Piecewise Deterministic Markov Processes for transdimensional sampling from flexible Bayesian survival models

Luke Hardcastle, Samuel Livingstone, Gianluca Baio
Dept. Statistical Science, University College London
luke.hardcastle.20@ucl.ac.uk

11th October 2024
Warwick Algorithms & Computationally Intensive Inference seminars

Talk structure

Part I

Polyhazard models

$$h(y) = \sum_{j=1}^{K} h_{D_j,\theta_j}(y \mid \theta_j, X_j)$$

Part II

Piecewise Exponential models

$$\log h(y) = \sum_{j=1}^{J} \alpha_j \mathbb{1}(y \in [s_{j-1}, s_j))$$

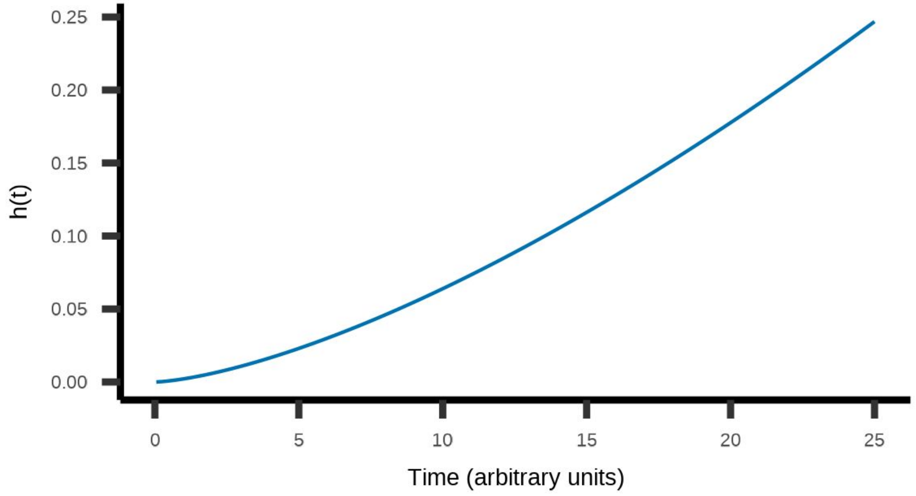
Survival analysis

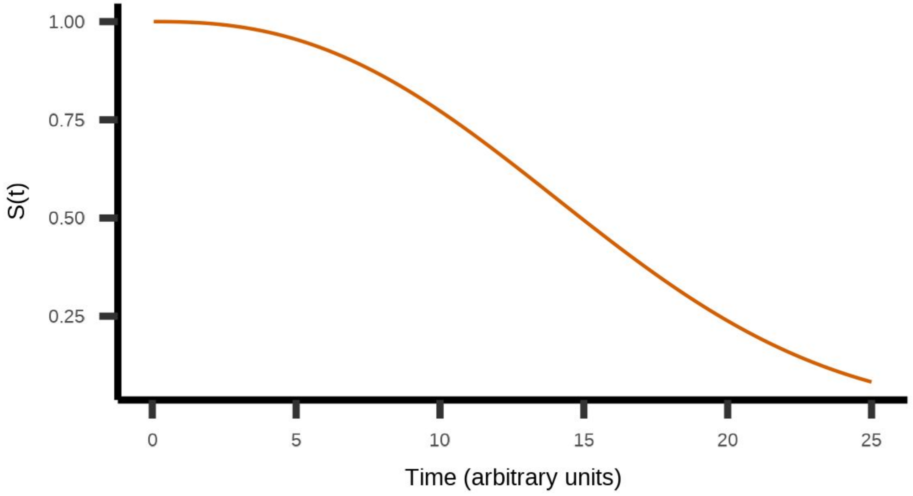
- Want to model the time, *Y*, to some event (e.g cancer progression, death)
- Observe partially right censored, iid data, $\mathcal{D} = (y_i, x_i, c_i)_{i=1}^{n}$
- Parametric likelihood

$$\mathcal{L}(\theta \mid \mathcal{D}) = \prod_i h_\theta(y_i)^{1-c_i} S_\theta(y_i)$$

$$h_\theta(y) = \lim_{\delta \to 0} \frac{\mathbb{P}_\theta(Y \leq y + \delta \mid Y > y)}{\delta}$$

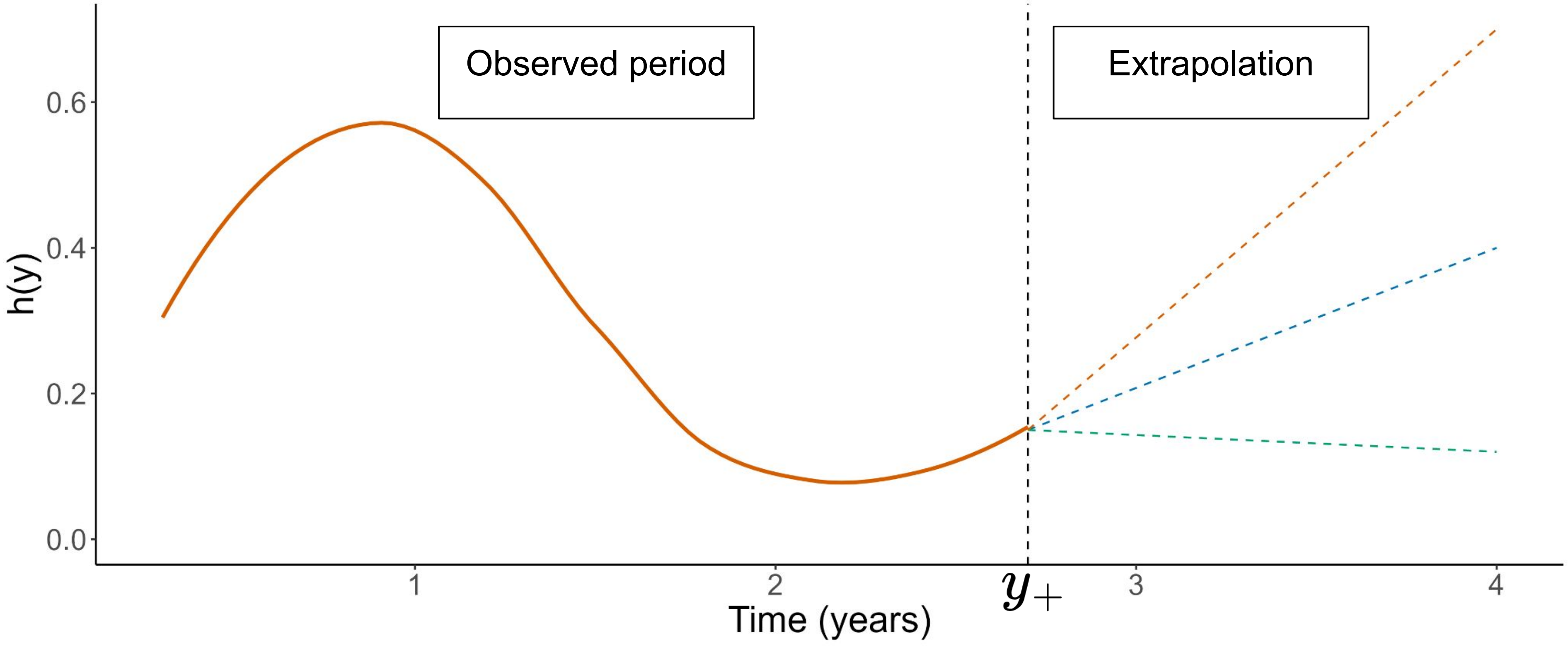
$$S_\theta(y) = \mathbb{P}_\theta(Y > y)$$





Jackson (2017)

Survival extrapolation



Polyhazard models

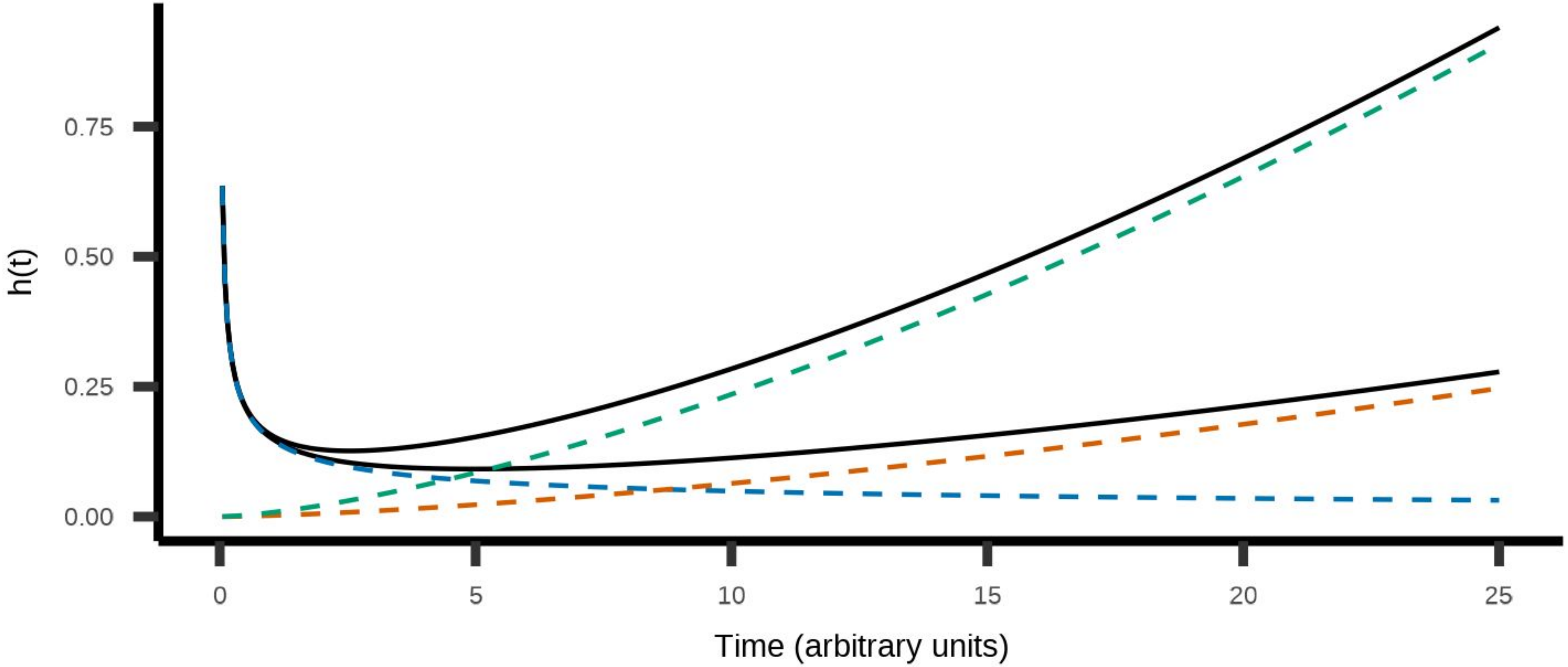
Polyhazard models

Louzada-Neto (1999)
Demiris et al. (2015)
Rutherford et al. (2020)

