive some new properties for all the models belonging to our proposal. To be precise, these models fit perfectly into the stochastic filtering setting, in particular we derive an expression for the optimal and the prediction filters. Thus, one can obtain information of an unobservable process, the signal, from the observation process. For instance, the structure of the construction generates tractable filters, in the sense that it lets us to derive expressions for some appealing statistics associated to them. Additionally, the construction falls into the duality framework described in the previous chapter. On other hand, the appropriate choice of measures that specifies the model needs to be treated separately. For this purpose, the last two sections of this chapter presents two remarkable models that fall into our proposal. The construction of AR(1) models, introduced by Pitt, Chatfield and Walker [37], and a novel set of processes derived from the Lancaster probabilities (*cf.* Lancaster [22] and Koudou [19]). This later class is related to the work of Diaconis, Khare and Saloff-Coste [6]. It is worth to mention that, this chapter will set a basis for the continuous-time case in Chapter 3.

2.1 Construction

Let $\mathbb{X}$ and $\mathbb{Y}$ be two subsets of $\mathbb{R}^d$, for $d \geq 1$, $\mathscr{X} = \sigma(\mathbb{X})$ and $\mathscr{Y} = \sigma(\mathbb{Y})$. Also, let π be a probability measure over $(\mathbb{X}, \mathscr{X})$ and, for $x \in \mathbb{X}$, let $\mathtt{f}(\cdot; x)$ be a probability measure over $(\mathbb{Y}, \mathscr{Y})$. When it is clear, for the sake of notation, we make no distinction between probability measures and their corresponding density function, assuming that they exist. Also, we assume that density functions are continuous with respect to the Lebesgue measure, letting clear that they can be continuous with respect to the count measure.

The model uses the measure π as the prior knowledge of a given phenomenon and, given some initial value x, the measure $\mathtt{f}(\cdot; x)$ represents a conditional distribution. Thus, one can define a joint distribution over the product space $(\mathbb{X}\times\mathbb{Y}, \mathscr{X}\times\mathscr{Y})$

as following

$$\int_B \int_A \mathtt{f}\,(u;v)\pi(v)\mathtt{d}v\mathtt{d}u, \tag{2.1}$$

for $A \in \mathscr{X}$ and $B \in \mathscr{Y}$. Clearly, the marginal distributions of (2.1) over the spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ are given by the probability measures π and

$$\mathsf{m}_\pi(\,\cdot\,) = \int_{\mathbb{X}} \mathtt{f}\,(\,\cdot\,;v)\pi(v)\mathrm{d}v,$$

respectively. Hereafter, we are assuming that the support of π coincides with the set $\{x \in \mathbb{X} : \mathtt{f}\,(\,\cdot\,;x) > 0\}$. Thus, Bayes' theorem allows us to obtain the posterior distribution for π that takes the form

$$\nu_0(\mathtt{d}x;y) = \frac{\mathtt{f}\,(\mathtt{d}y;x)\pi(\mathtt{d}x)}{\mathsf{m}_\pi(\mathtt{d}y)}, \tag{2.2}$$

for $x \in \mathbb{X}$, $y \in \mathbb{Y}$. Notice that, the measure (2.2) is well-defined because $\mathsf{m}_\pi(\mathtt{d}y) > 0$ for all $y \in \mathbb{Y}$. Hence, $\nu_0(\mathtt{d}x;y)$ is indeed a probability measure. In what follows, if ν_0 and π belong to the same family of distributions, then we say that π is conjugate with respect to $\mathtt{f}$.

Thus, once the measures π and $\mathtt{f}$ are specified, one can define an homogeneous discrete-time process, $X = (X_n)_{n\geq 0}$, with state space $(\mathbb{X}, \mathscr{X})$, driven by the transition probability function

$$\mathtt{k}(x_n, \mathtt{d}x_{n+1}) = \int_{\mathbb{Y}} \nu_0(\mathtt{d}x_{n+1};u)\mathtt{f}\,(u;x_n)\mathtt{d}u, \tag{2.3}$$

for $x_n, x_{n+1} \in \mathbb{X}$. Moreover, the transition (2.3) has π as invariant measure, *i.e.* $\mathtt{k}$ satisfies (1.3). The last statement follows by noticing that

$$\mathtt{k}(x_n, \mathtt{d}x_{n+1}) \;=\; \int_{\mathbb{Y}} \nu_0(\mathtt{d}x_{n+1};u)\left[\frac{\nu_0(\mathtt{d}x_n;u)\mathsf{m}_\pi(u)}{\pi(\mathtt{d}x_n)}\right]\mathtt{d}u,$$

which implies that

$$\begin{aligned}
\pi(\mathtt{d}x_n)\mathtt{k}(x_n, \mathtt{d}x_{n+1}) &= \pi(\mathtt{d}x_{n+1})\int_{\mathbb{Y}} \nu_0(\mathtt{d}x_n;u)\left[\frac{\nu_0(\mathtt{d}x_{n+1};u)\mathsf{m}_\pi(u)}{\pi(\mathtt{d}x_{n+1})}\right]\mathtt{d}u \\
&= \pi(\mathtt{d}x_{n+1})\int_{\mathbb{Y}} \nu_0(\mathtt{d}x_n;u)\mathtt{f}\,(u;x_{n+1})\mathtt{d}u \\
&= \pi(\mathtt{d}x_{n+1})\mathtt{k}(x_{n+1}, \mathtt{d}x_n).
\end{aligned}$$

As a consequence, the process X is reversible with invariant distribution π. Also, under the above construction, reversibility implies stationarity. In fact, integrating both sides of the last equation with respect to x_{n+1} leaves

$$\pi(\mathtt{d}x_n) = \int_{\mathbb{X}} \mathtt{k}(x_{n+1}, \mathtt{d}x_n)\pi(x_{n+1})\mathtt{d}x_{n+1},$$

for $x_n \in \mathbb{X}$. Therefore, as long as the space $\mathbb{X}$ is Polish, the transition probability function $\mathtt{k}$ and its invariant measure π characterize the law of a discrete-time reversible and stationary Markov process (See Section 1.3).

In general, knowing the transition probability kernel of a Markov process is a fundamental tool to make inferences. However, the expression of such kernel is not always tractable, it is here that the structure of the operator (2.3) eases this point. In fact, when one wants to define a transition for some process, one of the properties that one would look for is that it is computable, *i.e.* achieved at low cost. Having this in mind, given that (2.3) is built based on the product of two probability measures, it is seen that the construction generates computable probability kernels.

On other hand, similarly to the operator (2.3), one can define another transition probability kernel $\mathbb{Y}$-valued that share almost the same properties that $\mathtt{k}$. Indeed, let us define the operator

$$\mathtt{p}(z_n, \mathtt{d}z_{n+1}) = \int_{\mathbb{X}} \mathtt{f}\,(\mathtt{d}z_{n+1}; v)\nu_0(\mathtt{d}v; z_n)\mathtt{d}v, \tag{2.4}$$

for z_n, $z_{n+1} \in \mathbb{Y}$. It is straightforward to prove that the transition operator $\mathtt{p}$ satisfies the following equalities

$$\begin{aligned} \mathtt{m}_\pi(\mathtt{d}z_n)\mathtt{p}(z_n, \mathtt{d}z_{n+1}) &= \mathtt{m}_\pi(\mathtt{d}z_{n+1})\mathtt{p}(z_{n+1}, \mathtt{d}z), \\ \mathtt{m}_\pi(\mathtt{d}z_n) &= \int_{\mathbb{Y}} \mathtt{m}_\pi(z_{n+1})\mathtt{p}(z_{n+1}, \mathtt{d}z_n)\mathtt{d}z_{n+1}. \end{aligned}$$

Therefore, we can associate the well-defined transition probability function $\mathtt{p}$ to a time-homogeneous reversible Markov process, denoted by $Z = (Z_n)_{n\geq 0}$, with invariant measure $\mathtt{m}_\pi$.

Furthermore, the conditional probability structure used to define X and Z allows us to derive a relationship between them, this is precisely the duality introduced in Chapter 1. Indeed, considering the function $\mathtt{h} : \mathbb{X} \times \mathbb{Y} \to \mathbb{R}_+$ defined as the Radon-Nikodyn derivative between the posterior and prior distribution of the probability measure π, *i.e.*

$$\mathtt{h}(x;y) = \frac{\nu_0(x;y)}{\pi(x)} = \frac{\mathtt{f}(y;x)}{\mathtt{m}_\pi(y)}. \tag{2.5}$$

Then, using the fact that X is reversible and the definition of $\mathtt{h}$ it follows that

$$\begin{aligned}
\int_{\mathbb{X}} \mathtt{h}(v;y)\mathtt{k}(x,v)\mathtt{d}v &= \int_{\mathbb{X}} \frac{\nu_0(v;y)}{\pi(v)} \left[\frac{\pi(v)}{\pi(\mathtt{d}x)}\mathtt{k}(v,\mathtt{d}x)\right] \mathtt{d}v \\
&= \int_{\mathbb{X}} \frac{\nu_0(v;y)}{\pi(\mathtt{d}x)} \left(\int_{\mathbb{Y}} \nu_0(\mathtt{d}x;u)\mathtt{f}(u;v)\mathtt{d}u\right) \mathtt{d}v \\
&= \int_{\mathbb{Y}} \mathtt{h}(x;u) \left[\int_{\mathbb{X}} \mathtt{f}(u;v)\nu_0(v;y)\mathtt{d}v\right] \mathtt{d}u \\
&= \int_{\mathbb{Y}} \mathtt{h}(x;u)\mathtt{p}(y,\mathtt{d}u).
\end{aligned}$$

Denoting by $\mathbb{E}^x[\mathtt{g}(X_t)]$ the expected value of $\mathtt{g}(X_t)$, for any measurable function $\mathtt{g}$, given the initial value $X_0 = x$. Thus, the previous equality can be written as

$$\mathbb{E}^x[\mathtt{h}(X_t;y)] = \mathbb{E}^y[\mathtt{h}(x;Y_t)], \qquad x \in \mathbb{X},\ y \in \mathbb{Y}, t \in \mathbb{R}^+. \tag{2.6}$$

In this case, it is said that X is dual to Z with respect to the duality function $\mathtt{h}$. Since we require (2.6) holds for any arbitrary initial condition, Markov process duality is in fact a property of the transition kernels of two Markov processes (*cf.* Sturm and Swart [41]). The following proposition summarizes the above results and provides us an expression for the predictor operators of X and Y in terms of their corresponding $\mathtt{h}$-dual.

Proposition 1. *Let π be a probability measure over the measurable space $(\mathbb{X}, \mathscr{X})$ and $\{f(\cdot;x); x \in \mathbb{X}\}_{t\geq 0}$ be a probability model over the measurable space $(\mathbb{Y}, \mathscr{Y})$. Also, let X and Z be two Markov processes driven by the transition probability functions* (2.3) *and* (2.4)*, respectively. Then, X and Z are dual to each other with respect to the duality function* (2.5)*. Moreover,*

$$\int_{\mathbb{X}} \nu_0(v;z_0)k(v,dx)\,dv = \int_{\mathbb{Y}} \nu_0(dx;u)p(z_0,u)\,du, \tag{2.7}$$

and

$$\int_{\mathbb{Y}} f(u;x_0)p(u,dz)\,dz = \int_{\mathbb{X}} f(dy;v)k(x_0,v)\,dv, \tag{2.8}$$

for $x \in \mathbb{X}$, $y \in \mathbb{Y}$, *and where* $\nu_0(\cdot; y)$ *is given by* (2.2).

Proof. The first assertion was already proved. Now, for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$ it follows that

$$\begin{aligned}\int_{\mathbb{X}} \nu_0(v; z_0)\mathtt{k}(v, \mathtt{d}x)\mathtt{d}v &= \int_{\mathbb{X}} \mathtt{h}(v; z_0)\pi(v)\mathtt{k}(v, \mathtt{d}x)\mathtt{d}v \\ &= \pi(dx) \int_{\mathbb{X}} \mathtt{h}(v; z_0)\mathtt{k}(x, v)\mathtt{d}v \\ &= \int_{\mathbb{Y}} \nu_0(\mathtt{d}x; u)\mathtt{p}(z_0, u)\mathtt{d}u,\end{aligned}$$

where the second equality holds by reversibility and the last by duality. The proof of (2.8) is obtained similarly. □

Thus, duality allows us to calculate explicit formulas of the most complicated process using the transition of the simplest process. All processes derived from our proposal share this property.

2.2 Filtering

Let $\{(X_n, Y_n)\}_{n\geq 0}$ be a hidden Markov model, where $X = (X_n)_{n\geq 0}$ is driven by the transition probability kernel (2.3) and, given the signal, the sequence of observations $\{Y_n\}_{n\geq 0}$ are conditionally independent. The emission distribution, *i.e.* the law of $\{Y_n|X_n\}$, is chosen so that it matches with the probability measure $\mathtt{f}\,(\,\cdot\,; x)$. Then, denoting by $Y_{0:n} := (Y_0, ..., Y_n)$ and letting $\nu_n := \mathscr{L}(X_n|Y_{0:n})$, the exact or optimal filter, starting with $\nu_{-1} = \pi$, can be obtained through the recursion

$$\nu_n = \phi_{Y_n}(\psi(\nu_{n-1})), \qquad \text{for } n \geq 0, \tag{2.9}$$

where ϕ and ψ are known as the update and the predictor operators of the signal, respectively. That is to say,

$$\phi_y(\nu)(\mathtt{d}x) = \frac{\mathtt{f}\,(y; x)\nu(\mathtt{d}x)}{\mathsf{m}_\nu(y)}, \qquad \mathsf{m}_\nu(y) = \int_{\mathbb{X}} \mathtt{f}\,(\mathtt{d}y; x)\nu(x)\mathtt{d}x,$$

and

$$\psi(\nu)(\mathtt{d}x) = \int_{\mathbb{X}} \nu(v)\mathtt{k}(v, \mathtt{d}x)\mathtt{d}v.$$

for $x \in \mathbb{X}$. From the recursion (2.9) is clear that the measure ν_n depends on $y^{(n)} = (y_0, y_1, ..., y_n)$. Notice that, we denote intentionally the posterior distribution of π as the first step of the optimal filters, *i.e.* ν_0. Hence, for the sake of notation, we dropped the dependence of $y^{(n)}$ in ν_n, only making it explicit for ν_0 when such dependence is different of y_0, *i.e.* $\nu_0(\cdot; u)$.

On other hand, the prediction filters are obtained from the recursion $\psi(\nu_n)$ for $n \geq 0$. Clearly, $\psi(\nu_{-1}) = \pi$ since π is an invariant measure of the transition kernel $\mathtt{k}$. Moreover, in this case, the first step in the recursion, which turns out to be the predictor operator of the process X, takes the form

$$\psi(\nu_0)(\mathtt{d}x_1) = \int_{\mathbb{Y}} \nu_0(\mathtt{d}x_1; u)\mathtt{m}_{\nu_0}(u)\mathtt{d}u. \tag{2.10}$$

for $x_1 \in \mathbb{X}$. Here, the measure $\mathtt{m}_{\nu_0}(\mathtt{d}y)$ is the transition of the dual. This implies that we can compute (2.10) through an integral defined over the domain of the dual.

Thus, such duality exemplifies the effect that the conditional probability structure of the kernel $\mathtt{k}$ has on the calculation of the filters. Furthermore, based on the equality (2.10) one obtains

$$\begin{aligned} \mathtt{m}_{\psi(\nu_0)}(\mathtt{d}y_1) &= \int_{\mathbb{Y}} \mathtt{m}_{\nu_0}(u)\mathtt{m}_{\nu_0(u)}(y_1)\mathtt{d}u, \\ \nu_1(\mathtt{d}x_1) &= \int_{\mathbb{Y}} \omega(y^{(1)}, u)\phi_{y_1}(\nu_0(u))(\mathtt{d}x_1)\mathtt{d}u, \\ \psi(\nu_1)(\mathtt{d}x_2) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(1)}, u)\nu_0(\mathtt{d}x_2; u)\mathtt{d}u, \end{aligned}$$

where

$$\begin{aligned} \omega(y^{(1)}, u) &= \frac{\mathtt{m}_{\nu_0}(u)\mathtt{m}_{\nu_0(u)}(y_1)}{\mathtt{m}_{\psi(\nu_0)}(\mathtt{d}y_1)}, \\ \tilde{\omega}(y^{(1)}, u) &= \int_{\mathbb{Y}} \omega(y^{(1)}, u)\mathtt{p}(y_1, z, u)\mathtt{d}u, \\ \mathtt{q}(y_1, z, u) &= \int_{\mathbb{X}} \phi_{y_1}(\nu_0(u))(\mathtt{d}x_1)\mathtt{f}\,(z; x_1)\mathtt{d}x_1. \end{aligned}$$

for $x_0, x_1 \in \mathbb{X}$ and $y_0, y_1 \in \mathbb{Y}$. We used the following notation in the above equalities $\mathtt{m}_{\nu_0(u)}(y_1) = \int_{\mathbb{X}} \mathtt{f}\,(y_1; x)\nu_0(x; u)\mathtt{d}x$ and $\phi_{y_1}(\nu_0(u))(x_1) = \mathtt{f}(y_1; x_1)\nu_0(x_1; u)/\mathtt{m}_{\nu_0(u)}(y_1)$. The above equalities are obtained using (2.10) and applying Fubini's theorem to the recursion (2.9).

The next theorem provides an expression for the optimal an the prediction filters associated to X.

Theorem 1. *Let $\{(X_n, Y_n)\}_{n\geq 0}$ be a hidden Markov model, where the signal $X = \{X_n\}_{n\geq 0}$ is driven by the transition probability function k and; given the signal, the observations Y_n are conditionally independent, with emission distribution $\mathsf{f}(\cdot; x)$. If k is given by (2.3) then the optimal and prediction filters, starting at*

$$\tilde{\omega}(y_0, u) = m_{\nu_0}(u) \qquad \nu_0 = \phi_{y_0}(\pi),$$

are given by the following expressions,

$$\nu_{n+1}(\mathrm{d}x_{n+1}) = \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\phi_{y_{n+1}}(\nu_0(u))(\mathrm{d}x_{n+1})\,\mathrm{d}u, \tag{2.11}$$

$$\psi(\nu_n)(\mathrm{d}x_{n+1}) = \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\nu_0(\mathrm{d}x_{n+1}; u)\,\mathrm{d}u, \tag{2.12}$$

where

$$\begin{aligned}
m_{\psi(\nu_n)}(\mathrm{d}y_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u) m_{\nu_0(u)}(y_{n+1})\,\mathrm{d}u,\\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big) m_{\nu_0(u)}(y_{n+1})}{\mathsf{m}_{\psi(\nu_n)}(\mathrm{d}y_{n+1})},\\
\tilde{\omega}(y^{(n)}, u) &= \int_{\mathbb{Y}} \omega(y^{(n)}, k)\,\mathsf{q}(y_n, u, z)\,\mathrm{d}z\\
\mathsf{q}(y_n, u, z) &= \int_{\mathbb{X}} \phi_{y_n}(\nu_0(z))(\mathrm{d}x_n)\,\mathsf{f}(u; x_n)\,\mathrm{d}x_n,
\end{aligned}$$

for $n \geq 0$.

Proof. The proof of the theorem easily follows by induction. □

Let us notice that, if π is conjugate with respect to $\mathtt{f}(\cdot; x)$, then ν_0 and $\phi_y(\nu_0)$ belong to the same class of distributions as π. This implies that, ν_n and $\psi(\nu_n)$ are mixtures of distributions belonging to the class of distributions of π. In addition, the probability measures m_{ν_0} and q belong to the same class of distributions as m_π. Hence, one of the advantages of having a conjugate model is that it saves us the computation of the integrals corresponding to m_{ν_0} and q. Nonetheless, such property limits the choice of the signal.

The above result exemplifies the importance of the transition probability kernel of the signal in order to compute the filters. It is worth mentioning that, although the expression of the filters seems to be tractable, their computation is not simple, even for the conjugate case. This since it is still necessary to compute the weights ω and $\tilde{\omega}$.

Additionally, since the expressions in Theorem 1 of ν_{n+1} and $\psi(\nu_n)$ only depend on x_{n+1} through ν_0 and $\phi_{y_{n+1}}(\nu_0)$, respectively. The computation of statistics of the form $\mathbb{E}[\varphi(X_{t_n})|Y_t]$ and $\mathbb{E}[\varphi(X_{t_{n+1}})|Y_t]$ can be reduced to the computation of the corresponding statistics associated to $\phi_{y_{n+1}}(\nu_0(u))$ and ν_0, respectively. That is to say,

$$\mathbb{E}\left[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n+1)}\right] = \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\left(\int_{\mathbb{X}} \varphi(x_{n+1})\phi_{y_{n+1}}(\nu_0(u))(\mathtt{d}x_{n+1})\right)\mathtt{d}u, \tag{2.13}$$

$$\mathbb{E}\left[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n)}\right] = \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\left(\int_{\mathbb{X}} \varphi(x_{n+1})\nu_0(\mathtt{d}x_{n+1}; u)\right)\mathtt{d}u, \tag{2.14}$$

where ω and $\tilde{\omega}$ are given by the previous theorem. As before, if π is conjugate with respect to $\mathtt{f}$, then the above computation reduces to the computation of statistics of the form $\mathbb{E}[\varphi(X_{t_n})]$, where the expectation is taken with respect to a measure belonging to the same class of distribution of π. In this case, in order to compute the above statistics we still have to deal with the problem of computing the weights ω and $\tilde{\omega}$.

2.3 Examples

Now, we present a couple of examples that allow us to illustrate the construction proposed in this chapter. In fact, these examples are discrete-time versions of widely known continuous-time models, thus will ease the transition to such models.

2.3.1 Gamma-Poisson model

Let π and $\mathtt{f}(\cdot; x)$ be two probability measures with gamma(a, b) distribution and Poisson$(x\phi)$ distribution, for $\phi > 0$, respectively. In this case, an application of

Bayes' theorem results in the following

$$\nu_0(x;z) = \frac{(b+\phi)^{a+z}}{\Gamma(a+z)} x^{a+z-1} e^{-(b+\phi)x}.$$

i.e. ν_0 has gamma$(a+z, b+\phi)$ distribution. Hence, the distribution of π is conjugate with respect to $\mathtt{f}(\cdot;x)$. Then, one can define a reversible Markov process $X = \{X_n\}_{n\geq 0}$ driven by the well-defined one-step transition probability kernel,

$$\mathtt{k}(x_n, x_{n+1}) = e^{-x_n\phi-(b+\phi)x_{n+1}} \sum_{z=0}^{\infty} \frac{(x_n\phi)^z}{z!} \frac{(b+\phi)^{a+z}}{\Gamma(a+z)} x_{n+1}^{a+z-1},$$

for $x_n, x_{n+1} \in \mathbb{R}_+$. Moreover, the first moment associated to the operator $\mathtt{k}$ satisfies a linear relation in the mean, *i.e.*

$$\mathbb{E}[X_{n+1}|X_n = x] = (1-\rho)\mu + \rho x,$$

where $\rho = \phi/(b+\phi)$ and $\mu = \mathbb{E}[X]$. The above result can be obtained by computing the following $\mathbb{E}[\,\mathbb{E}[X_{n+1}|Y]\,|\,X_n]$. It is worth noticing that, given that $\rho \in [0,1]$, the conditional mean of $\{X_{n+1}|X_n = x\}$ is a convex combination between the expected value of the invariant distribution π and the initial value x. Additionally, the auto-correlation function for this model is given by ρ^r. The model that characterizes the process X is known as the gamma-Poisson model.

On other hand, let $Z = \{Z_n\}_{n\geq 0}$ be a reversible Markov process driven by the transition probability function

$$\mathtt{p}(z_n, z_{n+1}) = \frac{\Gamma(a+z_n+z_{n+1})}{\Gamma(a+z_n)z_{n+1}!} \left(\frac{b+\phi}{b+2\phi}\right)^{a+z_n} \left(\frac{\phi}{b+2\phi}\right)^{z_{n+1}},$$

for $z_n, z_{n+1} \in \mathbb{Z} \cup \{0\}$. Also, the process Z has invariant negative-binomial$(a+1, \phi/(b+\phi))$ distribution, *i.e.* its invariant measure has density probability function

$$\mathsf{m}_\pi(z) = \frac{\Gamma(a+z)}{\Gamma(a)z!} \left(\frac{b}{b+\phi}\right)^a \left(\frac{\phi}{b+\phi}\right)^z.$$

As we mentioned, conjugacy of π implies that the probability measure m_π and $\mathtt{q}$ belong to the same class of distributions, in this case they are negative-binomial distribution. Hence, the transition of the dual is negative-binomial measure rather than the integral of the product of two probability measure. In this sense, conjugacy saves an integral when we compute the filters.

Furthermore, as a result of the construction, X and Z are dual to each other with respect to to the duality function $\mathtt{h}$, which is define as the Radon-Nikodym between the measures ν_0 and π, *i.e.*

$$\mathtt{h}(x;z) = \frac{\Gamma(a)}{\Gamma(a+z)} \frac{(b+\phi)^{a+z}}{b^a} x^z e^{-\phi x},$$

for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$.

Now, consider the hidden Markov model where X is the signal, $Y = (Y_n)_{n\geq 0}$ is the observation process and $\mathtt{f}$ is the emission density. Starting with the measures ν_0 and $\tilde{\omega}(y_0, u)$, which are $\mathrm{Ga}(x_0; a + y_0, b + \phi)$ and $\mathrm{NB}(u; a + y_0 + 1, b/(b + 2\phi))$ distributions, respectively. The optimal and the prediction filters are given by

$$\begin{aligned}
\nu_{n+1}(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{\infty} \omega\big(y^{(n+1)}, u\big)\, \mathrm{Ga}\big(\mathtt{d}x_{n+1}; a + u + y_{n+1}, b + 2\phi\big), \\
\psi(\nu_n)(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{\infty} \tilde{\omega}(y^{(n)}, u)\, \mathrm{Ga}\big(\mathtt{d}x_{n+1}; a + u, b + \phi\big),
\end{aligned}$$

respectively, where

$$\begin{aligned}
\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \sum_{u=0}^{\infty} \tilde{\omega}(y^{(n)}, u)\, \mathrm{NB}\Big(\mathtt{d}y_{n+1}; a + u + 1, \frac{b}{b+2\phi}\Big), \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big)\mathrm{NB}\Big(y_{n+1}; a + u + 1, \frac{b}{b+2\phi}\Big)}{\mathtt{m}_{\psi(\nu_n)}(dy_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \sum_{k=0}^{\infty} \omega(y^{(n)}, k)\mathrm{NB}\Big(k; a + u + y_1 + 1, \frac{\phi}{b+3\phi}\Big).
\end{aligned}$$

for $n \geq 0$. Additionally, as we mentioned, one can compute statistics of the form (1.11), for instance the moment generating function of the optimal and prediction filters take the form

$$\begin{aligned}
\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}} \Big| \mathscr{Y}^{(n+1)}\Big] &= \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\Big(1 - \frac{\lambda}{b+2\phi}\Big)^{-(a+u+y_{n+1})} \mathtt{d}u, \\
\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}} \Big| \mathscr{Y}^{(n)}\Big] &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\Big(1 - \frac{\lambda}{b+\phi}\Big)^{-(a+u)} \mathtt{d}u,
\end{aligned}$$

where ω and $\tilde{\omega}$ are given as before. The above expressions are possible due to the conjugacy of π. Later we will see that non-conjugate cases have more complicated expressions. Notice that, the computation of both, the filters and their moment generating functions requires the computation of $\omega\big(y^{(n+1)}, u\big)$ and $\tilde{\omega}(y^{(n)}, u)$.

2.3.2 Generalized Poisson model

The generalized Poisson model, proposed by Anzarut, Mena, Nava and Prünster [?], is an extension of the gamma-Poisson model. Indeed, it replaces the gamma measure for any $\mathbb{R}_+$-valued probability measure. Specifically, let π be an absolutely continuous probability density function supported on $\mathbb{R}_+$, and as before let $\mathtt{f}(\cdot; x)$ be a Poisson$(x\phi)$ measure for some positive value ϕ. This implies that,

$$\nu_0(x; z) = \frac{x^z e^{-x\phi}\pi(x)}{\xi(z, \phi)}, \tag{2.15}$$

where $\xi(z, \phi) = \int_0^\infty u^z e^{-u\phi}\pi(u)\mathtt{d}u$, for $z \in \mathbb{N} \cup \{0\}$. The function (2.15) is defined as the Poisson weighted density. Moreover, the Laplace transform associated to measures with density function of the form (2.15) are given by,

$$\mathscr{L}_X(\lambda) = \frac{\xi(z, \phi + \lambda)}{\xi(z, \phi)}, \tag{2.16}$$

where $\mathscr{L}_X(\lambda) := \mathbb{E}[e^{-\lambda X}]$. Then, one can define a reversible Markov process $X = \{X_n\}_{n\geq 0}$ driven by the one-step transition probability kernel

$$\mathtt{k}(x_n, x_{n+1}) = \exp\{-\phi(x_{n+1} + x_n)\}\pi(x_{n+1}) \sum_{z=0}^{\infty} \frac{(x_{n+1}x_n\phi)^z}{z!\xi(z, \phi)},$$

for $x_n, x_{n+1} \in \mathbb{R}^+$ and with invariant distribution π. Additionally, the Laplace transform associated to the operator $\mathtt{k}$ takes the form

$$\mathscr{L}_{X_{n+1}|X_n}(\lambda) = \sum_{z=0}^{\infty} e^{-x\phi} \frac{(x\phi)^z}{z!} \frac{\xi(z, \phi + \lambda)}{\xi(z, \phi)}.$$

Clearly, the computation of the above operator is characterized by ξ, which depend of π. As a consequence, there is not a general expression for the first moment of the kernel $\mathtt{k}$, so there is no linear relation for the mean as in the gamma case.

On other hand, let $Z = \{Z_n\}_{n\geq 0}$ be a reversible Markov process driven by the following transition probability kernel,

$$\mathtt{p}(z_n, z_{n+1}) = \frac{\phi^{z_{n+1}}}{\xi(z_n,\phi) z_{n+1}!} \int_0^\infty x^{z_n+z_{n+1}} e^{-2x\phi} \pi(x) \mathtt{d}x,$$

for $z_n, z_{n+1} \in \mathbb{Z} \cup \{0\}$. Also, the invariant measure of the process Z takes the form $\mathtt{m}_\pi(z) = \xi(z,\phi)\,\phi^z/z!$. Furthermore, the process X and Z are dual to each with respect to the function $\mathtt{h}(x;z) = x^z e^{-x\phi}/\xi(z,\phi)$.

It is worth noticing that, given that in general π is not conjugate, so the expression of the filters can not be reduced to simpler expressions. Nevertheless, the computation of the filters is still tractable in the sense that the integral seen as a finite sum of positive summands can be truncated. In fact, the optimal and prediction filters are given by (2.11) and (2.12), where ν_0 is (2.15) and

$$\phi_{y_{n+1}}(\nu_0(u))(\mathtt{d}x_{n+1}) = \frac{\mathtt{f}(y_{n+1}; x_{n+1})\nu_0(x_{n+1}; u)}{\mathtt{m}_{\nu_0(u)}(y_{n+1})}.$$

Similarly, the statistic associated to the filters cannot be reduced to a simpler expression, so one has to replace directly the corresponding measures. Thus, these two examples show the advantages of using conjugate models. Additionally, to the best of our knowledge, the gamma case is the only one with a continuous-time version within this class.

2.3.3 Beta-binomial model

The beta-binomial model, found in Pitt, Chatfield and Walker [37], is constructed based on a thinning operation. Indeed, let π be a beta($\mathbf{a}$) measure and $\{\mathtt{f}(z_1;x); x \in (0,1)\}$ be a set of binomial($z_1; |\mathbf{z}|, x$) measures, where $\mathbf{a} = (a_1, a_2)$ and $|\mathbf{z}| = z_1 + z_2$ is fixed. Hence, the measure π is conjugate with respect to $\mathtt{f}$, which implies that

$$\nu_0(x;\mathbf{z}) = \frac{x^{a_1+z_1-1}(1-x)^{a_2+z_2-1}}{\mathsf{B}(\mathbf{a}+\mathbf{z})}, \qquad \text{for } x \in [0,1],$$

where B is the beta function, *i.e.* $\mathsf{B}(\mathbf{a}) = \Gamma(a_1)\Gamma(a_2)/\Gamma(|\mathbf{a}|)$, with $|\mathbf{a}| = a_1 + a_2$. Then, one an define a Markov processes $X = \{X_n\}_{n\geq 0}$ driven by the transition probability function,

$$\mathtt{k}(x_n, x_{n+1}) = \sum_{z_1=0}^{|\mathbf{z}|} \text{bin}(z_1; |\mathbf{z}|, x_n)\text{beta}(x_{n+1}; \mathbf{a} + \mathbf{z}),$$

for $x_n, x_{n+1} \in (0,1)$. Additional, the process X has invariant beta($\mathbf{a}$) distribution.

