Optimal filtering and the dual process

Omiros Papaspiliopoulos

ICREA & Universitat Pompeu Fabra

Joint work with
Matteo Ruggiero
Collegio Carlo Alberto & University of Turin

Outline

The filtering problem

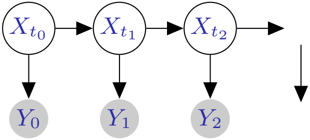
Computable filters

Linking filtering with duality

Examples

Extensions/Current work

The filtering problem



Hidden Markov model represented as a graphical model

Signal: latent Markov chain $X_{t_n} \in \mathcal{X}$, transition kernel $P_t(x, \mathrm{d}x')$ and initial distribution ν

Emission densities: Data $Y_n \in \mathcal{Y}$ with conditional density $f_x(y)$ (also consider non-dominated filtering models)

Goal: evaluate the filtering distributions $\nu_n(\mathrm{d}x) := \mathcal{L}(X_{t_n} \,|\, Y_0, \ldots, Y_n)$

Subcase: if $X_{t_0} = X_{t_1} = \cdots$, it's a classical Bayesian inference problem

Statistical applications

The filtering distributions $\nu_n(\mathrm{d}x) := \mathcal{L}(X_{t_n} \,|\, Y_0, \ldots, Y_n)$ are the backbone of all statistical estimation problems in this framework, such as:

- prediction of future signals, via $\mathcal{L}(X_{t_{n+k}} \,|\, Y_0, \ldots, Y_n)$
- derivation of smoothing distributions $\mathcal{L}(X_{t_{n-k}} \,|\, Y_0, \ldots, Y_n)$
- calculation of the marginal likelihood (in the dominated case)

$$L(y_1, \ldots, y_n) = \int_{\mathcal{X}} f_{x_{t_0}}(y_0)\mathrm{d}\nu(x_{t_0}) \prod_{i=1}^{n} \int_{\mathcal{X}} f_{x_{t_i}}(y_i)\mathrm{d}\mathcal{L}(x_{t_i} \,|\, y_0, \ldots, y_{i-1})$$

Filtering recursions

Mathematically, it's the solutions to the recursion

$$\nu_0 = \phi_{Y_0}(\nu)\,, \quad \nu_n = \phi_{Y_n}(\psi_{t_n - t_{n-1}}(\nu_n)), \quad n > 0,$$

where for probability measure ξ:

Update: $$\phi_y(\xi)(\mathrm{d}x) = \frac{f_x(y)\xi(\mathrm{d}x)}{p_\xi(y)}, \qquad p_\xi(y) = \int_{\mathcal{X}} f_x(y)\xi(\mathrm{d}x)$$

conjugate pair $(f_x, \mathcal{F}):\ \xi \in \mathcal{F} \implies \phi_y(\xi) \in \mathcal{F}$

Prediction: $$\psi_t(\xi)(\mathrm{d}x') = \int_{\mathcal{X}} \xi(\mathrm{d}x) P_t(x, \mathrm{d}x')$$

Note that *update* and *prediction* operators satisfy

$$\phi_y\left(\sum_{i=1}^n w_i \xi_i\right) = \sum_{i=1}^n \frac{w_i p_{\xi_i}(y)}{\sum_j w_j p_{\xi_j}(y)} \phi_y(\xi_i), \quad \psi_t\left(\sum_{i=1}^n w_i \xi_i\right) = \sum_{i=1}^n w_i\, \psi_t(\xi_i)$$

Finite-dimensional filters

A finite-dimensional filter is a solution to the previous recursion s.t. there exists a finite-dimensional family $\mathcal{F}_f$ of probability measures and

$$\nu \in \mathcal{F}_f \qquad \Rightarrow \qquad \phi_y(\nu), \psi_t(\nu) \in \mathcal{F}_f$$

so evolution of the mixtures can be described by successive transformations of a finite-dimensional parameter

Examples:

- "Nonparametric": any model where $\mathcal{X}$ is finite set (Baum filter)
- Parametric: linear Gaussian system (Kalman filter)