Idea: Sum simpler parametric hazards $\Longrightarrow$ flexible parametric model

e.g

$$h(y) = h_1(y) + h_2(y) + h_3(y)$$



Extending polyhazard models

Hazard distribution:

$$h(y) = \sum_{j=1}^{K} h_{D_j,\theta_j}(y \mid \theta_j, X_j)$$

Number of hazards:

Variable selection:

Extending polyhazard models

Hazard distribution:

$D_j \sim \text{Uniform}(\mathcal{H})$

$\mathcal{H} = \{\text{Weibull, Log-logistic,...}\}$

$$h(y) = \sum_{j=1}^{K} h_{D_j,\theta_j}(y \mid \theta_j, X_j)$$

Number of hazards:

$K \sim \text{Poisson}(\xi)_{>0}$

Variable selection:

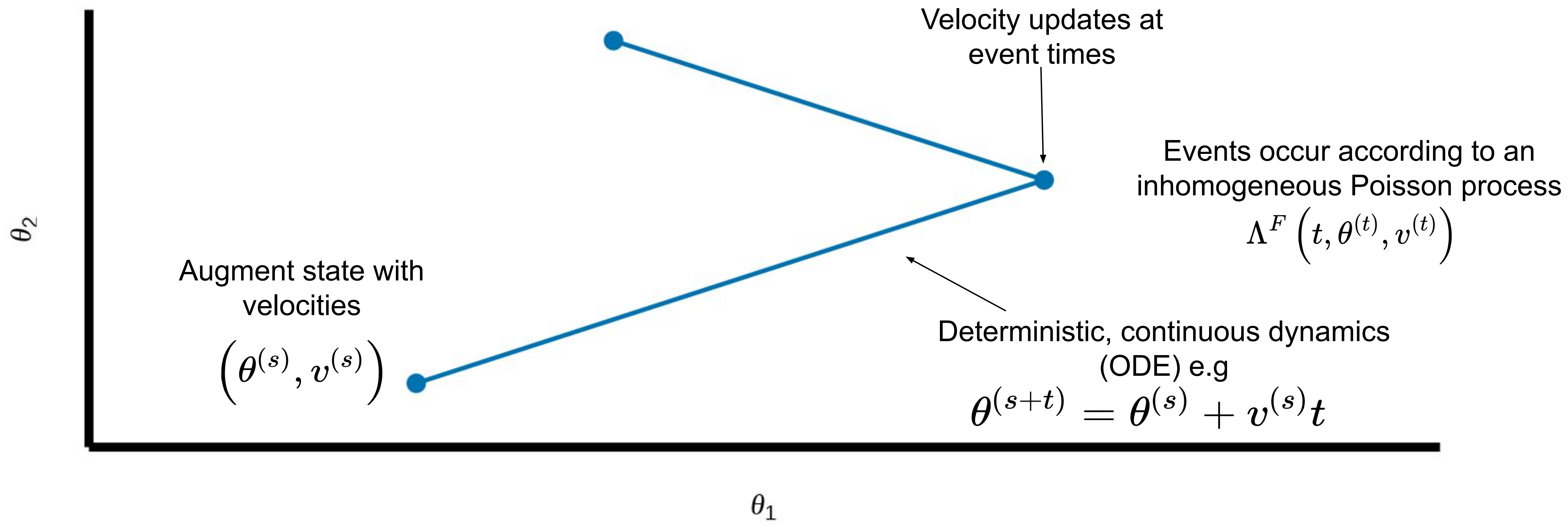
$$\pi_0(d\beta) \propto (1-\omega)p_0(d\beta) + \omega\delta_0(d\beta)$$

Gaussian slab

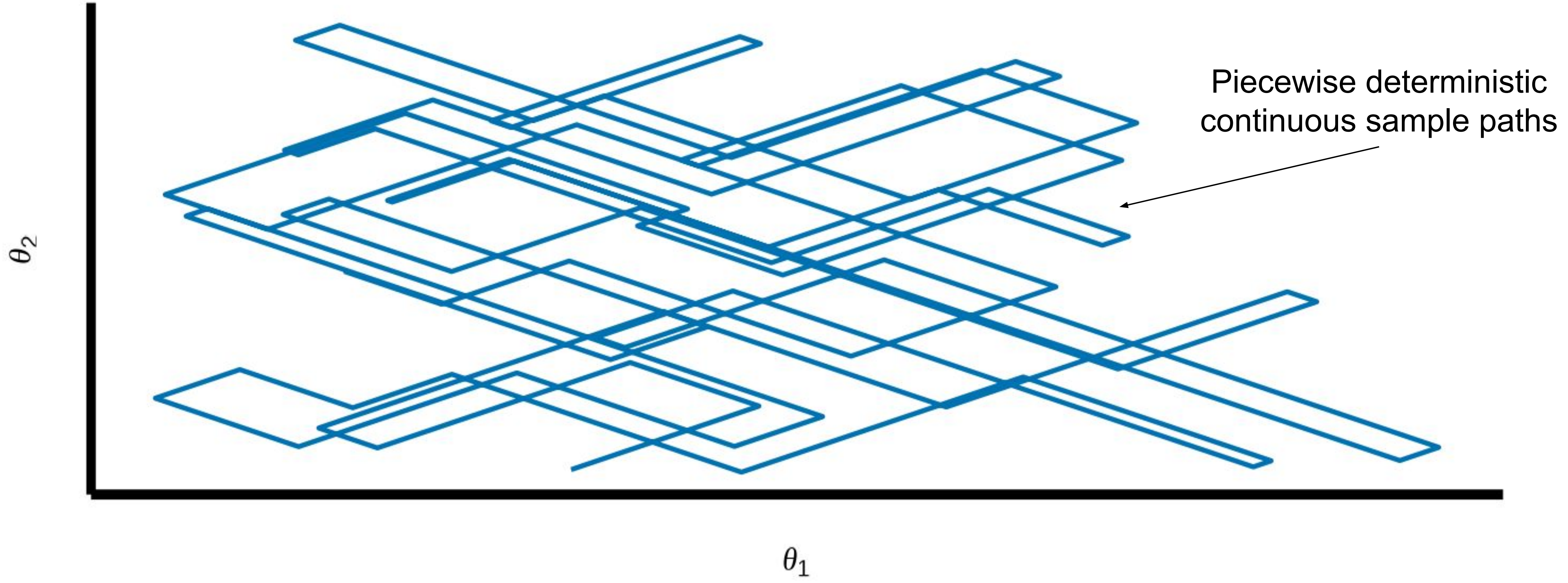
Dirac spike

Fearnhead et al. (2018)

Piecewise Deterministic Markov Processes

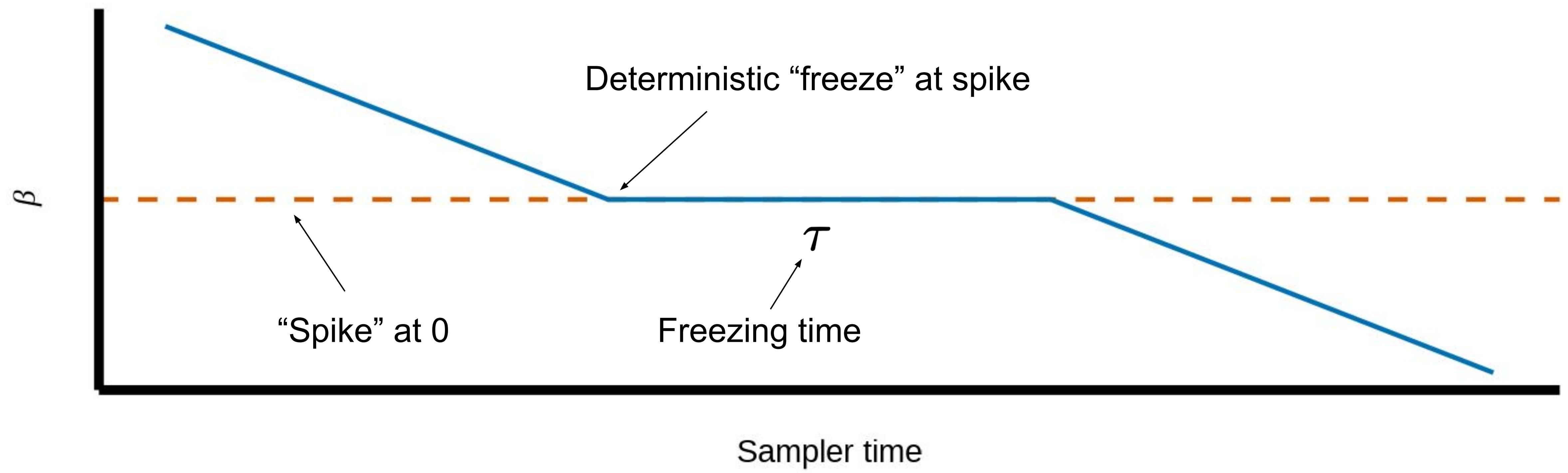


Piecewise Deterministic Markov Processes



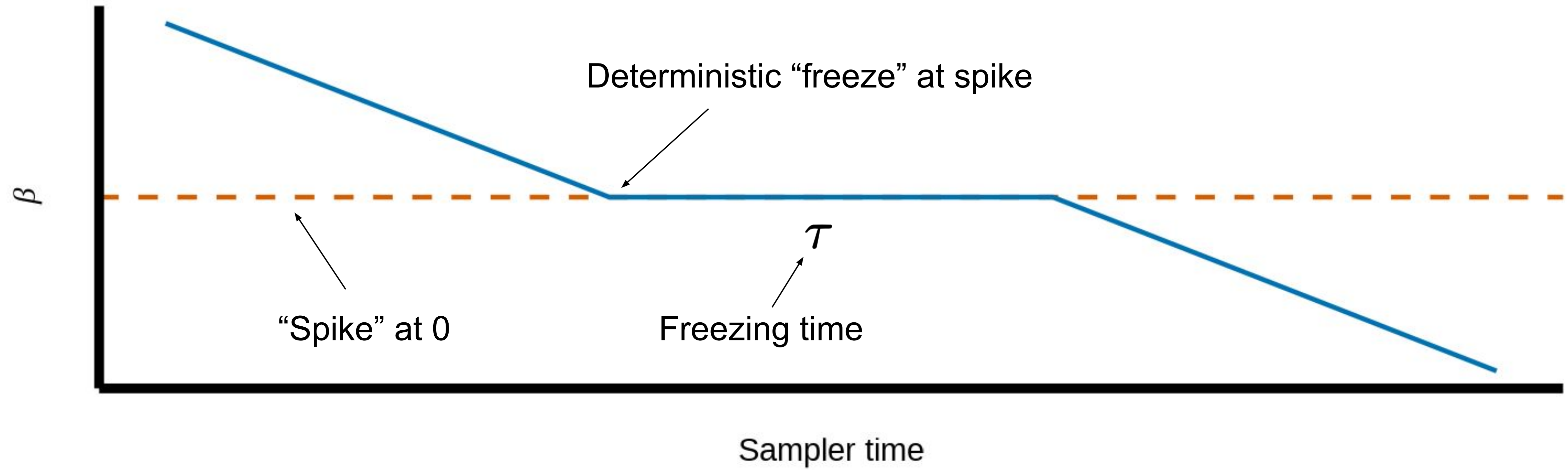
PDMPs for Variable selection

Chevalier et al. (2022)
Bierkens et al. (2023)