On other hand, we define the reversible Markov process $Z = \{Z_n\}_{n\geq 0}$ driven by the transition probability function,

$$\mathtt{p}(\mathbf{z}_n, \mathbf{z}_{n+1}) = \binom{|\mathbf{z}_{n+1}|}{z_1^{n+1}} \frac{\mathsf{B}(\mathbf{a} + \mathbf{z}_n + \mathbf{z}_{n+1})}{\mathsf{B}(\mathbf{a} + \mathbf{z}_n)}$$

for $\mathbf{z}_n = (z_1^n, z_2^n)$ and $\mathbf{z}_{n+1} = (z_1^{n+1}, z_2^{n+1})$ such that $|\mathbf{z}_n| = |\mathbf{z}_{n+1}|$ for all $n \geq 0$. Hence, the invariant distribution associated to $\mathtt{p}$ is given by

$$\mathsf{m}_\pi(z_1, |\mathbf{z}|) = \binom{|\mathbf{z}|}{z_1} \frac{\mathsf{B}(\mathbf{a} + \mathbf{z})}{\mathsf{B}(\mathbf{a})},$$

for $y \in \{0, 1, ..., n\}$. Thus, m_π and $\mathtt{p}$ are conjugate belonging to the class of beta-binomial distributions. Furthermore, for this model, the duality function takes the form

$$\mathtt{h}(x; \mathbf{z}) = \frac{\mathsf{B}(\mathbf{a})}{\mathsf{B}(\mathbf{a} + \mathbf{z})} x^{z_1}(1-x)^{z_2}.$$

This duality suggests a bivariate version of the moment duality attained for the Wright-Fisher model. In fact, such a model helps us to make clear that reversibility plays an important role in duality. Indeed, the dual to the Wright-Fisher model can be seen as a process that goes back in time.

Now, in order to provide an expression for the filters notice that ν_0 and $\tilde{\omega}(\mathbf{m}_0, \mathbf{u})$ are Beta($\mathbf{a} + \mathbf{m}_0$) and beta-binomial($|\mathbf{z}_{n+1}|, \mathbf{a} + \mathbf{z}_n$), respectively. Thus, the optimal and prediction filters take the form

$$\begin{aligned} \nu_{n+1}(\mathtt{d}x_{n+1}) &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2} \omega\big(\mathbf{m}^{(n+1)}, \mathbf{u}\big) \,\text{Beta}(x_{n+1}; \mathbf{a} + \mathbf{u} + \mathbf{m}_{n+1}), \\ \psi(\nu_n)(\mathtt{d}x_{n+1}) &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2} \tilde{\omega}(\mathbf{m}^{(n)}, \mathbf{u}) \,\text{Beta}(x_n; \mathbf{a} + \mathbf{u}), \end{aligned}$$

where

$$\begin{aligned}
\mathsf{m}_{\psi(\nu_n)}(\mathrm{d}\mathbf{m}_{n+1}) &= \sum_{u=0}^{\infty}\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u})\binom{|\mathbf{m}_{n+1}|}{m_{n+1}^1}\frac{\mathsf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})}{\mathsf{B}(\mathbf{a}+\mathbf{u})},\\
\omega\big(\mathbf{m}^{(n+1)},\mathbf{u}\big) &= \frac{\tilde{\omega}\big(\mathbf{m}^{(n)},\mathbf{u}\big)\binom{|\mathbf{m}_{n+1}|}{m_{n+1}^1}\frac{\mathsf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})}{\mathsf{B}(\mathbf{a}+\mathbf{u})}}{\mathsf{m}_{\psi(\nu_n)}(\mathrm{d}\mathbf{m}_{n+1})},\\
\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u}) &= \sum_{k=0}^{\infty}\omega(\mathbf{m}^{(n)},\mathbf{k})\binom{|\mathbf{k}|}{k_1}\frac{\mathsf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_n+\mathbf{k})}{\mathsf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_n)}.
\end{aligned}$$

for $n \geq 0$, with $\mathbf{m}_n = (m_n^1, m_n^2)$. As a consequence of the above result, statistics associated to the filters are give by the following expressions,

$$\begin{aligned}
\mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n+1)}\Big] &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\omega(\mathbf{m}^{(n+1)},\mathbf{u})\bigg(\int_0^1\varphi(x_{n+1})\mathrm{Beta}(x_{n+1};\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})\bigg)\mathrm{d}\mathbf{u},\\
\mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n)}\Big] &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u})\bigg(\int_0^1\varphi(x_{n+1})\mathrm{Beta}(x_{n+1};\mathbf{a}+\mathbf{u})\bigg)\mathrm{d}\mathbf{u}.
\end{aligned}$$

Since there is not a simple expression for the moment generating function of a beta distribution, the above result can be used to calculate, for instance, the moments associated to the filters.

2.4 Stationary first-order autoregressive models with exponential dispersion margin

The purpose of this section is to provide a mechanism for choosing the measures π and $\mathtt{f}$ used to construct the transition probabilities $\mathtt{k}$ in Section 1.1. Having this in mind, we will provide a brief review to a widely known model that falls within our construction. Such model, proposed by Jorgensen and Song [17], consists on stationary first-order autoregressive AR(1) processes with invariant distribution belonging to the class exponential dispersion distributions. These processes are built based on a thinning operation, which, given π, gives us an expression for the distribution associated to $\mathtt{f}$. Then, our proposal allows to find the dual associated to the mentioned processes. Moreover, assuming that the signal of a hidden Markov model is given by a stationary AR(1) model, we will give expressions for the optimal and the prediction filters.

2.4.1 Exponential dispersion distributions

The class of exponential dispersion distributions consists of the set of probability measures $\mathbb{R}^d$-valued, for some $d \geq 1$, whose density probability function has the following form,

$$\pi(x) = c(x;\tau) \exp\left\{x\vartheta - \tau * \kappa(\vartheta)\right\}, \tag{2.17}$$

where

$$\kappa(\vartheta) = \frac{1}{\tau} \log\left\{ \int c(x;\tau) e^{\vartheta x} \mathtt{d}x \right\},$$

for $\vartheta \in \Xi$, with $\Xi = \mathsf{int}\{\vartheta \in \mathbb{R}^d; \kappa(\vartheta) < \infty\}$ and Ξ is assumed to be non-empty. The class of random variables with density probability function of the form (2.17) are denoted by $ED(\vartheta;\tau)$, where ED stands for exponential dispersion. Additionally, assuming that $\tau \in \mathbb{R}_+$, any random variable X with exponential dispersion distribution is infinitely divisible. Clearly an exponential dispersion model is specified via the function $c(\cdot,\tau)$ or equivalently by specifying the function $\kappa(\vartheta)$. Moreover, the moment generating function of exponential dispersion distributions are given by,

$$\mathbb{E}\left[e^{\lambda X}\right] = \exp\left\{\tau\left[\kappa(\vartheta+\lambda) - \kappa(\vartheta)\right]\right\}, \qquad \text{for } \lambda \in \mathbb{R}.$$

where X denotes a random variable with $\mathrm{ED}(\vartheta;\tau)$ distribution. This implies, through simple differentiation that, the moments associated to any exponential dispersion distribution take the form,

$$\begin{aligned}
\mathbb{E}[X] &= \tau\kappa'(\vartheta) \\
\mathbb{E}[X^2] &= \tau\kappa''(\vartheta) + \tau^2[\kappa'(\vartheta)]^2 \\
\mathbb{E}[X^3] &= \tau\kappa^{(3)}(\vartheta) + 3\tau^2\kappa''(\vartheta)\kappa'(\vartheta) + \tau^3[\kappa'(\vartheta)]^3 \\
\mathbb{E}[X^4] &= \tau\kappa^{(4)}(\vartheta) + 4\tau^2\kappa^{(3)}(\vartheta)\kappa'(\vartheta) + 3\tau^2[\kappa''(\vartheta)]^2 + 6\tau^3\kappa''(\vartheta)[\kappa'(\vartheta)]^2 \\
&\quad +\tau^4[\kappa'(\vartheta)]^4,
\end{aligned}$$

and so on, provided the required moments of X exist. Also, denoting by $\mu = \mathbb{E}[X]$ and $\mu_r := \mathbb{E}[(X-\mu)^r]$ the following equations hold,

$$\begin{aligned} \mu_1 &= 0 \\ \mu_2 &= \tau\kappa''(\vartheta) \\ \mu_3 &= \tau\kappa^{(3)}(\vartheta) \\ \mu_4 &= \tau\kappa^{(4)}(\vartheta) + 3\tau^2[\kappa''(\vartheta)]^2, \end{aligned}$$

and so on. Then, the class of convolution-closed infinitely divisible families developed in Joe [16] that are not exponential dispersion models are those without moment generating functions. Nonetheless, a distribution without moment generating function may be considered by letting π in (2.17) with $\Xi = \{\vartheta_0\}$, *i.e.* consisting of a single point; without loss of generality $\vartheta_0 = 0$. Thus, the $ED(0;\tau)$ model corresponds to the class of distributions with characteristic functions ψ^τ, where ψ is a given characteristic function, and any $ED(0;\tau)$ distribution still satisfies some convolution formula. Moreover, in such case, the model is infinitely divisible if and only if $\tau \in \mathbb{R}^+$. With this convention, all distributions in Joe [16] belong to the class of convolution-closed infinitely divisible families (*cf.* Jorgensen and Song [17]).

2.4.2 Thinning operation

Letting Y and Z two independent random variables with $ED(\vartheta;\tau_1)$ and $ED(\vartheta;\tau_2)$ distribution, respectively. Then, the density probability function of $X := Y + Z$, denoted by π, is obtained in the following way

$$\begin{aligned} \pi(\mathtt{d}x) &= \exp\left\{x\vartheta - \tau\kappa(\vartheta)\right\}\int_{\mathbb{R}} c(x-z;\tau_2)c(z;\tau_1)\mathtt{d}z \\ &= c(x;\tau)\exp\left\{x\vartheta - \tau\kappa(\vartheta)\right\}, \end{aligned}$$

where $\tau = \tau_1 + \tau_2$. Thus, the class of exponential dispersion distributions are closed under convolutions. Such property is used to define a thinning operation. Indeed, the conditional probability of Z, given $X = x$, denoted by $G(\tau_1, \tau_2, x)$, takes the form

$$\begin{aligned} \mathtt{f}(z;x) &= \frac{\mathbb{P}[Z=\mathtt{d}z, Y=x-z]}{\pi(x)} \\ &= \frac{c(x-z;\tau_2)c(z;\tau_1)}{c(x;\tau)}, \end{aligned} \tag{2.18}$$

with $\tau=\tau_1+\tau_2$. The distribution $G(\tau_1,\tau_2,x)$ is called the contraction corresponding to $ED(\vartheta;\tau)$. This terminology comes from the fact that if $ED(\vartheta;\tau)$ is non-negative then $G(\tau_1,\tau_2,x)$ is concentrated in the interval $(0,x)$. It is worth to mention that, the density probability function $\mathtt{f}(\cdot;x)$ does not depend on ϑ. Furthermore, if the first moment of X exists and $\tau_1=\Theta\tau$ with $\Theta\in(0,1)$, then

$$\mathbb{E}\big[\mathbb{E}[Z|X]-\Theta X\big]=\mathbb{E}[Z]-\Theta\tau\kappa'(\vartheta)\equiv 0,$$

where the last equality holds because $\mathbb{E}[Z]=\tau\Theta\kappa'(\vartheta)$, and, from completeness arguments, it follows that $\mathbb{E}[Z|X]=\Theta X$. This result reinforces the notion of thinning. At this point, it is not straightforward to obtain an expression for the rest of the moments of exponential dispersion random variables.

For the model describe above, Bayes' theorem allows us to calculate the posterior distribution of π, denoted by ν_0, which is given by

$$\nu_0(x;z)=c(x-z;\tau_2)\exp\{(x-z)\vartheta-\tau_2\kappa(\vartheta)\}.$$

where $\tau_2=(1-\Theta)\tau$. Clearly, ν_0 is a $\mathrm{ED}(\vartheta;\tau_2)$ probability measure, or equivalently, the conditional distribution $\{X-Z|Z\}\overset{d}{=}\{X-Z\}$ has $\mathrm{ED}(\vartheta;\tau_2)$ distribution. The later result holds because $\{X-Z\}$ and Z are independent. As a result, the conditional distribution of $\{X|Z\}$ is a translation of $ED(\vartheta;\tau_2)$ distribution.

Then, one can define a time-homogeneous stationary Markov process $(\mathbb{X},\mathscr{X})$-valued, denoted by $X=\{X_n\}_{n\geq 0}$, with invariant distribution π, driven by the transition probability function,

$$\mathtt{k}(x_n,x_{n+1})=\int_{\mathbb{Y}}\nu_0(x_{n+1};z)\mathtt{f}(z;x_n)\mathtt{d}z, \qquad \text{for } x_n,x_{n+1}\in\mathbb{X}.$$

For this model, $\mathbb{X}\subset\mathbb{R}^d$, $\mathscr{X}=\mathscr{B}(\mathbb{R}^d)$, for $d\geq 1$, and $\mathbb{Y}\subset\mathbb{X}$. Moreover, the dynamics of the process associated to the above transition can be also written through the following stochastic equation,

$$X_n = A_n(X_{n-1}, \Theta) + \epsilon_n, \qquad n = 1, 2, ... \tag{2.19}$$

where the sequence $\{\epsilon_n\}_{n\geq 1}$ of random variables are i.i.d. with common $ED(\vartheta; \tau_2)$ distribution and, the conditional distribution of $\{A_n(X_{n-1}, \Theta)|X_{n-1} = x\}$ is given by $G(\tau_1, \tau_2, x)$. The random operator A_n defines a thinning operation and, we say that $A_n(X_{n-1}; \Theta)$ is the thinning of X by the proportion Θ.

Furthermore, the first moments associated to the transition kernel $\mathtt{k}$ are computed by using the conditional probability structure and the moments of exponential dispersion distributions, *i.e.*

$$\begin{aligned} \mathbb{E}[X_{n+1}|X_n] &= \mathbb{E}[\,\mathbb{E}[X_{n+1}|Z]\,|\,X_n] \\ &= \mathbb{E}[X_{n+1} - Z] + \mathbb{E}[\,Z\,|\,X_n] \\ &= (1-\Theta)\mu + \Theta X_n, \end{aligned}$$

where the second and third equality holds because $\{X_n - Y|Y\} \stackrel{d}{=} \{X_n - Y\} \sim ED(\vartheta; \tau_2)$, with $\mu := \mathbb{E}[X_n] = \tau k'(\vartheta)$ for all $n \geq 1$. Thus, the process X has a linear relation in its mean. Moreover, in order to compute the autocorrelation of the process, let us notice that

$$\begin{aligned} \mathbb{E}[X_{n+2}|X_n] &= \mathbb{E}[\,\mathbb{E}[X_{n+2}|X_{n+1}]\,|\,X_n] \\ &= \mathbb{E}[\,(1-\Theta)\mu + \Theta X_{n+1}\,|\,X_n] \\ &= (1-\Theta^2)\mu + \Theta^2 X_n. \end{aligned}$$

Following this recursion we can obtain, for $h \geq 1$, that

$$\mathbb{E}[X_{n+h}|X_n] = (1-\Theta^h)\mu + \Theta^h X_n,$$

Similarly,

$$\mathbb{E}[X_{n+h}X_n] = \mathbb{E}[X_n \mathbb{E}[X_{n+h}|X_n]] = \Theta^h \mathbb{E}[X_n^2] + (1-\Theta^h)\mu^2.$$

Hence, using the stationarity of the process X and the above equalities, the autocorrelation takes the form

$$\mathrm{Corr}(X_{n+h}, X_n) = \frac{\mathbb{E}[X_{n+h}X_n] - \mu^2}{\mathbb{E}[X_n^2] - \mu^2} = \Theta^h,$$

for $h \geq 1$. This provides with an extra parameter that could be used to fit a required autocorrelation. Note that, moments of order greater than one do not have a general expression given that the moments of $G(\tau_1, \tau_2, x)$ are unknown.

From the construction in Section 1.1, we known that the mechanism used in this model guarantees the existence of a dual process. The following section provides the form of the duality function and, also, the transition probabilities that drive the corresponding dual.

2.4.3 Duality of stationary AR(1) models

Let us define a time-homogeneous stationary Markov process $\{Z_n\}_{n\geq 0}$ driven by the transition probability kernel,

$$\mathtt{p}(z_n, z_{n+1}) = \int_{\mathbb{X}} \mathtt{f}(z_{n+1}; x)\nu_0(x; z_n)\mathrm{d}x, \qquad \text{for } z_n, z_{n+1} \in \mathbb{Y},$$

and invariant measure m_π, where m_π is a $\mathrm{ED}(\vartheta; \tau_1)$ distribution. Moreover, the first moment associated to the transition $\mathtt{p}$ also have a linear relation, *i.e.*

$$\begin{aligned}
\mathbb{E}[Z_{n+1}|Z_n] &= \mathbb{E}[\,\mathbb{E}[Z_{n+1}|X] \,|\, Z_n] \\
&= \mathbb{E}[\,\Theta X \,|\, Z_n] \\
&= \Theta(1-\Theta)\mu + \Theta Z_n,
\end{aligned}$$

where the third equality holds because $\{X - Z_n\} \sim \mathrm{ED}(\vartheta; \tau_2)$, with $\mu = \tau\kappa'(\vartheta)$. Then, the autocorrelation of the process Z can be obtained as following

$$\begin{aligned}
\mathbb{E}[Z_{n+2}|Z_n] &= \mathbb{E}[\,\Theta(1-\Theta)\mu + \Theta Z_{n+1} \,|\, Z_n] \\
&= \Theta(1-\Theta^2)\mu + \Theta^2 Z_n \\
\vdots \quad & \vdots \qquad \vdots \\
\mathbb{E}[Z_{n+h}|Z_n] &= \Theta(1-\Theta^h)\mu + \Theta^h Z_n,
\end{aligned}$$

and

$$\mathbb{E}[Z_{n+h}Z_n] = \Theta^h \mathbb{E}[Z_n^2] + \Theta^2(1-\Theta^h)\mu^2,$$

for $h \geq 1$. Thus, the stationarity of the process $\{Z_n\}_{n\geq 0}$ and the above equalities leads to an autocorrelation given by

$$\mathrm{Corr}(Z_{n+h}, Z_n) = \frac{\mathbb{E}[Z_{n+h}Z_n] - \Theta^2\mu^2}{\mathbb{E}[Z_n^2] - \Theta^2\mu^2} = \Theta^h.$$

Notice that, the autocorrelation of the processes $\{X_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$ are the same and only depend on Θ. Furthermore, the processes X and Z are dual to each other with respect to the duality function,

$$\mathtt{h}(x; z) = \frac{c(x - z; \tau_2)}{c(x; \tau)} \exp\{\tau_1\kappa(\vartheta) - z\vartheta\}.$$

for $x \in \mathbb{X}$ and $z \in \mathbb{Y}$. Thus, using a thinning operation, one can define two time-homogeneous reversible Markov processes with exponential dispersion distributions as invariant measures. These processes turn out to be dual to each other and, also, they share a linear mean property and the same autocorrelation function.

2.4.4 Filtering

Let $\{(X_n, Y_n)\}_{n\geq 0}$ be a hidden Markov model, where the signal $X = \{X_n\}_{n\geq 0}$ is a stationary AR(1) model, as described above; $\{Y_n\}_{n\geq 0}$ be a sequence of observations that are conditionally independent, given the signal; and with emission density (2.18). For this model, we have that

$$\nu_0(x_0) = c(x - y_0; \tau_2) \exp\big\{(x_0 - y_0)\vartheta - \tau_2\kappa(\vartheta)\big\}.$$

Let us notice that, the probability measure $\nu_0(x; z)$ as a function of $x - z$ is an $\mathrm{ED}(\vartheta; \tau_2)$, with $\tau_2 = \tau(1 - \Theta)$. This implies that, in general, π is not conjugate with respect to $\mathtt{f}$. Later, we will see that, as an exception, the normal model turns out to be conjugate model. From Theorem 1, we know that

$$\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) \quad = \quad \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\mathtt{p}(u, y_{n+1})\mathtt{d}u,$$

where

$$\begin{aligned}\omega\big(y^{(n+1)},u\big) &= \frac{\tilde{\omega}\big(y^{(n)},u\big)\mathtt{p}(u,y_{n+1})}{m_{\psi(\nu_n)}(dy_{n+1})},\\ \tilde{\omega}(y^{(n)},u) &= \int_{\mathbb{X}}\nu_n(dx_n)\mathtt{f}(u;x_n)dx_n,\end{aligned}$$

for $n \geq 0$. Thus, the optimal filters and their corresponding statistics can be computed from the following equalities

$$\begin{aligned}\nu_{n+1}(\mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}}\omega(y^{(n+1)},u)\phi_{y_{n+1}}(\nu_0(u))(\mathtt{d}x_{n+1})\mathtt{d}u,\\ \mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n+1)}\Big] &= \int_{\mathbb{Y}}\omega(y^{(n+1)},u)\bigg(\int_{\mathbb{X}}\varphi(x_{n+1})\phi_{y_{n+1}}(\nu_0(u))(\mathtt{d}x_{n+1})\bigg)\mathtt{d}u.\end{aligned}$$

respectively, for $n \geq 0$. On the other hand, it is known from (2.12) that, the predictor filters are given by the following expression

$$\psi(\nu_n)(\mathtt{d}x_{n+1}) = \int_{\mathbb{Y}}\tilde{\omega}(y^{(n)},u)\nu_0(\mathtt{d}x_{n+1};u)\mathtt{d}u,$$

for $n \geq 0$. Here, we use the notation assigned to the posterior distribution of π, *i.e.*

$$\nu_0(\mathtt{d}x_{n+1};u) = c(x_{n+1}-u;\tau_2)\exp\{(x_{n+1}-u)\vartheta-\tau_2\kappa(\vartheta)\}.$$

Hence, from the properties associated to exponential dispersion distributions, the moment generating function associated to the predictor filters takes the form,

$$\begin{aligned}\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] &= \int_{\mathbb{Y}}e^{\lambda u}\tilde{\omega}(y^{(n)},u)\bigg(\int_{\mathbb{X}}e^{\lambda(x_{n+1}-u)}\nu_0(x_{n+1};u)\mathtt{d}x_{n+1}\bigg)\mathtt{d}u\\ &= \exp\Big\{\tau_2\big[\kappa(\vartheta+\lambda)-\kappa(\vartheta)\big]\Big\}\int_{\mathbb{Y}}e^{\lambda u}\tilde{\omega}(y^{(n)},u)\mathtt{d}u\end{aligned}$$

Similarly, one can obtain expressions for the moments of the predictor filters. Let us emphazise that, the purpose of this section is to provide a mechanism for choosing the measures π and $\mathtt{f}$, which characterize our proposal. Moreover, given the nature of the construction, we were allowed to give an expression for the previous filters and its corresponding statistics.

2.4.5 Examples

The normal, Poisson, gamma, binomial and negative-binomial distributions were called as the univariate natural exponential families with quadratic variance function, Morris [31]. These families, together with the generalized hyperbolic secant, belong to the class of exponential dispersion distributions. Thus, we apply a thinning operation to these families in order to built reversible Markov processes. For each of these models, we find its dual and, its corresponding filters.

Poisson model

Let X be a random variable with Poisson(λe^{ϑ}) distribution, where $\lambda e^{\vartheta} > 0$, denoted by π. It is known that π belongs to the class of exponential dispersion distributions with $c(x;\tau) = \tau^x/x!$ and $k(\vartheta) = e^{\vartheta}$, where $\tau = \lambda$. Hence, letting $\tau_1 = \Theta\tau$ and $\tau_2 = (1-\Theta)\tau$, the contraction density function takes the form

$$\mathtt{f}(z;x) = \binom{x}{z}\Theta^z(1-\Theta)^{x-z},$$

for $z \in \{0,1,...,x\}$. Thus, if Z and Y are Poisson$(\tau_1 e^{\vartheta})$ and Poisson$(\tau_2 e^{\vartheta})$ distribution, respectively. Then, $X = Y + Z$ and the conditional $\{Z|X = x\}$ has binomial(x,Θ) distribution, for each $x \in \mathbb{Z}^+ \cup \{0\}$, denoted by $\mathtt{f}$. Hence, the distribution of $\{X|Z\}$ has density probability function

$$\nu_0(x;z) = e^{-\tau_2 e^{\vartheta}}\frac{\left(\tau_2 e^{\vartheta}\right)^{x-z}}{(x-z)!},$$

for $x \in \{z,...,\infty\}$. Note that, ν_0 is not a Poisson distribution as a function of x, which means that π is not conjugate with respect to $\mathtt{f}(\cdot;x)$. Nonetheless, clearly as a function of $x-z$ the measure ν_0 has Poisson distribution.

Now, define a time-homogeneous stationary Markov process $\{X_n\}_{n\geq 0}$ driven by the transition probability kernel,

$$\mathtt{k}(x_n,x_{n+1}) = e^{-\tau_2 e^{\vartheta}}(1-\Theta)^{x_n}\sum_{z=0}^{x_n\wedge x_{n+1}}\frac{\left(\tau_2 e^{\vartheta}\right)^{x_{n+1}-z}}{(x_{n+1}-z)!}\binom{x_n}{z}\left(\frac{\Theta}{1-\Theta}\right)^z,$$

for x_n, $x_{n+1} \in \mathbb{Z}^+ \cup \{0\}$, and the invariant measure π. For this model, the dynamics of $\{X_n\}_{n\geq 0}$ can be also written through the stochastic equation

$$X_n = A_n(X,\Theta) + \epsilon_n, \qquad \text{for } n \geq 1,$$

where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common Poisson$(\tau_2 e^{\vartheta})$ distribution and $A_n(X,\Theta)$ has binomial(X,Θ) distribution. In fact, the process $\{X_n\}_{n\geq 0}$ turns out to be a birth and death with immigration process.

On other hand, letting $\{Z_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{p}(z_n, z_{n+1}) = \frac{e^{-\tau_2 e^{\vartheta}}}{(\tau_2 e^{\vartheta})^{z_n}} \left(\frac{\Theta}{1-\Theta}\right)^{z_{n+1}} \sum_{v=z_n \vee z_{n+1}}^{\infty} \frac{\left(\tau_2 e^{\vartheta}\right)^v}{(v-z_n)!} \binom{v}{z_{n+1}} (1-\Theta)^v,$$

for $y_n,\ y_{n+1} \in \mathbb{Z}^+ \cup \{0\}$ such that $y_n \leq y_{n+1}$; and invariant Poisson$(\tau_1 e^{\vartheta})$ distribution. For this model, the duality function $\mathtt{h}$, defined as the Radon-Nikodyn derivative between the posterior and prior distribution of π, *i.e.*

$$\mathtt{h}(x;z) = \frac{x! e^{\tau_1 e^{\vartheta}}}{(x-z)!} \frac{(1-\Theta)^x}{(\tau_2 e^{\vartheta})^z},$$

guarantees the duality between the processes $\{X_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$.

For the sake of notation, we compute the filters of the signal assuming that $\vartheta = 0$. Starting with ν_0 and $\tilde{\omega}(y_0, u)$ which are given by a Po$(x_0 - y_0; \tau_2)$ distribution and the kernel $\mathtt{p}(y_0, u)$, respectively. The optimal and prediction filters are given by

$$\begin{aligned} \nu_{n+1}\big(\mathtt{d}x_{n+1}\big) &= \sum_{u=0}^{x_{n+1}} \omega\big(y^{(n+1)}, u\big) \frac{\mathrm{Bin}(y_{n+1}; x_{n+1}, \Theta)\mathrm{Po}\big(x_{n+1} - u; \tau_2\big)}{\mathtt{p}(u, y_{n+1})}, \\ \psi\big(\nu_n\big)(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \tilde{\omega}(y^{(n)}, u)\mathrm{Po}\big(x_{n+1} - u; \tau_2\big), \end{aligned}$$

where

$$\begin{aligned} \mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \sum_{u=0}^{\infty} \tilde{\omega}(y^{(n)}, u)\mathtt{m}_{\nu_0(u)}(y_{n+1}), \\ \omega\big(y^{(n+1)}, u\big) &= \frac{\mathtt{m}_{\nu_0(u)}(y_{n+1})\tilde{\omega}\big(y^{(n)}, u\big)}{\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1})}, \\ \tilde{\omega}(y^{(n)}, u) &= \sum_{x_n = u \vee u}^{\infty} \nu_n(dx_n)\mathrm{Bin}(u; x_n, \Theta). \end{aligned}$$

and $\nu_0(u)(x) = \text{Po}(x - u; \tau_2)$, with $n \geq 0$. Thus, the recursion requires to compute some infinite sums. However, since we are adding only positive terms, one can truncate such sums without losing accuracy in the result. Furthermore, the moment generating function associated to the prediction filters take the form

$$\mathbb{E}\Big[e^{\gamma X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] = \exp\left\{\tau_2(e^{\gamma} - 1)\right\}\sum_{u=0}^{\infty} e^{\gamma u}\tilde{\omega}(y^{(n)}, u),$$

where ω and $\tilde{\omega}$ are given before. For the optimal filters, the corresponding statistics are not that simple but it can be computed by replacing,

$$\phi_{y_{n+1}}(\nu_0(u))(x_{n+1}) = \frac{\text{Bin}(y_{n+1}; x_{n+1}, \Theta)\text{Po}\big(x_{n+1} - u; \tau_2\big)}{\mathtt{p}(u, y_{n+1})},$$

We would like to finish by noticing that, this model is not conjugate and, also, to our knowledge, the continuous version of this model represents the only Markov process whose invariant measure is not conjugate.

Normal model

The normal(μ, σ^2) distribution, denoted by π, with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$, belongs to the class of exponential dispersion distributions with $c(x; \tau) = (2\pi\tau)^{-1/2}\exp\left\{-x^2/2\tau\right\}$ and $\kappa(\vartheta) = \vartheta^2/2$, where $\tau = \sigma^2$ and $\vartheta = \mu/\tau$. Letting $\tau_1 = \Theta\tau$ and $\tau_2 = (1 - \Theta)\tau$ with $\Theta \in (0, 1)$, the contraction density function takes the form

$$\mathtt{f}(z; x) = \frac{1}{\sqrt{2\pi\tau_1\tau_2/\tau}}\exp\left\{-\frac{(z - x\tau_1/\tau)^2}{2\tau_1\tau_2/\tau}\right\},$$

for $z \in \mathbb{R}$. Note that $\mathtt{f}(\cdot; x)$ has normal$(x\Theta, \Theta(1-\Theta)\tau)$ distribution, for each $x \in \mathbb{R}$. Hence, the posterior distribution of π takes the form

$$\nu_0(x; z) = \frac{1}{\sqrt{2\pi\tau_2}}\exp\left\{-\frac{(x - z - \vartheta\tau_2)^2}{2\tau_2}\right\},$$

for $x \in \mathbb{R}$. This implies that, π is conjugate with respect to $\mathtt{f}(\cdot; x)$.

Now, let $\{X_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{k}(x_n, x_{n+1}) = \frac{1}{\sqrt{2\pi\tau(1 - \Theta^2)}}\exp\left\{-\frac{(x_{n+1} - x_n\Theta - \vartheta\tau_2)^2}{2\tau(1 - \Theta^2)}\right\},$$

for $x_n, x_{n+1} \in \mathbb{R}$; and invariant measure π. Moreover, the dynamics for this model can be also described by the following stochastic equation,

$$X_n = A_n(X, \Theta) + \epsilon_n, \qquad \text{for } n \geq 1,$$

where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common normal$(\vartheta\tau_2, \tau_2)$ distribution and $A_n(X, \Theta)$ has normal$(x\Theta, \Theta(1-\Theta)\tau)$ distribution.

On other hand, let $\{Z_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability kernel,

$$\mathtt{p}(z_n, z_{n+1}) = \frac{1}{\sqrt{2\pi\tau\Theta(1-\Theta^2)}} \exp\Big\{-\frac{[z_{n+1} - \Theta(z_n - \vartheta\tau_2)]^2}{2\tau\Theta(1-\Theta^2)}\Big\},$$

for $z_n, z_{n+1} \in \mathbb{R}$; and invariant distribution $\mathtt{m}_\pi$, which has normal$(\vartheta\tau_1, \tau_1)$ distribution. Clearly, conjugacy of π saves us the computation of the integral $\mathtt{p}$. Furthermore, the duality function $\mathtt{h}$ is given by

$$\mathtt{h}(x; z) = \frac{1}{2\pi\tau} \frac{1}{\sqrt{1-\Theta}} \exp\Big\{-\frac{\Theta x^2}{2\tau_2} + \frac{xz}{\tau_2} - \frac{z^2}{2\tau_2} - z\vartheta + \frac{\tau_1\vartheta^2}{2}\Big\}$$

This function guarantees the duality between the transition probability functions $\mathtt{k}$ and $\mathtt{p}$.