Discrete state space - Baum/Welch filter

The finite-dimensional filter is obtained by the definition of update/prediction operators & their action on mixtures (1):

$$\xi = \{\alpha_i, i \in \mathcal{M}\} \quad \Longrightarrow \quad \xi = \sum_{j \in \mathcal{M}} \alpha_j \delta_j$$

$$\psi(\delta_i) = \sum_{j \in \mathcal{M}} P_t(i,j)\delta_j \quad \Longrightarrow \quad \psi(\xi) = \sum_{i \in \mathcal{M}} \alpha_i \psi(\delta_i) = \sum_{j \in \mathcal{M}} \left(\sum_{i \in \mathcal{M}} \alpha_i P_t(i,j) \right) \delta_j$$

$$p_{\delta_i}(y) = f_i(y),\ \phi_y(\delta_i) = \delta_i \quad \Longrightarrow \quad \phi(\xi) = \sum_{j \in \mathcal{M}} \left(\frac{\alpha_j\, f_j(y)}{\sum_i \alpha_i\, f_i(y)} \right) \delta_j$$

Crucially, the complexity of these calculations are $\mathcal{O}(|\mathcal{M}|)$ for the update, but $\mathcal{O}(|\mathcal{M}|^2)$ for the prediction. Depending on the application, $|\mathcal{M}|$ might be constant or increase over time.

Other finite-dimensional filters?

- See for example, Runggaldier & Spizzichino (Bernoulli, 2001)

Motivating model: Cox-type process

Dynamic version of the standard conjugate Bayesian model for count data:

$$
\begin{aligned}
Y_n | X_{t_n} &\sim \text{Poisson}(X_{t_n}) \\
X_{t_n} & \quad \text{Markov chain with gamma marginal}
\end{aligned}
$$

For example

$$
\begin{aligned}
X_{t_n} &= U_n^2 \\
U_n &= a\, U_{n-1} + \beta\, \eta_n \quad \text{with stationary } N\left(0, \frac{\beta^2}{1-a^2}\right) \text{ when } |a| < 1
\end{aligned}
$$

More generally, we can relate the signal to the Feller/square root/CIR process/continuous-state branching process with immigration, e.g Kawazu & Watanabe (Theory Probab. Appl., 1971)

$$\mathrm{d}X_t \quad = \quad (\delta\sigma^2 - 2\gamma X_t)\mathrm{d}t + 2\sigma\sqrt{X_t}\mathrm{d}B_t\,, \quad \delta, \gamma, \sigma > 0$$

This process for $\delta \geq 2$ it has a stationary and limiting distribution, which is

$$\mathrm{gamma}(\delta/2, \gamma/\sigma^2)$$

For the economy of this talk, we focus on $\delta \geq 2$ (otherwise the boundary behaviour needs to be explicitly specified).

For this process, we have the following properties: the first is beautiful, the second incredible!

$$P_t(x, \mathrm{d}x') = \sum_{k\geq 0} \mathrm{Poisson}\left(k\,;\ \frac{\gamma}{\sigma^2}\frac{1}{e^{2\gamma t}-1}x\right) \mathrm{gamma}\left(k+\delta/2, \frac{\gamma}{\sigma^2}\frac{e^{2\gamma t}}{e^{2\gamma t}-1}\right).$$

$$\psi_t(\mathrm{gamma}(m+\delta/2,\theta)) =$$

$$\sum_{k=0}^{m} \mathrm{Bin}\left(k\,;\ m, \frac{\gamma}{\sigma^2}(\theta e^{2\gamma t}+\gamma/\sigma^2-\theta)^{-1}\right) \mathrm{gamma}\left(k+\delta/2, \frac{\gamma}{\sigma^2}\frac{\theta e^{2\gamma t}}{\theta e^{2\gamma t}+\gamma/\sigma^2-\theta}\right).$$

- When $\delta = 1$, these expressions can be obtained by elementary calculation, e.g. Genon-Catalot & Kessler (Bernoulli, 2004).
- The general case requires clever and lengthy calculations and is done in Chaleyat-Maurel & Genon-Catalot (SPA, 2006) - we return to this later.
- Note that the stationary density is obtained as invariant from the finite and as a limit from infinite mixture
- Note that we can obtain the infinite mixture from the finite by appropriate choice of θ and $m \to \infty$. In that limit $\text{gamma}(m + \delta/2, \theta) \to \delta_x$.
- Implications for filtering; recall that

$$f_y(x) \propto x^y e^{-x}$$

In this model, the filtering distributions evolve in the family of finite mixture of gamma distributions, but with a number of parameters that is increasing with n.

Computable filters

We will classify a filter as computable when ν_n is characterised in terms of a finite number of parameters that can be computed at a cost that grows polynomially with n.

