Duality-based inference for state-space models

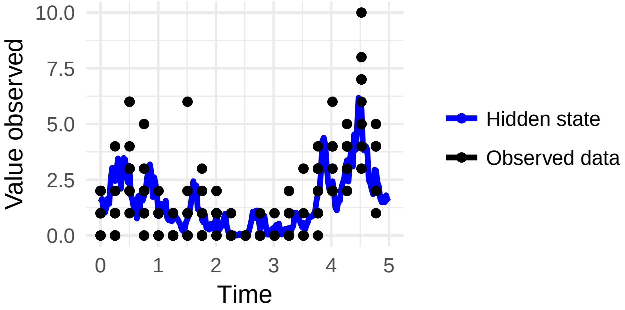
Guillaume KON KAM KING[1]
Andrea PANDOLFI[2]
Omiros PAPASPILIOPOULOS[2]
Marco PIRETTO[3]
Matteo RUGGIERO[4,5]

[1]Université Paris-Saclay, INRAE, MaIAGE, France
[2]Department of Decision Sciences, Bocconi University, Milano, Italy
[3]BrandDelta
[4]University of Torino, Italy
[5]Collegio Carlo Alberto, Torino, Italy

"Statistique au Sommet"
Rochebrune, 25 Mars 2024

A hidden Markov (diffusion) model: example

Cox-Ingersoll-Ross (CIR) process with Poisson observations
Inference taking into account **underlying dynamics.**



$$X_1 \sim \pi(x_1)$$
$$X_n|(X_{n-1} = x_{n-1}) \sim f^\theta(x_n|x_{n-1})$$
$$Y_n|(X_n = x_n) \sim g^\theta(y_n|x_n)$$

Inference targets:

$p(x_n|y_{1:n})$ (*filtering*)

$p(x_{1:n}|y_{1:n})$ (*smoothing*)

$p(\theta|y_{1:n}), \quad p(\theta, x_{1:n}|y_{1:n})$

Recursive relation for filtering

$$p(x_n|y_{1:n}) = \frac{g(y_n|x_n)\int_{\chi} p(x_{n-1}|y_{1:n-1})f(x_n|x_{n-1})\mathrm{d}x_{n-1}}{p(y_n|y_{1:n-1})}$$

Prediction: $\psi(\xi)(\mathrm{d}x) = \int_{\chi} \xi(x)f(x|x')\mathrm{d}x'$

Update: $\phi_{y_n}(\xi)(\mathrm{d}x) = \frac{g(y_n|x)\xi(\mathrm{d}x)}{\int_{\chi} g(y_n|x')\xi(x')\mathrm{d}x'}$

Filtering recurrence: $p(x_n|y_{1:n}) = \phi_{y_n}(\psi(p(x_{n-1}|y_{1:n-1})))$

Optimal Filtering: intractable prediction

Optimal filtering is **almost always an intractable problem**, notably because prediction is intractable in general :

$$\text{Prediction: } \psi(\xi)(\mathrm{d}x) = \int_{\chi} \xi(x) f(x|x')\mathrm{d}x' = \mathbb{E}\left[\xi(X_n)|X_{n-1} = x'\right]$$

- Cox-Ingersoll-Ross diffusion:

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

$$f(x|x') = \sum_{k\geq 0} \text{Poisson}(k|cx')\text{Gamma}(x|d+k, e)$$

- Wright-Fisher diffusion:

$$\mathrm{d}X_t = \frac{1}{2}\left(\theta_1(1 - X_t) - \theta_2 X_t\right)\,\mathrm{d}t + \sqrt{X_t(1 - X_t)}\,\mathrm{d}B_t$$

$$f(x|x') = \sum_{k\geq 0} q_k^{\theta} \sum_{l\geq 0}^{k} \text{Bin}(l|k, x')\text{Beta}(x|\theta_1, \theta_2)$$

with q_k^{θ} available only as an infinite series.