For this model, the probability measure π is conjugate with respect to $\mathtt{f}(\cdot; x)$. So, starting with ν_0 and $\tilde{\omega}(y_0, u)$ which are $\mathrm{N}(x_0; y_0 + \vartheta\tau_2, \tau_2)$ and $\mathrm{N}(y_0; \Theta x_0 + \Theta\vartheta\tau_2, \tau\Theta(1-\Theta^2))$ distributions, respectively. The optimal and prediction filters take the form

$$\begin{aligned}
\nu_{n+1}(\mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\mathrm{N}\Big(x_{n+1}; \frac{u + y_{n+1} + \vartheta\tau_2}{1+\Theta}, \frac{\tau_2}{1+\Theta}\Big)\mathtt{d}u, \\
\psi(\nu_n)(\mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\mathrm{N}(\mathtt{d}x_{n+1}; u + \vartheta\tau_2, \tau_2)\mathtt{d}u,
\end{aligned}$$

where

$$\begin{aligned}
\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\mathrm{N}\Big(y_{n+1}; \Theta[u - \vartheta\tau_2], \tau\Theta(1-\Theta^2)\Big)\mathtt{d}u, \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big)\mathrm{N}\big(y_{n+1}; \Theta u - \Theta\vartheta\tau_2, \tau\Theta(1-\Theta^2)\big)}{m_{\psi(\nu_n)}(dy_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \int_{\mathbb{Y}} \omega(y^{(n)}, k)\mathrm{N}\Big(u; \frac{\Theta(k + y + \vartheta\tau_2)}{1+\Theta}, \frac{\tau(\Theta - \Theta^2 + 2\Theta^3)}{1+\Theta}\Big)\mathtt{d}k.
\end{aligned}$$

This model is characterized since all the filters are given by normal measures. However, we decide to present such filters as mixtures of normal measures, because there is not a recursive way to express them in the formal case. Let us emphasize that, conjugacy of π saves us the computation of the integrals corresponding to $\mathtt{m}_\pi$, which are normal distributions. Additionally, the moment generating function of the optimal and prediction filters are given by

$$\begin{aligned}\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n+1)}\Big] &= \exp\left\{\frac{\lambda(y_{n+1}+\vartheta\tau_2)}{1+\Theta}+\frac{1}{2}\frac{\lambda^2\tau_2}{1+\Theta}\right\}\int_{\mathbb{Y}} e^{\frac{\lambda}{1+\Theta}u}\omega(y^{(n+1)},u)\mathtt{d}u,\\ \mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] &= \exp\left\{\lambda\vartheta\tau_2+\frac{1}{2}\lambda^2\tau_2\right\}\int_{\mathbb{Y}} e^{\lambda u}\tilde{\omega}(y^{(n)},u)\mathtt{d}u,\end{aligned}$$

where ω and $\tilde{\omega}$ are given before.

Gamma model

The gamma distribution belongs to the class of exponential dispersion distributions. In fact, letting π be a gamma(a,b) distribution, with $a,b>0$, we have that $c(x;\tau)=x^{\tau-1}/\Gamma(\tau)$ and $\kappa(\vartheta)=-\log(-\vartheta)$, for $\tau=a$ and $\vartheta=-b$. Thus, letting $\tau_1=\Theta\tau$ and $\tau_2=(1-\Theta)\tau$ with $\Theta\in(0,1)$, the contraction density function takes the form

$$\mathtt{f}(z;x)=\frac{\Gamma(\tau)}{\Gamma(\tau_1)\Gamma(\tau_2)}\left(1-\frac{z}{x}\right)^{\tau_2-1}\left(\frac{z}{x}\right)^{\tau_1-1}\frac{1}{x},$$

for $z\in(0,x)$. Clearly, the measure $\mathtt{f}$ as a function of z/x has beta (τ_1,τ_2) distribution. Hence, the posterior distribution of π is given by

$$\nu_0(x;z)=\frac{(-\vartheta)^{\tau_2}}{\Gamma(\tau_2)}(x-z)^{\tau_2-1}\exp\{\vartheta(x-z)\}$$

for $x\in(z,\infty)$. As for the Poisson model, the measure ν_0, as a function of x, has not gamma distribution. This implies that π is not conjugate with respect to $\mathtt{f}$. Nonetheless, as a function of $x-z$, ν_0 has gamma distribution.

Now, let $\{X_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{k}(x_n,x_{n+1})=\frac{\Gamma(\tau)(-\vartheta)^{\tau_2}e^{\vartheta x_{n+1}}}{\Gamma(\tau_1)\Gamma(\tau_2)\Gamma(\tau_2)x_n^{\tau-1}}\int_0^{x_n\wedge x_{n+1}}\big[\,(x_{n+1}-z)(x_n-z)\,\big]^{\tau_2-1}z^{\tau_1-1}e^{-\vartheta y}\mathtt{d}z,$$

for $x_n,x_{n+1}\in\mathbb{R}^+$; and invariant measure π. For this model, the dynamics of

$\{X_n\}_{n\geq 0}$ can be also written through the stochastic equation (2.19) with $A_n(X,\Theta) = A_n X$. That is to say,

$$X_n = A_n X_{n-1} + \epsilon_n, \qquad \text{for } n \geq 1,$$

where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common gamma$(\tau_2, -\vartheta)$ distribution and where A_n has beta(τ_1, τ_2) distribution.

On other hand, let $\{Z_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{p}(z_n, z_{n+1}) = \frac{\Gamma(\tau) z_{n+1}^{\tau_1 - 1}}{\Gamma(\tau_1)\Gamma(\tau_2)} \frac{(-\vartheta)^{\tau_2} e^{-\vartheta z_n}}{\Gamma(\tau_2)} \int_{z_n \vee z_{n+1}}^{\infty} \left(1 - \frac{z_{n+1}}{x}\right)^{\tau_2 - 1} \left(1 - \frac{z_n}{x}\right)^{\tau_1 - 1} \frac{e^{\vartheta x}}{x} \mathtt{d}x.$$

for $z_n, z_{n+1} \in \mathbb{R}_+$; and with invariant gamma$(\tau_1, -\vartheta)$ distribution. Furthermore, the duality function $\mathtt{h}$ for this model takes the form

$$\mathtt{h}(x; z) = \frac{\Gamma(\tau)(-\vartheta)^{-\tau_1}}{\Gamma(\tau_1)} \left(1 - \frac{z}{x}\right)^{\tau_1 - 1} \frac{e^{-z\vartheta}}{x^{\tau_2}}.$$

This function guarantees the duality between the processes $\{X_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$.

Then, starting at ν_0 which have gamma$(x_0 - y_0; -\vartheta, \tau_2)$ distribution. The optimal and the prediction filters are given by

$$\begin{aligned}
\nu_{n+1}\big(\mathtt{d}x_{n+1}\big) &= \int_0^{x_{n+1}} \omega\big(y^{(n+1)}, u\big) \frac{\mathrm{Beta}\Big(\frac{y_{n+1}}{x_{n+1}}; \tau\Theta, \tau(1-\Theta)\Big) \mathrm{Ga}\big(x_{n+1} - u; \tau_2, -\vartheta\big)}{\mathtt{p}(u, y_{n+1})} \mathtt{d}u, \\
\psi\big(\nu_n\big)\big(\mathtt{d}x_{n+1}\big) &= \int_0^{x_{n+1}} \tilde{\omega}(y^{(n)}, u) \mathrm{Ga}\big(x_{n+1} - u; \tau_2, -\vartheta\big) \mathtt{d}u,
\end{aligned}$$

where

$$\begin{aligned}
\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \int_0^{\infty} \tilde{\omega}(y^{(n)}, u) \mathtt{p}(u, y_{n+1}) \mathtt{d}u, \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big) \mathtt{p}(u, y_{n+1})}{\mathtt{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \int_u^{\infty} \nu_n(dx_n) \mathrm{Beta}\Big(\frac{u}{x}; \tau\Theta, \tau(1-\Theta)\Big) \mathtt{d}x.
\end{aligned}$$

with $\nu_0(u)(x) = \mathrm{Ga}(x - u; \tau_2, -\vartheta)$ and $n \geq 0$. Furthermore, the moment generating function of the prediction filters take the form

$$\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] = \Big(1-\frac{\lambda}{-\vartheta}\Big)^{\tau_2}\int_0^\infty e^{\gamma u}\tilde{\omega}(y^{(n)},u)\mathtt{d}u,$$

where ω and $\tilde{\omega}$ are given as before. For the optimal filters, the statistics are more elaborated but they can be computed by substituting

$$\phi_{y_{n+1}}(\nu_0(u))(x_{n+1}) = \frac{\text{Beta}\Big(\frac{y_{n+1}}{x};\tau\Theta,\tau(1-\Theta)\Big)\text{Ga}\big(x-u;\tau_2,-\vartheta\big)}{\mathtt{p}(u,y_{n+1})},$$

in the expression $\mathbb{E}[\varphi(X_{t_n})|Y_n]$, with $\varphi(x)=e^{ux}$ for $u\in\mathbb{R}$. Let us emphasize that this model is not the same as the gamma-Poisson model in Section 2.3.1. This exemplifies that there are many ways to define the distribution $\mathtt{f}$, which together with the measure π, characterizes the model.

Binomial model

Let π be a binomial(r,p) distribution, with $r\geq 1$ and $p\in(0,1)$. In this case, we have that $c(x;\tau)=\binom{\tau}{x}$ and $k(\vartheta)=\log(1+e^{\vartheta})$, where $\tau=r$ and $\vartheta=\log(p)-\log(1-p)$. Thus, the contraction density function takes the form

$$\mathtt{f}(z;x) = \frac{\binom{\tau_2}{x-z}\binom{\tau_1}{z}}{\binom{\tau}{x}},$$

for $z\in\{\max(0,x+\tau_1-\tau),...,\min(x,\tau_1)\}$. Note that, the measure $\mathtt{f}$ has hypergeometric distribution with parameters (τ,τ_1,x), where $x\in\{0,1,...,\tau\}$. Hence, the posterior distribution of π has density probability function

$$\nu_0(x;z) = \binom{\tau_2}{x-z}p^{x-z}(1-p)^{\tau_2-(x-z)},$$

for $x\in\{z,z+1,...,\tau_2\}$. Again for this model, ν_0 is not a binomial distribution as a function of x. However, as a function of $x-z$, the measure ν_0 has binomial distribution.

Now, let $\{X_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{k}(x_n, x_{n+1}) = \frac{(1-p)^{\tau_2}}{\binom{\tau}{x_n}} \sum_{z=\max(0, x_n-\tau_2)}^{\min(x_n,\tau_1,\tau_2)} \binom{\tau_2}{x_{n+1}-z}\binom{\tau_2}{x_n-z}\binom{\tau_1}{z} e^{\vartheta(x_{n+1}-z)},$$

for $x_n \in \{0,1,...,\tau\}$ and $x_{n+1} \in \{\min(\tau_1,\tau_2),...,\max(\tau_1,\tau_2)\}$; and with invariant measure π. For this model, the dynamics of $\{X_n\}_{n\geq 0}$ can be also written through the stochastic equation

$$X_n = A_n(X,\Theta) + \epsilon_n, \qquad \text{for } n \geq 1,$$

where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common binomial(τ_2, p) distribution and $A_n(X,\Theta)$ has hypergeometric(τ,τ_1,x) distribution.

On other hand, let $\{Z_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability kernel,

$$\mathtt{p}(z_n, z_{n+1}) = \binom{\tau_1}{z_{n+1}}\left(\frac{1-p}{p}\right)^{z_n}(1-p)^{\tau_2}\sum_{x=z_n\vee z_{n+1}}^{\tau_1\wedge\tau_2}\frac{\binom{\tau_2}{x-z_{n+1}}\binom{\tau_1}{x-z_n}}{\binom{\tau}{x}}\left(\frac{p}{1-p}\right)^{x},$$

for $z_n \in \{0,1,...,\tau_2\}$ and $z_{n+1} \in \{\max(0, z_n-\tau_2),...,\min(z_n,\tau_1,\tau_2)\}$; and with invariant binomial(τ_1,p) distribution. Furthermore, the duality function $\mathtt{h}$

$$\mathtt{h}(x;z) = \frac{\binom{\tau_2}{x-z}}{\binom{\tau}{x}}\left(\frac{1-p}{p}\right)^{z}(1-p)^{-\tau_1}.$$

As before, this function guarantees the duality between the processes $\{X_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$.

Now, starting at ν_0 which has binomial$(x_0-y_0;\tau_2,p)$ distribution. The optimal and the prediction filters are given by the following recursions

$$\begin{aligned}
\nu_{n+1}(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \omega\big(y^{(n+1)},u\big)\frac{\text{Hypgeo}\big(y_{n+1};\tau,\tau_1,x_{n+1}\big)\text{Bin}\big(x_{n+1}-u;\tau_2,p\big)}{\mathtt{p}(u,y_{n+1})},\\
\psi(\nu_n)(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \tilde{\omega}(y^{(n)},u)\text{Bin}\big(x_{n+1}-u;\tau_2,p\big),
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \sum_{u=0}^{\tau_2} \tilde{\omega}(y^{(n)}, u)\mathtt{p}(u, y_{n+1}), \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big)\mathtt{p}(u, y_{n+1})}{\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \sum_{x=u}^{\infty} \nu_n(dx_n)\mathrm{Hypgeo}\Big(u; \tau, \tau_1, x\Big).
\end{aligned}$$

with $\nu_0(u)(x) = \mathrm{Bin}(x-u; \tau_2, p)$ and $n \geq 0$. For this model, the moment generating function associated to the prediction filters takes the form

$$\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] = \Big(1 - p + pe^{\lambda}\Big)^{\tau_2} \sum_{u=0}^{\infty} e^{\gamma u}\tilde{\omega}(y^{(n)}, u),$$

where ω and $\tilde{\omega}$ are given before. On other hand, expressions for the optimal filters and their statistics can be computed form the equations (2.11) and (2.13).

Negative binomial model

The negative-binomial(r, p) distribution, denoted by π, with $n \geq 1$ and $p \in (0, 1)$, belong to the class exponential dispersion distributions. In this case, we have that $c(x; \tau) = \Gamma(x + \tau)/[\Gamma(\tau)x!]$ and $k(\vartheta) = -\log(1 - e^{\vartheta})$, for $\tau = r$ and $\theta = \log(p)$. Thus, letting $\tau_1 = \Theta\tau$ and $\tau_2 = (1 - \Theta)\tau$ with $\Theta \in (0, 1)$, the contraction density function takes the form

$$\mathtt{f}(z; x) = \binom{x}{z}\frac{\mathsf{B}(\tau_1 + z, \tau_2 + x - z)}{\mathsf{B}(\tau_1, \tau_2)}, \qquad \text{for } z \in \{0, 1, ..., x\},$$

where B denotes the beta function. This implies that $\mathtt{f}$ has beta-binomial(x, τ_1, τ_2) distribution, with $x \in \mathbb{N} \cup \{0\}$. Hence the posterior distribution of π has density probability function,

$$\nu_0(x; z) = \frac{\Gamma(x - z + \tau_2)}{\Gamma(\tau_2)(x - z)!}p^{x-z}(1 - p)^{\tau_2},$$

for $x \in \{z, z + 1, ...\}$. Again for this model, the measure ν_0 as a function of $x - z$ has negative-binomial distribution.

Now, let $\{X_n\}_{n \geq 0}$ a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{k}(x_n, x_{n+1}) = \frac{x_n! p^{x_{n+1}}(1-p)^{\tau_2}}{\Gamma(\tau_2)\mathsf{B}(\tau_1,\tau_2)} \sum_{z=0}^{x_n \wedge x_{n+1}} \frac{\Gamma(x_{n+1}-z+\tau_2)\mathsf{B}(\tau_1+z, \tau_2+x_n-z)}{(x_{n+1}-z)!(x_n-z)!z!} p^{-z},$$

for $x_n \in \mathbb{N} \cup \{0\}$, $x_{n+1} \geq x_n$; and with invariant measure π. For this model, the dynamics of $\{X_n\}_{n\geq 0}$ can also be written through the stochastic equation

$$X_n = A_n(X, \Theta) + \epsilon_n, \qquad \text{for } n \geq 1,$$

where $\{\epsilon_n\}_{n\geq 1}$ is a sequence of i.i.d. random variables with common negative-binomial(τ_2, p) distribution and $A_n(X,\Theta)$ has beta-binomial(x, τ_1, τ_2) distribution.

On other hand, let $\{Z_n\}_{n\geq 0}$ be a time-homogeneous stationary Markov process driven by the transition probability function,

$$\mathtt{p}(z_n, z_{n+1}) = \frac{(1-p)^{\tau_2}}{\Gamma(\tau_2)p^{z_n}} \sum_{x=z_n \wedge z_{n+1}}^{\infty} \frac{\mathsf{B}(\tau_1+z_{n+1}, \tau_2+x-z_{n+1})\Gamma(x-z_n+\tau_2)x!}{\mathsf{B}(\tau_1,\tau_2)(x-z_{n+1})!(x-z_n)!z_{n+1}!} p^x,$$

for $z_n \in \mathbb{N}\cup\{0\}$, $z_{n+1} \leq z_n$; and with invariant negative-binomial(τ_1, p) distribution. In this case, the duality function $\mathtt{h}$ takes the form

$$\mathtt{h}(x; z) = \frac{\Gamma(x-z+\tau_2)}{\Gamma(\tau_2)(x-z)!} \frac{\Gamma(\tau)x!}{\Gamma(x+\tau)} p^{-z}(1-p)^{-\tau_1},$$

As before the construction guarantees the duality between the processes $\{X_n\}_{n\geq 0}$ and $\{Z_n\}_{n\geq 0}$ with respect to the duality function $\mathtt{h}$.

Then, starting at ν_0 which has negative-binomial$(x_0 - y_0; \tau_2, p)$ distribution. The optimal and the prediction filters for this model are given by

$$\begin{aligned} \nu_{n+1}\big(\mathtt{d}x_{n+1}\big) &= \sum_{u=0}^{x_{n+1}} \omega\big(y^{(n+1)}, u\big) \frac{\text{Be-Bin}\Big(y_{n+1}; x_{n+1}, \tau_1, \tau_2\Big)\text{NB}\big(x_{n+1}-u; \tau_2, p\big)}{\mathtt{p}(u, y_{n+1})}, \\ \psi\big(\nu_n\big)(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \tilde{\omega}(y^{(n)}, u)\text{NB}\big(x_{n+1}-u; \tau_2, p\big), \end{aligned}$$

where

$$\begin{aligned}
\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \sum_{u=y_{n+1}}^{\infty} \tilde{\omega}(y^{(n)}, u)\mathtt{p}(u, y_{n+1}), \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big)\mathtt{p}(u, y_{n+1})}{\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \sum_{x=u}^{\infty} \nu_n(dx_n)\text{Be-Bin}\Big(u; x, \tau_1, \tau_2\Big).
\end{aligned}$$

with $\nu_0(u)(x) = \text{NB}(x-u; \tau_2, p)$ and $n \geq 0$. Furthermore, the moment generating function associated to the prediction filters takes the form

$$\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] = \Big(\frac{1-p}{1-pe^{\lambda}}\Big)^{\tau_2} \sum_{u=0}^{\infty} e^{\gamma u}\tilde{\omega}(y^{(n)}, u),$$

where ω and $\tilde{\omega}$ are given before. Then, as we mentioned, the optimal filters and their corresponding statistics can be computed from (2.11) and (2.13).

2.5 Lancaster probabilities

The construction of reversible Markov processes used along this chapter requires the specification of the probability measures π and $\mathtt{f}$. These measures allow two built transition probabilities via the joint distribution $\pi(\mathtt{d}x)\mathtt{f}(\mathtt{d}z; x)$. Hence, such joint distribution characterizes the construction that we proposed. Having this in mind, this section deals with the problem of defining a joint distribution, via Lancaster probabilities, that possesses some appealing properties. Hence, we begin defining and developing some properties of Lancaster properties. This will be done based on the work of Lancaster [22] and Koudou [19].

The Lancaster probabilities on $\mathbb{R}^2$ are a class of distributions satisfying a bi-orthonormal condition. In fact, Lancaster [21] states that, a bivariate distribution is completely characterized almost surely by its marginal distributions and its correlation matrix. In this case, such matrix is completely generated by the sets of orthogonal polynomials associated to their margins. It is worth to mention that, polynomial bi-orthogonality is a very strong condition, for instance it implies that moments of all degree exist. Such is the case of the class of Meixner distribution, which later we will see that they play an important role to define the Lancaster probabilities

2.5.1 Meixner distributions

Let us first recall some basis facts of orthogonal polynomials. Consider a general polynomial of degree n, denoted by P_n, given by

$$P_n(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0, \qquad a_n \neq 0,$$

where a_n is known as the leading coefficient of the polynomial P_n. The monic version of the polynomial P_n is obtained through a simple transformation of the form

$$\frac{P_n(x)}{a_n} = x^n + \frac{a_{n-1}}{a_n}x^{n-1} + \cdots + \frac{a_0}{a_n},$$

i.e. it is a version of P_n with leading coefficient equal to one. Hence, in what follows, we consider polynomials with leading coefficient equal to one. A system of polynomials $\{P_n(x)\}_{n\geq 0}$ is called an orthogonal system of polynomials with respect to some real positive measure π, if the following orthogonality relations hold

$$\int_{\mathbb{X}} P_n(x)P_m(x)\pi(x)\mathrm{d}x = b_n^2\delta_{n,m}, \qquad n, m \geq 0,$$

where $\mathbb{X}$ is the support of π and $\{b_n\}_{n\geq 0}$ are nonzero constants. If such constants are equal to one we say that the system is orthonormal.

Let X be a centered random variable possessing distribution G and moment generating functions ϕ, and let $P_n(x) = x^n + a_{n,1}x^{n-1} + ... + a_{n,n}$ be a set of orthogonal polynomials. If the function $\exp\{u(t)x\}/\phi(u(t))$ generates a system of orthogonal polynomials, *i.e.*

$$\mathtt{g}(t, x) = \frac{e^{xu(t)}}{\phi(u(t))} = \sum_{n=0}^{\infty} \frac{P_n(x)t^n}{n!}. \tag{2.20}$$

Then, we say that X belongs to the Meixner class and it is characterized by $u(t)$, where $u(t)$ is a power series in t having real coefficients, with $u(0) = 0$ and $u'(0) = 1$. Denoting by $v(u)$ the functional inverse of $u(t)$, *i.e.* $v(u(t)) \equiv t$, it follows that,

$$\frac{\mathrm{d}t}{\mathrm{d}u} = 1 - \lambda t - \kappa t^2 + \cdots .$$

Now, in order to compute the generating function of $b_m \equiv \mathbb{E}[P_m^2(X)]$, one can obtain the following equality

$$B(st) := \frac{\phi(u(s)+u(t))}{\phi(u(s))\phi(u(t))} = \int \mathtt{g}(x;s)\mathtt{g}(x;t)\mathtt{d}G(x) \equiv 1 + \sum_{m=1}^{\infty} \frac{b_m s^m t^m}{(m!)^2}.$$

The last equality holds because for $m \neq n$ the coefficient of $s^m t^n$ are zero by orthogonality.

On other hand, it is known that, any set of orthogonal polynomials in a centered variable where the polynomials have been normalized to have unit leading coefficients, there is a recurrence relation of the form,

$$\begin{aligned} P_{n+1}(x) &= (x+A_n)P_n(x) + C_n P_{n-1}(x), \qquad n = 1, 2, ...; \\ P_0(x) &= 1, \qquad P_1(x) = x, \end{aligned}$$

where A_n and C_n are constants, independent of x. Hence, multiplying both sides of the above equation by P_{n-1} and integrating with respect to x we obtain that

$$C_n b_{n-1} = -\mathbb{E}[xP_{n-1}(x)P_n(x)] = -\mathbb{E}[P_n^2(x) + P_n(x)Q_{n-1}(x)] = -b_n,$$

where $Q_{n-1}(x)$ is a polynomial in x of degree at most $(n-1)$. This implies that the constant C_n is equal to $-b_n/b_{n-1}$.

Denoting by $b_1 = \mathrm{Var}[X]$, deriving the equality (2.20) with respect to the real-valued function $u(t)$ one obtains

$$\frac{\mathtt{d}t}{\mathtt{d}u}\frac{\mathtt{d}}{\mathtt{d}t}\mathtt{g}(x;t) = \left(x - \frac{\phi'(u)}{\phi(u)}\right)\frac{e^{xu}}{\phi(u)} = (x - b_1 t)\,\mathtt{g}(x;t).$$

Hereafter, let $\psi'(u) = \phi'(u)/\phi(u)$. Then, replacing the expression of $(\mathtt{d}t/\mathtt{d}u)$ and the function $\mathtt{g}$ in terms of their polynomials, the above equality takes the form

$$(1 - \lambda t - \kappa t^2 + \cdots)\left(1 + \sum \frac{nt^{n-1}P_n(x)}{n!}\right) = (x - b_1 t)\left(1 + \sum \frac{t^n P_n(x)}{n!}\right). \qquad (2.21)$$

A comparison of the coefficients of $t^n/n!$ in (2.21) and the recurrence relation of $\{P_n\}_{n\geq 0}$ implies that

$$\frac{dt}{du} = 1 - \lambda t - \kappa t^2, \qquad (2.22)$$

where κ and λ constant. The equality (2.22) characterizes the class of Meixner distributions into six different families. Furthermore, such equality implies that the orthogonal polynomials satisfy the following relation,

$$P_{n+1} = (x + n\lambda)P_n + n((n-1)\kappa - b_1)P_{n-1},$$

for $n = 0, 1, 2, \ldots$, with $P_{-1}(x) = 0$ and $\kappa \leq 0$ because $C_n < 0$. Furthermore,

$$\frac{\mathtt{d}}{\mathtt{d}t} \log\{\phi(u)\} = \frac{b_1 t}{1 - \lambda t - \kappa t^2} \qquad \Rightarrow \qquad \psi(u) = \log\{\phi(u)\} = b_1\omega(t) + a,$$

where a is a constant and $\omega(t) = \int_0^t s/(1 - \lambda s - \kappa s^2)\mathtt{d}s$. Using the last equality one can derive some special cases within the Meixner class, namely

- Positive binomial distribution

  If $(1 - \lambda t - \kappa t^2) = (1 - pt)(1 + qt)$, with p, $q > 0$ such that $p + q = 1$, then

  $$\frac{\mathtt{d}u}{\mathtt{d}t} = \frac{q}{1 + qt} + \frac{p}{1 - pt} \qquad \Rightarrow \qquad U \equiv \exp\{u\} = \frac{1 + qt}{1 - pt}$$

  or equivalently, $t = (U - 1)/(q + pU)$. Thus, letting $b_1 = npq$ we obtain that

  $$\frac{\mathtt{d}\psi(u)}{\mathtt{d}U} = \frac{\mathtt{d}u}{\mathtt{d}U}\frac{\mathtt{d}\psi(u)}{\mathtt{d}u} = \frac{1}{U}b_1 t = \frac{npq}{U}\frac{U - 1}{q + pU}.$$

  whose solution is given by $\psi(u) = -np\log(U) + n\log(q + pU)$. Hence, the following function characterizes a binomial(n, p) distribution,

  $$\phi(u) = \left(\frac{1 - pt}{1 + qt}\right)^{np}\left(\frac{1}{1 - pt}\right)^{n}.$$

  Moreover, the orthogonal polynomials in this case are the Kravkut polynomials and its moment generating function is,

  $$\mathtt{g}(x; t) = \left(\frac{1 + qt}{1 - pt}\right)^{x + Np}(1 - pt)^N.$$

- Normal distribution

  If $(1 - \lambda t - \kappa t^2) = 1$ and letting $b_1 = 1$, then $u(t) = t$. Thus, $\psi'(u) = u$

whose solution is given by $\psi(u) = u^2/2$ with $\psi(0) = 0$. Hence, the following function characterizes a normal$(0,1)$ distribution,

$$\phi(u) = \exp\left\{\frac{1}{2}u^2\right\}.$$

Moreover, the orthogonal polynomials for the normal distribution are the Hermite polynomials and its moment generating function is,

$$\mathtt{g}(x;t) = \exp\left\{xt - \frac{1}{2}t^2\right\}.$$

- Poisson distribution

If $(1 - \lambda t - \kappa t^2) = 1 + t$, then $(\mathtt{d}u/\mathtt{d}t) = (1+t)^{-1}$ whose solution is $u = \log(1+t)$, or equivalently $t = e^u - 1$. Thus,

$$\frac{\mathtt{d}}{\mathtt{d}u}\psi(u) = b_1(e^u - 1) \qquad \Rightarrow \qquad \psi(u) = b_1(e^u - u) - \sigma^2,$$

with $\psi(0) = 0$. Hence, the following function characterizes a centered Poisson distribution with parameter b_1,

$$\phi(u) = e^{-\mu u}\exp\{\mu(e^u - 1)\}.$$

Moreover, the orthogonal polynomials for the Poisson distribution are the Poisson-Charlier polynomials and its moment generating function is,

$$\mathtt{g}(x;t) = (1+t)^{x+\mu}e^{-t\mu}$$

- Gamma distribution

If $(1 - \lambda t - \kappa t^2) = (1+t)^2$, then the solution of $(\mathtt{d}u/\mathtt{d}t)$ is $u = t/(1+t)$, or equivalently, $t = u/(1-u)$. Thus,

$$\frac{\mathtt{d}}{\mathtt{d}u}\psi(u) = b_1\frac{u}{1-u} \qquad \Rightarrow \qquad \psi(u) = -b_1(\log(1-u) + u),$$

with $\psi(0) = 0$. Hence, the following function characterizes a centered gamma distribution,

$$\phi(u) = e^{-b_1 u}(1-u)^{-b_1}.$$

Moreover, the orthogonal polynomials for the gamma distribution are the Laguerre polynomials and its moment generating function is,

$$\mathtt{g}(x;t) = (1+t)^{-b_1}\exp\left\{(x+b_1)\frac{t}{1+t}\right\}$$

- Negative binomial distribution

If $(1-\lambda t-\kappa t^2) = (1+pt)(1+qt)$ with $p+q=1$, then

$$\frac{\mathtt{d}u}{\mathtt{d}t} = \frac{q}{1+qt} - \frac{p}{1+pt} \qquad \Rightarrow \qquad U \equiv e^u = \frac{1+qt}{1+pt}$$

or equivalently, $t = (U-1)/(q-pU)$. Thus, letting $b_1 = npq$ we obtain that

$$\frac{\mathtt{d}\psi(U)}{\mathtt{d}U} = \frac{npq}{U}\frac{U-1}{q-pU} = np\left(\frac{1}{q-pU} - \frac{1}{U}\right)$$

This implies that, $\psi(u) = -n\log(q-pU) - np\log U$. Hence, the following function characterizes a binomial negative(n,p) distribution,

$$\phi(u) = \left(\frac{1+qt}{1+pt}\right)^{-np}\left(q - p\frac{1+qt}{1+pt}\right)^{-n}$$

Moreover, the orthogonal polynomials for the negative-binomial distribution are the Meixner polynomials and its moment generating function is,

$$\mathtt{g}(x;t) = \left(\frac{1+qt}{1+pt}\right)^{x+np}\left(\frac{q-p}{1+pt}\right)^{n}$$

- Hypergeometric distribution

If $(1-\lambda t-\kappa t^2)$ has two complex arguments, then its density function is proportional to the product of two gamma functions with complex argument.

Finally, a result derived in Eagleson [7] tell us that joint distributions, formed by taking convolutions of variables in a Meixner class with some held in common, posses the polynomial bi-orthogonal property.

2.5.2 Lancaster probabilities

Let μ and ν be probability measures on $\mathbb{R}$. Denote by (P_n) and (Q_n), respectively, the sequence of orthonormal polynomials with respect to μ and ν. These sequences are unique if we assume that the coefficients of x^n and z^n in (P_n) and (Q_n) are positive. Suppose there exists $\alpha > 0$ such that $\int e^{\alpha|x|}\mathtt{d}\mu(x) < \infty$ and $\int e^{\alpha|z|}\mathtt{d}\nu(z) < \infty$ (this condition guarantees the existence of the moments of all orders of μ and ν). Then (P_n) and (Q_n) are complete in $L^2(\mu)$ and $L^2(\nu)$, respectively. The bivariate probability distributions σ, with margins μ and ν such that if (X, Z) is a random variable with law σ then $\mathbb{E}[P_i(X)Q_j(Z)] = 0$ for $i \neq j$, are called the Lancaster probabilities with respect to μ and ν.

Denoting by $L(\mu, \nu)$ the set of Lancaster probabilities with margins μ and ν. The sequence $(\rho_n)_{n\geq 0}$ defined by $\rho_n = \mathbb{E}[P_n(X)Q_n(Z)]$ is then called a Lancaster sequence with respect to μ and ν, denoted by $S(\mu, \nu)$. This sequence characterizes σ in $L(\mu, \nu)$. The most interesting elements of $L(\mu, \nu)$ are those having a density of the form

$$\sigma(\mathtt{d}x, \mathtt{d}z) = \left[\sum_n \rho_n P_n(x)Q_n(z)\right]\mu(\mathtt{d}x)\nu(\mathtt{d}z), \tag{2.23}$$

for $x, z \in E \subset \mathbb{R}$. Note that, $L(\mu, \nu)$ is clearly a convex set. Thus, $L(\mu, \nu)$ is characterized by its extreme points. An extreme point α of a convex set C is an element of C for which there is no α_1, α_2 in C and no λ in $(0, 1)$ such that $\alpha_1 \neq \alpha_2$ and $\alpha = \lambda\alpha_1 + (1 - \lambda)\alpha_2$. The general problem is, given μ and ν, to find those extreme points, or equivalently, its extreme Lancaster sequences. Tyan, Derin and Thomas [42] showed that, if both μ and ν have unbounded supports, then the Lancaster sequences are necessarily moments sequences.