Special case are finite-dimensional filters for which the number of parameters does not grow with n, hence the cost of computing them grows linearly with n.

The CIR/gamma model is an example of a computable filter.

A first round of questions

- Are there other models for which computable filters can be devised?
- Is there a methodology that applies to all such filters as well as to the Kalman and Baum/Welch?
- What is their computational cost?

Models

Computable filters are available for dynamic versions of all standard conjugate finite-dimensional Bayesian models

- K-dimensional linear diffusion (Gaussian-Gaussian), Kalman filter
- Feller/CIR/CBI process (gamma-Poisson)
- 1-dimensional Wright-Fisher diffusion (beta-binomial)
- K-dimensional Wright-Fisher diffusion (Dirichlet-multinomial)

Later in the talk we mention results for infinite-dimensional models

Methodology

Our work reveals that the answer to the earlier questions relates the dual process.

Duality methods have a long history in Probability, dating back to the work of P. Lévy - see e.g. Jansen & Kurt (arXiv, 2013) for a recent review.

Widely applied to the study of interacting particle systems and population genetic models.

To solve the filtering problem, we require a dual that in general is given by two components: a deterministic process (dynamical system) and a (multidimensional) death process with countable state-space.

This approach - among others - yields duals that do not seem to have appeared before.

Computable filtering consists of filtering out the death process in a way akin to the Baum/Welch filter.

The dual is identified by studying the generator of the signal.

The plan

Sufficient conditions for computability

Local duality (generators)

Filtering algorithm

In the next section of the talk we will see the conditions in action in the aforementioned models

Conditions

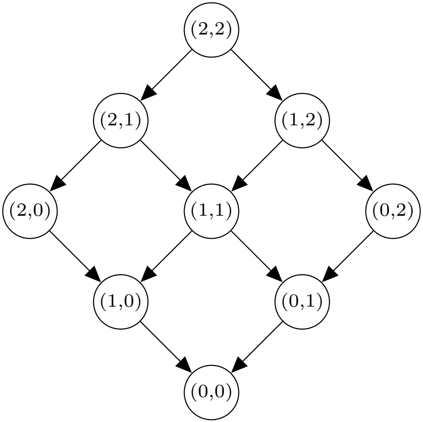
A1 (Reversibility): $\pi(\mathrm{d}x)P_t(x, \mathrm{d}x') = \pi(\mathrm{d}x')P_t(x', \mathrm{d}x)$.

Notation:

$$\mathcal{M} = \mathbb{Z}_+^K = \{\, \mathbf{m} = (m_1, \ldots, m_K) : m_j \in \mathbb{Z}_+,\ j = 1, \ldots, K \,\}.$$

$\mathbf{0}$; $\boldsymbol{e}_j$; $|\mathbf{m}| = \sum_i m_i$; product order on $\mathcal{M}$; $\mathbf{m} - \mathbf{i}$ for $\mathbf{i} \leq \mathbf{m}$

A 2D example for $\mathcal{M}$



A2 (Conjugacy): For $\Theta \subseteq \mathbb{R}^l,\ l \in \mathbb{Z}_+$, let $h : \mathcal{X} \times \mathcal{M} \times \Theta \to \mathbb{R}_+$ be such that $\sup_x h(x, \mathbf{m}, \theta) < \infty$ for all $\mathbf{m} \in \mathcal{M}, \theta \in \Theta$, and $h(x, \mathbf{0}, \tilde{\theta}) = 1$ for some $\tilde{\theta} \in \Theta$.

$$\mathcal{F} = \{h(x, \mathbf{m}, \theta)\pi(\mathrm{d}x),\ \mathbf{m} \in \mathcal{M}, \theta \in \Theta\}$$

assumed to be a family of probability measures s.t. there exist functions $t : \mathcal{Y} \times \mathcal{M} \to \mathcal{M}$ and $T : \mathcal{Y} \times \Theta \to \Theta$ with $\mathbf{m} \to t(y, \mathbf{m})$ increasing and such that

$$\phi_y(h(x, \mathbf{m}, \theta)\pi(\mathrm{d}x)) = h(x, t(y, \mathbf{m}), T(y, \theta))\pi(\mathrm{d}x)\,.$$

Note:

$$p_{h(x,\mathbf{m},\theta)\pi(\mathrm{d}x)}(y) =: c(\mathbf{m}, \theta, y) = \frac{f_x(y)h(x, \mathbf{m}, \theta)}{h(x, t(y, m), T(y, \theta))}$$

which does not depend on x.

