

Bayesian inference for dynamically evolving distributions

Omiros Papaspiliopoulos, Matteo Ruggiero and Dario Spanò

Abstract Filtering hidden Markov models amounts to perform sequential Bayesian inference on some dynamic parameter given observations collected at discrete time points. The filtering is exact if the sequence of posterior distributions are evaluated in closed form rather than approximated. Relying on the notion of duality for Markov processes, we provide general sufficient conditions for constructing a filter which sequentially evaluates the posterior distribution of finite-dimensional parameters on uncountable state spaces. We then exploit the application of such results to some specific models for extending the filtering strategies to some important classes of measure-valued diffusions, which describe dynamically evolving nonparametric distributions.

Key words: Sequential inference, optimal filtering, dual process, posterior distribution, Dirichlet process, gamma process.

1 The filtering problem

Consider a sequence of observations $\{Y_n, n \geq 0\}$, $Y \in \mathcal{Y}$, and an unobserved Markov chain $\{X_n, n \geq 0\}$, which we assume to be the discrete-time sampling of a homogeneous continuous-time Markov process X_t . We assume X_t has state-space $\mathcal{X}$, transition kernel $P_t(x, dx')$ and initial distribution $\nu(dx)$. The observations relate to the hidden signal by means of a family of conditional distributions $\mathcal{L}(y | x)$. If the model admits a common dominating measure for all values of the signal, so that densities $f_x(y)$ exist, we will say the model is *parametric*.

Omiros Papaspiliopoulos
ICREA & Department of Economics and Business, Universitat Pompeu Fabra
Ramón Trias Fargas 25-27, 08005, Barcelona, Spain. e-mail: omiros.papaspiliopoulos@upf.edu

Matteo Ruggiero
Collegio Carlo Alberto & Department of Economics and Statistics, University of Torino
C.so Unione Sovietica 220, 10134, Torino, Italy. e-mail: matteo.ruggiero@unito.it

Dario Spanò
Department of Statistics, University of Warwick
Coventry, CV4 7AL, UK. e-mail: D.Spano@warwick.ac.uk

The exact filtering problem is the closed form evaluation of the *filtering distributions* $\mathcal{L}(X_{t_n}|Y_0, \dots, Y_n)$, henceforth denoted $v_n(dx)$, that is the conditional distributions of the unobserved signal at time t_n given observations $Y_i = Y_{t_i}$ collected up to time t_n . Once these are available, a number of statistical applications can be carried out, such as the prediction of future observations, the derivation of smoothing distributions (the distribution of the signal given past and future observations) and the calculation of the marginal likelihood in dominated models.

For parametric models, deriving an optimal filter entails solving the recursion defined by $v_0 = \phi_{Y_0}(v)$ and $v_n = \phi_{Y_n}(\psi_{t_n-t_{n-1}}(v_{n-1}))$, for $n \in \mathbb{N}$. Here ϕ defines the *update operator*

$$\phi_v(v)(dx) = f_x(y)v(dx)/p_v(y), \quad p_v(y) = \int_{\mathcal{X}} f_x(y)v(dx),$$

which amounts to the application of Bayes theorem, with the filtering distribution $v(x)$ taking the role of the prior, and ψ the *prediction operator*

$$\psi_t(v)(dx') = vP_t(dx') = \int_{\mathcal{X}} v(dx)P_t(x, dx'), \quad (1)$$

which evaluates the distribution of the next step of the Markov chain initiated from v . The iteration of the two operators typically leads to intractable distributions, with notable exceptions being Baum-Welch filters, for finite state-space cases, and Kalman filters, for Gaussian state-space models. Previous results on parametric filtering include, among others, Sawitzki (1981), Ferrante and Vidoni (1998), Runggaldier and Spizzichino (2001), Chaleyat-Maurel and Genon-Catalot (2006).

In the nonparametric case, when a common dominating measure is not available, we regard the signal as infinite-dimensional or measure-valued. In this case the observations are samples from the signal itself, i.e., $Y_t | X_t \sim X_t$, and the update operator ϕ is not defined. In particular, if a discrete X_t is defined on an uncountable $\mathcal{X}$, optimal filtering must take into account the possibility that the support of X_t changes in time, as illustrated in Figure 1.

In Section 2 we derive sufficient conditions for the existence, in the parametric case, of a filter whose complexity grows polynomially in the number of observations. Such conditions rely on the existence of a certain type of *dual process* for the signal. In Section 3 we discuss some examples and briefly sketch how the results can be extended to the nonparametric case, at least for some classes of dynamically evolving distributions. The results presented here are based on Papaspiliopoulos and Ruggiero (2014) and Papaspiliopoulos, Ruggiero and Spanò (2014).

