“Bayesian Inference for Hawkes Processes”, Jakob Gulddahl Rasmussen, Methodology and Computing in Applied Probability, Springer, 2011, Link.

Bayesian inference in Hawkes Processes

Lucas Matheus Silva Pereira. July/2023

Hawkes process

Self-exciting marked point process

Events at random time instants

Events have types

Events spawn other processes

Conditional Intensity Function (CIF)

$$\lambda^*(t) = \frac{\mathbb{E}\left[N(dt)|\{(t_i, k_i)\}_{t_i<t}\right]}{dt}$$

- Average number of points in infinitesimal neighborhood of t
  - Considers all previous points and their marks
- Mark distribution:

$$\gamma^*(\kappa|t) = \gamma\left(\kappa|t, \{(t_i, k_i)\}_{t_i<t}\right)$$

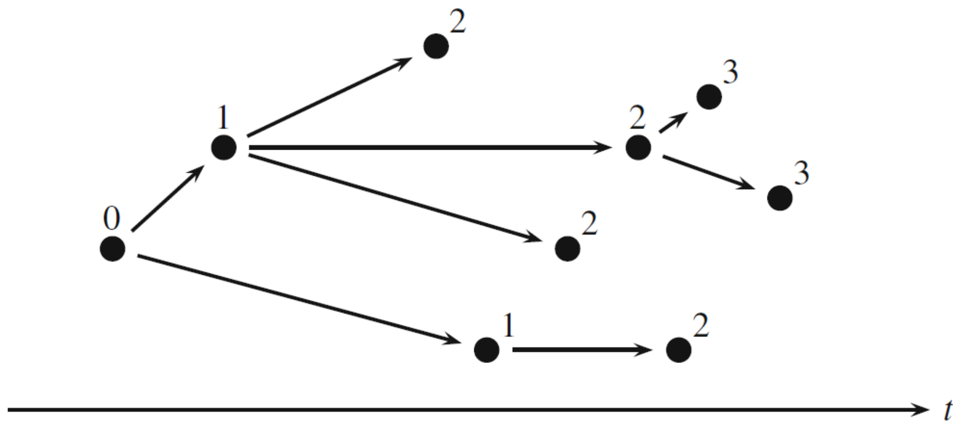
- Depends on:
  - Current instant
  - Past instant-mark pairs

Conditional Intensity Function (CIF)

- Specific CIF for Hawkes process:

$$\lambda^*(t) = \mu(t) + \sum_{t_i<t} \alpha(\kappa_i)\beta(t - t_i, \kappa_i)\,, t \in \mathbb{R}$$

- All **parameterized** and non-negative:
  - Immigrant intensity: $\mu(t) \in \mathbb{R}$
  - Total offspring intensity: $\alpha(\kappa) \in \mathbb{M}$
  - Normalized offspring intensity: $\beta(t, \kappa) \in \mathbb{R}$
  - Mark density: $\gamma^*(\kappa|t) \in \mathbb{R}$



Poisson Cluster Process

- Immigrants:
  - Follow Poisson process
  - Have marks
  - Create clusters of offspring
- Offspring create own marked Poisson process

Methods – Bayesian Inference

- Bayesian inference

$$\overbrace{p(\boldsymbol{\theta} \mid \boldsymbol{X})}^{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{X} \mid \boldsymbol{\theta})}^{\text{Likelihood}} \cdot \overbrace{p(\boldsymbol{\theta})}^{\text{Prior}}}{\underbrace{\int p(\boldsymbol{X} \mid \boldsymbol{\theta}^{*}) \cdot p(\boldsymbol{\theta}^{*})\, d\boldsymbol{\theta}^{*}}_{\text{Evidence}}}$$

- Denominator is computationally challenging

Methods - Markov Chain Monte Carlo

- MCMC to avoid cost
- Metropolis-within-Gibbs:

$\boxed{\text{Estimate } p(\boldsymbol{\theta} \mid \boldsymbol{X})}$ $\xrightarrow{\text{Gibbs}}$

$$p(\theta_1 \mid \theta_2, \dots, \theta_k, \boldsymbol{X})$$
$$p(\theta_2 \mid \theta_1, \dots, \theta_k, \boldsymbol{X})$$
$$\dots$$
$$p(\theta_k \mid \theta_1, \dots, \theta_{k-1}, \boldsymbol{X})$$