A3 (Duality): Let $r : \Theta \to \Theta$, $\lambda : \mathbb{Z}_+ \to \mathbb{R}_+$ be increasing, $\rho : \Theta \to \mathbb{R}_+$ be continuous, and consider a Markov process (M_t, Θ_t) with state-space $\mathcal{M} \times \Theta$ such that:

$$\mathrm{d}\Theta_t/\mathrm{d}t = r(\Theta_t), \quad \Theta_0 = \theta_0 \,,$$

when at $(M_t, \Theta_t) = (\mathbf{m}, \theta)$, the process jumps down to state $(\mathbf{m} - \boldsymbol{e}_j, \theta)$ with instantaneous rate

$$\lambda(|\mathbf{m}|)\rho(\theta)m_j \,.$$

and it is dual to X_t wrt functions h, i.e., $\forall x \in \mathcal{X}, \mathbf{m} \in \mathcal{M}, \theta \in \Theta, t \geq 0$:

$$\mathbb{E}^x[h(X_t, \mathbf{m}, \theta)] = \mathbb{E}^{(\mathbf{m},\theta)}[h(x, M_t, \Theta_t)]$$

When $K = 0$ or $l = 0$ the dual is just Θ_t or M_t

Remarks

Absorption/Ergodicity

- M_t can only jump to "smaller" states and $\mathbf{0}$ is absorbing

Duality functions

- Radon-Nikodym derivatives between measures that are conjugate to the emission density

Transition probabilities

$$p_{\mathbf{m},\mathbf{n}}(t;\theta) = \mathbb{P}[M_t = \mathbf{n} | M_0 = \mathbf{m}, \Theta_0 = \theta] , \quad \mathbf{n}, \mathbf{m} \in \mathcal{M}, \ \mathbf{n} \leq \mathbf{m} .$$

- analytic expressions exist (Proposition 2.1)

Duality and optimal filtering pt 1: propagation

Proposition

Under A1-A2-A3

$$\psi_t(h(x,\mathbf{m},\theta)\pi(\mathrm{d}x)) = \sum_{\mathbf{0}\leq\mathbf{i}\leq\mathbf{m}} p_{\mathbf{m},\mathbf{m}-\mathbf{i}}(t;\theta)h(x,\mathbf{m}-\mathbf{i},\Theta_t)\pi(\mathrm{d}x),$$

Proof.

$$\begin{aligned}\psi_t(h(x,\mathbf{m},\theta)\pi(\mathrm{d}x)) &= \int_{\mathcal{X}} h(x,\mathbf{m},\theta)\pi(\mathrm{d}x)P_t(x,\mathrm{d}x') = \int_{\mathcal{X}} h(x,\mathbf{m},\theta)\pi(\mathrm{d}x')P_t(x',\mathrm{d}x) \\ &= \pi(\mathrm{d}x')\mathbb{E}^{x'}[h(X_t,\mathbf{m},\theta)] = \pi(\mathrm{d}x')\mathbb{E}^{(\mathbf{m},\theta)}[h(x',M_t,\Theta_t)] \\ &= \sum_{\mathbf{n}\leq\mathbf{m}} p_{\mathbf{m},\mathbf{n}}(t;\theta)h(x',\mathbf{n},\Theta_t)\pi(\mathrm{d}x')\end{aligned}$$

□

Duality and optimal filtering pt 1: update

Proposition

For

$$\bar{\mathcal{F}}_f = \left\{ \sum_{\mathbf{m}\in\Lambda} w_{\mathbf{m}} h(x,\mathbf{m},\theta)\pi(dx) \,:\, \Lambda \subset \mathcal{M}\,,\, |\Lambda| < \infty\,,\, w_{\mathbf{m}} \geq 0\,,\, \sum_{\mathbf{m}\in\Lambda} w_{\mathbf{m}} = 1 \right\}.$$

and under A1-A2-A3, $\bar{\mathcal{F}}_f$ is closed under prediction and update:

$$\phi_y\left(\sum_{\mathbf{m}\in\Lambda} w_{\mathbf{m}} h(x,\mathbf{m},\theta)\pi(dx) \right) = \sum_{\mathbf{n}\in t(y,\Lambda)} \widehat{w}_{\mathbf{n}} h(x,\mathbf{n},T(y,\theta))\pi(dx)$$