Chevalier et al. (2022)
Bierkens et al. (2023)

PDMPs for Variable selection

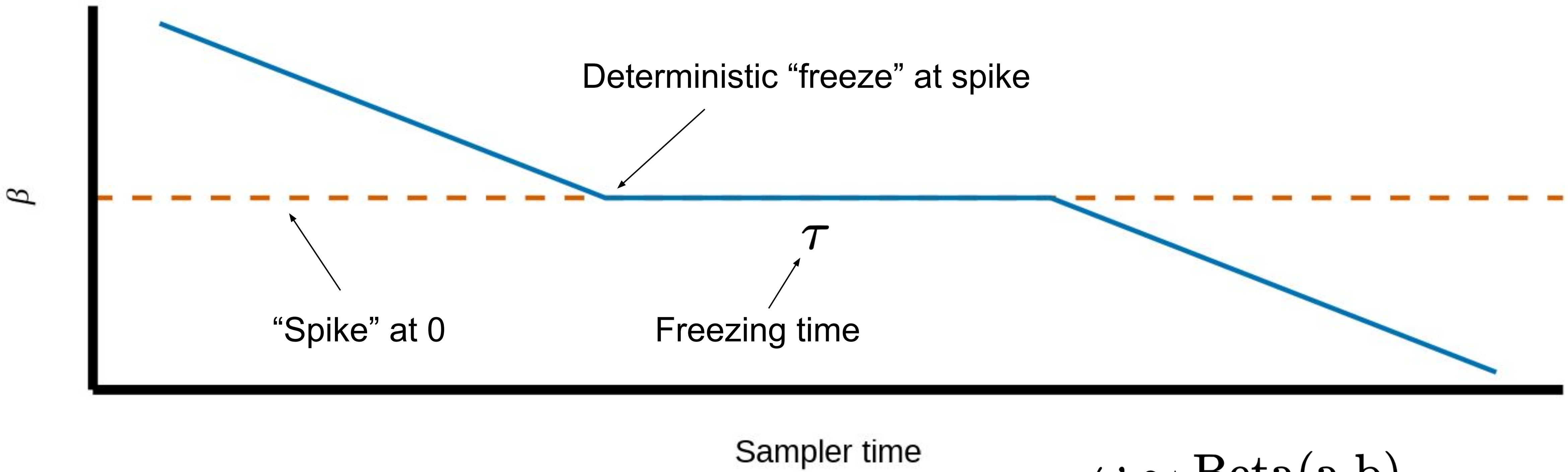


$$\tau \sim \mathrm{Exp}\left(\underbrace{\frac{1-\omega}{\omega}p_0(0)}\right)$$

Posterior ratio - likelihoods cancel by design

Chevalier et al. (2022)
Bierkens et al. (2023)

PDMPs for Variable selection



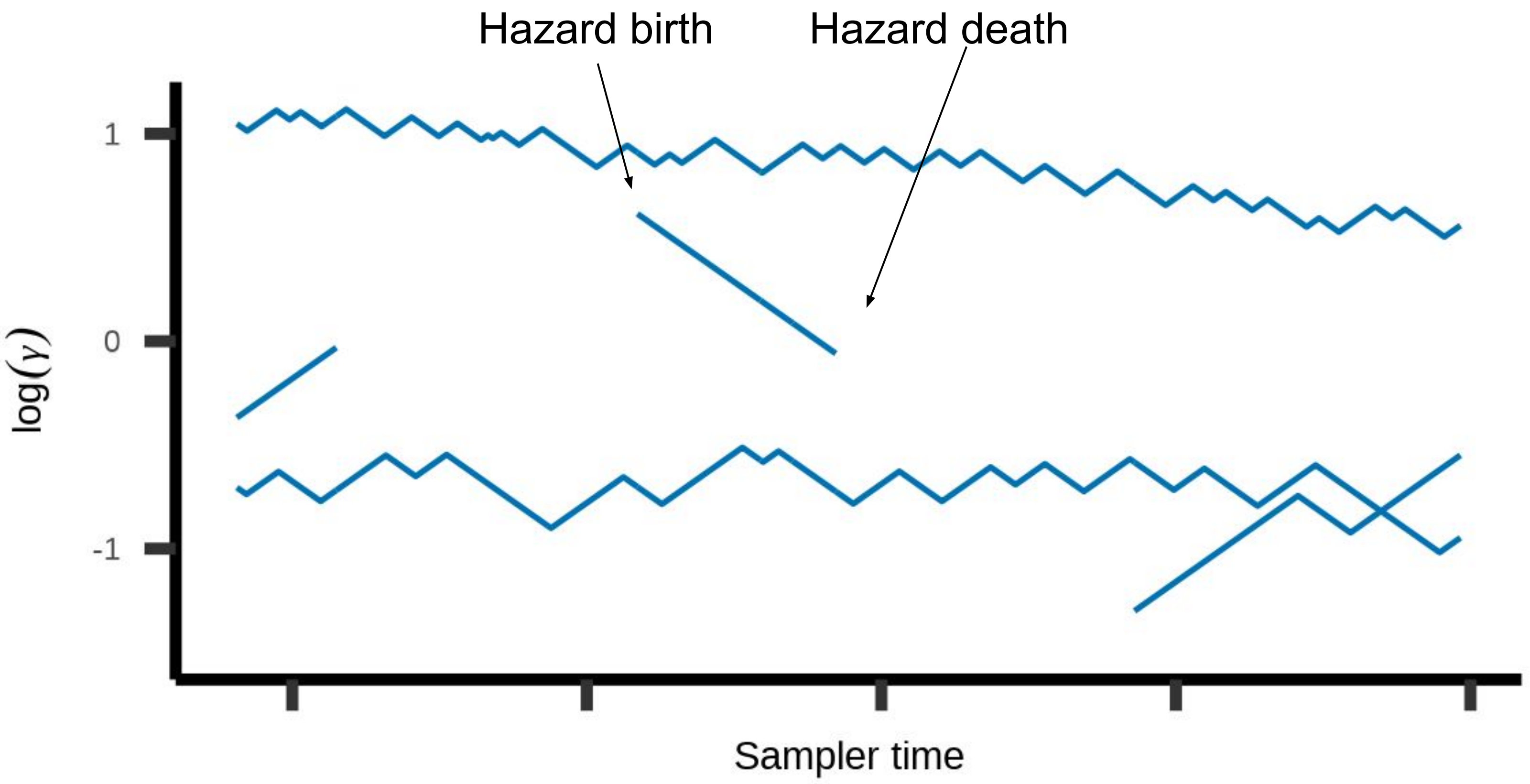
$$\omega \sim \mathrm{Beta(a,b)}$$

$$\tau \sim \mathrm{Exp}\left(\underbrace{\frac{1-\omega}{\omega} p_0(0)}_{}\right)$$

Posterior ratio - likelihoods cancel by design

Weights updated through continuous-time jump process

Birth-death process



- Births and deaths occur in continuous-time.
- Rates can be constructed to be bounded and generated by Poisson thinning.

Stephens (2000)
Zanella (2020)

Birth-death PDMPs

- Select birth and death rates that satisfy a detailed balance condition:

$$\Lambda^B(\theta)\pi(\theta, v)q(u) \; = \; \Lambda^D(\theta')\pi(\theta', v')q(u')$$

Density of proposed innovation

- E.g Traditional birth-death MCMC approach

$$\Lambda^B(\theta) \; = \; \Lambda_0 \qquad\qquad \Lambda^D(\theta') \; = \; \Lambda_0 \frac{\pi(\theta, v)\, q(u)}{\pi(\theta', v')\, q(u')}$$

Very hard to bound

Stephens (2000)
Zanella (2020)

Birth-death PDMPs

- Select birth and death rates that satisfy a detailed balance condition:

$$\Lambda^B(\theta)\pi(\theta, v)q(u) = \Lambda^D(\theta')\pi(\theta', v')q(u')$$

Density of proposed innovation

- E.g Traditional birth-death MCMC approach

$$\Lambda^B(\theta) = \Lambda_0 \qquad \Lambda^D(\theta') = \Lambda_0 \frac{\pi(\theta, v)\, q(u)}{\pi(\theta', v')\, q(u')}$$

Very hard to bound

- Instead use bounded balancing functions:

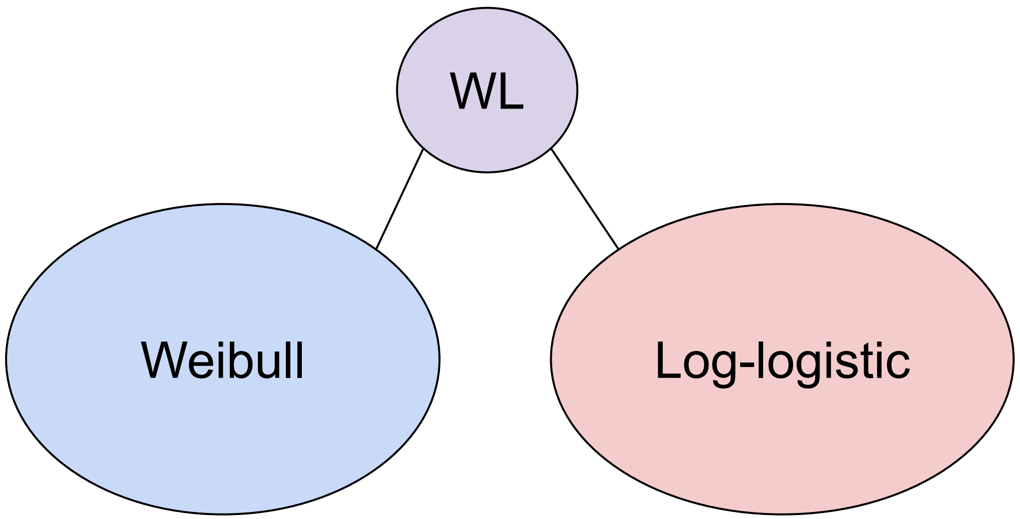
$$g(z) = zg(z^{-1}) \qquad \begin{aligned} \Lambda^B(\theta) &= g(z) \\ \Lambda^D(\theta') &= g(z^{-1}) \end{aligned} \qquad z = \frac{\pi(\theta', v')\, q(u')}{\pi(\theta, v)\, q(u)}$$