Optimal Filtering: tractable approaches and particle approximations

- In intractable cases, if **sampling from** $f(x|x')$ **is possible**, one can use a **particle approximation**:

$$
\begin{aligned}
x'_k &\sim \xi & k &= 1, \cdots, n \\
x_k | x'_k &\sim f(x_k | x'_k) & k &= 1, \cdots, n
\end{aligned}
$$

$$
\psi(\xi)(\mathrm{d}x) \approx \sum_{i=1}^{n} \delta_{x_k}(\mathrm{d}x)
$$

- Discrete state-space case is a **tractable exception**:

$$
\xi = \sum_{j \in \chi} \alpha_j \delta_j \qquad \psi(\xi)(i) = \sum_{j \in \chi} \xi(j) f(i|j)
$$

- We present **duality** as another approach to obtain tractable solutions

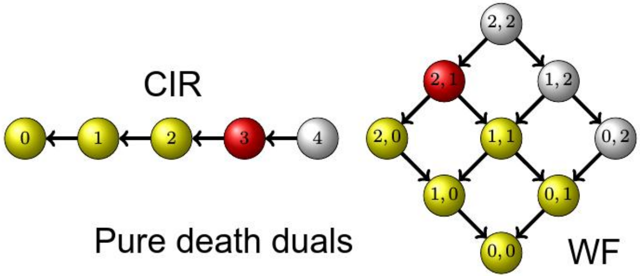
Dual Processes

What if there were another process D_n such that for a class of functions h:

$$\mathbb{E}\left[h(X_n, d) \mid X_0 = x\right] = \mathbb{E}\left[h(x, D_n) \mid D_0 = d\right] \quad ?$$

X_n, D_n are said to be in duality w.r.t. h.

Hope: Take h to (almost) be a filtering distribution and compute prediction using the more convenient dual process.



What is a good dual process ?

Prediction, now involving $\mathbb{E}\left[h(x, Dn) \mid D_0 = d\right]$, may become more tractable:

- **Pure death** dual process on a discrete state space: prediction involves a **finite sum**, tractable[1].
- **Birth-Death** dual process: prediction involves a **infinite sum**, **intractable** but easy to truncate/sample from.
- **Other duals ?**: some may be easier to handle/sample than the original process, this may even depend on the parameters/data at hand.

Remark: following Papaspiliopoulos and Ruggiero (2014), the dual is determined by the choice of h function.

[1] Filtering, smoothing and full Bayesian inference on $p(\theta, x_{1:n}|y_{1:n})$ described in Kon Kam King et al. (2021)

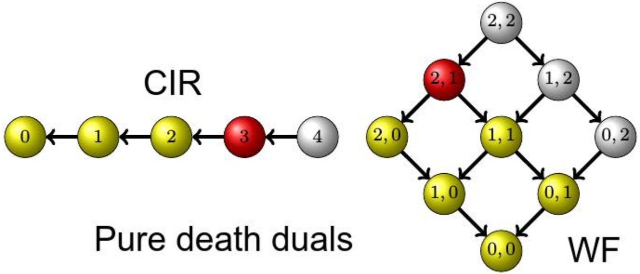
Dual process on a discrete space χ

Let $\xi(\mathrm{d}x) = h(x, d')\pi(\mathrm{d}x)$. Then:

$$\psi(h(x, d')\pi(\mathrm{d}x)) = \sum_{d \in \chi} p_D(d|d')h(x, d)\pi(\mathrm{d}x)$$

with p_D transition probabilities for the dual process D_n.