with

$$t(y,\Lambda) := \{\mathbf{n} : \mathbf{n} = t(y,\mathbf{m}), \mathbf{m}\in\Lambda\}$$

$$\widehat{w}_{\mathbf{n}} \propto w_{\mathbf{m}} c(\mathbf{m},\theta,y) \quad \textit{for} \quad \mathbf{n} = t(y,\mathbf{m})\,, \sum_{\mathbf{n}\in t(y,\Lambda)} \widehat{w}_{\mathbf{n}} = 1\,,$$

and

$$\psi_t\left(\sum_{\mathbf{m}\in\Lambda} w_{\mathbf{m}} h(x,\mathbf{m},\theta)\pi(dx) \right) = \sum_{\mathbf{n}\in G(\Lambda)} \left(\sum_{\mathbf{m}\in\Lambda,\mathbf{m}\geq\mathbf{n}} w_{\mathbf{m}} p_{\mathbf{m},\mathbf{n}}(t;\theta) \right) h(x,\mathbf{n},\theta_t)\pi(dx).$$

Remarks

Multiple sums

- number of terms of the sum over all $\mathbf{m} \geq \mathbf{n}$ grows with the dimension of $\mathcal{M}$

$G(\Lambda)$

- states that are accessible from the nodes in Λ, see 19

Creation of new components

- each update leaves the number fixed, but shifts indices, which are then filled up by the propagation

Mixture structure

- arises due to uncertainty about the state of the death process (time-discretisation) - weights relate to transition probabilities

Dual

- how can we find such dual??

Generators

(Care is needed with domains but will skip for economy of time)

The generator of a Markov process X_t, with semigroup operator P_t, is a linear operator $\mathcal{A}$ with domain denoted $\mathscr{D}(\mathcal{A})$, linked to the semigroup operator via the Kolmogorov backward equation

$$\frac{\partial}{\partial t} P_t f(x) = (\mathcal{A} P_t f)(x), \qquad f \in \mathscr{D}(\mathcal{A}),$$

where on the left hand side $P_t h(x)$ is differentiated in t for given x, whereas on the right hand side, $\mathcal{A}$ acts on $P_t h(x)$ as a function of x for given t

Local duality

(Care is needed with domains but will skip for economy of time)

If X_t solves an SDE on $\mathbb{R}^d$

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}B_t\,,$$

then its generator is

$$(\mathcal{A}f)(x) = \sum_{i=1}^{d} b_i(x)\frac{\partial f(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \qquad f \in \mathscr{D}(\mathcal{A}),$$

for $a_{i,j}(x) := (\sigma(x)\sigma(x)^T)_{i,j}$

If (M_t, Θ_t) is Markov process as in A3 its generator is

$$(Ag)(\mathbf{m},\theta) = \lambda(|\mathbf{m}|)\rho(\theta)\sum_{i=1}^{K} m_i[g(\mathbf{m}-\boldsymbol{e}_i,\theta)-g(\mathbf{m},\theta)]+\sum_{i=1}^{l} r_i(\theta)\frac{\partial g(\mathbf{m},\theta)}{\partial \theta}\,, \quad g \in \mathscr{D}(A),$$

A4 (Local duality): The function $h(x, \mathbf{m}, \theta)$ defined in A2 is such that $h(x, \mathbf{m}, \theta)$, as a function of x belongs to $\mathscr{D}(\mathcal{A})$ for all $(\mathbf{m}, \theta) \in \mathcal{M} \times \Theta$, as a function of $(\mathbf{m}, \theta)$ belongs to $\mathscr{D}(A)$ for all $x \in \mathcal{X}$, and $\forall x \in \mathcal{X}, \mathbf{m} \in \mathcal{M}, \theta \in \Theta$

$$(\mathcal{A}h(\cdot, \mathbf{m}, \theta))(x) = (Ah(x, \cdot, \cdot))(\mathbf{m}, \theta) \tag{1}$$

This implies the duality in A3.

(Proof: resolvents/Laplace transforms)

Filtering Algorithm

A5 (Initialisation): The initial distribution of the signal is
$\nu = h(x, \mathbf{m}_0, \theta_0)\pi(\mathrm{d}x) \in \mathcal{F}$, for some $\mathbf{m}_0 \in \mathcal{M}, \theta_0 \in \Theta$.