2 Duality and filtering

Let $\{V_t, t \geq 0\}$ on $\mathcal{V}$ and $\{W_t, t \geq 0\}$ on $\mathcal{W}$ be two Markov processes. These are said to be *dual* with respect to the function $q(v, w)$ if

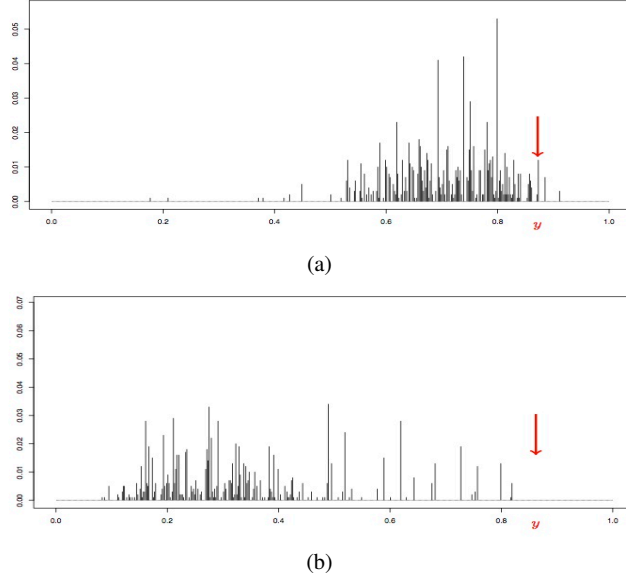


Fig. 1 Illustration of a nonparametric filtering problem: (a) X_0 is the unobserved starting state of a purely atomic measure-valued process, in the form of a discrete random measure, of which only the location $Y_0 = y$ is observed; (b) over an interval of length $t > 0$, X_0 has evolved into the discrete measure X_t represented; given $Y_0 = y$, the evaluation of the conditional law $\mathcal{L}(X_t | Y_0 = y)$ of the signal must take into account that Y_0 may no longer be an atom of X_t , as occurs in the illustrated sample path.

$$\mathbb{E}(q(V_t, w) | V_0 = v) = \mathbb{E}(q(v, W_t) | W_0 = w), \quad \forall v \in \mathcal{V}, w \in \mathcal{W}, t \geq 0.$$

With some additional restrictions, a sufficient condition for duality can be expressed in terms of the generators, through the condition

$$(\mathcal{A}q(\cdot, w))(v) = (Aq(v, \cdot))(w), \quad \forall v \in \mathcal{V}, w \in \mathcal{W},$$

where $\mathcal{A}$ and A are the generators of V_t and W_t respectively. In the following, for multi-indices $\mathbf{m} = (m_1, \dots, m_K) \in \mathcal{M} = \mathbb{Z}_+^K$ we will consider *duality functions* $h : \mathcal{X} \times \mathcal{M} \times \Theta \rightarrow \mathbb{R}_+$, $\Theta \subseteq \mathbb{R}^l$, where the first two components (X_t, M_t) will represent Markov processes and the third Θ_t a deterministic process. With slight abuse of notation we will write $h(x, \mathbf{m})$ instead of $h(x, \mathbf{m}, \theta)$ when Θ_t is absent or ancillary. We assume h satisfies $\sup_x h(x, \mathbf{m}, \theta) < \infty$ for all $\mathbf{m} \in \mathcal{M}, \theta \in \Theta, l \in \mathbb{Z}_+$. We also define the following notion of *conjugacy*, by assuming $\mathcal{F} = \{h(x, \mathbf{m}, \theta)\pi(dx), \mathbf{m} \in \mathcal{M}, \theta \in \Theta\}$ is a family of probability measures such that there exist functions $t : \mathcal{Y} \times \mathcal{M} \rightarrow \mathcal{M}$ and $T : \mathcal{Y} \times \Theta \rightarrow \Theta$ with $\mathbf{m} \rightarrow t(y, \mathbf{m})$ increasing and such that

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

Papaspiliopoulos and Ruggiero (2014) showed that when the signal X_t is reversible with respect to a probability measure π on $\mathcal{X}$ and has a dual process $(M_t, \Theta_t) \subset \mathcal{M} \times \Theta$ with generator