Pure death dual: finite number of d such that $f(d|d') \neq 0$



Particle filtering on the discrete dual space

In the case where an infinite number of states in χ can be reached from a state d', we can use a **particle approximation.**

- We assume here *(e.g. consequence of a convenient prior)* that:
$$\xi(\mathrm{d}x) = \sum_{d \in \chi} w_d h(x, d) \pi(\mathrm{d}x)$$
- We form an N-particle approximation:
$$\xi(\mathrm{d}x) \approx \frac{1}{N} \sum_{i=1}^{N} h(x, d_i') \pi(\mathrm{d}x), \qquad d_i' \sim \sum_{d \in \chi} w_d \delta_d$$
- Propagation of the particles via the dual process
$$d_i \sim p_D(d_i | d_i') \quad \Rightarrow \quad \psi(\xi)(\mathrm{d}x) \approx \frac{1}{N} \sum_{i=1}^{N} h(x, d_i) \pi(\mathrm{d}x)$$

Akin to particle approximation of the (discrete) mixing measure !

Can be cast into a Feynman-Kac process -> usual asymptotic guarantees

A variety of dual processes: CIR dualities

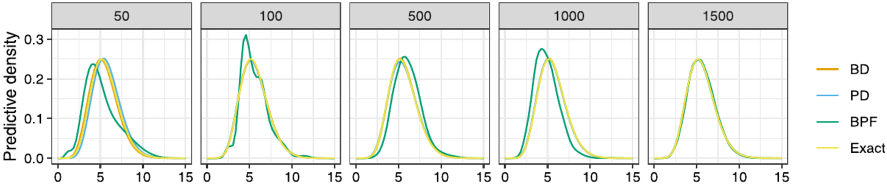
Let's consider a Cox-Ingersoll-Ross diffusion X_t with Poisson(X_t) observations.

- Pure death dual: **inhomogeneous** death process with instantaneous rate ($d \to d-1$): $\lambda_d \propto d\Theta(t)$
  - Simulation not trivial (time rescaling/thinning)
  - **Closed form** expression available for arbitrary time transition
- Birth-Death dual: **homogeneous** process jumping from:

$$d \to d+1 \text{ with rate } \lambda_d$$
$$d \to d-1 \text{ with rate } \mu_d$$

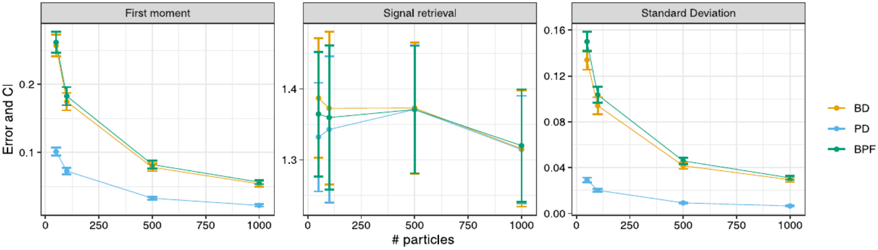
  - Easy to simulate (Gillespie algorithm), slow if high rate
  - **Closed form** expression for arbitrary time transition (linear BD process, Tavaré 2018).

Comparison of various CIR dual predictions



- BD: Birth-Death particle approx, PD: Pure Death particle approx, BPF: Bootstrap Particle Filter
- Pure death and Birth-Death duals with 100 particles on a par with Bootstrap Particle filtering 1500

Filtering performance comparison



- Pure Death particle approximation seems more efficient
- Would need to take timing into account
- Dataset/parameter value dependent

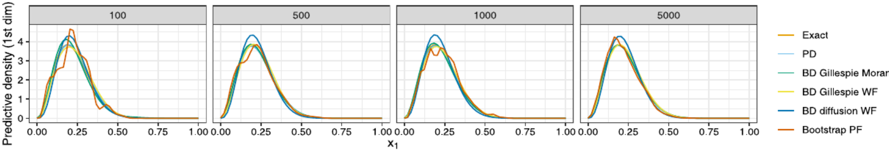
A variety of dual processes: Wright-Fisher dualities

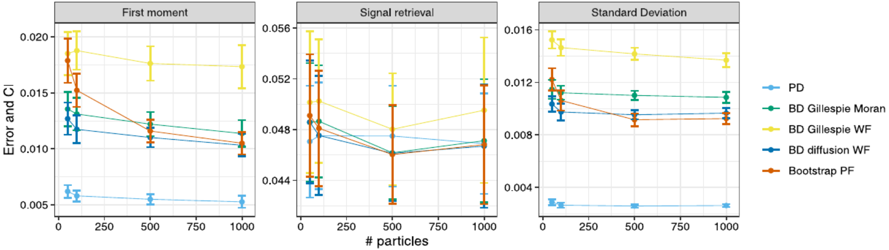
Let's consider a Wright-Fisher diffusion X_t with multinomial(X_t) observations.