- E.g

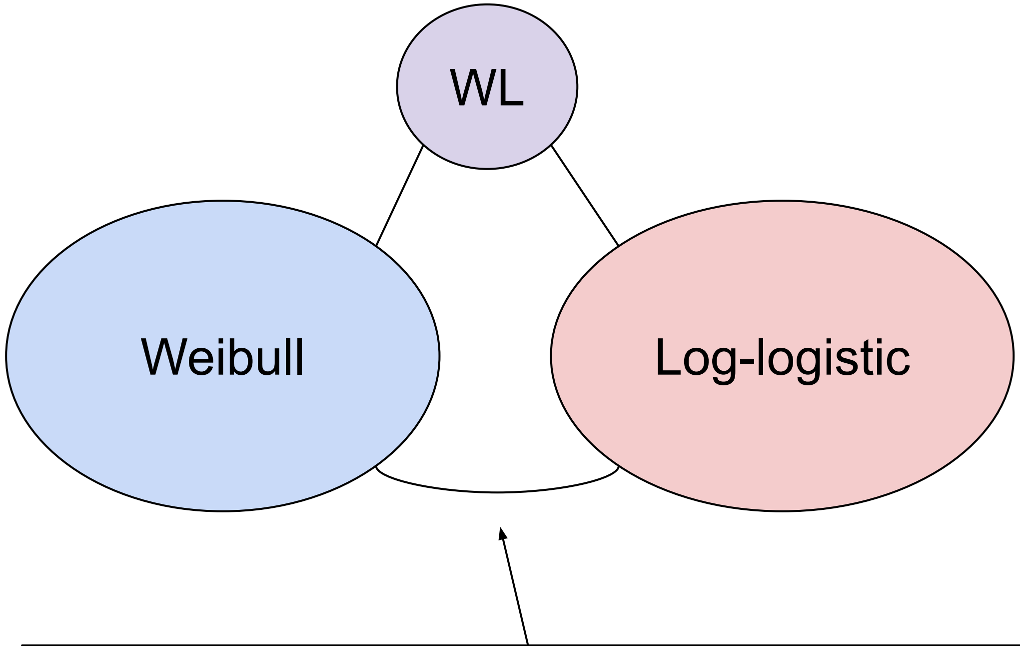
$$g_M(z) = \min(1,\, z) \qquad g_B(z) = \frac{z}{1+z}$$

Bounded - can be generated via Poisson thinning

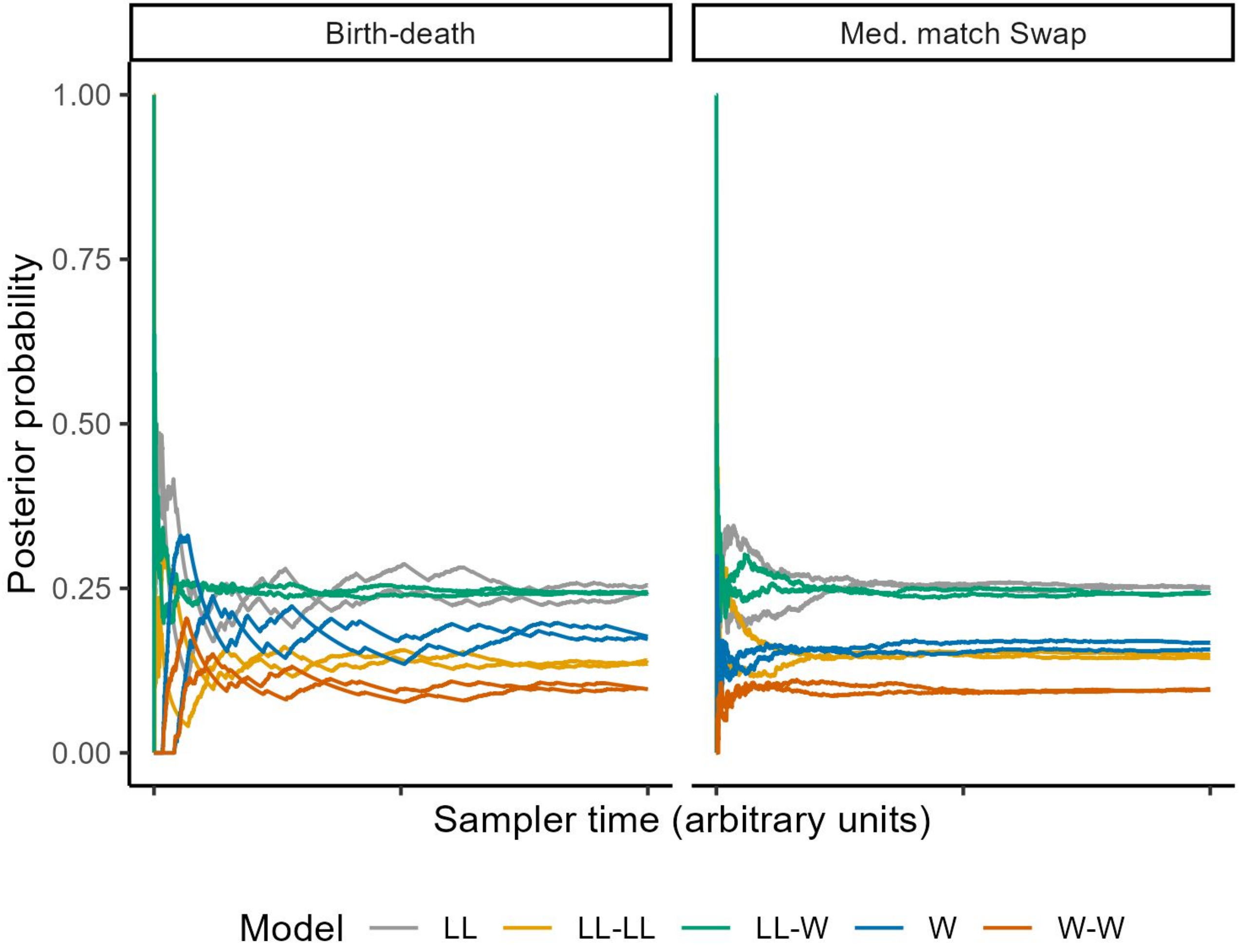
Birth-death-swap process



Birth-death-swap process

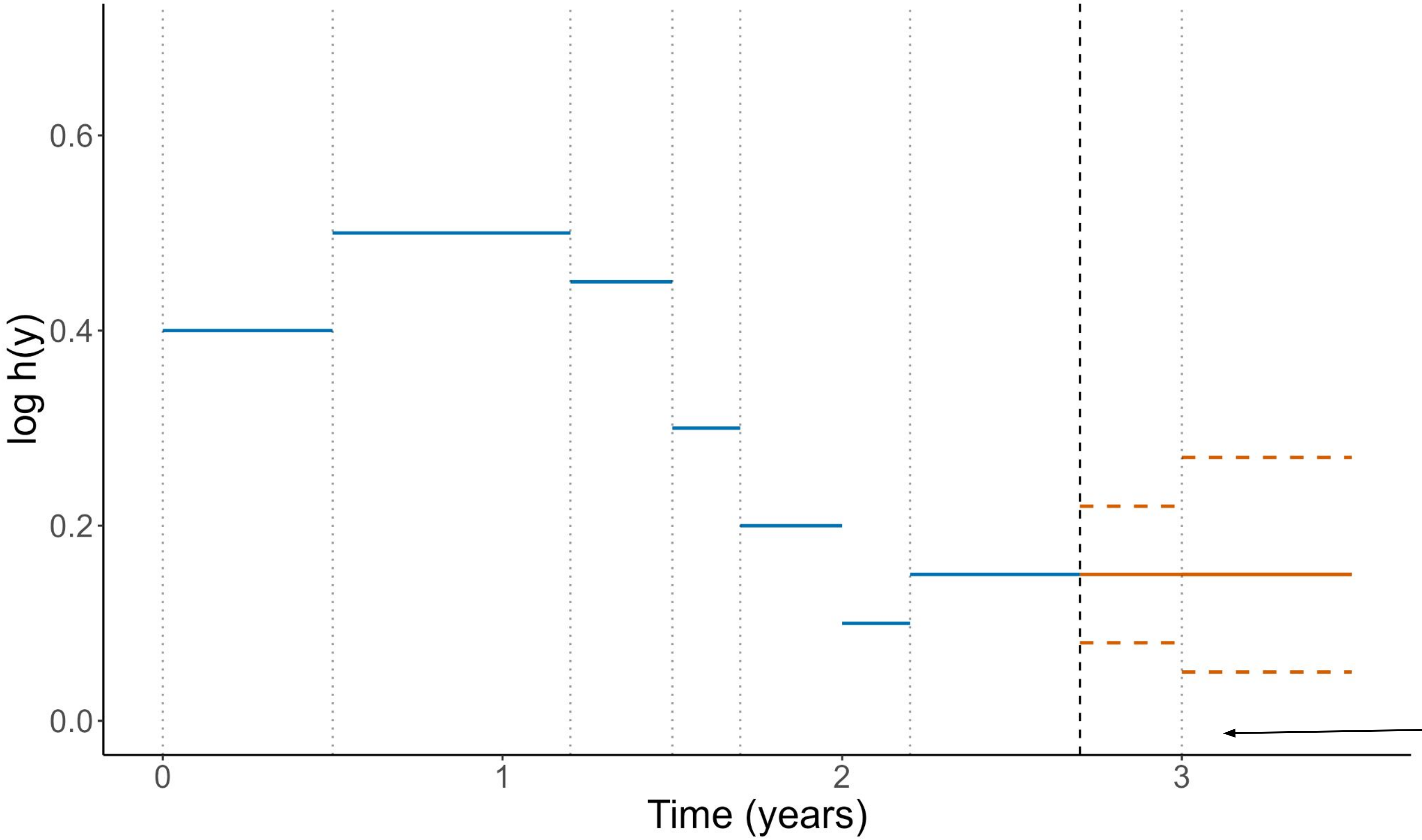


Distributions can be efficiently swapped in the sampler by matching their medians.