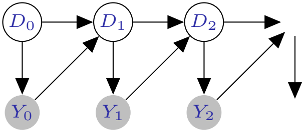


Figure: The partially observed Markov process dual to the hidden Markov model in Figure 3, where $D_i = (M_i, \Theta_i)$.

$$\mathcal{L}(X_{t_n}|Y_0, \ldots, Y_n) = \int h(x, t(Y_n, M_n), T(Y_n, \Theta_n))\pi(\mathrm{d}x)\mathrm{d}\mathcal{L}(D_n|Y_0, \ldots, Y_{n-1}) \,.$$

Complexity

$\mathcal{L}(D_n | Y_0, \ldots, Y_{n-1})$ has support on $\Lambda_n \times \{\theta_n\}$ with $\Lambda_{n+1} = G(t(Y_n, \Lambda_n))$ and weights that can be computed recursively.

The component probabilities at time n can be computed at a cost that is at most $\mathcal{O}(|\Lambda_n|^2)$, but Λ_n increases with n.

Proposition

Under the assumption that $t(y, \mathbf{m}) = \mathbf{m} + N(y)$, where $N : \mathcal{Y} \to \mathcal{M}$, we have that

$$|\Lambda_n| = G\left(\mathbf{m}_0 + \sum_{i=1}^{n} N(Y_i)\right) \leq \left(1 + \frac{d_n}{K}\right)^K$$

where $d_n = |\mathbf{m}_0 + \sum_{i=1}^{n} N(Y_i)|$.

When the observations follow a stationary process, d_n will be of order n, thus overall complexity of filtering $\mathcal{O}(n^{2K})$, where the constant depends on K but not n

Examples

I will focus on the properties of the signal and illustrate the dual. Clearly, for computable filtering we need the emission density to be conjugate to the family $\mathcal{F}$.

I will present results for:

- Linear diffusion signals - only deterministic dual
- CIR signals - both deterministic and death process dual

Due to time constraints I will skip results for:

- Multi-dimensional Wright-Fisher signals - only death process dual

Linear diffusion signals

SDE

$$\mathrm{d}X_t = -\frac{\sigma^2}{\alpha}(X_t - \gamma)\mathrm{d}t + \sqrt{2}\sigma \mathrm{d}B_t\,,$$

Invariant measure

$$\pi(\mathrm{d}x) \equiv \mathrm{Normal}(\mathrm{d}x; \gamma, \alpha)\,.$$

Generator

$$\mathcal{A} = (\sigma^2\gamma/\alpha - \sigma^2 x/\alpha)\frac{\mathrm{d}}{\mathrm{d}x} + \sigma^2\frac{\mathrm{d}^2}{\mathrm{d}x^2}$$

Duality functions

$$h(x,\mu,\tau) = \left(\frac{\alpha}{\tau}\right)^{1/2} \exp\left\{-\frac{(x-\mu)^2}{2\tau} + \frac{(x-\gamma)^2}{2\alpha}\right\} = \frac{\mathrm{Normal}(\mathrm{d}x;\mu,\tau)}{\pi(\mathrm{d}x)}$$

Linear diffusion signals

Kalman filter - in 1 line

A small calculation yields:

$$\mathcal{A}h(\cdot,\mu,\tau)(x) = \frac{\sigma^2}{\alpha}(\gamma-\mu)\frac{\partial}{\partial\mu}h(x,\mu,\tau) + 2\sigma^2(1-\tau/\alpha)\frac{\partial}{\partial\tau}h(x,\mu,\tau)\,.$$

Hence the dual is purely deterministic and described in terms of the ODEs:

$$\mathrm{d}\mu_t/\mathrm{d}t = \frac{\sigma^2}{\alpha}(\gamma-\mu_t)\mathrm{d}t\,, \quad \mathrm{d}\tau_t/\mathrm{d}t = 2\sigma^2(1-\tau_t/\alpha)\mathrm{d}t\,.$$

CIR signals

Some care is needed with domains
SDE, for $\sigma > 0, \gamma > 0, \delta \geq 2$

$$\mathrm{d}X_t = (\delta\sigma^2 - 2\gamma X_t)\mathrm{d}t + 2\sigma\sqrt{X_t}\mathrm{d}B_t$$