- Pure death dual (Kingman's coalescent): **homogeneous** death process with instantaneous rate ($d \rightarrow d-1$): λ_d
  - Easy to simulate, slow if high rate
  - **Closed form** expression available for arbitrary time transition (numerical stability is challenging, faster for **equal time steps**)
- Birth-Death dual: **homogeneous** process jumping from:

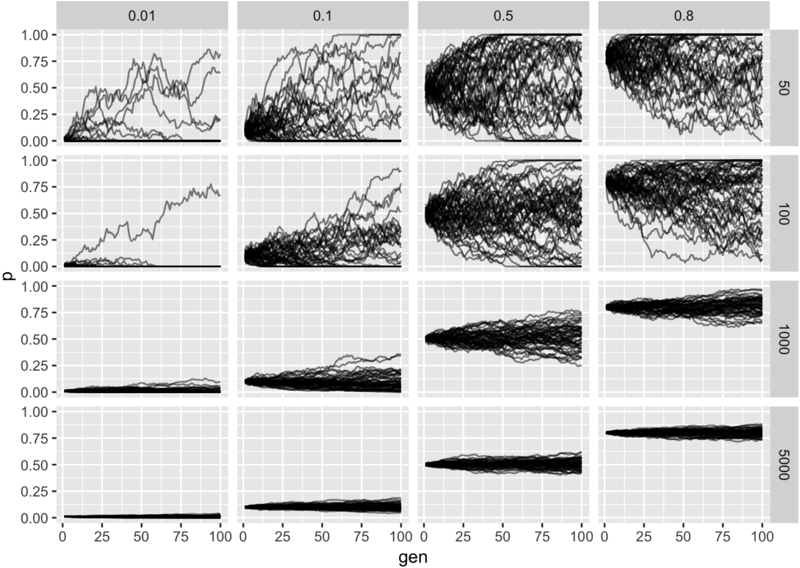
$$d \rightarrow d+1 \text{ with rate } \lambda_d$$
$$d \rightarrow d-1 \text{ with rate } \mu_d$$

  - Easy to simulate, slow if high rate
  - **No closed form** expression for arbitrary time transition.
  - May be approximated by a Wright-Fisher diffusion

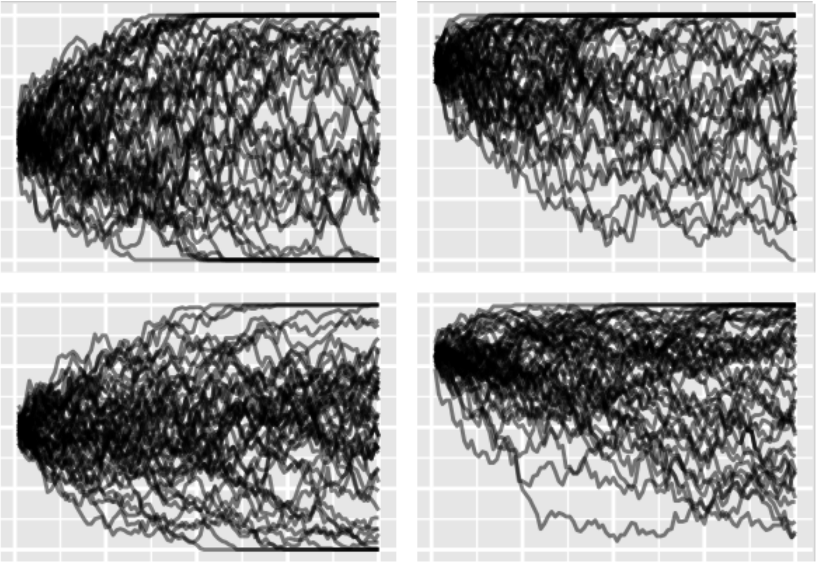

100
500
1000
5000
Predictive density (1st dim)
x_1
Exact
PD
BD Gillespie Moran
BD Gillespie WF
BD diffusion WF
Bootstrap PF