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

where $\mathbf{e}_i$ is the vector with the unit i th coordinate being the only non zero element and $r : \Theta \rightarrow \Theta$ is such that the differential equation

$$d\Theta_t/dt = r(\Theta_t), \quad \Theta_0 = \theta_0, \quad (2)$$

has a unique solution for all θ_0 , then (1) can be written

$$\psi_t(h(x, \mathbf{m}, \theta)\pi(dx)) = \sum_{\mathbf{0} \leq \mathbf{i} \leq \mathbf{m}} p_{\mathbf{m}, \mathbf{m}-\mathbf{i}}(t; \theta) h(x, \mathbf{m} - \mathbf{i}, \Theta_t) \pi(dx). \quad (3)$$

Here Θ_t is a deterministic process on Θ driven by the ODE (2) and M_t is a multidimensional pure death process on $\mathcal{M}$, with transition probabilities

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

available in closed form; see Proposition 2.1 in Papaspiliopoulos and Ruggiero (2014). Thus, under the stated conditions, (3) allows to express the time propagation of the signal as a finite sum even if the transition function in (1) has an infinite series representation. This implies that the conditional law of the signal can be written

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

where $\{D_n = (M_n, \Theta_n), n \geq 0\}$ is the chain embedded in (M_t, Θ_t) at the observation times. Thus, filtering X_{t_n} in the original model can be achieved by filtering D_n in the dual model. Since Θ_n evolves deterministically, optimal filtering for X_t reduces to filtering M_n , which has finite support and whose associated probabilities can be computed recursively.

3 Examples and nonparametric developments

3.1 Dirichlet case

Wright–Fisher diffusions are K dimensional processes which describe the evolution of K frequencies and have generator

$$\mathcal{A} = \frac{1}{2} \sum_{i=1}^K (\gamma_i - \vartheta x_i) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^K x_i (\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j}, \quad \vartheta = \sum_{i=1}^K \gamma_i,$$

where $\delta_{ij} = \mathbb{1}(i = j)$ and $\mathcal{A}$ acts on functions in $C^2(\Delta_K)$ on the $(K - 1)$ -dimensional simplex Δ_K . The diffusions with such operator are reversible with respect to the Dirichlet distribution $\pi_K(\mathbf{d}\mathbf{x}) \propto \mathbf{x}^{\gamma-1}\mathbf{d}\mathbf{x}$ (in multi-index notation), which is well known to be conjugate to Multinomial observations. It can be shown that the Wright–Fisher diffusion is dual, with respect to the function

$$h(\mathbf{x}, \mathbf{m}) = \frac{\Gamma(\vartheta + m)}{\Gamma(\vartheta)} \prod_{i=1}^K \frac{\Gamma(\gamma_i)}{\Gamma(\gamma_i + m_i)} \mathbf{x}^{\mathbf{m}}, \quad m = \sum_{i=1}^K m_i,$$

to a death process with generator obtained by setting $l = 0$, $\rho(\theta) \equiv 1$ and $\lambda(|\mathbf{m}|) = (\theta + m - 1)/2$ in (2). Here there is no deterministic component, and we write h with two arguments instead of three. Then, assuming a Multinomial likelihood for the observations, from the results of the previous section it follows that (3) holds with $h(\mathbf{x}, \mathbf{m})\pi(\mathbf{d}\mathbf{x})$ being a posterior Dirichlet distribution conditional on m observations, with multiplicities $\mathbf{m}$ for the K categories, and $p_{\mathbf{m}, \mathbf{m}-\mathbf{i}}(t)$ are the transition probabilities of the death process with generator described above.

The Fleming–Viot process with parent-independent mutation is characterized by the generator

$$\mathbb{A}\varphi(x) = \int_{\mathcal{Y}} x(\mathbf{d}y) M\left(\frac{\partial\varphi(x)}{\partial x(\cdot)}\right)(y) + \frac{1}{2} \int_{\mathcal{Y}} \int_{\mathcal{Y}} x(\mathbf{d}y)(\delta_{y'}(\mathbf{d}y) - x(\mathbf{d}y)) \frac{\partial^2\varphi(x)}{\partial x(\mathbf{d}y)\partial x(\mathbf{d}y')}$$

where $\partial\varphi(x)/\partial x(\mathbf{d}y) = \lim_{\varepsilon \rightarrow 0^+} \varepsilon^{-1}(\varphi(x + \varepsilon\delta_y) - \varphi(x))$, M is the *mutation operator*

$$(Mf)(y) = \frac{\theta}{2} \int_{\mathcal{Y}} [f(y') - f(y)] P_0(\mathbf{d}y'), \quad f \in B(\mathcal{Y}), \quad (4)$$

and $P_0 \in \mathcal{M}_1(\mathcal{Y})$ is the nonatomic mutant offspring distribution. Such Fleming–Viot process extends the Wright–Fisher diffusion and is reversible with respect to the Dirichlet process, which enjoys a conjugacy property. By elaborating on the results for Wright–Fisher diffusions and exploiting the relation between the two classes of processes, one can show that an infinite-dimensional version of (3) holds for this nonparametric case, yet still entailing a finite computation.