Indeed, suppose that μ and ν have unbounded support and that the convex support of μ is $\mathbb{R}$. Letting,

$$\gamma^{-1} = \limsup_{n\to\infty}\left|\frac{b_{2n}}{a_{2n}}\right| \tag{2.24}$$

where γ is set to be 0 if the lim sup is infinite. If $\rho \in S(\mu, \nu)$, then there exists a probability measure α_ρ on $[-\gamma, \gamma]$ such that

$$\rho_n = \frac{b_n}{a_n}\int_{-\gamma}^{\gamma} t^n \alpha_p(\mathtt{d}t), \qquad \forall n \in \mathbb{N}. \tag{2.25}$$

More precisely, denote by $C(\mu)$ and $C(\nu)$ the convex support of μ and ν, respectively.

- Let $C(\mu) = [A, \infty)$ and $C(\nu) = [B, \infty)$ where $A, B \in \mathbb{R}$. If $\rho \in S(\mu, \nu)$, then (2.25) holds with α_p on $[0, \gamma]$.

- Let $C(\mu) = [A, \infty)$ and $C(\nu) = [-\infty, B)$ where $A, B \in \mathbb{R}$. If $\rho \in S(\mu, \nu)$, then (2.25) holds with α_p on $[-\gamma, 0]$.

- If $C(\mu) = \mathbb{R}_+$ and $C(\nu) = \mathbb{R}$ or viceversa, then $\rho \in S(\mu, \nu) \Leftrightarrow \rho_n = 0$, $\forall n \geq 1$.

In the case where $\mu = \nu$ one has $\gamma = 1$. If the support of μ is unbounded and if $\rho \in S(\mu, \mu)$, there exists a probability measure α_p on $[-1, 1]$ such that (2.25) holds. In particular, if the convex support of μ is a half line, then α_p is supported by $[0, 1]$.

2.5.3 Duality and filtering

The construction of time-homogeneous Markov processes discrete in time are completely characterized by its invariant measure and its one-step transition probability function. See Section 1.3. Our proposal requires specifying the probability measures π and $\mathtt{f}$ to construct a bivariate distribution that characterizes the model. Thus, Lancaster probabilities give us an alternative way to construct this kind of models by defining a joint distribution of the form

$$\sigma(\mathtt{d}x, \mathtt{d}z) = \pi(\mathtt{d}x)\mathtt{f}(z; x) = \mathsf{m}_\pi(\mathtt{d}z)\nu_0(\mathtt{d}x; z).$$

for $x \in \mathbb{X}$ and $y \in \mathbb{Y}$, where σ is given by (2.23). It is clear that, the marginal distributions of σ over $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ are π and m_π, respectively. Thus, the construction implies that

$$\mathtt{f}(z; x) = \sum_n \rho_n P_n(x) Q_n(z) \mathsf{m}_\pi(\mathtt{d}z),$$

where $\{P_n\}_{n\geq 0}$ and $\{Q_n\}_{n\geq 1}$ are the orthonormal polynomials associated to π and m_π, respectively; and $\{\rho_n\}_{n\geq 0}$ is the Lancaster sequence of σ.

Then, one can define a time-homogeneous stationary Markov process over $(\mathbb{X}, \mathscr{X})$, denoted by $\{X_n\}_{n\geq 0}$, with invariant distribution π and one-step transition probability function given by,

$$\begin{aligned}\mathtt{k}(x_n, \mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}} \nu_0(\mathtt{d}x_{n+1}; z)\mathtt{f}(z; x_n)\mathtt{d}z \\ &= \int_{\mathbb{Y}} \sum_{j\geq 0} \rho_j P_j(\mathtt{d}x_{n+1})Q_j(z)\pi(\mathtt{d}x_{n+1})\Big[\sum_{k\geq 0} \rho_k P_k(x_n)Q_k(z)\mathsf{m}_\pi(z)\Big]\mathtt{d}z \\ &= \sum_{j\geq 0} \rho_j^2 P_j(x_n)P_j(x_{n+1})\pi(\mathtt{d}x_{n+1})\end{aligned}$$

where the last equality holds by the bi-orthogonality of $\{P_n\}_{n\geq 0}$ and $\{Q_n\}_{n\geq 0}$. Notice that, the expression of $\mathtt{k}$ proves that P_j is an eigenfunction for the eigenvalue ρ_j^2 of the operator $Tf(x) = \int_{\mathbb{X}} f(x)k(v,x)\mathtt{d}x$, for any measurable function f.

Again, using the above bi-orthogonal property, we derive an expression for the transition probabilities that drive the dual $\{Z_n\}_{n\geq 1}$. Indeed, one can built an homogeneous reversible Markov process $(\mathbb{Y}, \mathscr{Y})$-valued, with invariant measure m_π and one-step transition probability function,

$$\mathtt{p}(z_n, z_{n+1}) = \sum_{j\geq 0} \rho_j^2 Q_j(\mathtt{d}z_n)Q_j(\mathtt{d}z_{n+1})\mathsf{m}_\pi(\mathtt{d}z_{n+1}),$$

for $z_n, z_{n+1} \in \mathbb{Y}$.

Thus, as before, the processes X and Y are dual to each other with respect to the duality function defined as the Radon-Nikodyn derivative between the posterior and prior distribution of π, *i.e.*

$$\mathtt{h}(x; z) = \sum_{j\geq 0} \rho_j P_j(x)Q_j(z),$$

for $x \in \mathbb{X}$ and $z \in \mathbb{Y}$. Indeed, for this model, the duality equality is obtained via the bi-orthogonal property of the polynomials associated to π and m_π. Leaving us,

$$\begin{aligned}\mathbb{E}^x[\mathtt{h}(X_n, z)] &= \mathbb{E}^x\Big[\sum_{j\geq 0} \rho_j P_j(X_{n+1})Q_j(z)\Big] \\ &= \sum_{j\geq 0} \rho_j Q_j(z)\mathbb{E}^x\left[P_j(X_{n+1})\right] \\ &= \sum_{j\geq 0} \rho_j^3 Q_j(z)P_j(x) \\ &= \mathbb{E}^y[\mathtt{h}(x, Z_n)].\end{aligned}$$

Then, one can compute the predictor operator of $\{X_n\}_{n\geq 0}$ using the mentioned duality. That is to say,

$$\begin{aligned}\int_{\mathbb{X}} \nu_0(v;z)\mathtt{k}(v,\mathtt{d}x)\mathtt{d}v &= \int_{\mathbb{X}} \sum_{n\geq 0} \rho_n P_n(v) Q_n(z)\pi(v) \left[\sum_{k\geq 0} \rho_k^2 P_k(v) P_k(\mathtt{d}x)\pi(\mathtt{d}x)\right]\mathtt{d}v \\ &= \sum_{n\geq 0} \rho_n^3 P_n(x) Q_n(z)\pi(\mathtt{d}x).\end{aligned}$$

In fact, from the previous results, the above operator is the first step of the prediction filters, *i.e.* the law of $\{X_1|Y_0 = z\}$ denoted by $\psi(\nu_0)(\cdot)$.

Similarly, one can compute the predictor operator of $\{Z_n\}_{n\geq 0}$. In fact, the expression we get is very similar to the previous one,

$$\int_{\mathbb{Y}} \mathtt{f}(u;x)\mathtt{p}(u,\mathtt{d}z)\mathtt{d}u = \sum_{n\geq 0} \rho_n^3 P_n(x) Q_n(z)\mathtt{m}_\pi(\mathtt{d}z).$$

Clearly, the predictor operators of the Markov processes X and Z, which are defined using some Lancaster probability, possesses an expression based on the orthonormal polynomial rather than some transition function.

Furthermore, the computation of the filters can be obtained using the same bi-orthogonality of the model. In fact, we have that

$$\begin{aligned}\mathtt{m}_{\nu_0}(\mathtt{d}y_1) &= \int_{\mathbb{X}} \psi(\nu_0)(x_1)\mathtt{f}(y_1;x_1)\mathtt{d}x_1 \\ &= \sum_{n\geq 0} \rho_n^4 Q_n(y_0) Q_n(y_1)\mathtt{m}_\pi(\mathtt{d}y_1).\end{aligned}$$

Hence, the optimal filter $\{X_1|Y_0, Y_1\}$ takes the form

$$\nu_1(\mathtt{d}x_1) = \frac{\sum_{n\geq 0} \rho_n^3 P_n(x_1) Q_n(y_0)\pi(\mathtt{d}x_1)\left[\sum_{j\geq 0} \rho_j P_j(x_1) Q_j(y_1)\mathtt{m}_\pi(\mathtt{d}y_1)\right]}{\sum_{n\geq 0} \rho_n^4 Q_n(y_0) Q_n(y_1)\mathtt{m}_\pi(\mathtt{d}y_1)}.$$

Then, the predictor filter $\{X_2|Y_0, Y_1\}$ is given by

$$\psi(\nu_1)(\mathtt{d}x_2) = \frac{\sum_{n,j,i\geq 0} \rho_n^6 P_n(x_2) Q_j(y_0) Q_k(y_1)\mathtt{m}_\pi(\mathtt{d}y_1)\pi(\mathtt{d}x_2)\int_{\mathbb{X}} P_n(x_1)P_j(x_1)P_i(x_1)\pi(x_1)\mathtt{d}x_1}{\sum_{n\geq 0} \rho_n^4 Q_n(y_0) Q_n(y_1)\mathtt{m}_\pi(\mathtt{d}y_1)}.$$

Let us emphasize that, as we did in the previous section, the purpose of presenting this particular model is to provide a mechanism to specify the measures π and $\mathtt{f}$.

Nonetheless, expressions for the filters can be obtained following the above methodology. However, these are given in terms of orthogonal polynomials. Hence, this leads us to deal with the problem of handling polynomials. It is worth to mention that, there are examples where this is not complicated. For instance, below, we provide an alternative way to define the gamma-Poisson process of Section 1.3 via Lancaster probabilities. This will help to illustrate the above construction. Also, a Lancaster probability with beta margins is used to build the two Markov processes which are dual to each other.

The gamma-negative binomial case

Let π by a gamma$(a,1)$ distribution, with $a>0$. It is known that, the orthonormal polynomials with respect to π, denote by $\{P_n\}_{n\geq 0}$, are given by $P_n=\sqrt{n!/(a)_n}L_n^a$, with $(a)_n=a(a+1)\cdots(a+n-1)$, where $\{L_n^a\}_{n\geq 0}$ are the Laguerre polynomials defined by

$$L_n^a(x)=\sum_{k=0}^{n}\frac{(-1)^k(1+a)_n x^k}{k!(n-k)!(1+a)_k}$$

for $x\in(0,\infty)$. See Rainville (1971).

Also, let m_π be a negative-binomial(r,p) distribution, with $n>0$ and $p\in(0,1)$. The orthonormal polynomials with respect to m_π are the normalized Meixner polynomials, denoted by $\{Q_n\}_{n\geq 0}$, that take the form

$$Q_n(z)=\sqrt{\frac{p^n(r)_n}{n!}}\sum_{j=0}^{n}\frac{(-n)_j(-z)_j}{(r)_j j!}\Big(1-\frac{1}{p}\Big)^j.$$

See Erdelyi (1953).

Now, Koudou [19] proved that the Lancaster probability with marginals π and m_π has Lancaster sequences of the form $(t^n)_{n\geq 0}$, for $t\in\mathbb{R}$, if

$$g_t(x,z)=\sum_{n=0}^{\infty}P_n(x)M_n(z)t^n\geq 0,\qquad \text{for } (x,z)\in(0,\infty)\times\mathbb{N},$$

which is satisfied if and only if $0\leq t\leq\sqrt{p}$. Moreover, Koudou [19] also obtains the following expression,

$$g_{\sqrt{p}}(x,z) = \frac{x^z}{(a)_z}\frac{1}{(1-p)^{z+a}}\exp\left\{-\frac{xp}{1-p}\right\}.$$

Letting $\phi = p/(1-p)$ and $r = a$ then the Lancaster probability with gamma and negative-binomial margins takes the form,

$$\begin{aligned}\sigma(\mathtt{d}x,\mathtt{d}z) &= g_{\sqrt{p}}(x,z)\mathsf{m}_\pi(\mathtt{d}z)\pi(\mathtt{d}x)\\ &= \frac{x^z}{(a)_z}(1+\phi)^{z+a}e^{\{-x\phi\}}\binom{z+a-1}{z}\left(\frac{1}{1+\phi}\right)^a\left(\frac{\phi}{1+\phi}\right)^z\pi(\mathtt{d}x)\\ &= e^{x\phi}\frac{(x\phi)^z}{z!}\pi(\mathtt{d}x)\end{aligned}$$

for $(x,y)\in(0,\infty)\times\mathbb{N}$. Thus, the Lancaster probability σ with margins π and m_π, and Lancaster sequences of the form (t^n), for $t\in[0,\sqrt{p}]$, generates the gamma-Poisson model described previously. Moreover, the duality function turns out to be given by the function $g_{\sqrt{p}}$, with $p=\phi/(1+\phi)$. The filters for this model were presented in Section 1.3.

The beta-beta case

Buja (1990) developed an example of polynomial bi-orthogonality through a joint distribution called triangular bivariate beta distribution. This model considers random variables X ∼beta(a_1, a_2+1) and Z ∼beta(a_2, a_1+1). In this case, the Lancaster probability σ takes the form

$$\sigma(\mathtt{d}x,\mathtt{d}z) = \frac{a_1+a_2}{\mathsf{B}(a_1,a_2)}x^{a_1-1}z^{a_2-1},$$

for $x,z>0$, and $x+z<1$, and 0 otherwise. Recall that, the orthogonal polynomials associated to π and m_π are the Jacobi polynomials. Also, the Lancaster sequences associated to σ are given by

$$\rho_n = \frac{(-1)^n\sqrt{a_1a_2}}{\sqrt{(a_1+n)(a_2+n)}},$$

for $n\geq 0$. Thus, the bivariate distribution σ implies that

$$\mathtt{f}(\mathtt{d}z;x) = \frac{z^{a_2-1}(1-x)^{-a_2}}{\mathsf{B}(a_2,1)}, \qquad \text{for } z\in(0,1-x),$$

and

$$\nu_0(\mathtt{d}x; z) = \frac{x^{a_1-1}(1-z)^{-a_1}}{\mathsf{B}(a_1, 1)}, \qquad \text{for } x \in (0, 1-z).$$

Then, the construction used throughout this chapter allows us to built a reversible Markov process X driven by the transition probability function,

$$\mathtt{k}(x_n, x_{n+1}) \quad = \quad \frac{x_{n+1}^{a_1-1}(1-x_n)^{-a_2}}{\mathsf{B}(a_1,1)\mathsf{B}(a_2,1)} \int_0^{1-x_n} z^{a_2-1}(1-z)^{-a_1}\mathtt{d}z,$$

for $x_{n+1} \in (0, x_n)$. The expression for the transition of the dual turns out to be similar to $\mathtt{k}$. In addition, the duality function takes the form

$$\mathtt{h}(x; z) = \frac{\Gamma(a_1+1)\Gamma(a_2+1)}{\Gamma(a_1+a_2+1)} \frac{1}{(1-x)^{a_2}(1-z)^{a_1}}.$$

As the gamma-negative-binomial case, we present it in order to illustrate a mechanism to choose the measure π and $\mathtt{f}$. Nonetheless, expression for the filters can be obtained from Theorem 1.

Chapter 3

A continuous-time dual Markov construction for filtering problems

The theory of continuous-time Markov processes is without doubt one of the most important mechanisms when modeling random phenomena. Indeed, its application to different fields of science is overwhelming, ranging from stock price modeling to the evolution of genes (*cf.* Shiryaev [40]; Ewens [10]). The literature on the topic is vast covering, different characterizations, simulation algorithms and estimation methods. The study and application of such a theory, heavily lies on a tractable way to represent the dynamics modulating the evolution of the process, typically done via transition probabilities, infinitesimal generators, stochastic equations, etc. From an statistical viewpoint, an appealing feature is the availability of a transition probability function, that can be easily incorporated into estimation, simulation and prediction problems. However, the most popular characterization of continuous-time Markov processes is via stochastic equations or infinitesimal generators. Although, such approaches have yield a rich and elegant class of models, these posses several operational complications, e.g. a transition density is not always available for the solutions of stochastic differential equations (SDE).

Given all this, our proposal in this chapter, is to extend the results of Chapter 2 to the continuous-time case. In other words, we look for a construction of strictly stationary continuous-time Markov processes via their transition probabilities. In particular, we are concerned with the conditional structure, previously explained, decomposing the transition function. Furthermore, we extend the results concerned the $\mathtt{h}$-dual to the continuous-time case. This latter task is not immediate, as a modification to achieve Markovianity is necessary. In particular, we derive an expression for the infinitesimal generator of the dual, that in turn help us to identify it within a know class of Markov processes. As a consequence of this, we devise an alternative solution to the filtering problem using such a dual. This extends a well known duality relation of the Wright-Fisher diffusion model, with mutation and without selection, and the Kingman coalescent (*cf.* Griffiths and Spanò [12]), to models that fall in our proposal.

Also, in this chapter, we explore extensions of the construction of autoregressive and Lancaster probabilities models, seen in Chapter 2, to the continuous time case. Various examples, aimed at illustrating our findings, are presented.

3.1 Construction and filtering

Let $\mathbb{X}$ and $\mathbb{Y}$ be two subsets of $\mathbb{R}^d$, for $d \geq 1$, $\mathscr{X} = \sigma(\mathbb{X})$ and $\mathscr{Y} = \sigma(\mathbb{Y})$. Also, let π be a probability measure over the Polish space $(\mathbb{X}, \mathscr{X})$; $\mathtt{f}\,(\cdot; \Theta_t, x)$ be a probability measures over the Polish space $(\mathbb{Y}, \mathscr{Y})$, for every $x \in \mathbb{X}$; and $\Theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a non-decreasing measurable function. Here, we assume that the support of π coincides with the set $\{x \in \mathbb{X} : \mathtt{f}\,(\,\cdot\,; \Theta_t, x) > 0\}$. Let us notice that, the time dependence has been added to $\mathtt{f}$, this allows to preserve the reversibility and stationarity in the model. For the sake of notation, as we did for the discrete case, we will no make distinction between the probability measures and their corresponding densities, assuming that they exist. Also, we are assuming that density functions are continuous with respect to the Lebesgue measure, letting clear that they can be continuous with respect to the count measure.

Thus, one can define a joint distribution over the product space $(\mathbb{X} \times \mathbb{Y}, \mathscr{X} \times \mathscr{Y})$ as follows

$$\int_B \int_A \mathtt{f}\,(u; \Theta_t, v)\pi(v)\mathtt{d}v\mathtt{d}u, \tag{3.1}$$

for $A \in \mathscr{X}$, $B \in \mathscr{Y}$ and $t > 0$. Thus, the marginal distributions of (3.1) over the spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$ are given by the probability measures π and

$$\mathtt{m}_\pi(\,\cdot\,;\Theta_t) = \int_{\mathbb{X}} \mathtt{f}\,(\,\cdot\,;\Theta_t, v)\pi(v)\mathrm{d}v,$$

respectively. Hence, the posterior distribution for π, denoted by ν_0, obtained via Bayes' theorem, has density function,

$$\nu_0(\mathtt{d}x; y, \Theta_t) = \frac{\mathtt{f}\,(\mathtt{d}y; \Theta_t, x)\pi(\mathtt{d}x)}{\mathtt{m}_\pi(\mathtt{d}y; \Theta_t)}, \tag{3.2}$$

for $x \in \mathbb{X}$, $y \in \mathbb{Y}$. The probability measure (3.2) is well-defined since $\mathtt{m}_\pi(A; \Theta_t) > 0$ for all $A \in \mathscr{Y}$. Recall that, if ν_0 and π belong to the same family of distributions, then we say that π is conjugate with respect to $\mathtt{f}$.

Then, we define a time-homogeneous process continuous in time over $(\mathbb{X}, \mathscr{X})$, denoted by $X = (X_t)_{t\geq 0}$, driven by the transition probability function

$$\mathtt{k}_t(x_0, \mathtt{d}x) = \int_{\mathbb{Y}} \nu_0(\mathtt{d}x; u, \Theta_t)\mathtt{f}\,(u; \Theta_t, x_0)\mathtt{d}u, \tag{3.3}$$

for $x_0, x \in \mathbb{X}$. Assuming that there is a function Θ such that $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations, a standard process X is completely characterized by the transition kernel (3.3) and its marginal probability measure π. Moreover, the transition (3.3) satisfies the following equalities

$$\begin{aligned} \pi(\mathtt{d}x_0) &= \int_{\mathbb{X}} \mathtt{k}_t(x, \mathtt{d}x_0)\pi(x)\mathtt{d}x, \\ \pi(\mathtt{d}x_0)\mathtt{k}_t(x_0, \mathtt{d}x) &= \pi(\mathtt{d}x)\mathtt{k}_t(x, \mathtt{d}x_0). \end{aligned}$$

for $x_0 \in \mathbb{X}$. Therefore, the transition probability kernel $\mathtt{k}_t$ has invariant measure π and, if such kernel satisfies Chapman-Kolmogorov's equations then it characterizes the law of continuous-time reversible and stationary Markov process X. In particular, the existence of such process is guaranteed whenever $\mathbb{X}$ is a Polish space (see Section 1.3).

Now, let $\{(X_{t_n}, Y_n)\}_{n\geq 0}$ be a hidden Markov model, where the signal $\{X_{t_n}\}_{n\geq 0}$ is a discrete-time sampling of X and; given the signal, $\{Y_n\}_{n\geq 0}$ is a sequence of conditional independent observations. Here, without loss of generality, we will assume that observations are equally spaced in time, *i.e.* $t_{i+1} - t_i = \Delta$ for all $i \geq 0$ and some $\Delta > 0$. Thus, the emission density, *i.e.* the law of $\{Y_n | X_{t_n}\}$ for $n \geq 0$, is chosen in such a way that it matches with the probability measure $\mathtt{f}\,(\,\cdot\,;\Theta_\Delta, x)$. Hence, hereafter, we denote the emission distribution by $\mathtt{f}\,(\,\cdot\,;x)$, without the time dependence, but letting clear that such dependency does exist.

Thus, denoting by $Y_{0:n} := (Y_0, ..., Y_n)$ and letting $\nu_n := \mathscr{L}(X_{t_n}|Y_{0:n})$. The exact and the prediction filters, starting at $\nu_{-1} = \pi$ and $\psi(\nu_{-1}) = \pi$, respectively, are given by the recursions

$$\nu_n = \phi_{Y_n}(\psi(\nu_{n-1})) \qquad \text{and} \qquad \psi(\nu_n),$$

respectively, for $n \geq 0$. The above update and predictor operators are defined in Section 1.5. The later operator is with respect to the transition $\mathtt{k}_\Delta$. For the sake of notation, we have dropped the dependence of $y^{(n)}$ in ν_n, only making it explicit for ν_0 when such dependence is different of y_0, *i.e.* $\nu_0(\cdot; u)$.

The next theorem provides an expression for the mentioned filters.

Theorem 2. *Let $\{(X_{t_n}, Y_n)\}_{n\geq 0}$ be a hidden Markov model, where $\{X_{t_n}\}_{n\geq 0}$ is a a discrete-time sample of X, which is driven by the transition probability function $\mathtt{k}_t$; and $\{Y_n\}_{n\geq 0}$ is a sequence of observations, equally spaced in time, which, given the signal, are conditionally independent and; with emission density $\mathtt{f}\,(\cdot; x)$. If $\mathtt{k}_t$ is given by* (3.3) *then the optimal and prediction filters, starting at*

$$\tilde{\omega}(y_0, u) = m_{\nu_0}(u) \qquad \nu_0 = \phi_{y_0}(\pi),$$

are given by the following expressions,

$$\begin{aligned}
\nu_{n+1}(dx_{n+1}) &= \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\phi_{y_{n+1}}(\nu_0(u))(dx_{n+1})\,du, \\
\psi(\nu_n)(dx_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\nu_0(dx_{n+1}; u)\,du,
\end{aligned}$$

where

$$\begin{aligned}
m_{\psi(\nu_n)}(dy_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u) m_{\nu_0(u)}(y_{n+1})\,du, \\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big) m_{\nu_0(u)}(y_{n+1})}{m_{\psi(\nu_n)}(y_{n+1})}, \\
\tilde{\omega}(y^{(n)}, u) &= \int_{\mathbb{X}} \nu_n(dx_n) \mathtt{f}(u; x_n)\,dx_n,
\end{aligned}$$

for $n \geq 0$. Moreover, if π is conjugate with respect to $f(\cdot; x)$, then ν_0 and $\phi_y(\nu_0)$ also belong to the same class of distributions as π. As a consequence, the probability measures m_{ν_0} and $q(y_n, u, z)$ belong to the same class of distributions as $m_{\nu_{-1}}$, where

$$\begin{aligned} \tilde{\omega}(y^{(n)}, u) &= \int_{\mathbb{Y}} \omega(y^{(n)}, k)\, q(y_n, u, z)\, dz \\ q(y_n, u, z) &= \int_{\mathbb{X}} \phi_{y_n}(\nu_0(z))(dx_n) f(u; x_n)\, dx_n. \end{aligned}$$

Note that, the optimal and the prediction filters given above depend of x_{n+1} only through the measures $\phi_y(\nu_0)$ and ν_0, respectively. This allows us to derive tractable expression for statistics associated to such filters. To be precise, for any real-valued function φ, we have that

$$\begin{aligned} \mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n+1)}\Big] &= \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\bigg(\int_{\mathbb{X}} \varphi(x_{n+1})\phi_{y_{n+1}}(\nu_0(u))(\mathrm{d}x_{n+1})\bigg)\mathrm{d}u, \\ \mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n)}\Big] &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\bigg(\int_{\mathbb{X}} \varphi(x_{n+1})\nu_0(\mathrm{d}x_{n+1}; u)\bigg)\mathrm{d}u, \end{aligned}$$

where ω and $\tilde{\omega}$ are given by the previous theorem. For instance, the above equations can be used to calculate the moment generating function or the moments associated to the optimal and the prediction filters.

Let us emphasize that, the crucial point of our proposal is finding the appropriate choice of the probability model, that is to say π and $\mathtt{f}(\cdot; \Theta_t, x)$, such that the transition probabilities $\mathtt{k}_t$ satisfy Chapman-Kolmogorov's equations. A way to deal with the above problem is to aggregate some time dependence in discrete-time models. In this sense, we will present later in this chapter the corresponding continuous-time version of some models presented in Chapter 2.

3.2 Duality

As in the discrete case, the conditional probability structure used in our construction guarantees the existence of a dual process. This will allow to provide an alternative way to compute the filters. For such purpose, let $Z = \{Z_t\}_{t\geq 0}$ be a continuous-time process driven by the transition probability function,

$$\mathtt{p}_t(z_0, \mathtt{d}z) = \int_{\mathbb{X}} \mathtt{f}\,(\mathtt{d}z; \Theta_t, v)\nu_0(\mathtt{d}v; z_0, \Theta_t)\mathtt{d}v, \tag{3.4}$$

for z_0, $z \in \mathbb{Y}$. It is straightforward to prove that the operator (3.4) satisfies the following equalities

$$\begin{aligned} \mathtt{m}_\pi(\mathtt{d}z_0; \Theta_t)\mathtt{p}_t(z_0, \mathtt{d}z) &= \mathtt{m}_\pi(\mathtt{d}z; \Theta_t)\mathtt{p}_t(z, \mathtt{d}z_0), \\ \mathtt{m}_\pi(\mathtt{d}z_0; \Theta_t) &= \int_{\mathbb{Y}} \mathtt{m}_\pi(z; \Theta_t)\mathtt{p}_t(z, \mathtt{d}z_0)\mathtt{d}z. \end{aligned}$$

Note that, the probability measure $\mathtt{m}_\pi(\cdot; \Theta_t)$ defines an invariant measure associated to the transition kernel $\mathtt{p}_t$, which clearly depends of t. Hence, the above results suggest that the process Z is non-homogeneous in time. Nevertheless, the conditional probability structure used to define X and Z allows us to derive a relationship between their transition probabilities. Let us first, for the sake of notation, notice that $\mathtt{p}_\Delta(u, \cdot) = \mathtt{m}_{\nu_0(u)}(\cdot)$.

Now, consider the function $\mathtt{h}$ defined as the Radon-Nikodyn derivative between the posterior and prior probability of the probability measure π, *i.e.*

$$\mathtt{h}(x; z, \Theta_t) = \frac{\nu_0(x; z, \Theta_t)}{\pi(x)} = \frac{\mathtt{f}(z; \Theta_t, x)}{\mathtt{m}_\pi(z; \Theta_t)}. \tag{3.5}$$

Hence, using the fact that X is reversible and the definition of $\mathtt{h}$ it follows that,

$$\begin{aligned} \int_{\mathbb{X}} \mathtt{h}(v; z, \Theta_t)\mathtt{k}_t(x, v)\mathtt{d}v &= \int_{\mathbb{X}} \frac{\nu_0(v; z, \Theta_t)}{\pi(v)} \left[\frac{\pi(v)}{\pi(\mathtt{d}x)}\mathtt{k}_t(v, \mathtt{d}x)\right] \mathtt{d}v \\ &= \int_{\mathbb{X}} \frac{\nu_0(v; z, \Theta_t)}{\pi(\mathtt{d}x)} \left(\int_{\mathbb{Y}} \nu_0(\mathtt{d}x; u, \Theta_t)\mathtt{f}(u; \Theta_t, v)\mathtt{d}u\right) \mathtt{d}v \\ &= \int_{\mathbb{Y}} \mathtt{h}(x; u, \Theta_t) \left[\int_{\mathbb{X}} \mathtt{f}(u; \Theta_t, v)\nu_0(v; z, \Theta_t)\mathtt{d}v\right] \mathtt{d}u \\ &= \int_{\mathbb{Y}} \mathtt{h}(x; u, \Theta_t)\mathtt{p}_t(z, \mathtt{d}u). \end{aligned}$$

Equivalently,

$$\mathbb{E}^x[\mathtt{h}(X_t; z, \Theta_t)] = \mathbb{E}^z[\mathtt{h}(x; Z_t, \Theta_t)], \tag{3.6}$$

$x \in \mathbb{X}$, $y \in \mathbb{Y}$ and $t \in \mathbb{R}^+$. When two Markov processes satisfy the above equality,

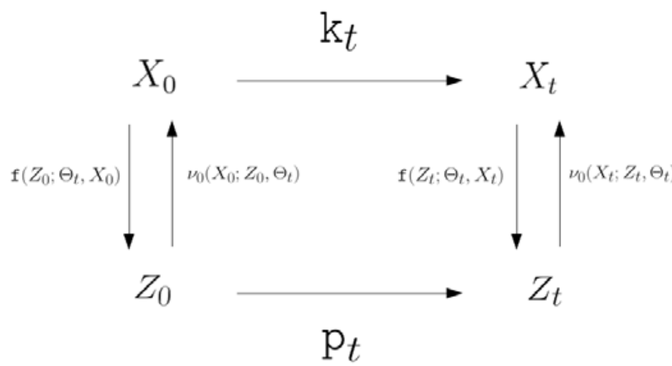

Figure 3.1: Duality between the prediction operators $\mathtt{k}$ and $\mathtt{p}$.

it is said that they are dual to each other with respect to the duality function $\mathtt{h}$. However, under the assumption that $\mathtt{k}_t$ satisfies Chapman-Kolmogorov property, the operator $\mathtt{p}_t$ does not necessarily satisfy such property. This implies that the equality (3.6) is just a property between two transitions kernels rather than two concrete processes. However, it will help us to guarantee the existence of a Markovian dual.

Before to prove the last statement, let us present a couple of equalities related to the predictor operators of the processes X and Z,

$$\begin{aligned}\int_{\mathbb{X}} \nu_0(v;z_0,\Theta_t)\mathtt{k}_t(v,\mathtt{d}x)\mathtt{d}v &= \int_{\mathbb{Y}} \nu_0(\mathtt{d}x;u,\Theta_t)\mathtt{p}_t(z_0,u)\mathtt{d}u,\\ \int_{\mathbb{Y}} \mathtt{f}(u;\Theta_t,x_0)\mathtt{p}_t(u,\mathtt{d}z)\mathtt{d}u &= \int_{\mathbb{X}} \mathtt{f}(\mathtt{d}z;\Theta_t,v)\mathtt{k}_t(x_0,v)\mathtt{d}v,\end{aligned}$$

for $x_0, x \in \mathbb{X}$ and $z_0, z \in \mathbb{Y}$. Thus, the construction allows us to compute the above operators in terms of its corresponding transition dual. Figure 3.1 describes the relationship obtained in the above equalities.

Assuming that we know the infinitesimal generator of the process X, denoted by $\mathscr{A}$, the equality (3.6) implies that,

$$\begin{aligned}\mathscr{A}\mathtt{h}(\,\cdot\,;z,\Theta_t)(x) &= \lim_{t\downarrow 0}\frac{1}{t}\left[\int_{\mathbb{X}} \mathtt{h}(v;z,\Theta_t)\mathtt{k}_t(x,v)\mathtt{d}v - \mathtt{h}(x;z,\Theta_t)\right]\\ &= \lim_{t\downarrow 0}\frac{1}{t}\left[\int_{\mathbb{Y}} \mathtt{h}(x;u,\Theta_t)\mathtt{p}_t(z,u)\mathtt{d}u - \mathtt{h}(x;z,\Theta_t)\right].\end{aligned}$$

Hence, one can guarantee that the infinitesimal generator $\mathscr{A}$ applied to $\mathtt{h}$ can be expressed as a function of (z,Θ_t), *i.e.* taking x as a constant. Thus, we can define an operator $\mathscr{G}$ that satisfies the following equality

$$\mathscr{A}\mathtt{h}(\cdot; z, \Theta_t)(x) = \mathscr{G}\mathtt{h}(x; \cdot, \Theta_t)(z). \tag{3.7}$$

As a result, there is a Markov process characterized by the infinitesimal generator $\mathscr{G}$. So, let us define an homogeneous Markov process $\tilde{Z}$ driven by a transition operator $\tilde{\mathtt{p}}_t$ associated to $\mathscr{G}$. Then, since (3.7) holds,

$$\mathbb{E}^x[\mathtt{h}(X_t; z, \Theta_t)] = \mathbb{E}^z[\mathtt{h}(x; \tilde{Z}_t, \Theta_t)], \tag{3.8}$$

also holds, where the expectation of the right-hand side is with respect to $\tilde{\mathtt{p}}_t$. See Jansen and Kurt [15].