First moment
Signal retrieval
Standard Deviation
Error and CI
particles
PD
BD Gillespie Moran
BD Gillespie WF
BD diffusion WF
Bootstrap PF

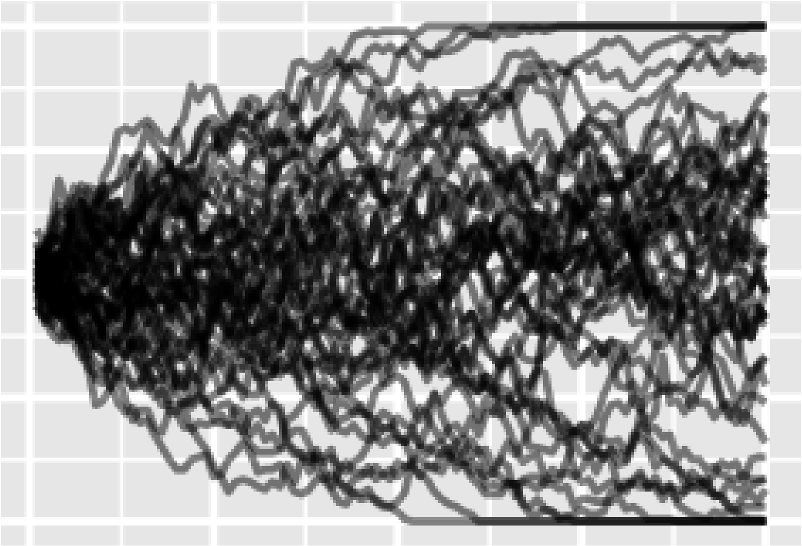
Wright-Fisher simulated data



Wright-Fisher simulated data



Wright-Fisher simulated data



Wright-Fisher simulated data

Summary

Prediction in hidden diffusion processes may be performed using a *dual* process:

- may lead to tractable prediction, filtering, smoothing, inference
- may provide good approximation strategies
- depends on the properties of the chosen dual:
  - easier to simulate than the original process
  - approximation on a different (dual) space

Main motivation: diffusion processes for which only Birth-Death dual exist

- Jacobi diffusion
- Wright-Fisher with selection

Thanks for your attention ! **Want further details ?**
Check out the ISBA-BNP series seminar of April 5th, 2023
`https://bnp-isba.github.io/webinars.html`

contact: guillaume.konkamking@inrae.fr

Boone, E. L., Merrick, J. R., and Krachey, M. J. (2014). A Hellinger distance approach to MCMC diagnostics. *Journal of Statistical Computation and Simulation*, 84(4):833–849.

Jansen, S. and Kurt, N. (2014). On the notion (s) of duality for Markov processes. *Probability surveys*, 11:59–120.

Jenkins, P. A. and Spano, D. (2017). Exact simulation of the Wright–Fisher diffusion. *The Annals of Applied Probability*, 27(3):1478–1509.

Kon Kam King, G., Papaspiliopoulos, O., and Ruggiero, M. (2021). Exact inference for a class of hidden Markov models on general state spaces. *Electronic Journal of Statistics*, 15(1):2832–2875.

Papaspiliopoulos, O. and Ruggiero, M. (2014). Optimal filtering and the dual process. *Bernoulli*, 20(4):1999–2019.

Papaspiliopoulos, O., Ruggiero, M., and Spano, D. (2016). Conjugacy properties of time-evolving Dirichlet and gamma random measures. *Electronic Journal of Statistics*, 10(2):3452–3489.