Piecewise Exponential models

Piecewise exponential models



Local log-hazard

Knots

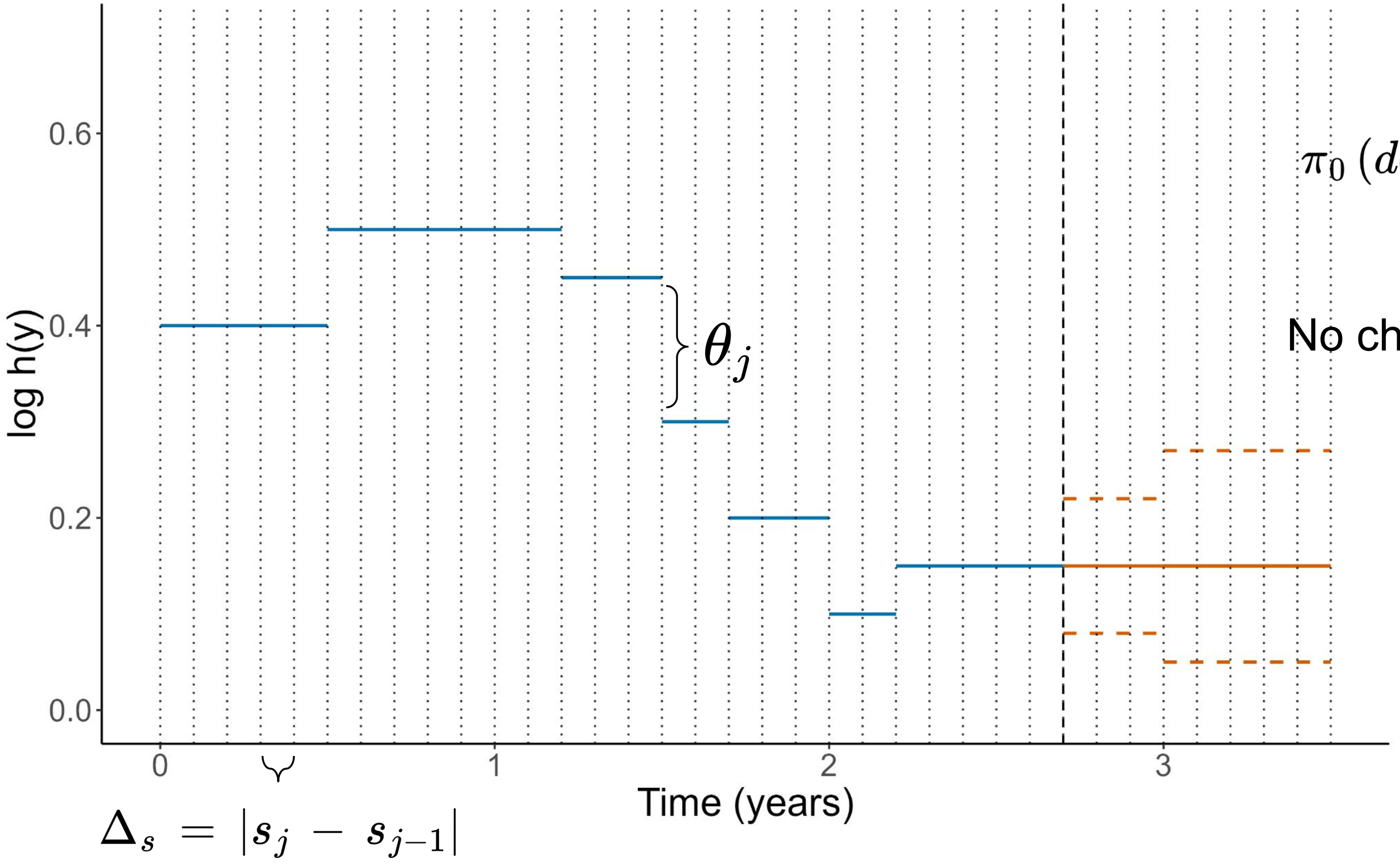
$$\log h(y) = \sum_{j=1}^{J} \alpha_j \mathbb{1}\left(y \in [s_{j-1}, s_j)\right)$$

Prior dependency for extrapolation, e.g

$$\alpha_j \sim \mathrm{Normal}\left(\alpha_{j-1}, \sigma^2\right)$$

Extrapolations very sensitive to choice of knot location

Piecewise exponential models



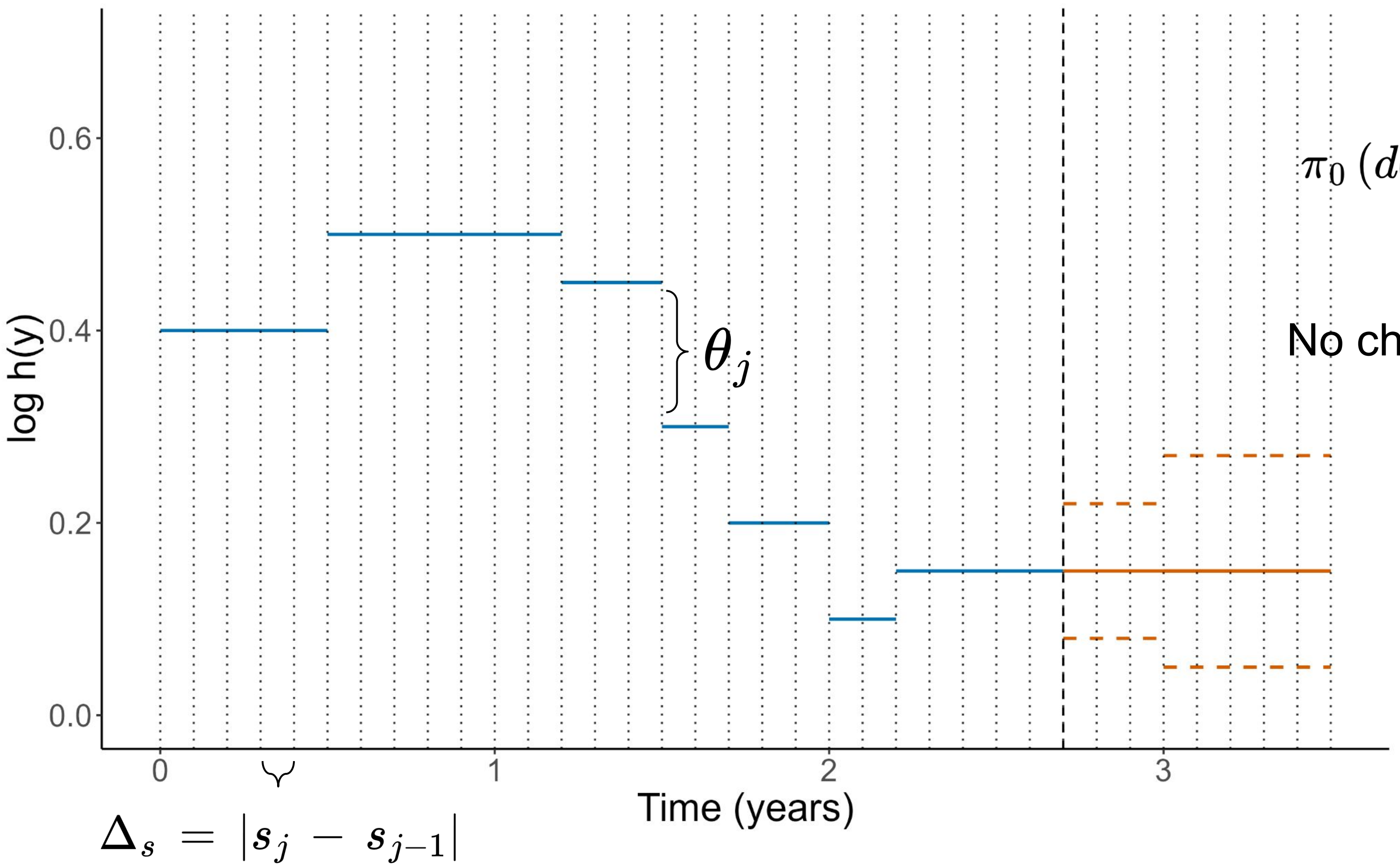
$$\theta_j = \alpha_j - \alpha_{j-1}$$

$$\pi_0(d\theta_j) \propto \omega\delta_0(d\theta_j) + (1-\omega)p_0(d\theta_j \mid \sigma)$$

No change in hazard

Change in hazard

Piecewise exponential models



$$\theta_j = \alpha_j - \alpha_{j-1}$$

$$\pi_0(d\theta_j) \propto \omega\delta_0(d\theta_j) + (1-\omega)p_0(d\theta_j \mid \sigma)$$

No change in hazard

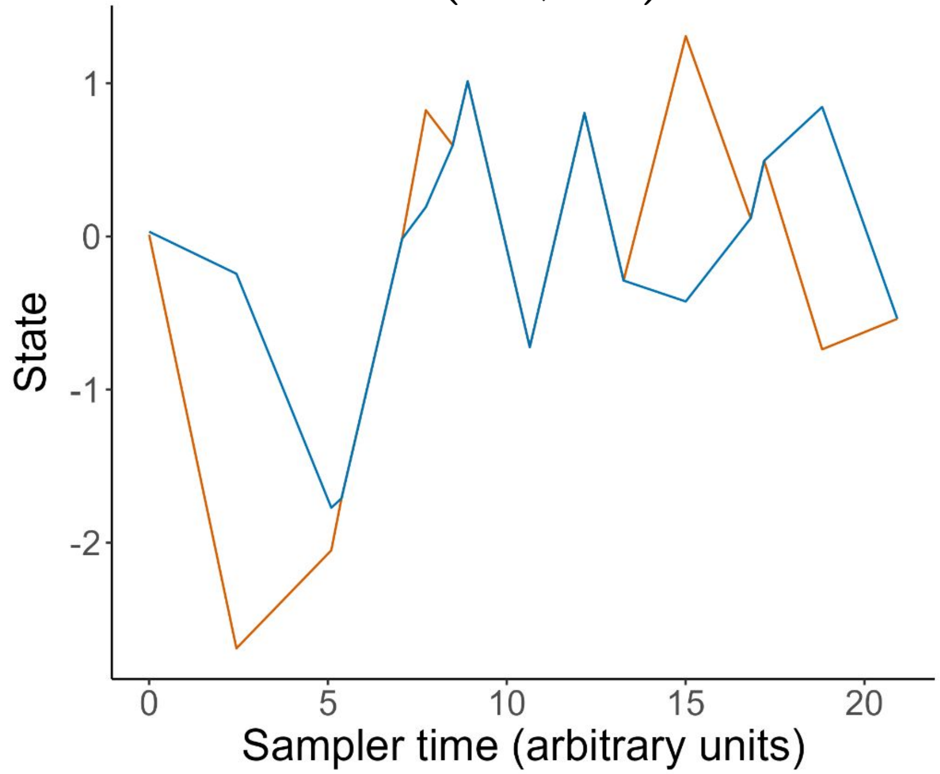
Change in hazard

Main idea: Can use a PDMP with sticky dynamics to sample on the distances between neighbouring hazards