Invariant measure

$$\pi(\mathrm{d}x) \equiv \mathrm{gamma}(\mathrm{d}x; \delta/2, \gamma/\sigma^2)\,.$$

Generator

$$\mathcal{A} = (\delta\sigma^2 - 2\gamma x)\frac{\mathrm{d}}{\mathrm{d}x} + 2\sigma^2 x \frac{\mathrm{d}^2}{\mathrm{d}x^2},$$

Duality functions

$$\begin{aligned} h(x,\mu,\tau) &= \frac{\Gamma(\delta/2)}{\Gamma(\delta/2+m)}\left(\frac{\gamma}{\sigma^2}\right)^{-\delta/2}\theta^{\delta/2+m}x^m \exp\{-(\theta-\gamma/\sigma^2)x\} \\ &= \frac{\mathrm{gamma}(\mathrm{d}x;\delta/2+m,\theta)}{\pi(\mathrm{d}x)} \end{aligned}$$

CIR signals

A direct calculation yields:

$$\begin{aligned}\mathcal{A}h(\cdot,m,\theta)(x) &= 2m\sigma^2\theta\, h(x,m-1,\theta)+\sigma^2(\delta+2m)(\theta-\gamma/\sigma^2)h(x,m+1,\theta)\\ &- \sigma^2[2m\theta+(\delta+2m)(\theta-\gamma/\sigma^2)]h(x,m,\theta)\end{aligned}$$

We consider a two-component process (M_t,Θ_t) with generator A as in (2), where

$$\lambda(m)=2m\sigma^2\,,\quad r(\theta)=2\sigma^2\theta(\gamma/\sigma^2-\theta)\,,\quad \rho(\theta)=\theta\,.$$

A direct calculation gives

$$\mathcal{A}h(\cdot,m,\theta)(x)=Ah(x,\cdot,\cdot)(m,\theta)\,.$$

CIR signals

Under this setting

$$\Theta_t = \frac{\gamma}{\sigma^2} \frac{\theta e^{2\gamma t}}{\theta e^{2\gamma t} + \gamma/\sigma^2 - \theta}$$

which implies that the transition probabilities for the death process simplify to binomial probabilities

$$p_{m,m-i}(t;\theta) = \text{Bin}\left(m - i\,;\, m, \frac{\gamma}{\sigma^2}(\theta e^{2\gamma t} + \gamma/\sigma^2 - \theta)^{-1}\right)$$

which in turns immediately yields the result

$$\psi_t(\text{gamma}(m + \delta/2, \theta)) =$$

$$\sum_{k=0}^{m} \text{Bin}\left(k\,;\, m, \frac{\gamma}{\sigma^2}(\theta e^{2\gamma t} + \gamma/\sigma^2 - \theta)^{-1}\right) \text{gamma}\left(k + \delta/2, \frac{\gamma}{\sigma^2} \frac{\theta e^{2\gamma t}}{\theta e^{2\gamma t} + \gamma/\sigma^2 - \theta}\right).$$

Numerical implementation - approximations

The framework lends itself to software implementation. User only needs to specify

- The h functions
- emission density
- solution to the ODE
- the two functions related to the rates of the death process

This feeds into a model independent code that carries all computations, regardless of the dimensions of X_t, Y_n, Θ_t, M_t to do

- Filtering
- Likelihood estimation
- Forward simulation - backward sampling

Unless the filter is finite-dimensional, the exact filtering algorithm is impractical with large number of observations, due to the polynomial increase in computational cost

However, most of the components can have negligible weights and simulations show this - even for non-stationary signals. The number of components with non-negligible weight to which the filter stabilises is under study. An implementation based on pruning has linear computational cost and simulations suggest very small error

We are also investigating the use of this machinery for approximate filtering

Infinite-dimensional signals

In Papaspiliopoulos, Ruggiero and Spanò (2013), we have obtained similar results for infinite-dimensional signals that correspond to dynamic versions of standard Bayesian nonparametric models. Note that in such models there is no common dominating measure for the emission distributions, hence there is no likelihood

- Fleming-Viot measure-valued diffusion (Dirichlet process prior)
- Dawson-Watanabe process (gamma process)

There is an interesting connection between the dual we use in these calculations and the famous Kingman's coalescent with mutation

Spectral decomposition of generators

- Classic results exist about generators with countable spectrum and expansions such that the Poisson-gamma mixture for the CIR
- This is related to computable filtering but the connection is not well understood