On the other hand, equality (3.7) allows to obtain an expression for the infinitesimal generator $\mathscr{G}$ applied to the duality function $\mathtt{h}$. This will helps us to identify it within a class of Markov processes. A simple way to do this is applying $\mathscr{A}$ to $\mathtt{h}$ and write it as a function of z. A similar approach was used in Chaleyat and Catalot [4], for the Wright-Fisher diffusion, and later replicated by Papaspiliopoulos and Ruggiero [33] for the Cox-Ingersoll-Ross and the Ornstein-Uhlenbeck diffusions. Let us emphasize that, our construction allows us to derive an expression for $\mathscr{G}$ applied to $\mathtt{h}$. This is done via the *backward* equation for Markov processes, *i.e.*

$$\mathscr{A}\mathtt{k}_t(x_0, \mathtt{d}x) = \frac{\partial}{\partial t}\mathtt{k}_t(x_0, \mathtt{d}x). \tag{3.9}$$

Then, the reversibility of X let us write the left-hand side of the equation (3.9) in the following form,

$$\begin{aligned}
\mathscr{A}\mathtt{k}_t(\,\cdot\,, \mathtt{d}x)(x_0) &= \mathscr{A}\left[\frac{\pi(\mathtt{d}x)}{\pi(\,\cdot\,)}\mathtt{k}_t(x,\,\cdot\,)\right](x_0) \\
&= \pi(\mathtt{d}x)\int_{\mathbb{Y}}\left[\mathscr{A}\mathtt{h}(\,\cdot\,; z, \Theta_t)(x_0)\right]\mathtt{f}\,(z; \Theta_t, x)\mathtt{d}z \\
&= \pi(\mathtt{d}x)\int_{\mathbb{Y}}\left[\mathscr{G}\mathtt{h}(x_0;\,\cdot\,, \Theta_t)(z)\right]\mathtt{f}\,(z; \Theta_t, x)\mathtt{d}z,
\end{aligned}$$

where the second equality holds by the definition of $\mathtt{k}_t$ and the last equality holds by the duality equation (3.7). Similarly, the right-hand side of the equation (3.9) takes the form

$$\frac{\partial}{\partial t}\mathtt{k}_t(x_0,\mathtt{d}x) = \pi(\mathtt{d}x)\Bigg(\int_{\mathbb{Y}}\Big[\frac{\partial}{\partial t}\mathtt{h}(x_0;z,\Theta_t)\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z + \int_{\mathbb{Y}}\Big[\mathtt{h}(x_0;z,\Theta_t)\Big\{\frac{\partial}{\partial t}\log\mathtt{f}\,(z;\Theta_t,x)\Big\}\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z\Bigg).$$

Hence, the *backward* equation (3.9) becomes

$$\int_{\mathbb{Y}}\Big[\mathscr{G}\mathtt{h}(x_0;\cdot,\Theta_t)(z)\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z = \int_{\mathbb{Y}}\Big[\frac{\partial}{\partial t}\mathtt{h}(x_0;z,\Theta_t)\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z + \int_{\mathbb{Y}}\mathtt{h}(x_0;z,\Theta_t)\Big[\frac{\partial}{\partial t}\log\mathtt{f}\,(z;\Theta_t,x)\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z.$$

If we separate by terms and we factor $\mathtt{f}\,(z;\Theta_t,x)$, then we can obtain an expression for $\mathscr{G}$ applied to $\mathtt{h}$ as a function of z. However, the second term in the right-hand side of the above equation is also a function of x, and it is necessary to express it only as a function of z.

To deal with this issue notice that, the equality (3.2) allows us to write $\{\log\mathtt{f}\,(z,\Theta_t,x)\}$ as the sum of two functions, one depends on (x,z,Θ_t) and the other one depends on (z,Θ_t), *i.e.*

$$\frac{\partial}{\partial\Theta_t}\log\mathtt{f}\,(z;\Theta_t,x) = \mathtt{g}_1(z,\Theta_t) + \mathtt{g}_2(x;z,\Theta_t), \tag{3.10}$$

for some functions $\mathtt{g}_1$ and $\mathtt{g}_2$. Hence, the goal is to find functions $\mathtt{g}$, $\tilde{\mathtt{g}}_2$, such that

$$\int_{\mathbb{Y}}\Big[\mathtt{h}(x_0;z,\Theta_t)\mathtt{g}_2(x;z,\Theta_t) - \mathtt{h}(x_0;\mathtt{g}(z),\Theta_t)\tilde{\mathtt{g}}_2(z,\Theta_t)\Big]\mathtt{f}\,(z;\Theta_t,x)\mathtt{d}z = 0. \tag{3.11}$$

These functions exist because the equality (3.7) holds. This result is summarized in the next theorem.

Theorem 3. *Let π be a probability measure over $(\mathbb{X},\mathscr{X})$; $\mathtt{f}\,(\cdot;\Theta_t,x)$ be a probability measure over $(\mathbb{Y},\mathscr{Y})$, for every $x\in\mathbb{X}$; and $\Theta:\mathbb{R}^+\to\mathbb{R}^+$ be a non-decreasing measurable function. Consider the Markov process X, which is driven by the transition probability functions* (3.3) *and with invariant measure π. Assuming that the infinitesimal generator of X is known, denoted by $\mathscr{A}$, then there exists a Markov process $\tilde{Z}=\{\tilde{Z}_t\}_{t\geq 0}$ which is dual to X with respect to the dual function* (3.5).

Moreover, the infinitesimal generator of $\tilde{Z}$, denoted by $\mathscr{G}$, satisfies,

$$\begin{aligned}\mathscr{G}\boldsymbol{h}(x;\cdot,\Theta_t)(y) &= \frac{\partial\Theta_t}{\partial t}\Big\{\big[\boldsymbol{g}_1(z,\Theta_t)\big]\boldsymbol{h}(x;z,\Theta_t)+\big[\tilde{\boldsymbol{g}}_2(z,\Theta_t)\big]\boldsymbol{h}(x;\boldsymbol{g}(z),\Theta_t)\Big\}\\ &\quad \frac{\partial\Theta_t}{\partial t}\Big[\frac{\partial}{\partial\Theta_t}\boldsymbol{h}(x;z,\Theta_t)\Big],\end{aligned} \tag{3.12}$$

where $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$ are obtain by (3.10)*; $\boldsymbol{g}$ and $\tilde{\boldsymbol{g}}_2$ are obtain by* (3.11).

It is worth noticing that, the infinitesimal generator $\mathscr{G}$ is completely characterized by $\mathtt{f}(\cdot;x)$. Moreover, expression (3.12) suggests that the dual $\tilde{Z}$ must be subordinated by Θ in order to be Markovian. If the duality function does not depend on t, as in the Wright-Fisher diffusion, then the dual is not subordinated by Θ. In such cases, the computation of the transition probabilities, associated to a given infinitesimal generator, is simpler.

Additionally, equality (3.8) leads to the following result that will help to give an alternative way to compute the filters.

Proposition 2. *Let π be a probability measure over $(\mathbb{X},\mathscr{X})$; $\boldsymbol{f}(\cdot;\Theta_t,x)$ be a probability measure over $(\mathbb{Y},\mathscr{Y})$, for every $x\in\mathbb{X}$; and $\Theta:\mathbb{R}^+\to\mathbb{R}^+$ be a non-decreasing measurable function. Assuming that, the operator $\boldsymbol{k}_t$ defined by* (3.3) *satisfies Chapman-Kolmogorov's property. Then, the reversible Markov process X driving by* (3.3) *and with invariant measure π has a Markovian $\boldsymbol{h}$-dual, denoted by $\tilde{Z}$ driven by $\tilde{\mathtt{p}}_t$. Moreover,*

$$\int_{\mathbb{X}}\nu_0(v;z_0,\Theta_t)\boldsymbol{k}_t(v,\mathrm{d}x)\,\mathrm{d}v=\int_{\mathbb{Y}}\nu_0(\mathrm{d}x;u,\Theta_t)\tilde{\mathtt{p}}_t(z_0,u)\,\mathrm{d}u, \tag{3.13}$$

for all $t\in\mathbb{R}^+$, $x\in\mathbb{X}$, $z_0\in\mathbb{Y}$, and where $\nu_0(\cdot;z,\Theta_t)$ is given by (3.2).

Proof. The first assertion was already proved. Now, for $x\in\mathbb{X}$, $z_0\in\mathbb{Y}$ and $t\geq 0$ it follows that

$$\begin{aligned}\int_{\mathbb{X}}\nu_0(v;z_0,\Theta_t)\mathtt{k}_t(v,\mathtt{d}x)\mathtt{d}v &= \int_{\mathbb{X}}\mathtt{h}(v;z_0,\Theta_t)\pi(v)\mathtt{k}_t(v,\mathtt{d}x)\mathtt{d}v\\ &= \pi(dx)\int_{\mathbb{X}}\mathtt{h}(v;z_0,\Theta_t)\mathtt{k}_t(x,v)\mathtt{d}v\\ &= \int_{\mathbb{Y}}\nu_0(\mathtt{d}x;u,\Theta_t)\tilde{\mathtt{p}}_t(z_0,u)\mathtt{d}u,\end{aligned}$$

where the second equality holds by reversibility and the last by duality. □

This proposition allows us to compute the predictor operator of X via the transition probability function of its corresponding dual. Papaspiliopoulos and Ruggiero [33] obtained an expression similar to (3.13) for some particular models. In fact, their result is based on the assumption that the signal is reversible and has a Markovian dual, which is a particular death process. In addition, they use a family of distributions defined by $\{\mathtt{h}(x; y, \Theta_t)\pi(\mathtt{d}x); y \in \mathbb{Y}\}$ as the posterior distribution of the signal, which makes the distribution π conjugate with respect to the emission density.

Furthermore, equality (3.13) helps to derive an alternative computation of the filters. Since the left-hand side of (3.13) turns out to be the first step in the recursion that compute the optimal filters.

3.3 Examples

This section presents some appealing models that fall into our proposal and, for each of these we derive the recursion for their filters. Then, we obtain an expression for the infinitesimal generator's dual. First, we present a continuous-time version of the gamma-Poisson model, which is conjugate and turns out to be a re-parametrization of the Cox-Ingersoll-Ross (CIR) diffusion. Then, we will see that the Wright-Fisher diffusion belongs to the class of Markov processes that we propose. For this model, the dual of such diffusion is a version of Kingman's Coalescent (see for instance Griffiths and Spanò [12]).

3.3.1 Gamma-Poisson process

Let π be a $\mathrm{Ga}(a, b)$ distribution and $\mathtt{f}\,(y; \Theta_t, x)$ be a $\mathrm{Po}(x\Theta_t)$ distribution, for each $x \in \mathbb{R}^+$. Hence the posterior distribution of X is $\mathrm{Ga}(a+y, b+\Theta_t)$ and, its transition probability function takes the form

$$\mathtt{k}_t(x_0, x) = \frac{\exp\{-[\,\Theta_t(x_0+x)+bx\,]\}}{(\Theta_t+b)^{-\frac{a+1}{2}}\Theta_t^{\frac{a-1}{2}}} \times \left(\frac{x}{x_0}\right)^{\frac{a-1}{2}} \times \mathtt{I}_{a-1}\left(2\sqrt{x_0 x\Theta_t(b+\Theta_t)}\right),$$

for $x_0,\ x \in \mathbb{R}^+$, where $\mathtt{I}_\nu(\cdot)$ denotes the modified Bessel function of the first kind with argument ν. In this case, if $\Theta_t = b/(e^{ct}-1)$, with $c>0$, then $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations (*cf.* Mena and Walker [30]). Moreover, the process X is the only strong solution to the following stochastic differential equation

$$\mathtt{d}X_t = c\left(\frac{a}{b} - X_t\right)\mathtt{d}t + \sqrt{\frac{2c}{b}X_t}\mathtt{d}W_t,$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. This process is known as Gamma-Poisson model and constitutes a re-parameterization of the stationary version of the Cox-Ingersoll-Ross (CIR) model (*cf.* Cox, Ingersoll and Ross [5]).

For this model, it is straightforward to obtain the transition probability function of the process Z, that takes the form

$$\mathtt{p}_t(z_0, z) = \frac{\Gamma(a+z_0+z)}{z!\Gamma(a+z_0)}\left(\frac{b+\Theta_t}{b+2\Theta_t}\right)^{a+z_0}\left(\frac{\Theta_t}{b+2\Theta_t}\right)^{z}.$$

for $z_0,\ z \in \mathbb{Z}^+ \cup \{0\}$. The above operator is a binomial-negative kernel, denoted by $\mathrm{NB}(z; a+z_0, (b+\theta_t)/(b+2\theta_t))$.

In order to compute the filters, we consider observations equally spaced. Thus, starting with the measures ν_0 and $\tilde{\omega}(y_0, u)$ which has $\mathrm{Ga}(x_0; a+y_0, b+\Theta_\Delta)$ and $\mathrm{NB}(u; a+y_0+1, b/(b+2\Theta_\Delta))$ distributions, respectively. The optimal and the prediction filters are given by

$$\begin{aligned}
\nu_{n+1}(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{\infty}\omega\big(y^{(n+1)}, u\big)\,\mathrm{Ga}\big(\mathtt{d}x_{n+1}; a+u+y_{n+1}, b+2\Theta_\Delta\big),\\
\psi(\nu_n)(\mathtt{d}x_{n+1}) &= \sum_{u=0}^{\infty}\tilde{\omega}(y^{(n)}, u)\,\mathrm{Ga}\big(\mathtt{d}x_{n+1}; a+u, b+\Theta_\Delta\big),
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \sum_{u=0}^{\infty}\tilde{\omega}(y^{(n)}, u)\,\mathrm{NB}\Big(\mathtt{d}y_{n+1}; a+u+1, \frac{b}{b+2\Theta_\Delta}\Big),\\
\omega\big(y^{(n+1)}, u\big) &= \frac{\tilde{\omega}\big(y^{(n)}, u\big)\mathrm{NB}\Big(y_{n+1}; a+u+1, \frac{b}{b+2\Theta_\Delta}\Big)}{\mathsf{m}_{\psi(\nu_n)}(dy_{n+1})},\\
\tilde{\omega}(y^{(n)}, u) &= \sum_{k=0}^{\infty}\omega(y^{(n)}, k)\mathrm{NB}\Big(k; a+u+y_1+1, \frac{\Theta_\Delta}{b+3\Theta_\Delta}\Big).
\end{aligned}$$

for $n \geq 0$. Note that, since π is conjugate with respect to $\mathtt{f}$, the above filters are mixtures of gamma distributions. Moreover, the weights are given by mixtures of negative-binomial distributions. Hence, conjugacy reduces the computational effort required in order to compute the filters. In particular, this model requires the computation of infinite sums, however each term in every sum is positive. This allows us to truncate such sums obtaining an exact result.

Additionally, one can compute statistics associated to the above filters. Indeed, the moment generating function of the optimal and the prediction filters take the form

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda X_{t_{n+1}}}\middle|\mathscr{Y}^{(n+1)}\right] &= \int_{\mathbb{Y}} \omega(y^{(n+1)}, u)\left(1 - \frac{\lambda}{b + 2\Theta_\Delta}\right)^{-(a+u+y_{n+1})} \mathtt{d}u, \\
\mathbb{E}\left[e^{\lambda X_{t_{n+1}}}\middle|\mathscr{Y}^{(n)}\right] &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)}, u)\left(1 - \frac{\lambda}{b + \Theta_\Delta}\right)^{-(a+u)} \mathtt{d}u,
\end{aligned}$$

where ω and $\tilde{\omega}$ are given before. That is to say, these statistics are mixture of moment generating functions associated to gamma distributions.

On other hand, define $\mathtt{h}$ as the Radon-Nikodyn derivative between the posterior and prior distribution of π. Then, an expression for the infinitesimal generator's dual is obtained by noticing that

$$\frac{\partial}{\partial t} \log \mathtt{f}\,(z; \Theta_t, x) = -x\frac{\partial \Theta_t}{\partial t} + \frac{z}{\Theta_t}\frac{\partial \Theta_t}{\partial t}.$$

Hence, from the above equality, the functions $\mathtt{g}_1$ and $\mathtt{g}_2$ in Theorem 3.1 are given by $\mathtt{g}_1(z, \Theta_t) = (z/\Theta_t)(\partial\Theta_t/\partial t)$ and $\mathtt{g}_2(x, z, \Theta_t) = -x(\partial\Theta_t/\partial t)$. Also, we have that $\partial\Theta_t/\partial t = \frac{c}{b}(b + \Theta_t)\Theta_t$, which implies

$$\begin{aligned}
\sum_{u=0}^{\infty} \mathtt{h}(x_0; u, \Theta_t) x \mathtt{f}\,(u; \Theta_t, x) &= \sum_{u=0}^{\infty} \mathtt{h}(x_0; u, \Theta_t)(u+1)\left[e^{-\Theta_t x}\frac{(x\Theta_t)^{u+1}}{(u+1)!}\right] \\
&= \sum_{u=0}^{\infty} \mathtt{h}(x_0; u-1, \Theta_t) u \left[e^{-\Theta_t x}\frac{(x\Theta_t)^{u}}{u!}\right],
\end{aligned}$$

which means that $\tilde{\mathtt{g}}_2(z, \Theta_t) = -(c/b)(b + \Theta_t)z$ and $\mathtt{g}(z) = z - 1$. Therefore, the infinitesimal generator of the $\mathtt{h}$-dual to X applied to $\mathtt{h}$ is given by,

$$\begin{aligned}\mathscr{G}\mathtt{h}(x;\cdot,\theta_t)(z) &= \frac{c}{b}(b+\Theta_t)z\mathtt{h}(x;z,\Theta_t) - \frac{c}{b}(b+\Theta_t)zh(x;z-1,\Theta_t) \\ &+\frac{c}{b}\Theta_t(b+\Theta_t)\frac{\partial}{\partial\Theta_t}\mathtt{h}(x;z,\Theta_t).\end{aligned}$$

The above operator suggest that the dual is a death process subordinated by a deterministic one, the later driven by the ordinary differential equation $\mathtt{d}\Theta_t = (c/b)(b+\Theta_t)\Theta_t\mathtt{d}t$.

Let us notice that, a similar duality was used in Papaspiliopoulos and Ruggiero [33], however their duality function, unlike ours, does not depend on t. Nonetheless, they found an expression for the transition probability function associated to the operator $\mathscr{G}$. Here we notice that, this transition probability does not seem to satisfy the duality equation, which they assumed in their work.

3.3.2 Wright-Fisher diffusion

The Wright-Fisher diffusion with no selection and parent mutation has a widely known eigenvalue expansion for the transition probability function (*cf.* Griffiths and Spanò [12]). Such expansion allows us to build the Wright-Fisher diffusion via the construction described in this chapter. Thus, let π be a beta($\mathbf{a}$) distribution, with $\mathbf{a}=(a_1,a_2)$, $a_1,a_2>0$ and $|\mathbf{a}|=a_1+a_2$. Also, for each $x\in[0,1]$, let $\mathtt{f}\,(\cdot;\Theta_t,x)$ be a bivariate probability measure given by

$$\mathtt{f}\,(\mathbf{m};\Theta_t,x) = \mathrm{Bin}\big(m_1;|\mathbf{m}|,x\big)\mathtt{q}_{|\mathbf{m}|}^{|\mathbf{a}|}(t) \qquad \text{for } \mathbf{m}=(m_1,m_2)\in\mathbb{Z}_+^2,$$

where $\mathrm{Bin}(m_1;|\mathbf{m}|,x)$ stands for a Binomial($|\mathbf{m}|,x$) distribution and, $\mathtt{q}_{|\mathbf{m}|}^{|\mathbf{a}|}(t)$ is the transition probability function of a death process with entrance boundary of infinity, with death rates $|\mathbf{m}|(|\mathbf{a}|+|\mathbf{m}|-1)/2$, $|\mathbf{m}|\geq 1$. In fact, such transition takes the form

$$\mathtt{q}_{|\mathbf{m}|}^{|\mathbf{a}|}(t) = \sum_{j=|\mathbf{m}|}^{\infty}\rho_j^{|\mathbf{a}|}(t)(-1)^{j-|\mathbf{m}|}\frac{(2j+|\mathbf{a}|-1)\big(|\mathbf{m}|+|\mathbf{a}|\big)_{(j-1)}}{|\mathbf{m}|!\big(j-|\mathbf{m}|\big)!},$$

with $\rho_j^{|\mathbf{a}|}(t)=\exp\big\{-\frac{1}{2}j(|\mathbf{a}|+j-1)t\big\}$. In fact, $\mathtt{q}_{|\mathbf{m}|}^{|\mathbf{a}|}(t)$ turns out to be the transition probability function of a death process with entrance boundary of infinity and, with death rates $|\mathbf{m}|(|\mathbf{a}|+|\mathbf{m}|-1)/2$, $|\mathbf{m}|\geq 1$. Hence, the posterior distribution of π in this model is given by a beta$(\mathbf{a}+\mathbf{m})$ distribution, noticing that such posterior does not depend of the time t. Then, let $X=(X_t)_{t\geq 0}$ be a Markov process driven by the transition probability function,

$$\mathtt{k}_t(x_0,\mathtt{d}x)=\sum_{|\mathbf{m}|=0}^{\infty}\mathtt{q}_{|\mathbf{m}|}^{|\mathbf{a}|}(t)\sum_{m_1=0}^{|\mathbf{m}|}\text{Beta}(\mathtt{d}x;\mathbf{a}+\mathbf{m})\text{Bin}\big(m_1;|\mathbf{m}|,x_0\big),$$

for x_0, $x\in[0,1]$. Moreover, it is known that the infinitesimal generator of X takes the form

$$\mathscr{A}=\frac{1}{2}\big(a_1-|\mathbf{a}|x\big)\frac{\partial}{\partial x}+\frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}.$$

See for instance, Griffiths and Spanó [12].

Thus, starting at ν_0 which has Beta$(\mathbf{a}+\mathbf{m}_0)$ distribution, the optimal and the prediction filters are given by

$$\begin{aligned}\nu_{n+1}(\mathtt{d}x_{n+1}) &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\omega\big(\mathbf{m}^{(n+1)},\mathbf{u}\big)\,\text{Beta}(x_{n+1};\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1}),\\ \psi(\nu_n)(\mathtt{d}x_{n+1}) &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u})\,\text{Beta}(x_{n+1};\mathbf{a}+\mathbf{u}),\end{aligned}$$

where $\mathbf{m}_n=(m_n^1,m_n^2)$ and

$$\begin{aligned}\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}\mathbf{m}_{n+1}) &= \sum_{u=0}^{\infty}\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u})\,q_{|\mathbf{m}_{n+1}|}^{|\mathbf{a}|}(t)\binom{|\mathbf{m}_{n+1}|}{m_{n+1}^1}\frac{\mathbf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})}{\mathbf{B}(\mathbf{a}+\mathbf{u})},\\ \omega\big(\mathbf{m}^{(n+1)},\mathbf{u}\big) &= \frac{\tilde{\omega}\big(\mathbf{m}^{(n)},\mathbf{u}\big)\,q_{|\mathbf{m}_{n+1}|}^{|\mathbf{a}|}(t)\binom{|\mathbf{m}_{n+1}|}{m_{n+1}^1}\frac{\mathbf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})}{\mathbf{B}(\mathbf{a}+\mathbf{u})}}{\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}\mathbf{m}_{n+1})}\\ \tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u}) &= \sum_{k=0}^{\infty}\omega(\mathbf{m}^{(n)},k)\mathtt{q}_{|\mathbf{k}|}^{|\mathbf{a}|}(t)\binom{|\mathbf{k}|}{k_1}\frac{\mathbf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_n+\mathbf{k})}{\mathbf{B}(\mathbf{a}+\mathbf{u}+\mathbf{m}_n)}.\end{aligned}$$

for $n\geq 0$. Let us notice that, since π is conjugate with respect to $\mathtt{f}$, the about filters are mixtures of beta distributions. Also, conjugacy implies that, instead of compute the integrals corresponding to m_{ν_0} and $\mathtt{q}$, in Theorem 2.1, we compute the expression of proportional beta-binomial distributions.

Additionally, as before, one can compute statistics associated to the optimal and the prediction filters as follows

$$\begin{aligned}\mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n+1)}\Big] &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\omega(\mathbf{m}^{(n+1)},\mathbf{u})\Bigg(\int_0^1\varphi(x_{n+1})\mathrm{Beta}(x_{n+1};\mathbf{a}+\mathbf{u}+\mathbf{m}_{n+1})\Bigg)\mathtt{d}\mathbf{u},\\ \mathbb{E}\Big[\varphi(X_{t_{n+1}})\Big|\mathscr{Y}^{(n)}\Big] &= \sum_{\mathbf{u}\in\mathbb{Z}_+^2}\tilde{\omega}(\mathbf{m}^{(n)},\mathbf{u})\Bigg(\int_0^1\varphi(x_{n+1})\mathrm{Beta}(x_{n+1};\mathbf{a}+\mathbf{u})\Bigg)\mathtt{d}\mathbf{u},\end{aligned}$$

where ω and $\tilde{\omega}$ are given before. Hence, these statistics are given by mixtures of the corresponding statistics of beta distributions.

On the other hand, let $\mathtt{h}$ be the Radon-Nikodyn derivative between the posterior and prior distribution of π. This implies that,

$$\mathtt{h}(x;\mathbf{m}) = \frac{\mathrm{B}(\mathbf{a})}{\mathrm{B}(\mathbf{a}+\mathbf{m})}x^{m_1}(1-x)^{m_2}.$$

It is worth noticing that, since the posterior distribution of π does not depend of t, the dual function $\mathtt{h}$ does not depend of t either. Then, applying the infinitesimal generator $\mathscr{A}$ to $\mathtt{h}$ and rearranging the terms we obtain

$$\begin{aligned}\mathscr{G}\mathtt{h}(x;\cdot,\cdot)(\mathbf{m}) &= \frac{|\mathbf{a}|+|\mathbf{m}|-1}{2}\Big[m_1\,\mathtt{h}(x;\mathbf{m}-\mathbf{e}_1)+m_2\mathtt{h}(x;\mathbf{m}-\mathbf{e}_2)\Big]\\ &\quad -\frac{|\mathbf{m}|(|\mathbf{m}|+|\mathbf{a}|-1)}{2}\mathtt{h}(x;\mathbf{m}). \qquad (3.14)\end{aligned}$$

where $\mathbf{e}_1=(1,0)$ and $\mathbf{e}_2=(0,1)$. Thus, if we associate $\mathscr{G}$ to a time-homogeneous Markov process, as it was done by Chaleyat and Catalot [4]). Then, the equality (3.14) suggests that the $\mathtt{h}$-dual is a death process that jumps down from state $\mathbf{m}$ to $(\mathbf{m}-\mathbf{e}_j)$ with instantaneous rate

$$\frac{|\mathbf{a}|+|\mathbf{m}|-1}{2}m_j$$

for $j=1,2$. As a matter of fact, the transition probability function of such death processes is given by

$$\tilde{\mathtt{p}}_t(\mathbf{m},\mathbf{n}) = \mathtt{q}^{\mathbf{a}}_{|\mathbf{m}|,|\mathbf{n}|}(t)\frac{\binom{m_1}{n_1}\binom{m_2}{n_2}}{\binom{|\mathbf{m}|}{|\mathbf{n}|}}$$

where

$$\mathfrak{q}^{a}_{m,n}(t) = \sum_{j=n}^{m} \rho^{a}_{j}(t)(-1)^{j-n} \frac{(2j+a-1)(n+a)_{(j-1)}(n)_{\llcorner j \lrcorner}}{n!(j-n)!(m+a)_{(j)}}$$

with $(n)_{(j)} = n(n+1)\cdot(n+j-1)$ and $(n)_{\llcorner j \lrcorner} = n(n-1)\cdots(n-j+1)$. Clearly, the above transition requires the computation of finite sums, which implies that this model has finite dimensional filters. Furthermore, the multivariate version of the Wright-Fisher diffusion resemblances the above model and, its dual can be found in Papaspiliopoulos and Ruggiero [33].

Let us emphasize that, the existence of the above dual is guaranteed by the conditional probability structure used in our proposal. This extends the work of Chaleyat and Catalot [4] which, using a different approach, found such duality for the Wright-Fisher diffusion.

3.4 Continuous-time stationary models with exponential dispersion margin

This section deals with the problem of specifying the probability measures π and $\mathtt{f}$, by applying a thinning operation over a exponential dispersion distribution. Unlike the discrete case, here we need to aggregate a real-valued function to the model, this will be done through the contraction distribution (see Section 2.4). Then, we will check if the transition probabilities that we propose satisfy Chapman-Kolmogorov's property. For those models that satisfy such property, we derive their corresponding dual and filters.

3.4.1 Thinning operation

Let $X = (X_t)_{t\geq 0}$ be a continuous-time reversible process over $(\mathbb{X}, \mathscr{X})$, where $\mathbb{X} \subset \mathbb{R}^d$ for $d \geq 1$ and $\mathscr{X} = \mathscr{B}(\mathbb{X})$, with invariant measure π, which belong to the class of exponential dispersion distributions. This implies that the density probability function associated to π has the form

$$\pi(x) = c(x;\tau)\exp\left\{x\vartheta - \tau * \kappa(\vartheta)\right\},$$

where $\kappa(\vartheta) = \log\big\{\int c(x;\tau)e^{\vartheta x}\mathtt{d}x\big\}/\tau$, for $\vartheta \in \Xi = \mathtt{int}\{\vartheta \in \mathbb{R}^d; \kappa(\vartheta) < \infty\}$ and Ξ is assumed to be non-empty. Using a thinning operation to define a contraction of π, denoted by $G(\tau_1(t), \tau_2(t), x)$, one can define a probability model $\{\mathtt{f}(\cdot;\Theta_t, x); x \in \mathbb{R}\}$, such that

$$\mathtt{f}(z;\Theta_t, x) = \frac{c\big(x - z; \tau_2(t)\big)c\big(z; \tau_1(t)\big)}{c(x;\tau)},$$

where $\tau_1(t) = \tau\Theta_t$ and $\tau_2(t) = \tau(1 - \Theta_t)$ for all $t \geq 0$, with $\Theta : \mathbb{R}_+ \to (0, 1)$. Thus, Bayes' theorem allows us to compute the posterior distribution of π as following

$$\nu_0(x; z, \Theta_t) = c\big(x - z; \tau_2(t)\big) \exp\big\{(x - z)\vartheta - \tau_2(t)\kappa(\vartheta)\big\}$$

Clearly, the probability measure ν_0 as a function of $x - z$ has $\mathrm{ED}(\vartheta; \tau_2(t))$ distribution. Then, following the procedure of the previous sections, the process X is driven by the transition probability function

$$\mathtt{k}_t(x_0, \mathtt{d}x) = \int_{\mathbb{Y}} \nu_0(x; z, \Theta_t)\mathtt{f}(z; \Theta_t, x_0)\mathtt{d}z,$$

for $x_0, x \in \mathbb{X}$. As we mentioned before, we need to check if the above transition operator satisfies Chapman-Kolmogorov's property. In particular, we look for an expression for the moment generating function associated to the above operator, which in this case is given by

$$\mathscr{L}_{X_t|X_0=x_0}(\lambda) \quad = \quad \exp\Big\{\tau_2(t)\big[\kappa(\vartheta + \lambda) - \kappa(\vartheta)\big]\Big\} \int_{\mathbb{Y}} e^{\lambda z}\mathtt{f}(z; \Theta_t, x_0)\mathtt{d}z,$$

for any $\lambda \in \mathbb{R}$. However, according to our knowledge, there is not a general expression for the moment generating function associated to the measure $\mathtt{f}$. In fact, there is not a general expression for the moments associated to $\mathtt{f}$, *i.e.* $\int_{\mathbb{Y}} z^r \mathtt{f}(z; \Theta_t, x)\mathtt{d}z$ for any $x \in \mathbb{X}$ and $r \geq 2$. Nonetheless, it is known, from the discrete case, that there is a linear relation in the mean associated to $\mathtt{k}$, *i.e.*

$$\mathbb{E}[X_t|X_0 = x] = \tau_2(t)\kappa'(\vartheta) + x\Theta_t.$$

Hence, for the first moment we have to check if

$$\mathbb{E}\Big[X_{t+s}\big|X_0 = x\Big] = \mathbb{E}\Big[\mathbb{E}[X_{t+s}|X_s]\big|X_0 = x\Big]. \tag{3.15}$$

holds, for $s, t \geq 0$. Having this in mind, let us derive the right-hand side of the above equation as follows

$$\begin{aligned}\mathbb{E}\Big[\mathbb{E}[X_{t+s}|X_s]\,\big|X_0 = x\Big] &= \mathbb{E}\Big[\tau_2(t)\kappa'(\vartheta) + \Theta_t X_s\big|X_0 = x\Big] \\ &= \tau[1-\Theta_t]\kappa'(\vartheta) + \Theta_t\big\{\tau[1-\Theta_s]\kappa'(\vartheta) + x\Theta_s\big\} \\ &= \tau[1-\Theta_t\Theta_s]\kappa'(\vartheta) + x\Theta_t\Theta_s.\end{aligned}$$

Thus, the equation (3.15) holds if and only if $\Theta_{t+s} = \Theta_t\Theta_s$, whose solution is given by $\Theta_t = e^{-ct}$ for $c > 0$. It is worth to mention that, even if the equation (3.15) holds, it does not imply that $\mathtt{k}_t$ is a Markovian kernel. Nevertheless, it gives us an expression for the real-valued function Θ. Later, we will present a couple of Markovian models and, also, some non-Markovian.