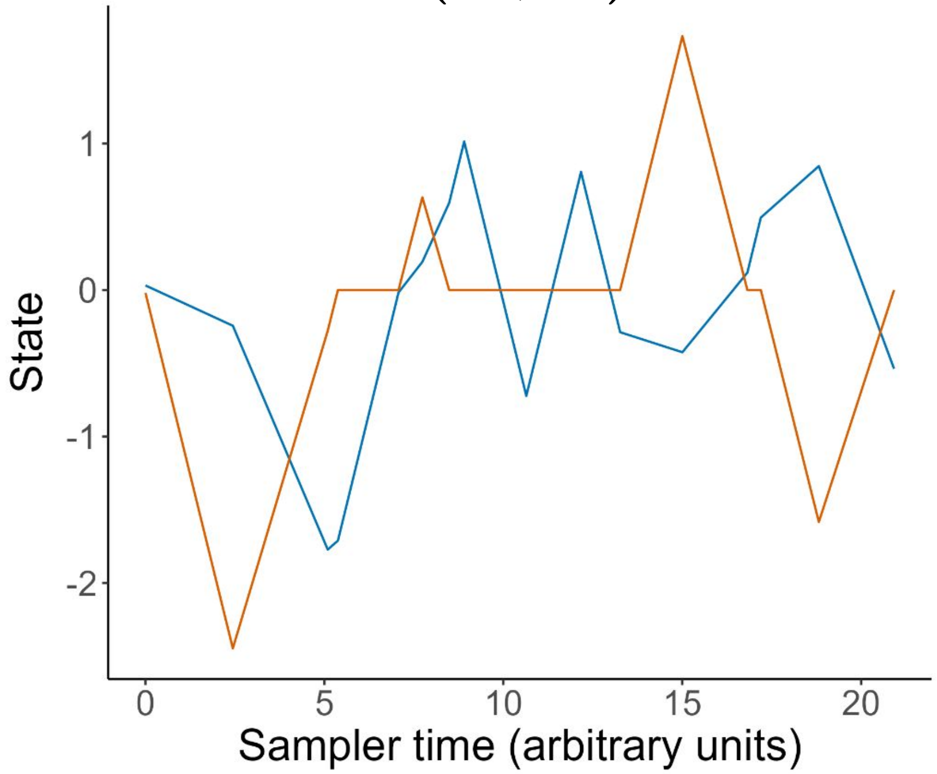
Prior on (ω, σ) allows extrapolations to be informed by observed volatility

Split-merge PDMPs

(α_1, α_2)



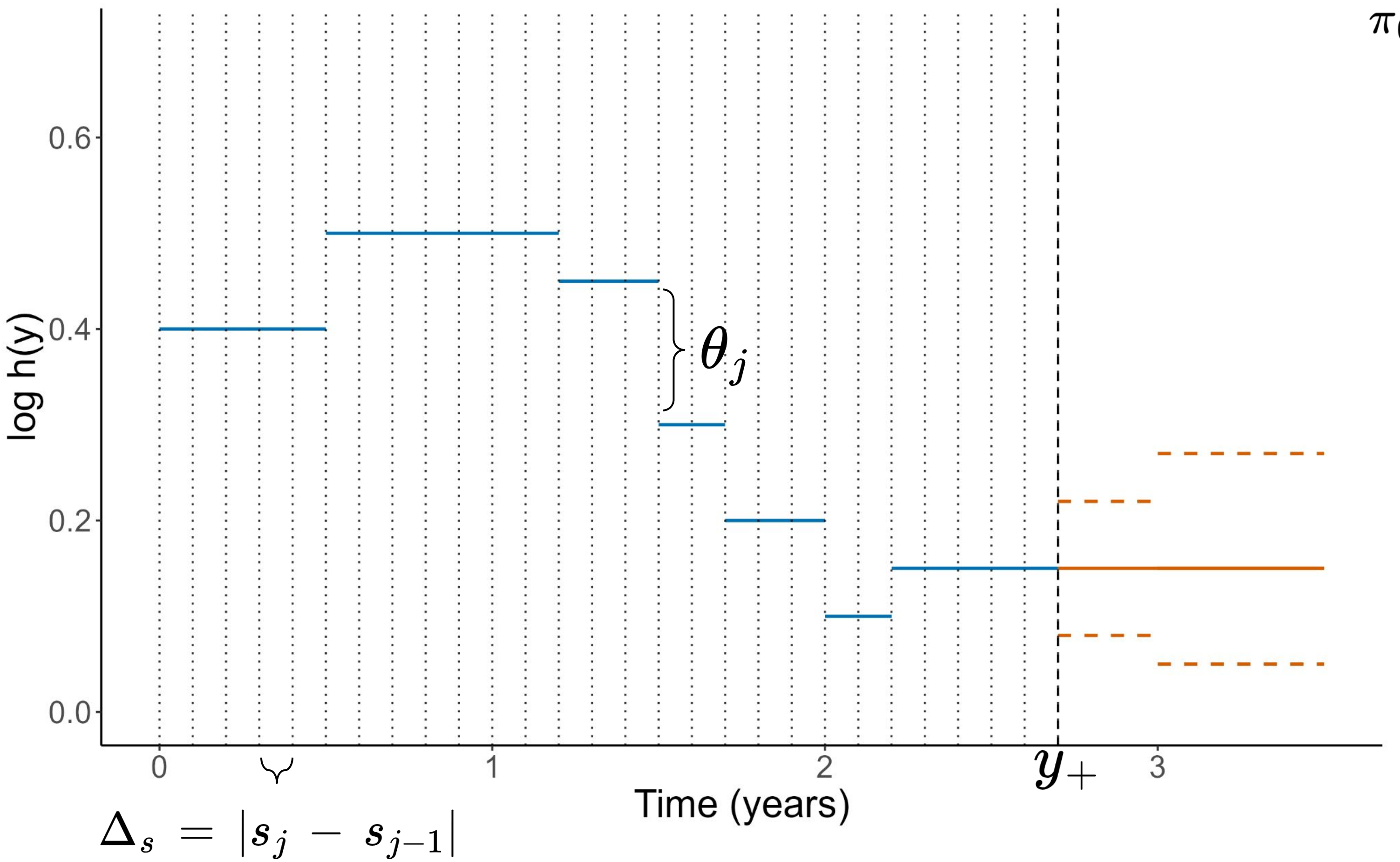
(θ_1, θ_2)



$$\sum_{j=1}^{J} \alpha_j \mathbb{1}\left(y \in [s_{j-1}, s_j)\right)$$

$$\theta_j = \alpha_j - \alpha_{j-1}$$

Piecewise exponential models



$$\pi_0\left(d\theta_j\right) \propto (1-\omega)\delta_0\left(d\theta_j\right) + \omega p_0\left(d\theta_j \mid \sigma\right)$$

- Still some sensitivity to specification of knot locations.
- Consider the limit:

$$\Delta_s, \, \omega \to 0,$$

- Implies the prior:

$$\{s_j\}_{j=1}^{J} \sim PP\left(\gamma, \, (0, y_+)\right)$$

- Hyperprior structure:

$$\gamma = \omega\Gamma,$$
$$\text{Fixed}, \; \Gamma > \gamma$$
$$\omega \sim Beta\left(a, b\right)$$

Sticky PDMPs with infinite knot locations

Challenge: Infinite number of candidate knots mean unsticking rate cannot be directly computed

$$\{s_j\}_{j=1}^{J} \sim PP\left(\gamma,\, (0, y_+)\right)$$
$$\gamma = \omega\Gamma,$$
$$\omega \sim Beta\left(a, b\right)$$

Sticky PDMPs with infinite knot locations

Challenge: Infinite number of candidate knots mean unsticking rate cannot be directly computed

$$\{s_j\}_{j=1}^{J} \cup \{r_k\}_{k=1}^{K} = \{m_l\}_{l=1}^{K+J} \sim PP\left(\Gamma, (0, y_+)\right)$$

Unthinned

Thinned

$$\{s_j\}_{j=1}^{J} \sim PP\left(\gamma,\, (0, y_+)\right)$$
$$\gamma = \omega\Gamma,$$
$$\omega \sim Beta\left(a, b\right)$$

Sticky PDMPs with infinite knot locations

Challenge: Infinite number of candidate knots mean unsticking rate cannot be directly computed

$$\{s_j\}_{j=1}^{J} \cup \{r_k\}_{k=1}^{K} = \{m_l\}_{l=1}^{K+J} \sim PP\left(\Gamma, (0, y_+)\right)$$

Unthinned → $\{s_j\}_{j=1}^{J}$

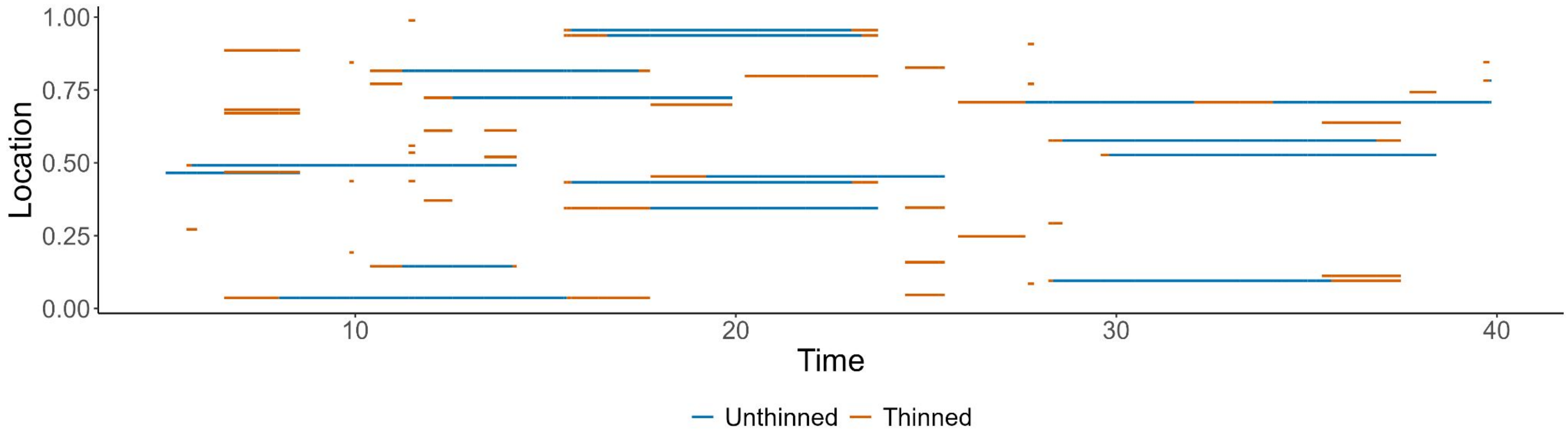
Thinned → $\{r_k\}_{k=1}^{K}$

1. Sample $\pi\left(\theta,\, \{m_l\}_{l=1}^{K+J} \mid \omega, \sigma, K+J, \mathcal{D}\right)$ using sticky PDMP dynamics
2. Sample $K \sim \text{Poisson}\big((1-\omega)\Gamma y_+\big)$
3. Sample $\{r_k\}_{k=1}^{K} \overset{iid}{\sim} \text{Uniform}\big((0, y_+)\big)$