3.2 Gamma case

The Cox–Ingersoll–Ross model (Cox, Ingersoll and Ross, 1985), which is reversible and ergodic with respect to a gamma $G(\alpha, \beta)$ distribution, is a diffusion with generator

$$\mathcal{B} = \frac{1}{2}(\alpha - \beta x) \frac{d}{dx} + \frac{1}{2}x \frac{d^2}{dx^2},$$

acting on $C^2([0, \infty))$ functions which vanish at infinity. This can be shown to be dual to a two-component process (M_t, Θ_t) , taking values in $\mathbb{Z}_+ \times \mathbb{R}_+$, with generator

$$Bg(m, \theta) = 2m(\beta + \theta)(g(m-1, \theta) - g(m, \theta)) - 2\theta(\beta + \theta) \frac{\partial}{\partial \theta} g(m, \theta),$$

with respect to the duality function

$$h(x, m, \theta) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + m)} \left(\frac{\beta + \theta}{\beta} \right)^\alpha (\beta + \theta)^m x^m e^{-\theta x}.$$

The dual process (M_t, Θ_t) is composed by a death process M_t on $\mathbb{Z}_+$, which, given $\Theta_t = \theta$, decreases from m to $m-1$ at rate $2m(\beta + \theta)$ and is absorbed at zero, and a deterministic process Θ_t , whose associated ODE has solution

$$\Theta_t = \beta e^{-2\beta t} / (1 + \beta/\theta_0 - e^{-2\beta t}), \quad \Theta_0 = \theta_0.$$

Note that $h(x, m, \theta)G(\alpha, \beta) = G(\alpha + m, \beta + \theta)$, which is a posterior Gamma distribution, conditional on n Poisson observations whose sum is m . Then, assuming Poisson likelihood for the observations, we have

$$\psi_t(G(\alpha + m, \beta + \theta)) = \sum_{i=0}^m \text{Bin}(m-i; m, p(t)) G(\alpha + m-i, \beta + \theta_t),$$

where $p(t) = \beta(\theta_0 - (\beta + \theta_0)e^{2\beta t})^{-1}$. Furthermore, such relation can be extended to the infinite-dimensional case for the Dawson-Watanabe measure-valued branching diffusion with generator

$$\mathbb{B}\varphi(x) = \int_{\mathcal{Y}} x(dy) M \left(\frac{\partial \varphi(x)}{\partial x(\cdot)} \right) (y) + \frac{1}{2} \int_{\mathcal{Y}} x(dy) \frac{\partial^2 \varphi(x)}{\partial x(dy)^2},$$

with M as in (4), which extends the CIR process. This follows by exploiting the infinite divisibility property of the gamma process, which is the reversible measure, together with the results of Section 3.1 and the well-known representation of a Dirichlet process as a normalised gamma random measure.

References

- CHALEYAT-MAUREL, M. AND GENON-CATALOT, V. (2006). Computable infinite-dimensional filters with applications to discretized diffusion processes. *Stoch. Proc. Appl.* **116**, 1447–1467.
- COX, J.C., INGERSOLL, J.E. AND ROSS, S.A. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.
- FERRANTE, M. AND VIDONI, P. (1998). Finite dimensional filters for nonlinear stochastic difference equations with multiplicative noises. *Stoch. Proc. Appl.* **77**, 69–81.
- PAPASPILIOPOULOS, O. AND RUGGIERO, M. (2014). Optimal filtering and the dual process. *Bernoulli*, in press.
- PAPASPILIOPOULOS, O., RUGGIERO, M. AND SPANÒ, D. (2014). Optimal nonparametric filtering and the dual process. *Preprint*.
- RUNGALDIER, W. AND SPIZZICHINO, F. (2001). Sufficient conditions for finite dimensionality of filters in discrete time: A Laplace transform-based approach. *Bernoulli* **7**, 211–221.
- SAWITZKI, G. (1981). Finite-dimensional filter systems in discrete time. *Stochastics* **5**, 107–114.