Before that, we will derive an expression for the autocorrelation function of the process X, which follows by noticing that

$$\mathbb{E}[X_{t+s}X_t] = \Theta_t^s\mathbb{E}\big[\,X_t^2\big] + [1-\Theta_t^s]\mu^2,$$

where $\mathbb{E}[X_t] = \mu$, for $t \geq 0$, and $h > 0$. Hence, using the stationarity of the process $\{X_t\}_{t\geq 0}$ we have that,

$$\mathrm{Corr}(X_{t+s}, X_t) = \frac{\mathbb{E}[X_{t+s}X_t] - \mu^2}{\mathbb{E}[X_t^2] - \mu^2} = \Theta_t^s.$$

Clearly, this property holds whether the process is or is not Markovian.

On other hand, the calculation of the optimal and the prediction filters are given by the Theorem 2 by replacing the following measures,

$$\begin{aligned}\nu_0(\mathtt{d}x_{n+1}; u) &= c(x_{n+1} - u)\exp\{(x_{n+1} - u)\vartheta - \tau_2(t)\kappa(\vartheta)\}, \\ \phi_{y_{n+1}}(\nu_0(u))(\mathtt{d}x_{n+1}) &= \frac{\mathtt{f}(y_{n+1}; x_{n+1})\nu_0(x_{n+1}; u)}{\mathtt{m}_{\nu_0(u)}(y_{n+1})}.\end{aligned}$$

As we can see, the filters can not be reduced to simpler expressions. The same happens for statistics associated to the optimal filters, because it implies the computation of statistics associated to $\phi_{y_{n+1}}(\nu_0(u))$. However, for the prediction filter, one can obtain a general result for its moment generation function, which is given by

$$\mathbb{E}\left[e^{\gamma X_{t_{n+1}}}\middle|\mathscr{Y}^{(n)}\right] = \exp\left\{\tau_2(t)\left[\kappa(\vartheta+\lambda)-\kappa(\vartheta)\right]\right\}\int_{\mathbb{Y}} e^{\gamma u}\tilde{\omega}(y^{(n)},u)\mathtt{d}u.$$

Thus, using a thinning operation on a exponential dispersion distribution helps us to solve the problem of finding the appropriate choice of π and $\mathtt{f}$. Although, it remains to prove the Markovianity of the model.

3.4.2 Duality

As a result of the model, it is known that there exists a transition probability kernel, which is dual to $\mathtt{k}_t$, given by

$$\mathtt{p}_t(z_0,\mathtt{d}z) = \int_{\mathbb{X}} \mathtt{f}(z;\Theta_t,x)\nu_0(x;z_0,\Theta_t)\mathtt{d}x,$$

for $z_0, z \in \mathbb{Y}$. Clearly, as the measure $\mathtt{f}$ was defined as a contraction of π, the space $\mathbb{Y}$ turns out to be a subset of $\mathbb{X} \subset \mathbb{R}^d$, for $d \geq 1$. Also, as we did for $\mathtt{k}_t$, we can obtain a linear relation in the mean associated to the transition probabilities $\mathtt{p}_t$, *i.e.*

$$\mathbb{E}[Z_t|Z_0 = z] = \Theta_t[1-\Theta_t]\mu + \Theta_t z,$$

for $t \geq 0$. This allows us to check if the first moment associated to $\mathtt{p}_t$ satisfies Chapman-Kolmogorov's property. To this end, notice that

$$\begin{aligned}
\mathbb{E}\big[\,\mathbb{E}[Z_{t+s}|Z_s]\,\big|Z_0 = x\big] &= \mathbb{E}[\Theta_t\tau_2(t)\kappa'(\vartheta) + \Theta_t Z_s|Z_0 = z] \\
&= \tau\Theta_t[1-\Theta_t]\kappa'(\vartheta) + \Theta_t\{\tau\Theta_s[1-\Theta_s]\kappa'(\vartheta) + \Theta_s z\} \\
&= \tau\Theta_t[1-\Theta_t+\Theta_s-\Theta_s^2]\kappa'(\vartheta) + \Theta_t\Theta_s z.
\end{aligned}$$

Replacing the function Θ obtained from the first moment associated to $\mathtt{k}_t$, *i.e.* $\Theta_t = e^{ct}$ for $c > 0$, one can prove that the first moment associated to $\mathtt{p}_t$ does not satisfy Chapman-Kolmogorov's property.

On other hand, the autocorrelation function associated to the transition $\mathtt{p}_t$ has the same form as the discrete-time process, *i.e.*

$$\mathrm{Corr}(Z_{t+h}, Z_t) = \Theta_t^h,$$

for $h > 0$. This tells us that the dual inherits almost all properties that in the discrete case except the Markovianity. Nevertheless, the duality between the transitions $\mathtt{k}$ and $\mathtt{p}$ is still fulfilled, where the duality function takes the form

$$\mathtt{h}(x; z, \Theta_t) = \frac{c(x - z; \tau_1(t))}{c(x; \tau)} \exp\{\tau_1(t)\kappa(\vartheta) - z\vartheta\},$$

for $x \in \mathbb{X}$ and $z \in \mathbb{Y}$. Since there is not a general expression for the infinitesimal generator associated to $\mathtt{k}$, we are not able to derive an expression for the infinitesimal generator's dual. However, we present a couple of Markovian models, namely the Poisson-binomial model and the normal model. For these models, we present the corresponding filters and, also, the expression for their infinitesimal generator's dual.

3.4.3 Poisson-Binomial model

Let X be a re-parametrization of the Poisson-Binomial process, developed in Mena and Walker [30], which is a reversible death and birth process with invariant Poisson(λ) distribution, denoted by Po(λ). For this model, $\mathtt{f}(\cdot; \Theta_t, x)$ has Bin(x, Θ_t) distribution, for each $x \in \mathbb{Z}^+ \cup \{0\}$, where Bin$(n, p)$ stands for a binomial(n, p) distribution. Hence, the marginal distribution over $(\mathbb{Y}, \mathscr{Y})$ is Poisson$(\lambda\Theta_t)$ and the posterior distribution of π is

$$\nu_0(x; z, \Theta_t) = e^{-\lambda(1-\Theta_t)} \frac{[\lambda(1 - \Theta_t)]^{x-z}}{(x - z)!},$$

for $x \in \{z, ..., \infty\}$. Clearly, the probability measure ν_0 as a function of x does not belong to the same class of distribution than π, *i.e.* π is not conjugate with respect to $\mathtt{f}$. Although, as a function of $x - z$ the measure ν_0 has Po$(\lambda(1 - \Theta_t))$ distribution. Thus, the transition kernel associated to X takes the form

$$\mathtt{k}_t(x_0, x) = e^{-\lambda(1-\Theta_t)} \sum_{u=0}^{x_0 \wedge x} \frac{[\lambda(1 - \Theta_t)]^{x-u}}{(x - u)!} \binom{x_0}{u} \Theta_t^u (1 - \Theta_t)^{x_0 - u},$$

for x_0, $x \in \mathbb{Z}^+ \cup \{0\}$. Moreover, if $\Theta_t = e^{-ct}$, with $c > 0$, then the transition operator $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations (*cf.* Mena and Walker [30]). Additionally, one can prove that the X is birth and death process with infinitesimal generator $Q = \{q_{i,j}\}$, where

$$q_{ij} = \begin{cases} -c(i+\lambda) & , i = j, \\ c\lambda & , i = j+1, \\ cx & , i = j-1. \end{cases}$$

Note that, the process X is a $M/M/\infty$ queue with birth and death parameters $\lambda_n = c\lambda$ and $\mu_n = cn$ (*cf.* Schoutens [39]). On the other hand, the transition associated to the process Z is given by

$$\mathsf{p}_t(z_0, z) = e^{-\lambda(1-\Theta_t)} \sum_{v=z_0 \wedge z}^{\infty} \frac{\left[\lambda(1-\Theta_t)\right]^{v-z_0}}{(v-z_0)!} \binom{v}{z} \Theta_t^z (1-\Theta_t)^{v-z}$$

for z_0, $z \in \mathbb{Z}^+ \cup \{0\}$. As we can see, given that π is not conjugate, the expression of p_t is not tractable in comparison with conjugate cases.

Now, starting at ν_0 which has $\text{Po}(x_0 - y_0; (1-\Theta_\Delta)\lambda)$ distribution. The optimal and prediction filters take the form

$$\begin{aligned} \nu_{n+1}(\mathsf{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \omega\big(y^{(n+1)}, u\big) \frac{\text{Bin}(y_{n+1}; x_{n+1}, \Theta_\Delta)\text{Po}\big(x_{n+1} - u; \lambda_\Delta\big)}{\mathsf{m}_{\nu_0(u)}(y_{n+1})}, \\ \psi\big(\nu_n\big)(\mathsf{d}x_{n+1}) &= \sum_{u=0}^{x_{n+1}} \tilde{\omega}(y^{(n)}, u)\text{Po}\big(x_{n+1} - u; \lambda_\Delta\big), \end{aligned}$$

where $\lambda_\Delta := \lambda(1-\Theta_\Delta)$ and

$$\begin{aligned} \mathsf{m}_{\psi(\nu_n)}(\mathsf{d}y_{n+1}) &= \sum_{u=0}^{\infty} \tilde{\omega}(y^{(n)}, u)\mathsf{m}_{\nu_0(u)}(y_{n+1}), \\ \omega\big(y^{(n+1)}, u\big) &= \frac{\mathsf{m}_{\nu_0(u)}(y_{n+1})\tilde{\omega}\big(y^{(n)}, u\big)}{\mathsf{m}_{\psi(\nu_n)}(\mathsf{d}y_{n+1})}, \\ \tilde{\omega}(y^{(n)}, u) &= \sum_{x_n = u \vee u}^{\infty} \nu_n(dx_n)\text{Bin}(u; x_n, \theta_\Delta). \end{aligned}$$

for $n \geq 0$ and $\Delta > 0$. Note that, $\nu_0(u)(x) = \text{Po}(x - u; \lambda_\Delta)$. Also, since π is not conjugate, the optimal filters ν_{n+1} require the computation of the integral $\mathsf{m}_{\nu_0(u)}(y_{n+1})$. This implies more computational resources in every recursion. Nonetheless, as before, this model only requires to compute sums of positive terms which allows us to truncate such sums, without losing accuracy in the computation.

In this model, statistics associated to the optimal filters turn out to be mixture of statistics associated to,

$$\frac{\text{Bin}(y_{n+1}; x_{n+1}, \Theta_\Delta)\text{Po}\big(x_{n+1} - u; \lambda_\Delta\big)}{\mathtt{m}_{\nu_0(u)}(y_{n+1})}.$$

In addition, statistics associated to the predictor filters have simpler expressions, *e.g.* the moment generating function takes the form

$$\mathbb{E}\Big[e^{\gamma X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] = \exp\big\{\lambda_\Delta(e^\gamma - 1)\big\}\sum_{u=0}^{\infty} e^{\gamma u}\tilde{\omega}(y^{(n)}, u)\mathtt{d}u,$$

where ω and $\tilde{\omega}$ are given as before. That is to say, the above operators turn out to be the Laplace transform associated to the weights $\tilde{\omega}$.

On the other hand, Theorem 3 tells us that there exists a dual to X and also allows us to derive an expression for its infinitesimal generator. This can be done by noticing that

$$\frac{\partial}{\partial t}\log \mathtt{f}\,(z; \Theta_t, x) = \frac{z}{\Theta_t}\frac{\partial \Theta_t}{\partial t} - \Big(\frac{x-z}{1-\Theta_t}\Big)\frac{\partial \Theta_t}{\partial t}.$$

Thus, the functions of the Theorem 3 are given by $\mathtt{g}_1(z, \theta_t) = z/\Theta_t$ and $\mathtt{g}_2(x, z, \Theta_t) = (x - z)/(1 - \Theta_t)$ so

$$\begin{aligned}\sum_{u=0}^{x}\mathtt{h}(x_0; u, \Theta_t)\frac{x-u}{1-\Theta_t}\mathtt{f}\,(u; \Theta_t, x) &= \sum_{u=0}^{x-1}\mathtt{h}(x_0; u, \Theta_t)\frac{u+1}{\Theta_t}\mathtt{f}\,(u+1, \Theta_t, x)\\ &= \sum_{u=0}^{x}\mathtt{h}(x_0; u+1, \Theta_t)\frac{u}{\Theta_t}\mathtt{f}\,(u, \Theta_t, x),\end{aligned}$$

which means that $\tilde{\mathtt{g}}_2(u, \Theta_t) = u/\Theta_t$ and $\mathtt{g}(u) = u + 1$. Therefore, the infinitesimal generator of the $\mathtt{h}$-dual takes the form

$$\mathscr{G}\mathtt{h}(x; \cdot, \Theta_t)(z) = \frac{\partial \Theta_t}{\partial t}\frac{\partial}{\partial \Theta_t}\mathtt{h}(x; z, \Theta_t) + \frac{z}{\Theta_t}\frac{\partial \Theta_t}{\partial t}\mathtt{h}(x; z, \Theta_t) - \frac{z}{\Theta_t}\frac{\partial \Theta_t}{\partial t}\mathtt{h}(x; z+1, \Theta_t).$$

The above operator suggests that the dual $\tilde{Z}$ is a pure birth process subordinated by a deterministic one, driven by the following ordinary differential equation $\mathtt{d}\Theta_t =$

$-c\Theta_t \mathtt{d}t$.

Let us emphasize that, the above model was defined without any conjugacy assumption. Moreover, the dual of the above Poisson-binomial model, to the best of our knowledge, is new in the literature.

3.4.4 Ornstein-Uhlenbeck model

Let $X = \{X_t\}_{t\geq 0}$ be a continuous-time process with invariant $\mathrm{N}(\gamma, \alpha)$ distribution. For this model, the probability model $\mathtt{f}(\cdot; \Theta_t, x)$ defines a set of $\mathrm{N}(x\Theta_t, \alpha\Theta_t)$ distributions, where $\Theta : \mathbb{R}^+ \to (0,1)$. Thus, the probability measure ν_0 has normal$(z + \gamma(1-\Theta_t), (1-\Theta_t)\alpha)$ distribution, which implies that π is conjugate with respect to $\mathtt{f}$. In fact, this model turns out to be a conjugate model within the class of processes defined via the above thinning operation. Hence, the transition probabilities associated to X are given by

$$\mathtt{k}(x_0, x) = \frac{1}{\sqrt{2\pi\alpha(1-\Theta_t^2)}} \exp\left\{ - \frac{\left[x - x_0\Theta_t - \gamma(1-\Theta_t)\right]^2}{2\alpha(1-\Theta_t^2)} \right\},$$

for $x_0, x \in \mathbb{R}$, *i.e.* the process X has normal transition kernel. Moreover, the transition $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations if and only if $\Theta_t = e^{-ct}$ for $c > 0$. Additionally, one can prove that X is characterized as the only strong solution of the following stochastic differential equation,

$$\mathtt{d}X_t = -c(X_t - \gamma)\mathtt{d}t + \sqrt{2c\alpha}\mathtt{d}W_t$$

where $W = (W_t)_{t\geq 0}$ is a standard Brownian motion. See Mena and Walker [30]. This process constitutes a re-parameterization of the Ornstein-Uhlenbeck model.

For this model, it is straightforward to obtain the transition probability kernel of the process Z, that takes the form

$$\mathtt{p}_t(z_0, z) = \frac{1}{\sqrt{2\pi\alpha\Theta_t(1-\Theta_t^2)}} \exp\left\{ - \frac{\left[z - \Theta_t(z_0 - \gamma(1-\Theta_t))\right]^2}{2\alpha\Theta_t(1-\Theta_t^2)} \right\},$$

for z_0, $z \in \mathbb{Z}^+ \cup \{0\}$. Clearly, the above operator defines a normal kernel which does not satisfy Chapman-Kolmogorov's equations. However, it is used to prove the existence of a Markov dual.

As in the discrete case the filters are given by mixtures of normal distribution, in fact they turn out to be the same except for the additional time dependence. Indeed, starting at ν_0 and $\tilde{\omega}(y_0,u)$, which has $\mathrm{N}(x_0; y_0+\gamma(1-\Theta_\Delta), \alpha(1-\Theta_\Delta))$ and $\mathrm{N}(y_0; \Theta_\Delta x_0 + \Theta_\Delta\gamma(1-\Theta_\Delta), \alpha\Theta_\Delta(1-\Theta_\Delta^2))$ distributions, respectively. The optimal and the prediction filters are given by the following recursions

$$\begin{aligned}
\nu_{n+1}(\mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}} \omega(y^{(n+1)},u)\mathrm{N}\Big(x_{n+1}; \frac{u+y_{n+1}+\gamma(1-\Theta_\Delta)}{1+\Theta_\Delta}, \frac{\alpha(1-\Theta_\Delta)}{1+\Theta_\Delta}\Big)\mathtt{d}u,\\
\psi(\nu_n)(\mathtt{d}x_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)},u)\mathrm{N}(\mathtt{d}x_{n+1}; u+\gamma(1-\Theta_\Delta), \alpha(1-\Theta_\Delta))\mathtt{d}u,
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{m}_{\psi(\nu_n)}(\mathtt{d}y_{n+1}) &= \int_{\mathbb{Y}} \tilde{\omega}(y^{(n)},u)\mathrm{N}\Big(y_{n+1}; \Theta_\Delta\big[u-\gamma(1-\Theta_\Delta)\big], \alpha\Theta_\Delta\big[1-\Theta_\Delta^2\big]\Big)\mathtt{d}u,\\
\omega\big(y^{(n+1)},u\big) &= \frac{\tilde{\omega}\big(y^{(n)},u\big)\mathrm{N}\big(y_{n+1}; \Theta_\Delta u - \Theta_\Delta\gamma(1-\Theta_\Delta), \alpha\Theta_\Delta(1-\Theta_\Delta^2)\big)}{m_{\psi(\nu_n)}(dy_{n+1})},
\end{aligned}$$

and

$$\begin{aligned}
\tilde{\omega}(y^{(n)},u) &= \int_{\mathbb{Y}} \omega(y^{(n)},k)\tilde{\mathtt{p}}(y_n,u,k)dk,\\
\tilde{\mathtt{p}}(y_n,u,k) &= N\Big(u; \frac{\Theta_\Delta(k+y+\gamma(1-\Theta_\Delta))}{1+\Theta_\Delta}, \frac{\alpha(\Theta_\Delta-\Theta_\Delta^2+2\Theta_\Delta^3)}{1+\Theta_\Delta}\Big).
\end{aligned}$$

Additionally, the moment generating function of the optimal and the prediction filters are given by

$$\begin{aligned}
\mathbb{E}\Big[e^{\lambda X_{t_n}}\Big|\mathscr{Y}^{(n)}\Big] &= \exp\Big\{\frac{\lambda(y_n+\gamma(1-\Theta_\Delta))}{1+\Theta_\Delta} + \frac{1}{2}\frac{\lambda^2\tau_2}{1+\Theta_\Delta}\Big\}\int_{\mathbb{Y}} e^{\left(\frac{\lambda}{1+\Theta_\Delta}\right)u}\omega(y^{(n)},u)\mathtt{d}u,\\
\mathbb{E}\Big[e^{\lambda X_{t_{n+1}}}\Big|\mathscr{Y}^{(n)}\Big] &= \exp\Big\{\lambda\gamma(1-\Theta_\Delta) + \frac{1}{2}\lambda^2\alpha(1-\Theta_\Delta)\Big\}\int_{\mathbb{Y}} e^{\lambda u}\tilde{\omega}(y^{(n)},u)\mathtt{d}u.
\end{aligned}$$

for $n \geq 0$. Clearly, the conjugacy of π allows us to derive the above expressions.

On other hand, to derive an expression for the infinitesimal generator of the dual, consider the duality function

$$\mathtt{h}(x; z, \Theta_t) = \frac{1}{\sqrt{(1-\Theta_t)}}\exp\Big\{-\frac{(z-x\Theta_t)^2}{2\alpha\Theta_t(1-\Theta_t)} + \frac{(z-\gamma\Theta_t)^2}{2\alpha\Theta_t}\Big\}.$$

In order to derive an expression for the infinitesimal generator's dual, we apply $\mathscr{A}$

to $\mathtt{h}$ and re-write as a function of (z, Θ_t), *i.e.*

$$\mathscr{A}\mathtt{h}(x; z, \Theta_t) = -c(x-\gamma)\frac{\partial}{\partial x}\mathtt{h}(x; z, \Theta_t) + c\alpha\frac{\partial^2}{\partial x^2}\mathtt{h}(x; z, \Theta_t).$$

Clearly, it is necessary to calculate the following derivatives

$$\begin{aligned}\frac{\partial}{\partial x}\mathtt{h}(x; z, \Theta_t) &= \Big(\frac{z - x\Theta_t}{\alpha(1-\Theta_t)}\Big)\mathtt{h}(x; z, \Theta_t),\\ \frac{\partial^2}{\partial x^2}\mathtt{h}(x; z, \Theta_t) &= \Big[\Big(\frac{z - x\Theta_t}{\alpha(1-\Theta_t)}\Big)^2 - \frac{\Theta_t}{\alpha(1-\Theta_t)}\Big]\mathtt{h}(x; z, \Theta_t).\end{aligned}$$

Then, after some algebra, one can obtain the following expression for the generator infinitesimal of the dual, denoted by $\mathscr{G}$, applied to $\mathtt{h}$,

$$\mathscr{G}\mathtt{h}(x; z, \Theta_t) = \big(\frac{z}{\Theta_t} - \gamma\big)\frac{\partial\Theta_t}{\partial t}\frac{\partial}{\partial z}\mathtt{h}(x; z, \Theta_t) - \alpha\frac{\partial\Theta_t}{\partial t}\frac{\partial}{\partial\Theta_t}\mathtt{h}(x; z, \Theta_t),$$

where

$$\begin{aligned}\frac{\partial}{\partial z}\mathtt{h}(x; z, \Theta_t) &= \Big(\frac{x - z - \gamma(1-\Theta_t)}{\alpha(1-\Theta_t)}\Big)\mathtt{h}(x; z, \Theta_t),\\ \frac{\partial^2}{\partial z^2}\mathtt{h}(x; z, \Theta_t) &= \Big[\Big(\frac{x - z - \gamma(1-\Theta_t)}{\alpha(1-\Theta_t)}\Big)^2 - \frac{1}{\alpha(1-\Theta_t)}\Big]\mathtt{h}(x; z, \Theta_t),\end{aligned}$$

and $(\partial\Theta_t/\partial t) = -ce^{-ct}$ for any $c > 0$. Therefore, the expression of $\mathscr{G}\mathtt{h}$ suggest that the dual can be associated to a diffusion with drift and diffusion coefficients given by $-c(z - \gamma\Theta_t)$ and $2c\alpha\Theta_t$, respectively. Hence, as the drift and diffusion coefficients are Θ_t dependent, it is not straightforward to obtain an expression for the transition kernel associated to $\mathscr{G}$. Even more, this implies that the dual is subordinate by Θ_t as the Theorem 2 indicates it.

3.4.5 Some non-Markovian examples

The class of exponential dispersion models has six natural exponential families, namely the normal, the Poisson, the gamma, the binomial, the negative-binomial and the hyperbolic secant distributions. See Morris [31]. For these families, we already known that for the normal and the Poisson cases, there exist a real-valued function Θ such that the transition probability kernel $\mathtt{k}_t$, defined via the thinning operation developed in this section, satisfies Chapman-Kolmogorov's property. On other hand, the gamma, the binomial and the negative-binomial cases, we found that Chapman-Kolmogorov's property is not fulfilled. In order to illustrate such statement, below, we present these examples.

Gamma case

Consider a probability measure π with gamma(τ, b) distribution. Letting $\tau_1(t) = \Theta_t \tau$ and $\tau_2(t) = (1-\Theta_t)\tau$ with $\Theta : \mathbb{R}_+ \to (0,1)$, the contraction associated to π takes the form

$$\mathtt{f}(z;\Theta_t,x) = \frac{\Gamma(\tau)}{\Gamma(\tau_1(t))\Gamma(\tau_2(t))}\left(1-\frac{z}{x}\right)^{\tau_2(t)-1}\left(\frac{z}{x}\right)^{\tau_1(t)-1}\frac{1}{x}, \qquad \text{for } z \in (0,x).$$

Hence, the probability measure $\mathtt{f}$ as a function of z/x has beta$(\tau_1(t), \tau_2(t))$ distribution and, the probability measure ν_0 is given by

$$\nu_0(x;\Theta_t,z) = \frac{b^{\tau_2(t)}}{\Gamma(\tau_2(t))}(x-z)^{\tau_2(t)-1}\exp\{-b(x-z)\}$$

for $x \in (z,\infty)$. Thus, the measure ν_0 as a function of $x-z$ has gamma$(\tau_2(t), b)$. For this model, the continuous-time reversible process $X = \{X_t\}_{t\geq 0}$ is driven by the transition probability function,

$$\mathtt{k}_t(x_0,x) = \frac{\Gamma(\tau)b^{\tau_2(t)}e^{-bx}}{\Gamma(\tau_1(t))\Gamma(\tau_2(t))\Gamma(\tau_2(t))x_0^{\tau(t)-1}}\int_0^{x_0\wedge x}\big[(x-z)(x_0-z)\big]^{\tau_2(t)-1}z^{\tau_1(t)-1}e^{bz}\mathtt{d}z,$$

for $x_0, x \in \mathbb{R}^+$; and has invariant measure π. Moreover, the first moment associated to $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's property if and only if $\Theta_t = e^{-ct}$, with $c > 0$. However, this does not imply that X is a Markov process.

In fact, using the second moment of the process, one can prove that the operator $\mathtt{k}_t$ is not a Markovian kernel. Indeed, denoting by $\mathbb{E}^x$ the conditional expected valued of $\{X_t|X_0 = x\}$, we have that

$$\mathbb{E}^x[X_{t+s}^2] = \frac{\tau}{b^2}(1-\Theta_{t+s})\big[\tau(1-\Theta_{t+s})+1\big] + \frac{2\tau}{b}(1-\Theta_{t+s})\Theta_{t+s}x + \frac{\Theta_{t+s}(1+\tau\Theta_{t+s})}{\tau+1}x^2,$$

for $t, s > 0$. In this case, Chapman-Kolmogorov's equations are equivalent to prove that $\mathbb{E}^x[X_{t+s}^2] = \mathbb{E}^x[\mathbb{E}[X_{t+s}^2|X_s]]$. Hence, the right-hand side of the last equality is equal to

$$\begin{aligned}
\mathbb{E}^x[\mathbb{E}[X_{t+s}^2|X_s]] &= \frac{\tau(1-\Theta_t)[\tau(1-\Theta_t)+1]}{b^2} + 2\left(\frac{\tau}{b}\right)^2 \Theta_t(1-\Theta_t)(1-\Theta_s) \\
&+ \frac{\Theta_t+\tau\Theta_t^2}{\tau+1}\frac{\tau}{b^2}(1-\Theta_s)[\tau(1-\Theta_s)+1] \\
&+ 2\frac{\tau}{b}\Theta_t\Theta_s\left[(1-\Theta_t)+\frac{(1+\tau\Theta_t)}{\tau+1}(1-\Theta_s)\right]x \\
&+ \frac{\Theta_t(1+\tau\Theta_t)}{\tau+1}\frac{\Theta_s(1+\tau\Theta_s)}{\tau+1}x^2.
\end{aligned}$$

If we focus on the coefficient of x^2 in both sides of Chapman-Kolmogorov's equations one can see that

$$\Theta_{t+s}\left(\frac{1+\tau\Theta_{t+s}}{\tau+1}\right) = \Theta_t\Theta_s\left(\frac{1+\tau\Theta_t}{\tau+1}\right)\left(\frac{1+\tau\Theta_s}{\tau+1}\right).$$

Then, replacing the equality obtained for the first moment associated to $\mathtt{k}_t$, *i.e.* $\Theta_{t+s} = \Theta_t\Theta_s$, in the above equality we obtain that

$$(\Theta_t - 1)(\Theta_s - 1) = 0 \qquad \Rightarrow \qquad \Theta_t = 1, \qquad \forall t \geq 0.$$

Therefore, it does not exists a function $\Theta : \mathbb{R}^+ \to (0,1)$ such that the process X is a Markov process. A similar result is obtained for the binomial and binomial-negative case.

Binomial case

Consider a probability measure π with $\text{Bin}(\tau, p)$ distribution, $\tau_1(t) = \Theta_t\tau$ and $\tau_2(t) = (1-\Theta_t)\tau$; where $\Theta : \mathbb{R}_+ \to (0,1)$. Hence, the contraction associated to π takes the form

$$\mathtt{f}(z;\Theta_t,x) = \frac{\binom{\tau_2}{x-z}\binom{\tau_1}{z}}{\binom{\tau}{x}}, \qquad \text{for } z \in \{\max(0, x+\tau_1-\tau), ..., \min(x,\tau_1)\}.$$

The probability measure $\mathtt{f}$ has hypergeometric$(n, n\rho_t, x)$ distribution and, the probability measure ν_0 is given by

$$\nu_0(x;\Theta_t,z) = \binom{\tau_2(t)}{x-z}p^{x-z}(1-p)^{\tau_2(t)-(x-z)},$$

for $x \in \{z, z+1, ..., \tau_2(t)\}$. Thus, the measure ν_0 as a function of $x - z$ has $\text{Bin}(\tau_2(t), p)$. For this model, the continuous-time reversible process $X = \{X_t\}_{t \geq 0}$ is driven by the transition probability kernel,

$$\mathtt{k}_t(x_0, x) = \frac{(1-p)^{\tau_2(t)}}{\binom{\tau}{x_0}} \sum_{z=\max\{0, x_0 - \tau_2(t)\}}^{\min\{x_0, \tau_1(t), \tau_2(t)\}} \binom{\tau_2(t)}{x-z} \binom{\tau_2(t)}{x_0 - z} \binom{\tau_1(t)}{z} \left(\frac{p}{1-p}\right)^{x_0 - z},$$

for $x_0 \in \{0, 1, ..., \tau\}$ and $x \in \{\min\{\tau_1(t), \tau_2(t)\}, ..., \max\{\tau_1(t), \tau_2(t)\}\}$; and has invariant measure π. Moreover, the first moment associated to $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations if and only if $\Theta_t = e^{-ct}$, with $c > 0$.

Then, as we did for the gamma case, we will use the second moment of the process to prove that X is not a Markov process. In fact, one can obtain that

$$\begin{aligned}\mathbb{E}^x[X_{t+s}^2] &= \tau p(1-p)(1-\Theta_{t+s}) + \tau^2 p^2 (1-\Theta_{t+s})^2 + \tau\Theta_{t+s}(1-\Theta_{t+s})\left(2p + \frac{1}{n-1}\right)x \\ &\quad + \Theta_{t+s}\left(\frac{\tau\Theta_{t+s} - 1}{\tau - 1}\right)x^2\end{aligned}$$

for $t, s > 0$. Hence,

$$\begin{aligned}\mathbb{E}^x\Big[\mathbb{E}\big[X_{t+s}^2 | X_s\big]\Big] &= \tau p(1-p)(1-\Theta_t) + \tau^2 p^2(1-\Theta_t)^2 + \tau^2 p\theta_t(1-\Theta_t)(1-\Theta_s)\left(2p + \frac{1}{\tau-1}\right) \\ &\quad + \Theta_t\left(\frac{\tau\Theta_t - 1}{\tau - 1}\right)\left[\tau p(1-p)(1-\Theta_s) + \tau^2 p^2(1-\Theta_s)^2\right] \\ &\quad + \tau\Theta_t\Theta_s\left(2p + \frac{1}{\tau-1}\right)\left[(1-\Theta_t) + (1-\Theta_s)\left(\frac{\tau\Theta_t - 1}{\tau - 1}\right)\right]x \\ &\quad + \Theta_t\Theta_s\left(\frac{\tau\Theta_t - 1}{\tau - 1}\right)\left(\frac{\tau\Theta_s - 1}{\tau - 1}\right)x^2.\end{aligned}$$

If we focus on the coefficient of x^2 for both sides of Chapman-Kolmogorov's equations one can see that

$$\Theta_{t+s}\left(\frac{\tau\Theta_{t+s} - 1}{\tau - 1}\right) = \Theta_t\Theta_s\left(\frac{\tau\Theta_t - 1}{\tau - 1}\right)\left(\frac{\tau\Theta_s - 1}{\tau - 1}\right).$$

Then, replacing $\Theta_{t+s} = \Theta_t\Theta_s$, in the above equality we obtain that

$$(\Theta_t - 1)(\Theta_s - 1) = 0 \qquad \text{if and only if} \qquad \Theta_t = 1 \qquad \forall t \geq 0.$$

Clearly, this result is very similar to the gamma and the binomial-negative cases.