(Steps 2–3: Update thinned points)

$$\{s_j\}_{j=1}^{J} \sim PP\left(\gamma,\, (0, y_+)\right)$$
$$\gamma = \omega\Gamma,$$
$$\omega \sim Beta\left(a, b\right)$$

Infinite knot locations



1. Sample $\pi\left(\theta,\, \{m_l\}_{l=1}^{K+J} \mid \omega, \sigma, K+J, \mathcal{D}\right)$ using sticky PDMP dynamics

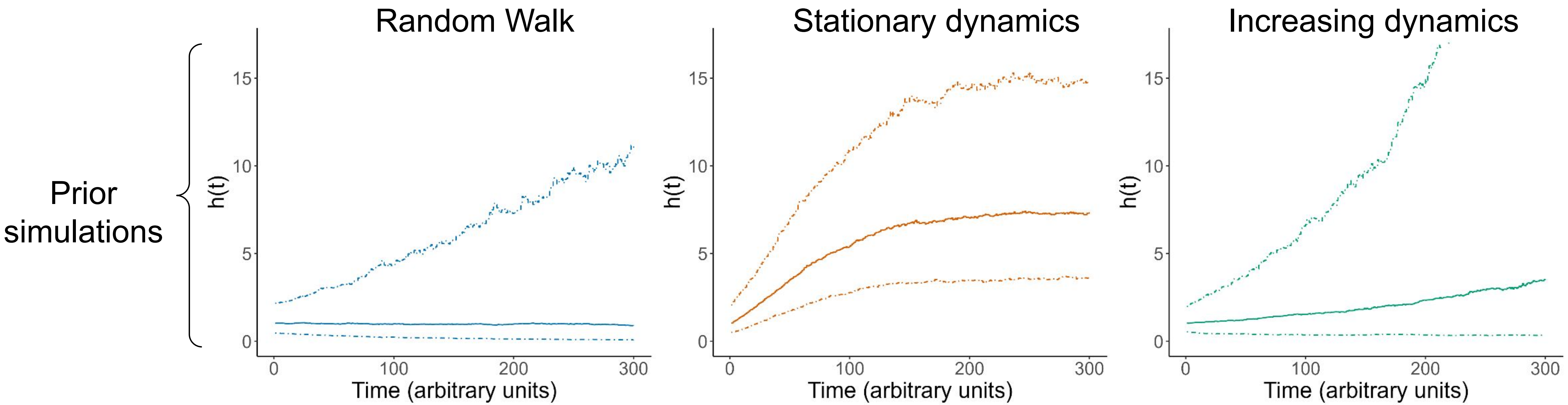
2. Sample $K \sim \text{Poisson}\big((1-\omega)\Gamma y_+\big)$

3. Sample $\{r_k\}_{k=1}^{K} \overset{iid}{\sim} \text{Uniform}\big((0, y_+)\big)$

Update thinned points (steps 2–3)

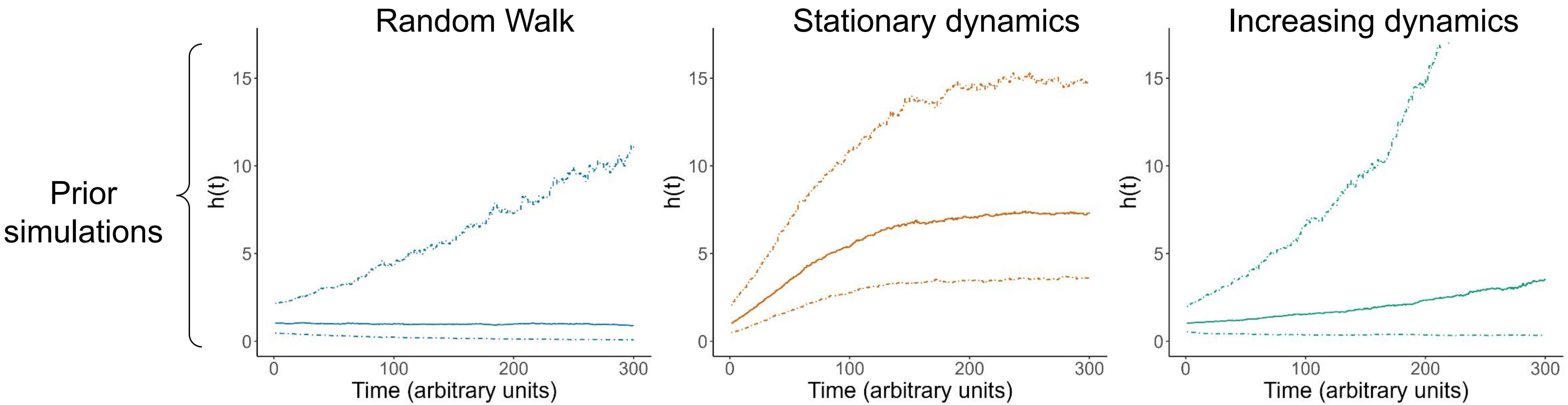
Aalen & Gjessing (2004)
Roberts & Sangalli (2010)

Discretised diffusion priors



Aalen & Gjessing (2004)
Roberts & Sangalli (2010)

Discretised diffusion priors



Assume *a priori* that local log-hazards $\{\alpha_j\}_{j=1}^{J}$ are discretisations of an SDE:

$$d\alpha_t = \mu(\alpha_t)dt + \sigma(\alpha_t)dW_t$$

Drift coefficient

Volatility coefficient, assume: $\sigma(\alpha_t) = 1$

Livingstone et al. (2024)

Discretisation schemes

Standard choice for discretising a diffusion, the ***Euler-Maruyama*** discretisation,

$$\alpha_j = \alpha_{j-1} + \sigma^2 \mu\left(\alpha_{j-1}\right) + \theta_j \qquad \theta_j \sim \text{Normal}\left(0, \sigma^2\right)$$

Step-size interpreted in relation to the drift - difficult to elicit priors.

Lipschitz condition for drift coefficient

✗ $\mu\left(\alpha_j\right) = a - b\exp\left(\alpha_{j-1}\right)$

Livingstone et al. (2024)

Discretisation schemes

Standard choice for discretising a diffusion, the ***Euler-Maruyama*** discretisation,

$$\alpha_j = \alpha_{j-1} + \sigma^2 \mu\left(\alpha_{j-1}\right) + \theta_j \qquad \theta_j \sim \mathrm{Normal}\left(0, \sigma^2\right)$$

Step-size interpreted in relation to the drift - difficult to elicit priors.

Lipschitz condition for drift coefficient

❌ $\mu\left(\alpha_j\right) = a - b \exp\left(\alpha_{j-1}\right)$

Alternatively can use ***Barker*** discretisation:

$$\alpha_j = \alpha_{j-1} + \theta_j$$

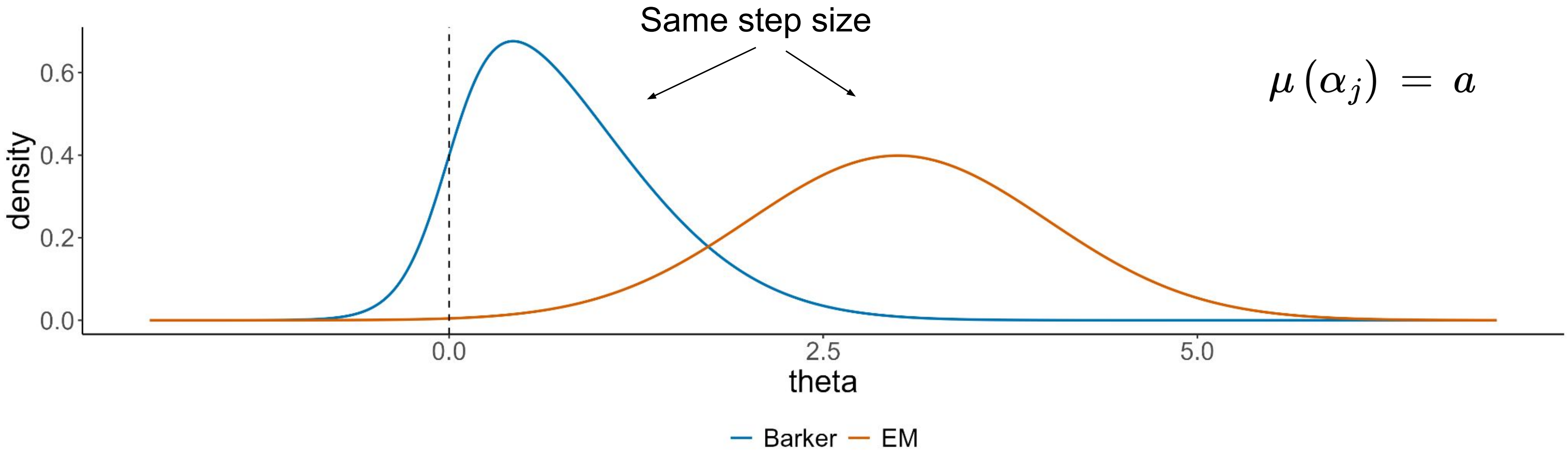
$$\pi_0\left(\theta_j\right) \propto \sigma \underbrace{\left(1 + \tanh\left[\mu\left(\alpha_{j-1}\right)\theta_j\right]\right)}_{\text{Skewing term}} \phi\left(\theta_j\right)$$