Metropolis-Hastings

Methods - CIF

- Dataset x with n points: $(t,\kappa)_n$
- Posterior:

$$p(\mu,\alpha,\beta,\gamma \mid x) \propto p(x \mid \mu,\alpha,\beta,\gamma) \cdot p(\mu,\alpha,\beta,\gamma)$$

- Likelihood:

$$p(x \mid \mu,\alpha,\beta,\gamma) = \left(\prod_{i=1}^{n} \lambda^*(t_i)\gamma^*(\kappa_i \mid t_i)\right) \exp\left(-\Lambda^*(T)\right)$$

$$\Lambda^*(T) = \int_0^t \lambda^*(s)ds = \int_0^t \mu(s)ds + \sum_{t_i<t} \alpha(\kappa_i)\, B(t-t_i,\kappa_i)$$

Methods – Poisson cluster

- Estimate both $(\mu_k, \alpha_k, \beta_k, \gamma_k)$ and branching structure Y
- Multiple marked Poisson processes: immigrants and their marked offspring
- Likelihood:

$$p(x \mid y, \mu, \alpha, \beta, \gamma) = P(\boldsymbol{I} \mid y, \mu, \gamma)\left(\prod_{j=1}^{n} p\left(O_j \mid y, \alpha, \beta, \gamma\right)\right)$$

- Conditioned on the branching structure
- Independence between processes

Example 1 - unmarked

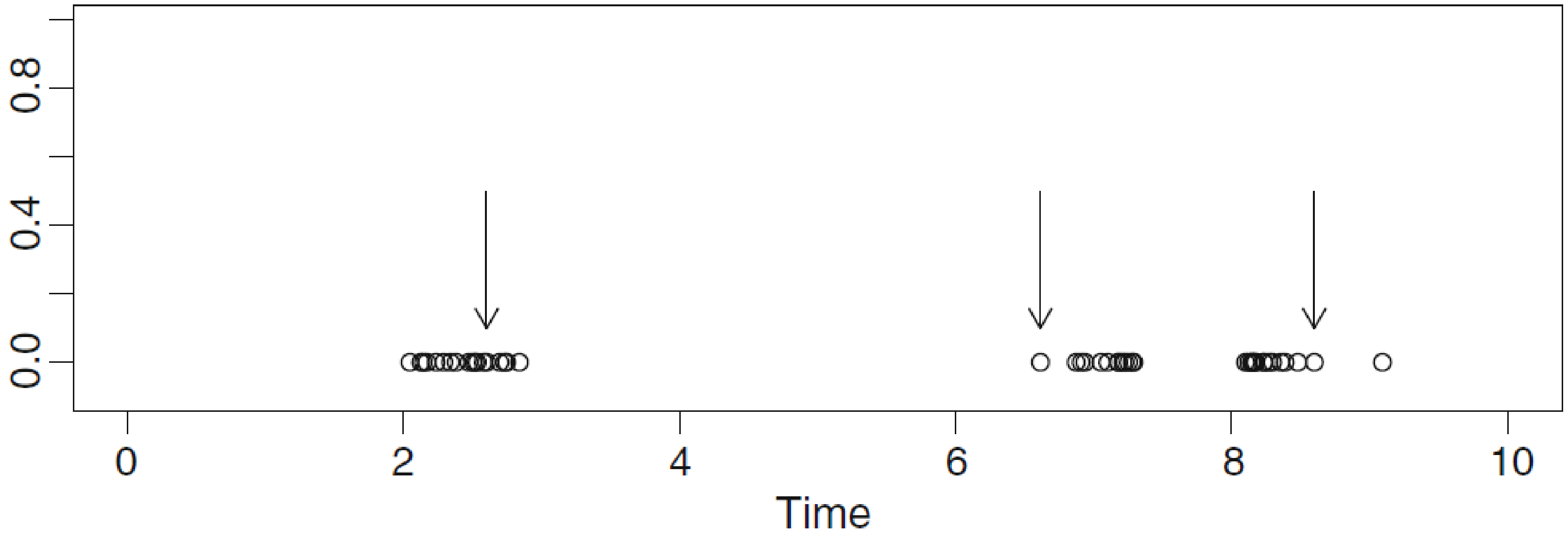
- Formulation:

$$\mu(t) = \mu_1 \mathbf{1}(t \geq 0)$$

$$\alpha(\kappa) = \alpha_1$$

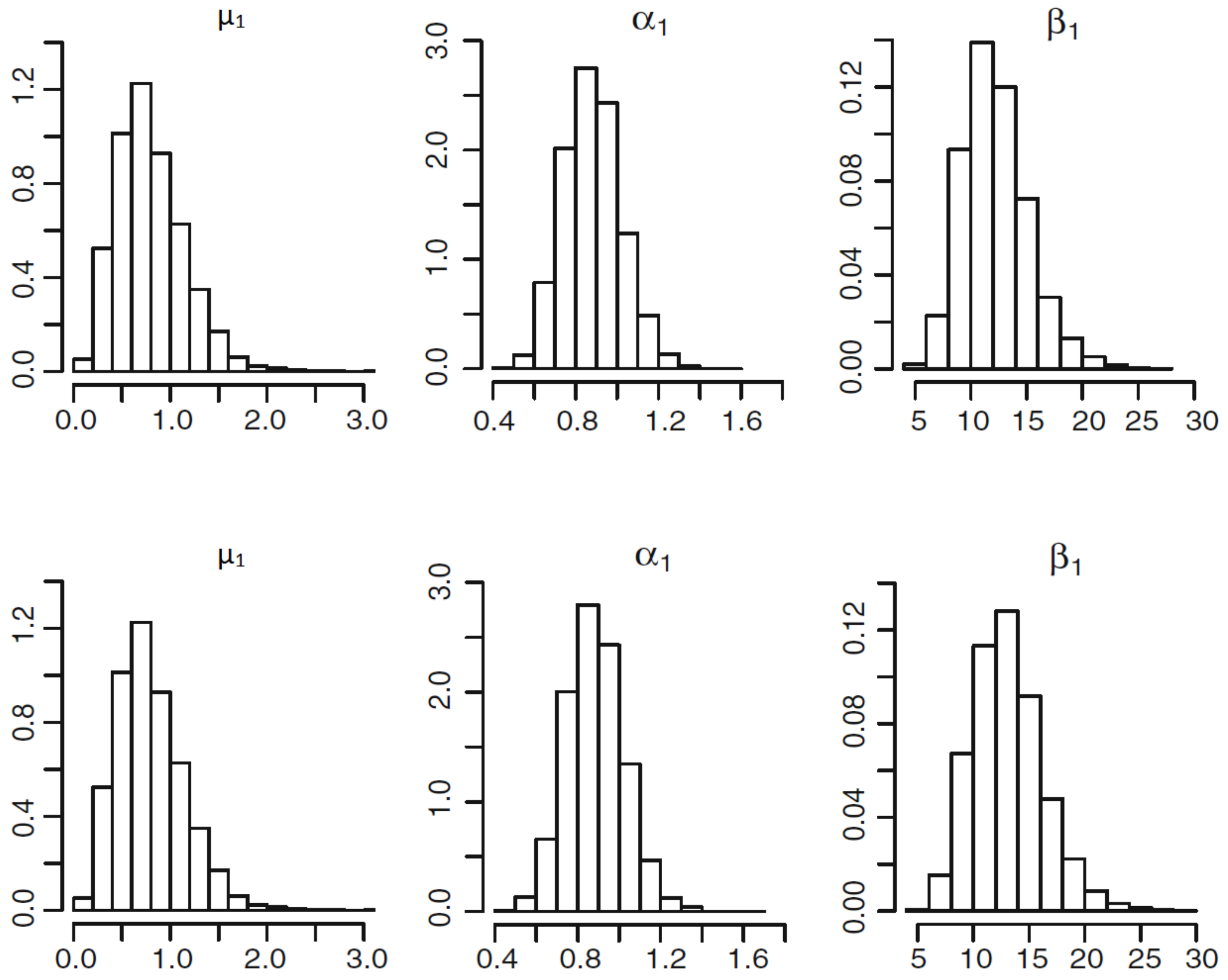
$$\beta(t, \kappa) = \beta_1 \exp(-\beta_1 t)$$