Negative binomial case

Let π be a probability measure with negative-binomial(τ, p) distribution, $\tau_1(t) = \Theta_t \tau$ and $\tau_2(t) = (1 - \Theta_t)\tau$; where $\Theta : \mathbb{R}_+ \to (0,1)$. In this case, we have that the contraction of π is given by

$$\mathtt{f}(z; \Theta_t, x) = \binom{x}{z} \frac{\mathsf{B}(\tau_1(t) + z, \tau_2(t) + x - z)}{\mathsf{B}(\tau_1(t), \tau_2(t))}, \qquad \text{for } z \in \{0, 1, ..., x\}.$$

Hence, as before, the measure ν_0 as a function of $x - z$ has negative-binomial$(\tau_2(t), p)$. For this model, the continuous-time reversible process $X = \{X_t\}_{t \geq 0}$ is driven by the transition probability kernel,

$$\mathtt{k}_t(x_0, x) = \frac{x_0! p^x (1-p)^{\tau_2(t)}}{\Gamma(\tau_2(t)) \mathsf{B}(\tau_1(t), \tau_2(t))} \sum_{z=0}^{x_0 \wedge x} \frac{\Gamma(x - z + \tau_2(t)) \mathsf{B}(\tau_1(t) + z, \tau_2(t) + x_0 - z)}{(x-z)!(x_0 - z)! z!} p^{-z},$$

for $x_0 \in \mathbb{N} \cup \{0\}$ and $x \geq x_0$; and has invariant measure π. Moreover, the first moment associated to $\mathtt{k}_t$ satisfies Chapman-Kolmogorov's equations if and only if $\Theta_t = e^{-ct}$, with $c > 0$.

Then, as the gamma case, we will use the second moment of the process to prove that X is not a Markov process. In fact, one can obtain that

$$\begin{aligned}
\mathbb{E}^x\left[X_{t+s}^2\right] &= \frac{\tau p(1 - \Theta_{t+s})}{(1-p)^2} + \left(\frac{\tau p(1 - \Theta_{t+s})}{1-p}\right)^2 + \tau \Theta_{t+s}(1 - \Theta_{t+s}) \left(\frac{2p}{1-p} + \frac{1}{\tau + 1}\right) x \\
&\quad + \frac{\Theta_{t+s} + \tau \Theta_{t+s}^2}{\tau + 1} x^2,
\end{aligned}$$

for $t, s > 0$. Thus,

$$\begin{aligned}
\mathbb{E}^x\Big[\mathbb{E}\big[X_{t+s}^2 \big| X_s\big]\Big] &= \frac{\tau p(1 - \Theta_t)}{(1-p)^2} + \frac{\tau^2 p}{1-p} \Theta_t (1 - \Theta_t)(1 - \Theta_s) \left(\frac{2p}{1-p} + \frac{1}{\tau + 1}\right) \\
&\quad + \left(\frac{\tau p(1 - \Theta_t)}{1-p}\right)^2 + \frac{\Theta_t + \tau \Theta_t^2}{\tau + 1} \left[\frac{\tau p(1 - \Theta_s)}{(1-p)^2} + \left(\frac{\tau p(1 - \Theta_s)}{1-p}\right)^2\right] \\
&\quad + \frac{\Theta_t + \tau \Theta_t^2}{\tau + 1} \left[\tau \Theta_s (1 - \Theta_s) \left(\frac{2p}{1-p} + \frac{1}{\tau + 1}\right)\right] x \\
&\quad + \tau \Theta_t \Theta_s (1 - \Theta_t) \left(\frac{2p}{1-p} \frac{1}{\tau + 1}\right) x + \frac{\Theta_t + \tau \Theta_t^2}{\tau + 1} \left(\frac{\Theta_s + \tau \Theta_s^2}{\tau + 1}\right) x^2.
\end{aligned}$$

If we focus on the coefficient of x^2 in both sides of Chapman-Kolmogorov's equations one can see that

$$\frac{\Theta_{t+s} + \tau\Theta_{t+s}^2}{\tau + 1} = \frac{\Theta_t + \tau\Theta_t^2}{\tau + 1}\frac{\Theta_s + \tau\Theta_s^2}{\tau + 1}.$$

Then, replacing $\Theta_{t+s} = \Theta_t\Theta_s$, in the above equality we obtain that

$$(\Theta_t - 1)(\Theta_s - 1) = 0 \qquad \text{if and only if} \qquad \Theta_t = 1 \qquad \forall t \geq 0.$$

Therefore, it does not exists a function $\Theta : \mathbb{R}^+ \to (0, 1)$ such that the process X is a Markov process.

3.5 Continuous-time Lancaster probabilities

Similar to the previous section, we use Lancaster probabilities to deal with the problem of specifying the measures that characterizes our construction. For this purpose, let us remember the definition of Lancaster probabilities. Let $\sigma(\mathtt{d}x, \mathtt{d}z)$ be a bivariate distribution with margins π and m_π defined over the measurable spaces $(\mathbb{X}, \mathscr{X})$ and $(\mathbb{Y}, \mathscr{Y})$, respectively. Also, let $\{P_n\}_{n\geq 0}$ and $\{Q_n\}_{n\geq 0}$ be orthonormal polynomials associated to π and m_π, respectively. Hence, if $\mathbb{E}[P_i(X)Q_j(Z)] = 0$ for $i \neq j$, then σ is known as Lancaster probability. Moreover, we are interested on bivariate probabilities of the form,

$$\sigma(\mathtt{d}x, \mathtt{d}z) = \sum_{j\geq 0} \rho_j P_n(x) Q_n(z) \pi(\mathtt{d}x) \mathsf{m}_\pi(\mathtt{d}z).$$

where $\rho_n = \mathbb{E}[P_n(x)Q_n(z)]$ characterizes the above bivariate distribution, known as the Lancaster sequence of σ. Thus, Lancaster probabilities allows us to consider the measure $\mathtt{f}(z; x) = \sigma(\mathtt{d}x, \mathtt{d}z)/\pi(\mathtt{d}x)$. Then, as we did for the discrete case, we define a reversible Markov process driven by the transition probability function

$$\mathtt{k}(x_n, \mathtt{d}x_{n+1}) = \sum_{j\geq 0} \rho_n^2 P_n(x_n) P_n(x_{n+1}) \pi(\mathtt{d}x_{n+1}), \tag{3.16}$$

for $x_n, x_{n+1} \in \mathbb{X}$. Hence, in order to preserve the stationarity of X in the continuous case, the time dependence should be aggregated to the Lancaster sequence. In fact, we notice that there is a connection with the spectral expansion of the transition probability functions associated to Markov processes.

3.5.1 Spectral expansion of the transition density

The introduction of this chapter argued that if we model any random phenomenon via some stochastic equation, then we do not always have an expression for the transition density that describes the dynamics of it. The spectral expansion of the transition probabilities has been a useful tool for such operational complications. In this section, we focus on two classes of Markov processes, namely birth and death processes and diffusion processes.

Birth and death processes

In the analysis of birth and death processes, a prominent role is played by a sequence of polynomials $\{P_n\}_{n\geq 0}$, called birth and death polynomials. They are determined uniquely by the recurrence relation

$$-xP_n(x) = \mu_n P_{n-1}(x) - (\lambda_n + \mu_n)P_n(x) + \lambda_n P_{n+1}(x), \qquad n \geq 0, \tag{3.17}$$

together with $P_{-1} = 0$ and $P_0(x) = 1$. The space state for birth and death processes is given by $S = \{-1, 0, 1, 2, ...\}$. In fact, these processes are characterized by their birth and death rates λ_i, μ_i, for $i \in S$, where -1 is an absorbing state. Ignoring this state if $\mu_0 = 0$ and 0 becomes in this case a reflecting state. Karlin and McGregor [18] proved the the transition function can be represented as

$$\mathtt{p}_{i,j}(t) = \mathbb{P}[X_t = j | X_0 = i] = \pi_j \int_0^\infty e^{-xt} P_i(x) P_j(x) d\phi(x), \tag{3.18}$$

for $i, j = 0, 1, 2, ...$, and $t > 0$, where ϕ is a positive Borel measure with total mass 1 and with support on the nonnegative real axis; ϕ is called the spectral measure of $\mathtt{p}$. Taking $t = 0$ in (3.18) one easily sees that the polynomials $\{P_n\}_{n\geq 0}$ are orthogonal with respect to ϕ. Note that, the dependence on t in the right-hand side of (3.18) is restricted entirely to the monotone decreasing exponential term e^{-tx}. Also, the dependence on i and j in (3.18) is factored as the product $P_i(x)P_j(x)$.

Moreover, if $\mu_0 = 0$, we have $P_i(0) = 1$ for $i \geq 0$, and thus the limiting stationary distribution, when it exists, is equal to

$$\mathtt{p}_j = \frac{\pi_j}{\sum_{k=0}^\infty \pi_k} = \lim_{t\to\infty} \mathtt{p}_{i,j}(t) = \pi_j P_i(0) P_j(0) \phi(\{0\}) = \pi_j \phi(\{0\}).$$

The symmetry relation $\pi_i \mathtt{p}_{i,j}(t) = \pi_j \mathtt{p}_{j,i}(t)$ also follows directly from (3.18).

Diffusion processes

Consider an arbitrary diffusion $\{X_t\}_{t\geq 0}$ with drift coefficient $\mu(x)$ and diffusion coefficient $\sigma^2(x)$. Hence, its infinitesimal generator, denoted by $\mathscr{A}$, takes the form

$$\mathscr{A}g(x) \quad = \mu(x)\tfrac{\mathtt{d}}{\mathtt{d}x}g(x) + \tfrac{1}{2}\sigma^2(x)\tfrac{\mathtt{d}^2}{\mathtt{d}x^2}g(x),$$

for $x \in \mathbb{X}$ and g in the domain of $\mathscr{A}$. Moreover, if the diffusion $\{X_t\}_{t\geq 0}$ is stationary, then the invariant measure π of the process, also known as speed measure, is obtained by solving the differential equation,

$$\frac{\mathtt{d}}{\mathtt{d}x}\left[\frac{1}{2}\sigma^2(x)\pi(x)\right] = \mu(x)\pi(x),$$

whose solution is given by,

$$\pi(x) = \frac{2c}{\sigma^2(x)} \exp\left\{\int_{x_0}^{x} \frac{2\mu(v)}{\sigma^2(v)}\mathtt{d}v\right\},$$

where c is a positive constant chosen so that π is a probability measure, and x_0 is an arbitrarily chosen state.

Denote by $\mathscr{L}^2(\mathbb{X}, \pi)$ the space of real-valued functions on $\mathbb{X}$ that are square integrable with respect to the density π. Let f and g be functions in the domain of $\mathscr{A}$. Then, upon integration by parts, one obtains the following property

$$\langle \mathscr{A}f, g\rangle_\pi = \langle f, \mathscr{A}g\rangle_\pi,$$

where the inner product $\langle \cdot, \cdot\rangle_\pi$ is defined by

$$\langle f, g\rangle_\pi = \int_{\mathbb{X}} f(x)g(x)\pi(x)\mathtt{d}x.$$

Consider the case in which $\mathbb{X}$ is a closed and bounded interval. Thus, if ω is an eigenfunction of $\mathscr{A}$ corresponding to an eigenvalue γ, *i.e.* $\mathscr{A}\omega = \gamma\omega$, then $u(t, x) = e^{\gamma t}\omega(x)$ solves the backward equation

$$\frac{\partial}{\partial t}u(t, x) = e^{\gamma t}\gamma\omega(x) = e^{\gamma t}\mathscr{A}\omega = \mathscr{A}u(t, x).$$

Assuming that the set of eigenvalues, counting multiplicities, is countable, say $\gamma_0, \gamma_1, \gamma_2, ...$ with corresponding eigenfunctions $\omega_0, \omega_1, \omega_2, ...$ of unit length, that is $\langle \omega_n, \omega_n \rangle_\pi^{1/2} = 1$. It is known that, eigenfunctions corresponding to distinct eigenvalues are orthogonal with respect to the inner product $\langle \cdot, \cdot \rangle_\pi$. Also if there is more than one linearly independent eigenfunction for a single eigenvalue, then these eigenfunctions can be orthogonalized by the Gram-Schmidt procedure. So $\omega_0, \omega_1, \omega_2, ...$ can be taken orthonormal.

Equivalently, the same collection of polynomials, which generates an orthogonal basis of $\mathscr{L}_2(\mathbb{X}, \pi)$, can be obtained via the following sequence,

$$\omega_n(x) \quad = \tfrac{1}{\pi(x)} \tfrac{\mathtt{d}^n}{\mathtt{d}x^n} \left\{ \pi(x) \left[\tfrac{1}{2} \sigma^2(x) \right]^n \right\}.$$

In particular, if $\mu(x)$ and $\sigma^2(x)$ (but not necessarily $\sigma(x)$) are polynomials, then $\omega_n(x)$ is a polynomial of degree n. In which case, it follows

$$\|\omega_n\|_\pi^2 = (-1)^n \left(\frac{\mathtt{d}^n}{\mathtt{d}x^n} \omega_n(x) \right) \int_{\mathbb{X}} \left[\frac{1}{2} \sigma^2(x) \right]^n \pi(\mathtt{d}x).$$

Note that, the n-th derivative of $\omega_n(x)$ is a constant, being an n-degree polynomial, and equals the leading coefficient of ω_n times $n!$. Moreover, since $\|\omega_n\|_\pi^2$ and $\int_{\mathbb{X}} \left[\frac{1}{2} \sigma^2(x) \right]^n \pi(\mathtt{d}x)$ are positive, the sign of the leading coefficient of ω_n is $(-1)^n$. Hereafter, we consider each ω_n normalized, *i.e.* write ω_n instead of $\omega_n / \|\omega_n\|_\pi$. Thus, the collection of orthonormal polynomials $\{\omega_n\}_{n \geq 0}$ turns out to be eigenfunctions of the infinitesimal generator $\mathscr{A}$, *i.e.*

$$\mathscr{A} \omega_n = -\gamma_n \omega_n$$

for some non-negative constant γ_n.

Furthermore, if the set of finite linear combinations of eigenfunctions is complete in $\mathscr{L}^2(\mathbb{X}, \pi)$, then each $f \in \mathscr{L}^2(\mathbb{X}, \pi)$ has a Fourier expansion of the form

$$f = \sum_{n=0}^{\infty} \langle f, \omega_n \rangle_\pi \omega_n.$$

Hence, considering the linear combination defined by

$$u_f(t, x) = \sum_{n=0}^{\infty} e^{\gamma_n t} \langle f, \omega_n \rangle_\pi \omega_n(x),$$

one obtains that $u_f(t,x)$ satisfies the backward equation with initial condition $u_f(0,x) = f(x)$. However, the semigroup associated to $\{X_t\}_{t\geq 0}$, denoted by $\mathtt{k}_t$, also satisfies the backward equation and the same initial condition. So, if there is uniqueness for a sufficiently large class of initial functions f, then we get

$$\int_{\mathbb{X}} f(x)\mathtt{k}_t(x_0,x)dx = \int_{\mathbb{X}} f(x)\sum_{n=0}^{\infty} e^{\gamma_n t}\omega_n(x_0)\omega_n(x)\pi(x)\mathtt{d}x.$$

Therefore, under the above assumptions, the transition probability function admits a density given by

$$\mathtt{p}_t(x_0,dx) = \sum_{n=0}^{\infty} e^{-\gamma_n t}\omega_n(x_0)\omega_n(x)\pi(\mathtt{d}x). \tag{3.19}$$

It is worth noticing that, the above expression of the transition density is the same that we obtain for the process built based on Lancaster probabilities with $\rho_n = e^{\gamma_n t}$.

3.5.2 Duality

The transition expansions (3.18) and (3.19) suggest that birth an death processes and diffusions can be constructed via Lancaster probabilities, by considering Lancaster sequences given by $\rho_n(t) = e^{-\gamma_n t}$, for $n \geq 0$. That is to say,

$$\sigma_t(\mathtt{d}x,\mathtt{d}z) = \sum_{n\geq 0} e^{-\gamma_n t} P_n(\mathtt{d}x)Q_n(\mathtt{d}y)\pi(\mathtt{d}x)\mathsf{m}_\pi(\mathtt{d}z),$$

for $x \in \mathbb{X}$ and $z \in \mathbb{Y}$. Clearly, the measure σ has margins π and m_π, and they are characterized by the orthonormal polynomials $\{P_n\}_{n\geq 0}$ and $\{Q_n\}_{n\geq 0}$, respectively.

Thus, we can define a time-homogeneous reversible Markov process $\{X_t\}_{t\geq 0}$ with state space $(\mathbb{X},\mathscr{X})$, invariant distribution π and driven by the transition probability function

$$\mathtt{k}_t(x_0,\mathtt{d}x) = \sum_{j\geq 0} \rho_j^2(t)P_j(x_0)P_j(x)\pi(\mathtt{d}x)$$

As for the discrete case, the expression of $\mathtt{k}_t$ implies that P_j is an eigenfunction for the eigenvalue ρ_j^2 of the operator $Tf(x_0) = \int_{\mathbb{X}} f(x)\mathtt{k}_t(x_0,x)\mathtt{d}x$, for any measurable function f. On other hand, let $Z = \{Z_t\}_{t\geq 1}$ be stochastic process over the space

$(\mathbb{Y}, \mathscr{Y})$ driven by the transition probability kernel,

$$\mathtt{k}_t(z_0, \mathtt{d}z) = \sum_{j\geq 0} \rho_j^2(t) Q_j(z_0) Q_j(z) \mathsf{m}_\pi(\mathtt{d}z; \Theta_t),$$

for $z_0, z \in \mathbb{Y}$. Then, letting us define the duality function $\mathtt{h}$ as the Radon-Nikodyn derivative between the posterior and prior distribution of π, *i.e.*

$$\mathtt{h}(x; z, \Theta_t) = \sum_{j\geq 0} \rho_j(t) P_j(x) Q_j(z),$$

for $x \in \mathbb{X}$ and $z \in \mathbb{Y}$. The processes $\{X_t\}_{t\geq 0}$ and $\{Z_t\}_{t\geq 0}$ are dual with respect to the duality function $\mathtt{h}$. This implies that,

$$\mathbb{E}^x[\mathtt{h}(X_t; z, \Theta_t)] = \sum_{j\geq 0} \rho_j^2(t) Q_j(y) P_j(x) = \mathbb{E}^z[\mathtt{h}(x; Z_t, \Theta_t)],$$

and the predictor operator of $\{X_t\}_{t\geq 0}$ takes the form,

$$\int_{\mathbb{X}} \nu_0(x; z, \Theta_t) \mathtt{k}_t(x_0, x) \mathtt{d}x = \sum_{n\geq 0} \rho_n^3(t) P_n(x_0) Q_n(z) \pi(\mathtt{d}x_0).$$

Let us emphasize that, the use of orthonormal polynomials do not simplify the computation of the filters or some other properties associated to the transition kernel $\mathtt{k}_t$. Nonetheless, we use Lancaster probabilities as a mechanism to specify the model that characterizes our construction. For illustrative purposes, we will present the case where the Lancaster probability has gamma and negative-binomial margins.

3.5.3 Gamma-binomial-negative model

Let π by the probability measure with gamma$(a, 1)$ distribution, for $a > 0$. The orthonormal polynomials with respect to π, denote by $\{P_n\}_{n\geq 0}$, are given by $P_n = \sqrt{n!/(a)_n} L_n^a$, with $(a)_n = a(a+1)\cdots(a+n-1)$, where $\{L_n^a\}_{n\geq 0}$ are the Laguerre polynomials. Also, let m_π be a negative-binomial(r, p) distribution, with $n > 0$ and $p \in (0, 1)$. It is known that, the orthonormal polynomials with respect to m_π are the normalized Meixner polynomials, denoted by $\{Q_n\}_{n\geq 0}$. Then, consider the Lancaster probability σ with marginals π and m_π, and Lancaster sequences $\{\rho_n\}_{n\geq 0}$, given by

$$\sigma(\mathtt{d}x, \mathtt{d}z) = \sum_{n=0}^{\infty} \rho_n P_n(x) Q_n(y) \pi(\mathtt{d}x) \mathsf{m}_\pi(\mathtt{d}z).$$

An extreme Lancaster sequence for the above bivariate distribution is determined by $\rho_n = (\sqrt{p})^n$. Hence, considering the function $p : \mathbb{R}^+ \to (0,1)$ such that

$$\rho_n(t) = e^{-\gamma_n t} = \left[\sqrt{p(t)}\right]^n \qquad \Rightarrow \qquad p(t) = e^{-ct}, \qquad \text{for } c > 0.$$

This implies that, $\gamma_n = cn/2$. Moreover, letting $\phi_t = p(t)/(1-p(t)) = (e^{ct}-1)^{-1}$, with $c > 0$, and $r = a$, then the Lancaster probability takes the form,

$$\sigma(\mathtt{d}x, \mathtt{d}z) = e^{x\phi_t} \frac{(x\phi_t)^z}{z!} \pi(\mathtt{d}x),$$

for $(x, z) \in (0, \infty) \times \mathbb{N}$. It is straightforward to see that the transition probability kernel defined from σ coincides with the kernel of the gamma-Poisson process, with $b = 1$. Furthermore, it is known that the orthogonal polynomials $\{P_n\}_{n \geq 0}$ satisfy the following identity,

$$\begin{aligned} \mathtt{k}_t(x_0, \mathtt{d}x) &= \sum_{n=0}^{\infty} e^{-\gamma_n t} P_n(x_0) P_n(x) \pi(x) \mathtt{d}x \\ &= \frac{\exp\{-[\,\phi_t(x_0 + x) + x\,]\}}{(\phi_t + 1)^{-(a+1)/2} \phi_t^{(a-1)/2}} \left(\frac{x}{x_0}\right)^{\frac{a-1}{2}} \mathrm{I}_{a-1}\left(2\sqrt{x_0 x \phi_t (1 + \phi_t)}\right), \end{aligned}$$

for x_0, $x \in \mathbb{R}^+$, and where $\mathrm{I}_\nu(\cdot)$ denotes the modified Bessel function of the first kind with argument ν. Therefore, this model exemplifies the use of Lancaster probabilities in order to specify the measures that characterize our construction.

3.5.4 Wright-Fisher diffusion

The Wright-Fisher diffusion with parent mutation has been widely used to model genetic evolution, such diffusion is characterized as the only solution to the following stochastic differential equation,

$$\mathtt{d}X_t = \frac{1}{2}\left[a_1(1 - X_t) + a_2 X_t\right]\mathtt{d}t + \sqrt{X_t(1 - X_t)}\mathtt{d}W_t$$

where W is a standard Brownian motion and $a_1, a_2 > 0$. Moreover, the above equation determines a stationary process with invariant measures

$$\pi(\mathrm{d}x) = \frac{x^{a_1-1}(1-x)^{a_2-1}}{B(a_1,a_2)}.$$

In this case, the orthonormal polynomials associated to the beta distribution are known as the Jacobi polynomials, which are obtained via the recursion (3.19). In fact, such polynomials satisfy

$$P_m(x) = c_m(a_1)_m \sum_{\ell=0}^{m} \binom{m}{\ell} \frac{(m+a-1)_\ell}{(a_1)_\ell}(-x)^\ell,$$

where

$$c_m = \frac{1}{\|P_m\|_\pi} = \sqrt{\frac{a+2m-1}{a+m-1}\frac{(a)_m}{(a_1)_m(a_2)_m}\frac{1}{m!}},$$

with $a = a_1 + a_2$. Thus, the transition probability function associated to the process $\{X_t\}_{t\geq 0}$ has an expression in terms of the Jacobi polynomials of the form,

$$\mathtt{k}_t(x_0, \mathtt{d}x) = \left[\sum_j e^{-\gamma_n t} P_n(x_0) P_n(x)\right] \pi(\mathtt{d}x),$$

for $x_0, x \in (0,1)$ and $t \geq 0$, with $\gamma_m = \frac{1}{2}m(m+a-1)$. Furthermore, Ethier and Griffiths [9] derived an equivalent expression for the transition of the process,

$$\mathtt{k}_t(x_0, \mathtt{d}x) = \sum_{m=0}^{\infty} \mathtt{q}_m^a(t) \sum_{k=0}^{m} \mathrm{Beta}(x; a_1+k, a_2+m-k)\mathrm{Bin}(k; m, x). \tag{3.20}$$

where

$$\mathtt{q}_m^a(t) = \sum_{j=m}^{\infty} e^{-\gamma_m t}(-1)^{j-m} \frac{(2j+a-1)(m+a)_{(j-1)}}{m!(j-m)!}.$$

In fact, the $\{q_m^a(t)\}$ are the transition probability kernel of a death process with an entrance boundary of infinity, and death rates γ_m. The death process represents the number of non-mutant ancestral linages back in times in the coalescent process with mutation. The number of linages decreases from m to $m-1$ from coalescent rate $\binom{k}{2}$ or mutation at rate $ma/2$. If there is no mutation, $\{\mathtt{q}_m^a(t)\}$ are transition functions of number of edges in a Kingman coalescent tree. The expansion (3.20) is derived from a two-dimensional dual death process which looks back in time in the Wright-Fisher diffusion.

On other hand, one can obtain that the measure m_π takes the form

$$\mathsf{m}_\pi(\mathbf{m}) = \mathsf{q}^{|\mathbf{a}|}_{|\mathbf{m}|}(t)\binom{|\mathbf{m}|}{m_1}\frac{\mathrm{B}(\mathbf{a}+\mathbf{m})}{\mathrm{B}(\mathbf{a})}$$

Furthermore, the transition operator of the $\mathtt{h}$-dual, which is a non-homogeneous process, is given by

$$\mathsf{p}_t(\mathbf{m},\mathbf{n}) = \mathsf{q}^{|\mathbf{a}|}_{|\mathbf{n}|}(t)\binom{|\mathbf{n}|}{n_1}\frac{\mathrm{B}(\mathbf{a}+\mathbf{m}+\mathbf{n})}{\mathrm{B}(\mathbf{a}+\mathbf{m})}$$

for $\mathbf{m} = (m_1, m_2)$ and $\mathtt{n} = (n_1, n_2)$, where the binomial coefficient and the fraction represent a beta-binomial$(|\mathbf{n}|, \mathbf{a} + \mathbf{m})$ distribution. It is worth to mention that, the orthogonal polynomials associated to the beta-binomial distribution are given by the Hahn polynomials. In this case, the construction suggest that the Hahn and Jacobi polynomials satisfy a bi-orthonormal property. The proof of the last statement remains to be made. It represents one of the authors' current work.

Chapter 4

A dual construction of measure-valued reversible Markov processes

This chapter uses the conditional probability inherent on exchangeable observations to build the transition probabilities that drive a class of measure-valued Markov processes. Indeed, we use a similar structure, as the one used in the previous chapters, but without the parametric simplification. That is to say, in the full extend of exchangeable random variables that leads us to infinite dimensional spaces. For this purpose, we will borrow the construction of random probability measures from the Bayesian nonparametric literature. These measures, commonly referred as nonparametric priors distributions, will be invariant measures of the signal processes. In order to carry out our proposal, we require mathematical tractability of the posterior distributions associated to such prior distributions. Unfortunately, in general, such expressions are typically more involved. Here is where the duality of previous chapters will help us to derive new properties associated to such measure-valued processes, via the predictor operator of exchangeable sequences. Indeed, this later operator will play an important rol in the nonparametric version of our proposal.

Then, using the projective properties mentioned in Section 1.6, we will provide a mechanism to define a duality function that leads to the duality of Chapter 1 for the nonparametric case. In the same way, we will be able to calculate the optimal and the prediction filters associated to a hidden Markov model in which the signal is a nonparametric prior.

As before, the class of processes that we propose is divided into discrete-time and continuous-time cases. Hence, as in Chapter 3, the continuous model requires that the afore transition probabilities satisfy Chapman-Kolmogorov's property. Having this in mind, the Fleming-Viot measure-valued diffusion turns out to be a remarkable example that falls within our proposal. So, we will used the mechanism developed by Walker, Hatjispyros and Nicoleris [43] to build a class of continuous-time measure-valued Markov processes. Keeping the structure used throughout this thesis, we will provide an alternative approach to built continuous-time nonparametric models, via dependent random measures. In fact, these later models turn out to be continuous-time versions of stationary Feller-Harris processes.

4.1 Nonparametric priors

For the sake of completeness, this section presents a brief review about nonparametric priors, dragged from Lijoi and Prünster [26]. In particular, we focus on suitable transformations of completely random measures. Hence, let us begin by defining such measures. Denote by $\mathcal{M}_{\mathbb{X}}$ the space of boundedly finite measures on $(\mathbb{X}, \mathscr{X})$, this implies that for any μ in $\mathcal{M}_{\mathbb{X}}$ and any bounded set $A \in \mathscr{X}$ one has $\mu(A) < \infty$. Moreover, $\mathscr{M}_{\mathbb{X}}$ stands for the corresponding Borel σ-algebra on $\mathcal{M}_{\mathbb{X}}$.

Definition 6. *Let μ be a measurable mapping from $(\Omega, \mathscr{F}, \mathbb{P})$ into $(\mathcal{M}_{\mathbb{X}}\mathscr{M}_{\mathbb{X}})$ and such that for any $A_1, ..., A_n \in \mathscr{X}$, with $A_i \cap A_j = \emptyset$ for any $i \neq j$, the random variables $\mu(A_1), ..., \mu(A_n)$ are mutually independent. Then μ is termed completely random measure (CRM).*

An important property of CRMs is their almost sure discreteness, which means that their realization are discrete measures with probability one. Hence, they can always be represented as the sum of two components: a completely random measure $\sum_{i=1}^{\infty} J_i \delta_{X_i}$, where both the positive jumps J_i's and the $\mathbb{X}$-valued locations X_i's are random, and a measure with random masses at fixed locations. Here, without loss of generality, we will consider CRMs with no fixed points of discontinuity. Thus μ is characterized by the Lévy-Khintchine representation which states that

$$\mathbb{E}\left[e^{-\int_{\mathbb{X}} f(x)\mu(dx)}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{X}} \left[1 - e^{-sf(x)}\right] \nu(ds, dx)\right\},$$

where $f : \mathbb{X} \to \mathbb{R}$ is a measurable function such that $\int |f| d\mu < \infty$ almost surely. Assuming that ν is a measure on $\mathbb{R}^+ \times \mathbb{X}$ such that

$$\int_B \int_{\mathbb{R}^+} \min\{s, 1\} \nu(ds, dx) < \infty,$$

for any $b \in \mathscr{X}$. The measure ν is referred as the Lévy intensity of μ. It will be often useful to separate the jump and the location part of ν by writing it as

$$\nu(ds, dx) = \rho_x(ds)\alpha(dx), \tag{4.1}$$

where α is a measure on $(\mathbb{X}, \mathscr{X})$ and, ρ is a transition kernel on $\mathbb{X} \times \mathscr{B}(\mathbb{R}^+)$. If $\rho_x = \rho$ for any x, then the distribution of the jumps of μ is independent of their locations and both ν and μ are termed homogeneous.

4.1.1 Normalizing completely random measure

The idea of normalization was inspired by Ferguson's construction of the Dirichlet process and, it has been widely used to define a class of nonparametric priors. An important generalization of this approach is found in Lijoi and Prünster [26], which consists in normalizing completely random measures. Mathematically, such priors are defined as following.

Definition 7. *Let μ be a CRM on $\mathbb{X}$ such that $0 < \mu(\mathbb{X}) < \infty$ almost surely. Then, the random probability measure $p = \mu/\mu(\mathbb{X})$ is termed normalized random measure with independent increments (NRMI).*

Both, finiteness and positiveness of $\mu(\mathbb{X})$, are required for the normalization to be well-defined. Such conditions, in terms of the Lévy intensity of the CRM, are $\rho_x(\mathbb{R}^+) = \infty$ for every $x \in \mathbb{X}$ and $0 < \alpha(\mathbb{X}) < \infty$. Note that, one can build a NRMI by just providing a Lévy intensity. Thus NRMIs generate a very large class of nonparametric priors. Hereafter, we will say that the NMRI is homogeneous (non-homogeneous) if the underlying CRM (or Lévy intensity) is homogeneous (non-homogeneous).

Let us recall that, an homogeneous CRM whose Lévy intensity is given by $\nu(\mathtt{d}s, \mathtt{d}x) = [e^{-s}/s]\mathtt{d}s\alpha(\mathtt{d}x)$ is a gamma measure with parameter measure α on $\mathbb{X}$. Hence, if $\mathbb{X}$ is a Polish space then $p = \mu/\mu(\mathbb{X})$ has the same distribution than the

Dirichlet process on $\mathbb{X}$ with parameter α. Furthermore, the Dirichlet process is characterized within the above class of priors by its conjugacy property (*cf.* James, Lijoi and Prünster [14]). This property makes the Dirichlet process the most used within the class of nonparametric priors. As a matter of fact, such property will also facilitate the construction of transition probabilities that drive a measure-valued Markov process with the Dirichlet process as invariant measure.