Need not be Lipschitz

Corresponds directly to expected jump size

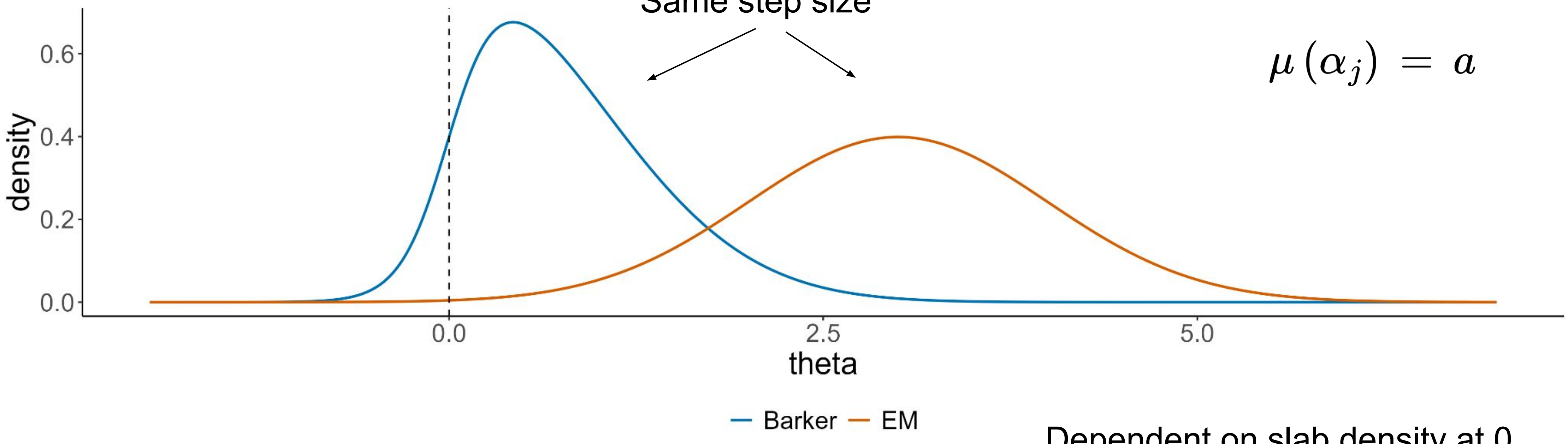
Std. Normal density

Discretisation schemes



$$\mu\left(\alpha_j\right) = a$$

Discretisation schemes



$$\mu\left(\alpha_j\right) = a$$

Dependent on slab density at 0

Recall unsticking rate given by

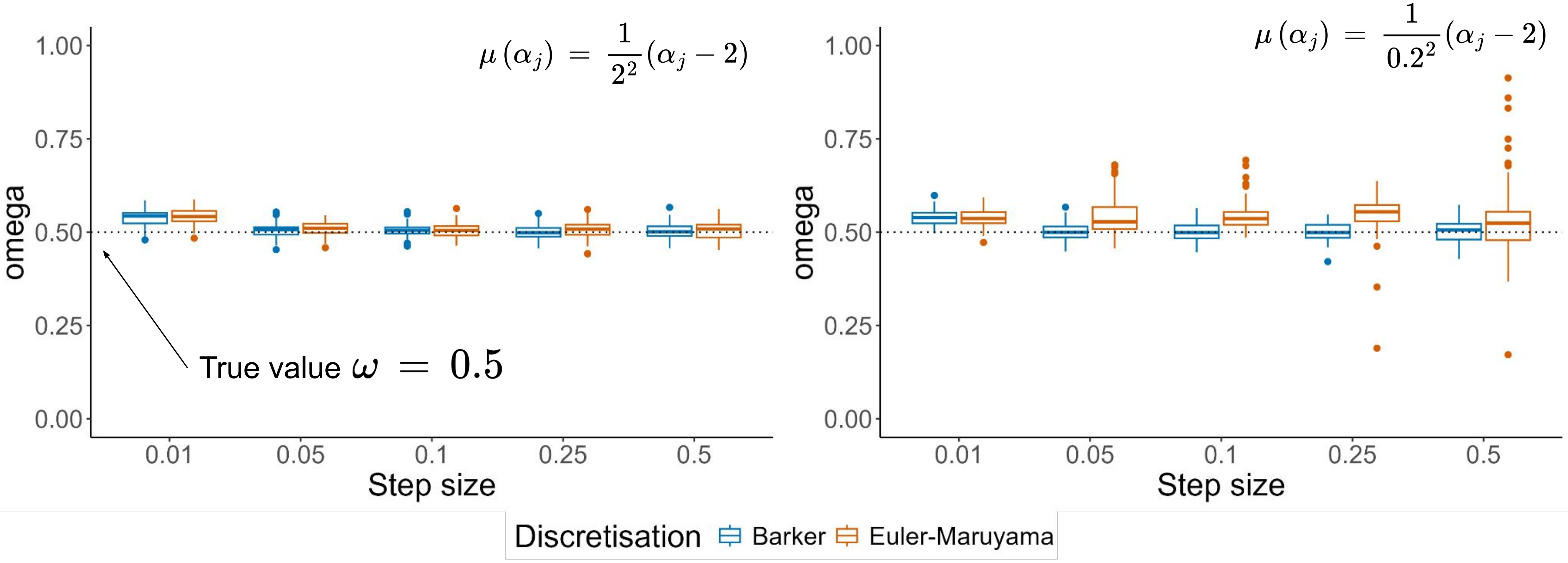
$$\Lambda\left(t\right) = \frac{\omega}{1-\omega} p_0(0 \mid \theta_t, \sigma)$$

$$p_0^B\left(0 \mid \theta_t, \sigma\right) = p_0^B\left(0 \mid \sigma\right) \geq p_0^{EM}\left(0 \mid \theta_t, \sigma\right) \quad \Longrightarrow$$

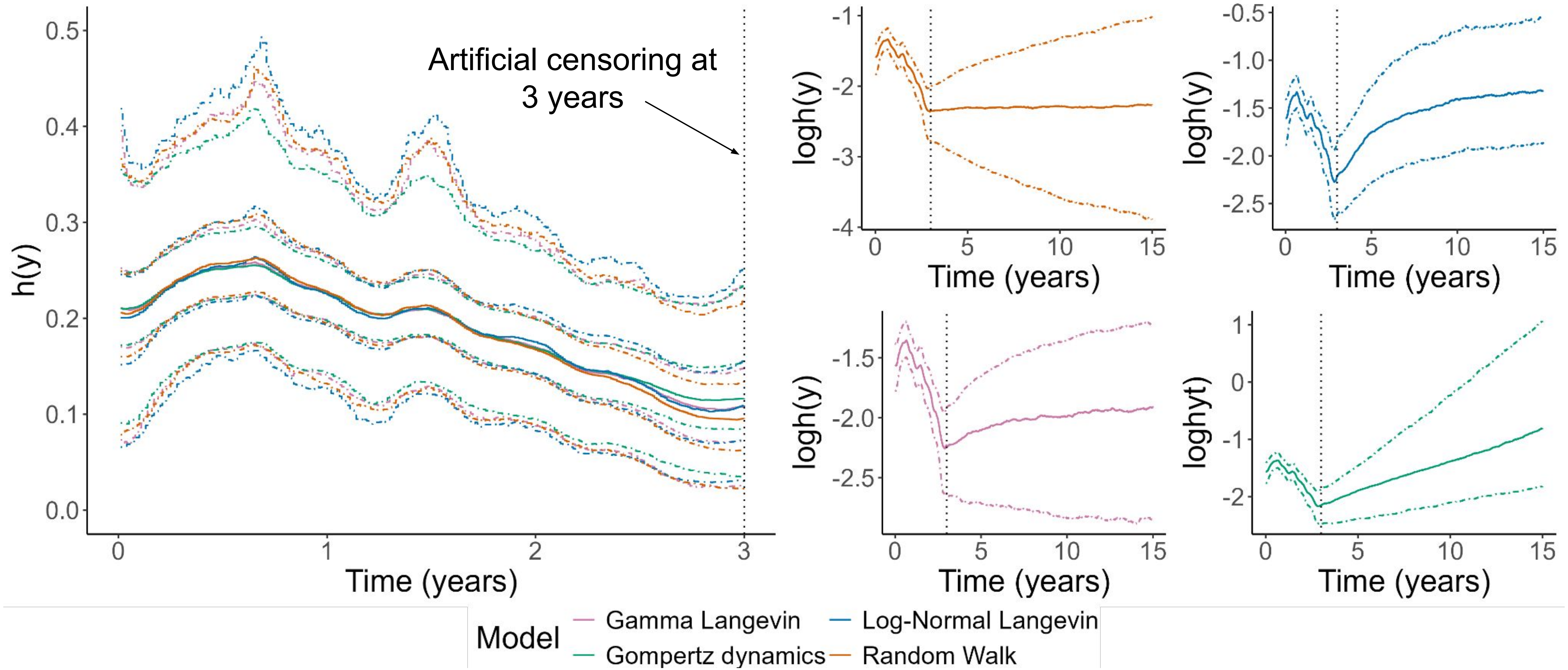
We can expect faster mixing under the Barker parameterisation

Discretisation schemes

Sampling from prior with for fixed number of PDMP iterations and set values for σ



Application to Colon cancer data



References

Bierkens, J., Grazzi, S., Meulen, F. V. D., & Schauer, M. (2023). Sticky PDMP samplers for sparse and local inference problems. *Statistics and Computing*, *33*(1), 8.

Chevallier, A., Fearnhead, P., & Sutton, M. (2023). Reversible jump PDMP samplers for variable selection. *Journal of the American Statistical Association*, *118*(544), 2915-2927.

Fearnhead, P., Bierkens, J., Pollock, M., & Roberts, G. O. (2018). Piecewise Deterministic Markov Processes for Continuous-Time Monte Carlo. *Statistical Science*, *33*(3), 386-412.

Jackson, C., Stevens, J., Ren, S., Latimer, N., Bojke, L., Manca, A., & Sharples, L. (2017). Extrapolating survival from randomized trials using external data: a review of methods. *Medical decision making*, *37*(4), 377-390.

Livingstone, S., Nüsken, N., Vasdekis, G., & Zhang, R. Y. (2024). Skew-symmetric schemes for stochastic differential equations with non-Lipschitz drift: an unadjusted Barker algorithm. *arXiv preprint arXiv:2405.14373*.

Louzada-Neto, F. (1999). Polyhazard models for lifetime data. *Biometrics*, *55*(4), 1281-1285.

Roberts, G. O., & Sangalli, L. M. (2010). Latent diffusion models for survival analysis. *Bernoulli*, *16*(2), 435-458.

Rutherford, M. J., Lambert, P. C., Sweeting, M. J., Pennington, R., Crowther, M. J., & Abrams, K. R. (2020). NICE DSU technical support document 21: flexible methods for survival analysis. *Decision Support Unit, ScHARR, University of Sheffield*.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of statistics*, 40-74.

Zanella, G. (2020). Informed proposals for local MCMC in discrete spaces. *Journal of the American Statistical Association*, *115*(530), 852-865.

Discussion

PDMP samplers are useful tools for transdimensional sampling problems in Bayesian inference.

$$h(y) = \sum_{j=1}^{K} h_{D_j,\theta_j}(y \mid \theta_j, X_j)$$

- Extended polyhazard models to account for structural uncertainty through Bayesian model averaging.
- Incorporated birth-death processes with existing PDMP samplers + developed new event generation methods.

Hardcastle, L., Livingstone, S., Baio, G. ***Averaging polyhazard models using Piecewise deterministic Monte Carlo with applications to data with long-term survivors.*** *2024, arxiv:2406.14182 (Under review)*

$$\log h(y) = \sum_{j=1}^{J} \alpha_j \mathbb{1}(y \in [s_{j-1}, s_j))$$

- Extended existing samplers based on sticky dynamics to determine knot locations.
- Introduced new prior structure based on Barker discretisation of underlying diffusion.

Hardcastle, L., Livingstone, S., Baio, G. ***Piecewise exponential models, Piecewise Deterministic Markov Processes and survival extrapolation.*** *2024+, Coming Soon!*