Now, we will summarize the concept of exchangeable partition probability function that will help us present an expression for the predictor operator associated to discrete nonparametric priors.

4.1.2 Exchangeable partition probability function

The nature of realizations of discrete random probability measures leads to analyze the partition structure among the observations that they generate. Hence, consider a sequence of observations $\{Y_i\}_{i=1}^n$ sampled from a discrete random measure p. Also, define Ψ_n to be a random partition of the integers $\{1, ..., n\}$ such that any two integers i and j belong to the same set in Ψ_n if and only if $Y_i = Y_j$. Then, let $k \in \{1, ..., n\}$ and suppose $\{C_1, ..., C_k\}$ is a partition of $\{1, ..., n\}$ into k sets C_i. Thus, to an exchangeable sequence with a.s. discrete de Finetti's measure, correspond a distribution of Ψ_n. That is,

$$\mathbb{P}\big[\,\Psi_n = \{C_1, ..., C_k\}\,\big] = \Pi_k^{(n)}(n_1, ..., n_k). \tag{4.2}$$

This characterizes the probability of observing k distinct species with frequencies $(n_1, ..., n_k)$ in n draws from a population. The set $\{\Pi_k^{(n)} : 1 \leq k \leq n, n \geq 1\}$ is termed as exchangeable partition probability function (EPPF). Thus $\Pi_k^{(n)}$ is a symmetric function of its arguments and, hence, the following addition rule holds,

$$\Pi_k^{(n)}(n_1, ..., n_k) = \Pi_{k+1}^{(n+1)}(n_1, ..., n_k, 1) + \sum_{j=1}^{k} \Pi_k^{(n+1)}(n_1, ..., n_j + 1, ..., n_k).$$

Pitman [35] proved that every non-negative symmetric function satisfying this addition rule is the EPPF of some exchangeable sequence.

Furthermore, if $Y^{(n)} := (Y_1, ..., Y_n)$ contains k distinct values $Y_1^*, ..., Y_k^*$ and n_j

of them are equal to Y_j^* one has,

$$\mathbb{P}[Y_{n+1} = \text{"new"}|Y^{(n)}] = \frac{\Pi_{k+1}^{(n+1)}(n_1, ..., n_k, 1)}{\Pi_k^{(n)}(n_1, ..., n_k)}, \tag{4.3}$$

$$\mathbb{P}[Y_{n+1} = Y_j^*|Y^{(n)}] = \frac{\Pi_k^{(n+1)}(n_1, ..., n_j + 1, ..., n_k)}{\Pi_k^{(n)}(n_1, ..., n_k)}. \tag{4.4}$$

In particular, if p is a NRMI with non-atomic parameter measure α, the associated EPPF is given by

$$\Pi_k^{(n)}(n_1, ..., n_k) = \frac{1}{\Gamma(n)} \int_0^\infty u^{n-1} e^{-\psi(u)} \left\{ \prod_{j=1}^{k} \int_{\mathbb{Y}} \tau_{n_j}(u|y) \alpha(\mathtt{d}y) \right\} \mathtt{d}u,$$

where ψ is the Laplace exponent of a completely random measure μ, with $p = \mu/\mu(\mathbb{Y})$, and; $\tau_m(u|y) := \int_{\mathbb{R}^+} s^m e^{-us} \rho_y(\mathtt{d}s)$, for any $m \geq 1$. Hence, one can deduce the system of predictive distributions of Y_{n+1}, given $Y^{(n)}$, *i.e.*

$$\mathbb{P}\big[\, Y_{n+1} \in \mathtt{d}y_{n+1}|Y^{(n)} \,\big] = w_k^{(n)} Q(\mathtt{d}y_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_{j,k}^{(n)} \delta_{Y_j^*}(\mathtt{d}y_{n+1}),$$

where $Q = \alpha/\alpha(\mathbb{X})$ and, $\{w_k^{(and)}\}$ and $\{w_{j,k}^{(n)}\}$ are random weights depending on the corresponding Lévy measure, obtained from(4.3) and (4.4). A more extensive class of nonparametric priors, known as Poisson-Kingman models, posses a similar structure for their predictor operators than those corresponding to NRMIs.

4.1.3 Poisson-Kingman model and Gibbs-type priors

Denote by $J_{(1)} \geq J_{(2)} \geq \cdots$ the ranked jumps of the CRM, set $T = \sum_{i\geq 1} J_{(i)}$ and assume that the probability distribution of the total mass T is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$. Then, define the weights $w_{(i)} = J_{(i)}/T$ and denote by $S^* = \big\{(p_1, p_2, ...) : p_1 \geq p_2 \geq \cdots \geq 0, \sum_{i\geq 1} p_i = 1\big\}$ the set of all sequences of ordered non-negative real numbers that sum up to 1.

Definition 8. *Let $P_{\rho,t}$ be the conditional distribution of the sequence $\{w_{(i)}\}_{i\geq 1}$, given $T = t$, and; let η be a probability distribution on $\mathbb{R}^+$. The distribution*

$$\int_{\mathbb{R}^+} P_{\rho,t}\, \eta(dt) \tag{4.5}$$

on S^*, denoted by PK(ρ,η), is termed Poisson-Kingman distribution with Lévy intensity ρ and mixing distribution η.

The discrete random probability measure $P=\sum_{i\geq 1} w_{(i)}\delta_{X_i}$, where the $w_{(i)}$'s follow a PK(ρ,η) distribution, is termed PK(ρ,η) random probability measure. If η coincides with the probability distribution of T, denoted by PK(ρ), then the random probability measures are equivalent to homogeneous NRMIs.

An expression for the EPPF of a general PK(ρ,η) was obtained in Pitman [36] but it is difficult to evaluate. Having this in mind, Gnedin and Pitman [11] proposed a class of random probability measures within the PK(ρ,η) models, known as Gibb-type priors. Denote by $\Delta_{n,k} := \left\{(n_1,...,n_k) : n_i \geq 1, \sum_{i=1}^k n_i = n\right\}$.

Definition 9. *Let $P=\sum_{i\geq 1} w_i\delta_{X_i}$ be a discrete random probability measure, where the locations and the weights are independent and, $X_i \overset{i.i.d.}{\sim} Q$ with Q a non-atomic probability measure on $\mathbb{X}$. Then P is termed Gibbs-type random probability measure if, for all $1\leq k\leq n$ and for any $(n_1,...,n_k)\in\Delta_{n,k}$ its EPPF can be represented as*

$$\Pi_k^{(n)}(n_1,...,n_k) = V_{n,k}\prod_{j=1}^k (1-\sigma)_{n_j-1} \tag{4.6}$$

for some $\sigma\in[0,1)$. The random partition of $\mathbb{N}$ determined by (4.6) is termed Gibbs-type random partition.

As a consequence, the predictor operator associated to exchangeable sequences with de Finetti's measure within the class of Gybb-type priors takes the form

$$\mathbb{P}[X_{n+1}\in dx|X^{(n)}] = \frac{V_{n+1,k+1}}{V_{n,k}}Q(dx) + \frac{V_{n+1,k}}{V_{n,k}}\sum_{j=1}^k (n_j-\sigma)\delta_{X_j^*}(dx)$$

The prediction rule can be seen as resulting from a two step procedure: X_{n+1} is either "new" or "old" with probability depending on n and k but not on the frequencies n_i's. Given X_{n+1} is "new", it is sampled from Q. Given X_{n+1} is "old" it will coincide with a particular X_j^* with probability $(n_j-\sigma)/(n-k\sigma)$.

4.2 Discrete-time model

The aim of this section is to build a set of transition probabilities that drive measure-valued stationary Markov processes, where the invariant distributions are given by the priors of the previous section. It is worth to mention that, Walker, Hatjispyros and Nicoleris [43] developed a similar construction, within the nonparametric approach, for the case in which the invariant measure is given by the Dirichlet process. Hence, our proposal generalizes their methodology to more general nonparametric priors and also provides new properties associated to the model.

Denoting by $\mathscr{P}_{\mathbb{Y}}$ the set of all probability measures on the measurable space $(\mathbb{Y}, \mathscr{Y})$, where $\mathbb{Y}$ is a Polish space. Also, consider a random probability measure μ whose distribution, denoted by Π, belongs to the class of discrete nonparametric priors presented in the previous section. Thus, our construction will be based on the following joint distribution

$$\mathbb{P}\big[\,\mathtt{d}\mu, \mathtt{d}Y\,\big] = \mu(\mathtt{d}Y)\Pi(\mathtt{d}\mu).$$

Hence, denoting by $Q(\cdot) := \int_{\mathscr{P}_{\mathbb{Y}}} \mu(\cdot)\Pi(\mathtt{d}\mu)$, the conditional distribution of $\{\mu|Y\}$, denoted by $\Pi(\,\cdot\,|Y)$, is given by

$$\Pi(\mathtt{d}\mu|Y) = \frac{\mathbb{P}\big[\,\mathtt{d}\mu, \mathtt{d}Y\,\big]}{Q(Y)}.$$

Then, we define a reversible Markov process $\mathscr{P}_{\mathbb{Y}}$-valued driven by the transition probability function

$$\mathtt{k}(\mu, \mathtt{d}\nu) = \int_{\mathbb{Y}} \Pi(\mathtt{d}\nu|y)\mu(\mathtt{d}y), \tag{4.7}$$

and invariant measure Π. As in the parametric case, the reversibility follows from structure of the operator $\mathtt{k}$. Furthermore, we build a discrete-time reversible Markov process $\mathbb{Y}$-valued driven by the transition probability function

$$\mathtt{p}(y_n, \mathtt{d}y_{n+1}) = \int_{\mathscr{P}_{\mathbb{Y}}} \mu(\mathtt{d}y_{n+1})\Pi(\mathtt{d}\mu|y_n).$$

and invariant measure Q. Since the objective of this chapter generalizes the results from the parametric case, one can guess that the operators $\mathtt{k}$ and $\mathtt{p}$ are dual to each other with respect to a function. However, due to the complexity of deal with infinite dimensional spaces, this issue will be discussed later.

Similar to the above mechanism, we can define a set of transition probabilities by considering the joint distribution

$$\mathbb{P}\left[\, \mathtt{d}\mu, \mathtt{d}Y_1, ..., \mathtt{d}Y_n \,\right] = \prod_{j=1}^{n} \mu(\mathtt{d}Y_j)\Pi(\mathtt{d}\mu).$$

Thus, denoting by $\Pi(\,\cdot\,|Y^{(n)})$ the conditional distribution of μ, given $Y^{(n)} = (Y_1, ..., Y_n)$, we define a transition probability function on $\mathscr{P}_{\mathbb{Y}}$ as following

$$\mathtt{k}(\mu, \mathtt{d}\nu) = \int_{\mathbb{Y}^{(n)}} \Pi(\mathtt{d}\nu|y_1, ..., y_n) \prod_{j=1}^{n} \mu(\mathtt{d}y_j). \tag{4.8}$$

The above operator $\mathtt{k}$ describes the dynamics of a reversible Markov process μ and, in the same from as the operator (4.7), has Π as invariant measure. On the other hand, assuming that $\{Y_n\}_{n\geq 1}$ is an exchangeable sequence with de Finetti's measure Π, then the joint distribution of $(Y_1, ..., Y_n)$ has the following integral representation

$$\mathbb{P}\left[\, Y_1 \in B_1, ..., Y_n \in B_n \,\right] = \int_{\mathscr{P}_{\mathbb{Y}}} \prod_{i=1}^{n} \mu(B_i)\Pi(d\mu),$$

where $\mu \sim \Pi$ and $(B_1, ..., B_n) \in \mathscr{X}^{(n)}$, for $n \geq 1$. Thus the marginal distribution of every random variable Y is given by Q. Moreover, the predictor operator associated to such a sequence takes the form

$$\mathtt{p}(\mathtt{d}y^{(n)}, y_{n+1}) := \int_{\mathscr{P}_{\mathbb{Y}}} \mu(\mathtt{d}y_{n+1})\Pi(\mathtt{d}\mu|y_1, ..., y_n). \tag{4.9}$$

where $\mu, \nu \in \mathscr{P}_{\mathbb{Y}}$. If μ is defined via the normalization of a subordinator, then such a operator takes the form

$$\mathtt{p}(y^{(n)}, \mathtt{d}y_{n+1}) = w_k^{(n)} Q(\mathtt{d}y_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_{j,k}^{(n)} \delta_{y_j^*}(\mathtt{d}y_{n+1}),$$

where $(y_1^*, ..., y_k^*)$ are the distinct values within the sequence $(y_1, ..., y_n)$ and, $w_k^{(n)}$ and $w_{j,k}^{(n)}$ are random weights.

To illustrate the above construction, let us present the case where Π is given by the widely known distribution of a Dirichlet process. Thus, if Π is a Dirichlet process with parameter α, then $\Pi(\,\cdot\,|y)$ is also a Dirichlet process with updated parameter $\alpha + \delta_y$. Moreover, letting $\theta = \alpha(\mathbb{Y})$ the operator (4.7) in this case takes the form

$$\mathtt{p}(y_n, \mathtt{d}y_{n+1}) = \frac{\theta}{\theta+1} Q(\cdot) + \frac{1}{\theta+1}\delta_{y_n}.$$

Similarly, the predictor operator (4.9) is given by

$$\mathtt{p}(y_{n+1}|\mathtt{d}y^{(n)}) = \frac{\theta}{\theta+n} Q(\cdot) + \frac{n}{\theta+n}\sum_{j=1}^{n}\delta_{y_j}.$$

It is worth noticing that, one of the main differences between these two operators is that the former can be trivially associated to a Markov process and the later has not share such historical dependency.

4.3 Duality and filtering

Using the projective properties of Section 1.6 one can prove that our proposal leads to the duality of Section 1.4 between the models defined in the previous section. To be precise, the projective properties of Π and duality properties of these projections. That is to say, consider a measurable partition $A = (A_1, ..., A_k)$ of $\mathbb{Y}$ we have that $(\mu_t(A_1), ..., \mu_t(A_k))$ is distributed according to $\Pi^{\alpha(A)}$, where α is the parameter of Π and $\alpha(A) = (\alpha(A_1), ..., \alpha(A_k))$. This allows us to define the duality function $\mathtt{h} : \mathscr{P}_{\mathbb{Y}} \times \mathbb{Y}^{(n)} \to \mathbb{R}$ given by

$$\mathtt{h}(\mu; Y^{(n)}) = \frac{\Pi^{\alpha(A)}(\mu|Y^{(n)})}{\Pi^{\alpha(A)}(\mu)} = \frac{\prod_{j=1}^{n}\mu(Y_j)}{\mathbb{P}[Y^{(n)} \in \mathtt{d}y^{(n)}]}.$$

Then, the projective properties of Π guarantees the existence of a duality between the transition probabilities of a measure-valued process and the predictor operator of an exchangeable sequence of random variables, *i.e.*

$$\mathbb{E}^{\mu_0}[\mathtt{h}(\mu; y^{(n)})] = \mathbb{E}^{y^{(n)}}[\mathtt{h}(\mu_0; Y_{n+1})],$$

where $\mu_0 \in \mathscr{P}_{\mathbb{Y}}$ and $y^{(n)} \in Y \in \mathbb{Y}^{(n)}$. Clearly, the expected value of the left-hand side of the above equation is with respect to $\mathtt{k}$ and the one in the right-hand side is with respect to the predictor operator $\mathtt{p}$. The proof of the mentioned equality follows from the parametric results. An immediate result derived from the above duality is the calculation of the predictor operator associated to $\mathtt{k}$. In particular, if the random probability measure with distribution Π is a NMRI, then

$$\begin{aligned}\int_{\mathscr{P}_{\mathbb{Y}}} \Pi(\mu|y^{(n)})\mathtt{k}(\mu,\mathtt{d}\nu)\mathtt{d}\mu &= \int_{\mathbb{Y}} \Pi(\nu|y_{n+1})\mathtt{p}(y^{(n)},y_{n+1})\mathtt{d}y_{n+1} \\ &= \int_{\mathbb{Y}} \Pi(\nu|y_{n+1})\Big[w_k^{(n)}Q(y_{n+1}) + \frac{1}{n}\sum_{j=1}^{k} w_{j,k}^{(n)}\delta_{y_j^*}(y_{n+1})\Big]\mathtt{d}y_{n+1} \\ &= w_k^{(n)}\Pi(\mathtt{d}\nu) + \frac{1}{n}\sum_{j=1}^{k} w_{j,k}^{(n)}\Pi(\nu|y_j^*).\end{aligned}$$

The first equality is true because of the duality between $\mathtt{k}$ and $\mathtt{p}$. Clearly, the above result can be generalize to those measures belonging to the class of Gibb-type priors, this is true due to the structure of their predictor operator.

Let us emphasize that, one of issues in the nonparametric context is the availability of a simple expression for the posterior distribution of Π. As a consequence, the Dirichlet process, which is conjugate, becomes the first option within the class of discrete nonparametric priors. Hence, if Π turns out to be the distribution of the Dirichlet process we have that

$$\int_{\mathscr{P}_{\mathbb{Y}}} \Pi(\mu|y^{(n)})\mathtt{k}(\mu,\mathtt{d}\nu)\mathtt{d}\mu = \frac{\theta}{\theta+n}\Pi(\mathtt{d}\nu) + \frac{n}{\theta+n}\sum_{j=1}^{n}\Pi(\nu|y_j).$$

This result allows us to exemplify the advantages that the conditional probability structure used throughout this thesis gives, and, in particular, the duality obtained from it.

Furthermore, consider a hidden Markov model where the signal μ has distribution Π within the class of discrete nonparametric priors. In this case, the observation process Y consists of an exchangeable sequence and, the emission density $\mathtt{f}(\cdot;\mu) = \mu(\cdot)$ turns out to be random. Then, using the projective properties of Π used in this section allows us to obtain an expression for the optimal and the prediction filters associated to the afore signal via Theorem 1 in Chapter 2. So that we will have an expression for every measurable partition of $\mathbb{Y}$.

4.4 Continuous-time model

The final section of this thesis is focused on build a class of continuous-time measure-valued reversible Markov process. This will be done by adding a time dependence to the operators (4.7) and (4.8). In fact, we will focus on two different approaches. The first one is motivated by the construction of the Fleming-Viot diffusion, which takes the number of observations from an exchangeable sequence random. The other approach will be done via dependent random measures. As a result, we will obtain a continuous-time version of the stationary Feller-Harris model, presented in Anzarut and Mena [2].

4.4.1 Fleming-Viot measure-valued diffusion

Ethier and Griffiths [9] provided an expression for the transition function for a particular Fleming-Viot measure-valued diffusion process. Later Walker, Hatjispyros and Nicoleris [43] established a comprehensive construction of such process by extending ideas of the Gibbs sampler construction. Hence, let us present the idea of such construction. First, we recall that the invariant measure of the Fleming-Viot diffusion is given by the Dirichlet process. Earlier in this chapter we built a discrete-time transition probability function with the Dirichlet process, denoted by Π, as its invariant measure. Such operator was defined as following

$$\mathtt{k}(\mu, \mathtt{d}\nu) = \int_{\mathbb{Y}^{(n)}} \Pi(\mathtt{d}\nu | y^{(n)}) \prod_{j=1}^{n} \mu(\mathtt{d}y_i),$$

for $\mu, \nu \in \mathscr{P}_{\mathbb{Y}}$. In order to preserve the stationarity and reversibility of the process that drives the above operator, we need to aggregate a time dependence the expression $\prod_{j=1}^{n} \mu(\mathtt{d}y_i)$. This leaves us two options: let μ be a dependent random measure or, letting the sample size n be a function of the time. Here we deal with the later case. In fact, it is known that, the sample size for the Fleming-Viot diffusion is driven by a death process. To be precise, let $d_n(t) = \mathbb{P}[D_t = n]$, where D_t is a death process, $D_0 = \infty$ a.s., and with rate $\lambda_n = (1/2)n(n-1+\theta)$ for some $\theta > 0$. Denote by $d_n(t)$ the probability that the number of samples being used is n for the transition with time t. Thus, the transition function of the Fleming-Viot process is given by

$$\mathtt{k}_t(\mu, \mathtt{d}\nu) = \sum_{n=0}^{\infty} d_n(t) \int_{\mathbb{Y}^{(n)}} \Pi(\mathtt{d}\nu | y^{(n)}) \prod_{j=1}^{n} \mu(\mathtt{d}y_i) \tag{4.10}$$

For this model, Chapman-Kolmogorov's property was proved by Walker, Hatjispyros and Nicoleris [43] via the trajectorial properties associated to the above death process. On other hand, the h- dual associated to such diffusion was found in Papaspiliopoulos, Ruggiero and Spanó [34], this was accomplished based on the mentioned projective properties of the signal and duality properties of these projections. Moreover, conditional to the event $\{D_t = n\}$ the predictor operator of the exchangeable sequence coincides with the discrete-time version, *i.e.*

$$\mathbb{P}\left[Y_{n+1} \in B_{n+1} | Y^{(n)} \in \mathtt{d}y^{(n)}, D_t = n\right] = \int_{\mathscr{P}_{\mathbb{Y}}} \mu(\mathtt{d}y_{n+1}) \Pi\big(\mathtt{d}\mu | y^{(n)}\big)$$

This implies that, denoting by $\mathtt{p}_t$ the unconditioned predictor operator and replacing the discrete-time expression one obtains

$$\mathtt{p}_t(y^{(n)}, \mathtt{d}y_{n+1}) = \sum_{n=0}^{\infty} d_n(t) \Big[\frac{\theta}{\theta+n} Q(\mathtt{d}y_{n+1}) + \frac{n}{\theta+n} \sum_{j=1}^{n} \delta_{y_j}(\mathtt{d}y_{n+1})\Big]$$

At this point, it must be clear that, the duality between the operators $\mathtt{k}_t$ and $\mathtt{p}_t$ is preserved for every partition of the space $\mathbb{Y}$. This happens because the construction entirely depends on the structure of the model and not on the state space, which in the nonparametric case is the space of all probability measures. Hence, the duality function in this case is also given by

$$\mathtt{h}(\mu; Y^{(n)}) = \frac{\Pi^{\alpha(A)}(\mu | Y^{(n)})}{\Pi^{\alpha(A)}(\mu)} = \frac{\prod_{j=1}^{n} \mu(Y_j)}{\mathbb{P}[Y^{(n)} \in \mathtt{d}y^{(n)}]}.$$

where the measure α is the parameter associated to the distribution Π. Also, for this particular case if we associate the Fleming-Viot diffusion to the stochastic filtering, then it is possible, by considering the above projections, to calculate the filters associated to the model. Indeed, the conjugacy allows us to solve the mentioned issue.

Furthermore, if we replace the Dirichlet process by another random probability measure on the operator $\mathtt{k}_t$, then the construction allows us to build a measure-valued Markov process for any other random measure. Moreover, each of these processes have a dual that provides more information about the model. Nonetheless, besides the Fleming-Viot case, the tractability of the filters depends on the tractability of the corresponding posterior distribution or the availability of the predictor operator of a random probability measure.

4.4.2 Continuous-time stationary Feller-Harris process

An alternative to the above model, that serves to build the transition of a class of continuous-time measure valued Markov process, is found by considering dependent random measures on operators of the form (4.8). Such measures consist in random probability measures with an additional dependence structure. Roughly speaking, assuming that the covariate is given by $\mathbb{R}_+$, μ is a dependent random measure if μ_t is a random probability measure for every $t \geq 0$. Thus, letting μ_t a stationarity dependent random measure with distribution Π, one can consider the following transition probability function

$$\mathtt{k}(\mu_0, \mu_t) = \int_{\mathbb{Y}^{(n)}} \Pi(\mu_t | y^{(n)}) \prod_{j=1}^{n} \mu_0(\mathtt{d}y_j),$$

where $\mu_0, \mu_t \in \mathscr{P}_{\mathbb{Y}}$ and $n \geq 1$. Notice that, the stationarity of μ_t is in the sense that its marginal law does not depend on the time. On other hand, if Y is an exchangeable sequence with de Finetti's measure Π, then

$$Y_n \sim Q := \mathbb{E}^{\Pi}[\, \mu_t \,] = \int_{\mathscr{P}_{\mathbb{Y}}} \mu_t \, \Pi(\mathtt{d}\mu_t), \qquad \forall n \geq 1.$$

Additionally, if μ_t is almost-surely discrete then the predictor operator of the exchangeable sequence for two consecutive times has the following form

$$\mathbb{P}[Y_{n+1} \in A | Y_n] = (1 - \epsilon) Q(A) + \epsilon \delta_{Y_n}(A),$$

for all $n \geq 1$ and $\epsilon \in (0, 1)$. Note that, the stationarity of μ_t implies that Q does not depend on t, so the time dependence on the above operator is given through ϵ. In particular, considering $\epsilon = e^{-ct}$, for $c > 0$, one can define the following transition kernel

$$\mathtt{p}_t(y, B) = (1 - e^{-ct})Q(B) + e^{-ct}\delta_y(B), \qquad \forall t \geq 0 \text{ and } B \in \mathscr{Y}, \tag{4.11}$$

An immediate property associated to the above kernel is that it satisfies Chapman-Kolmogorov's property. The class of continuous-time processes with transition of the form (4.11) was introduced in Anzarut and Mena [2], where more properties related with them were presented. Then, again using some projective properties, we consider a measurable partition $A = (A_1, ..., A_k)$ of $\mathbb{Y}$ such that $(\mu_t(A_1), ..., \mu_t(A_k))$ is distributed according to $\Pi^{\alpha(A)}$, where α is the parameter of Π and $\alpha(A) = (\alpha(A_1), ..., \alpha(A_k))$, so

$$\mathtt{h}(\mu_t; Y) = \frac{\Pi_Y^{\alpha(A)}(\mu_t)}{\Pi^{\alpha(A)}(\mu_t)}.$$

Thus, the transition kernels $\mathtt{k}$ and $\mathtt{p}_t$ are dual to each other with respect to the duality function $\mathtt{h}$.

Furthermore, using the fact that the infinitesimal generator associated to $\mathtt{p}_t$, denoted by $\mathscr{G}$, takes the form

$$\mathscr{G}f(y) = c\left[\int_{\mathbb{Y}} f(u)Q(du) - f(y)\right],$$

one can obtain an expression for the infinitesimal generator of the dual applied to $\mathtt{h}$ and, then proceed as we did for the parametric case. Hence,

$$\begin{aligned}
\mathscr{G}\mathtt{h}(\mu_t;)(y) &= c\left[\int_{\mathbb{Y}} \mathtt{h}(\mu_t; u)Q(du) - \mathtt{h}(\mu_t; y)\right] \\
&= c\left[\int_{\mathbb{Y}} \mu_t(u)\mathtt{d}u - \frac{\mu_t(y)}{Q(y)}\right] \\
&= c\left[1 - \frac{\mu_t(y)}{Q(y)}\right] = \mathscr{A}\mathtt{h}(\mu_t; y),
\end{aligned}$$

where $\mathscr{A}$ is the infinitesimal generator of the $\mathtt{h}$-dual to Y. In this context, there are still many issues to solve, however we decide to present this model in order to illustrate future work.

Bibliography

[1] Al-Osh, M.A. and Aly, E.A.A. (1992). First-order autoregressive time series with negative binomial and geometric marginals. Commun. Statist, **A 21**, 2483-2492.

[2] Anzarut, M. and Mena, R. H. (2016). A Harris process to model stochastic volatility. arXiv:1605.05382.

[3] Cappé, O., Moulines, E. and Rydén, T. (2005). Inference in hidden Markov models. Springer, New York.

[4] Chaleyat-Maurel, M. and Genon-Catalot, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. Stochastic Process. Appl. **116**, 1447-1467.

[5] Cox, J. C., Ingersoll Jr, J. E. and Ross, S. A. (1985). A theory of the term structure of interest rates. Econometrica: Journal of the Econometric Society, 385-407.

[6] Diaconis, P., Khare, K. and Saloff-Coste, L. (2008) Gibbs sampling, exponential families and orthogonal polynomials. Statist Scien. **23** 151-178.

[7] Eagleson, G. K. (1964). Polynomial expansions of bivariate distributions. Ann. Math. Statist. **35**, 1208-1215.

[8] Ethier, S. and Kurtz, T. (1986) Markov processes: characterization and convergence. New York: John Wiley & Sons.

[9] Ethier, S. and Griffiths, R. C. (1993) The transition function of a Fleming-Viot process. The Annals of Probability. 1571-1590.

[10] Ewens, W. J. (2012) Mathematical population genetics 1: theoretical introduction (Vol. 27). Springer Science & Business Media.

[11] Gnedin, A. and Pitman, J. (2005). Exchangeable Gibbs partitions and Stirling triangles. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) **325**, 83–102.

[12] Griffiths, r. C. and Spanò, D. (2010). Diffusion processes and coalescent trees. In Probability and Mathematical Genetics. London Mathematical Society Lecture Note Series **378**, 358-379.

[13] Jacobs, P.A. and Lewis, P.A.W. (1977). A mixed autoregressive moving average exponential sequence and point process (EARMA 1,1). Adv. Appl. Prob. **9**, 87-104.

[14] James, L. F., Lijoi, A. and Prünster, I. (2006). Conjugacy as a distinctive feature of the Dirichlet process. Scand. J. Statist. **33**, 105-120.

[15] Jansen, S. and Kurt, N. (2012). On the notion(s) of duality for Markov processes. Probab. Surv. **11**, 59-120.

[16] Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. J. Appl. Probab. **33**, No.3, 664-677.

[17] Jorgensen, B. and Song, P. X. (1998). Stationary time series models with exponential dispersion model margins. J. Appl. Probab. **35**, No.1, 78-92.

[18] Karlin, S. and McGregor, J.L. (1957) The differential equations of birth and death processes, and the Stieltjes moment problem. Trans. Amer. Math. Soc. **85**, 489-546.

[19] Koudou, A. E. (1998). Lancaster bivariate probability distributions with Poisson, negative binomial and gamma margins. Test. **7**, 95-110.

[20] Lancaster, H. O. (1958). The structure of bivariate distributions. Ann. Math. Statist. **29**, No.3, 719-736.

[21] Lancaster, H. O. (1963). Correlations and canonical forms of bivariate distributions. Ann. Math. Statist. **34**, 532-538.

[22] Lancaster, H. O. (1975). Joint probability distributions in the Meixner class. J. Roy. Statist. Soc. **37**, No.3, 434-443.

[23] Lawrence, A. J. (1982). The innovation distribution of a gamma distributed autoregressive process. Scand. J. Statist. **9**, 234-236.

[24] Lawrence, A. J. and Lewis, P. A. W. (1977). An exponential moving average sequence and point process (EMA 1). J. Appl. Probab **14**, 98-113.

[25] Lewis, P. A. W., McKenzie, E. and Hugus, D. K. (1989). Gamma process. Commun. Statist.-Stoch. Models **5**, 1-30.

[26] Lijoi, A. and Prünster, I. (2009). Models beyond the Dirichlet processes. In Bayesian nonparametrics. 80-136. Cambridge University Press, Cambridge.

[27] McKenzie, E. (1986). Autoregressive moving average processes with negative-binomial and geometric marginal distribution. Adv. Appl. Probab. **18**, 679-705.

[28] McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. Adv. Appl. Probab. **20**, 822-835.

[29] Mena, R. and Walker, S. G. (2007). On stationary version of the generalized hyperbolic ARCH model. AISM, **59** 325-348.

[30] Mena, R. and Walker, S. G. (2009). On a construction of Markov models in continuous time. METRON Inter. J. Statist. **62**, 303-323.

[31] Morris, C. N. (1982). Natural exponential families with quadratic variance functions. Ann. Statist. **10** 65-80.

[32] Anzarut, M., Mena, R.H., Nava, C. R., and Prünster, I. (2016). Poisson driven stationary Markov models. Journal of Business & Economic Statistics.

[33] Papaspiliopoulos, O. and Ruggiero, M. (2014). Optimal filtering and the dual process. Bernoulli. **20**, No.4, 1999-2019.

[34] Papaspiliopoulos, O., Ruggiero, M. and Spanò, D. (2016). Filtering hidden Markov measures. arXiv:1411.4944.

[35] Pitman, J. (1995). Exchangeable and partially exchangeable random partitions. Probab. Theo. Rel. Fields. **102**, 145-158.

[36] Pitman, J. (2003). Poisson Kingman partitions. Technical report **625**.

[37] Pitt, M.K., Chatfield, C. and Walker, S.G. (2002). Constructing first order stationary autoregressive models via latent processes. Scand J. Statist. **29** 657-663.

[38] Pitt, M.K. and Walker, S.G. (2005). Constructing stationary time series models using auxiliary variables with applications. J. Amer. Statist. Assoc. **100** 554-564.

[39] Schoutens, W. (2000). Stochastic processes and orthogonal polynomials. Springer, New York.

[40] Shiryaev, A. N. (1999). Essentials of stochastic finance: facts, models, theory (Vol. 3). World scientific.

[41] Sturm, A. and Swart, J.M. (2015). Monotone and additive Markov process duality. arXiv preprint arXiv:1510.06284.

[42] Tyan, S., Derin, H. and Thomas, J. B. (1976). Two necessary conditions on the representation of bivariate distributions by polynomials. Ann. Statist. **4**, No.1, 216-222.

[43] Walker, S. G., Hatjispyros, S. J. and Nicoleris, T. (2007). A Fleming-viot process and Bayesian nonparametrics. The annals of applied probability. **17**, No.1, 67-80.