- Points generated in time interval $[0, 10)$



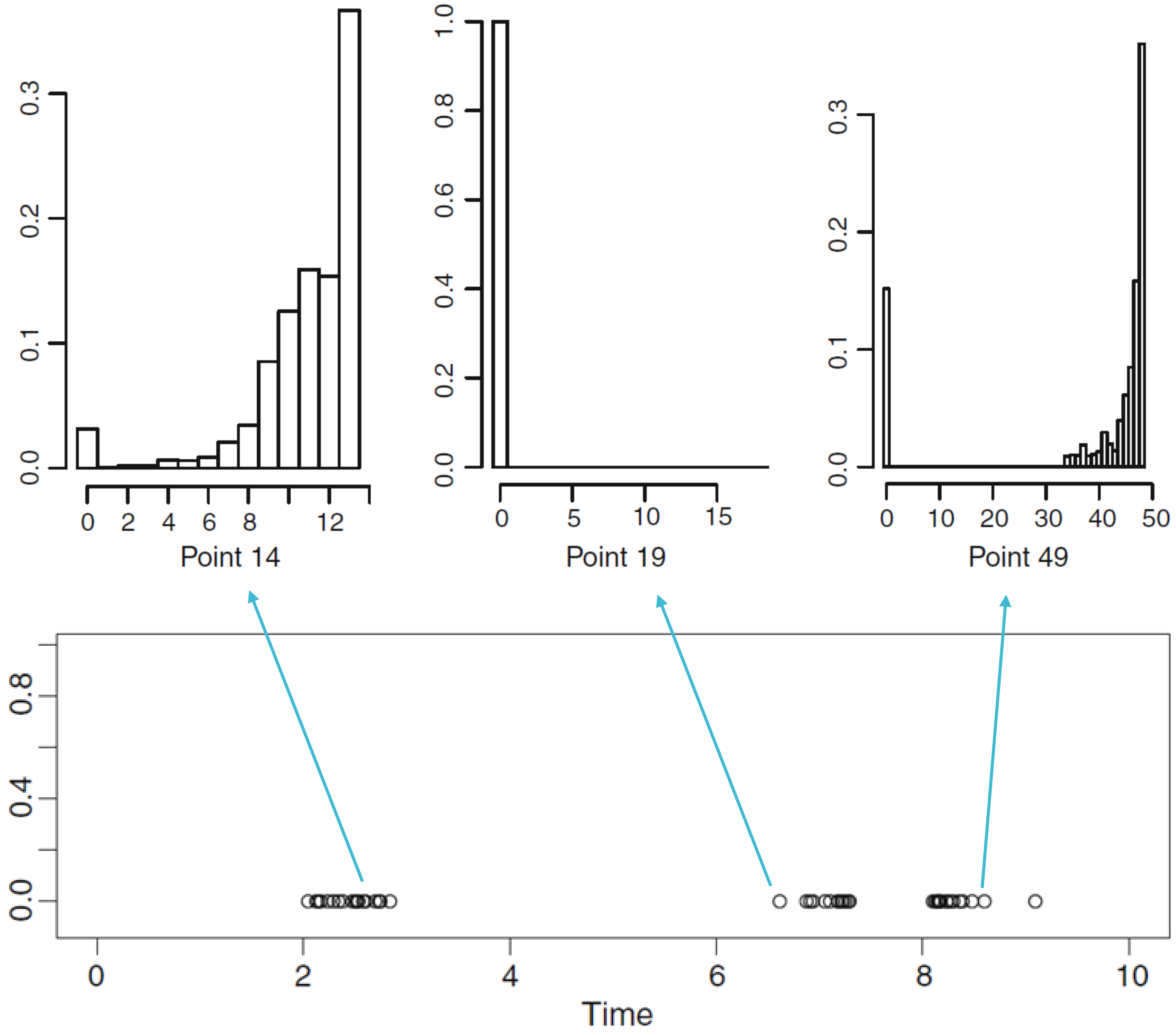
Example 1 - unmarked

- Posteriors
  - Top - CIF
  - Bottom - cluster process
- Same results
- MRE of 30% in parameter means



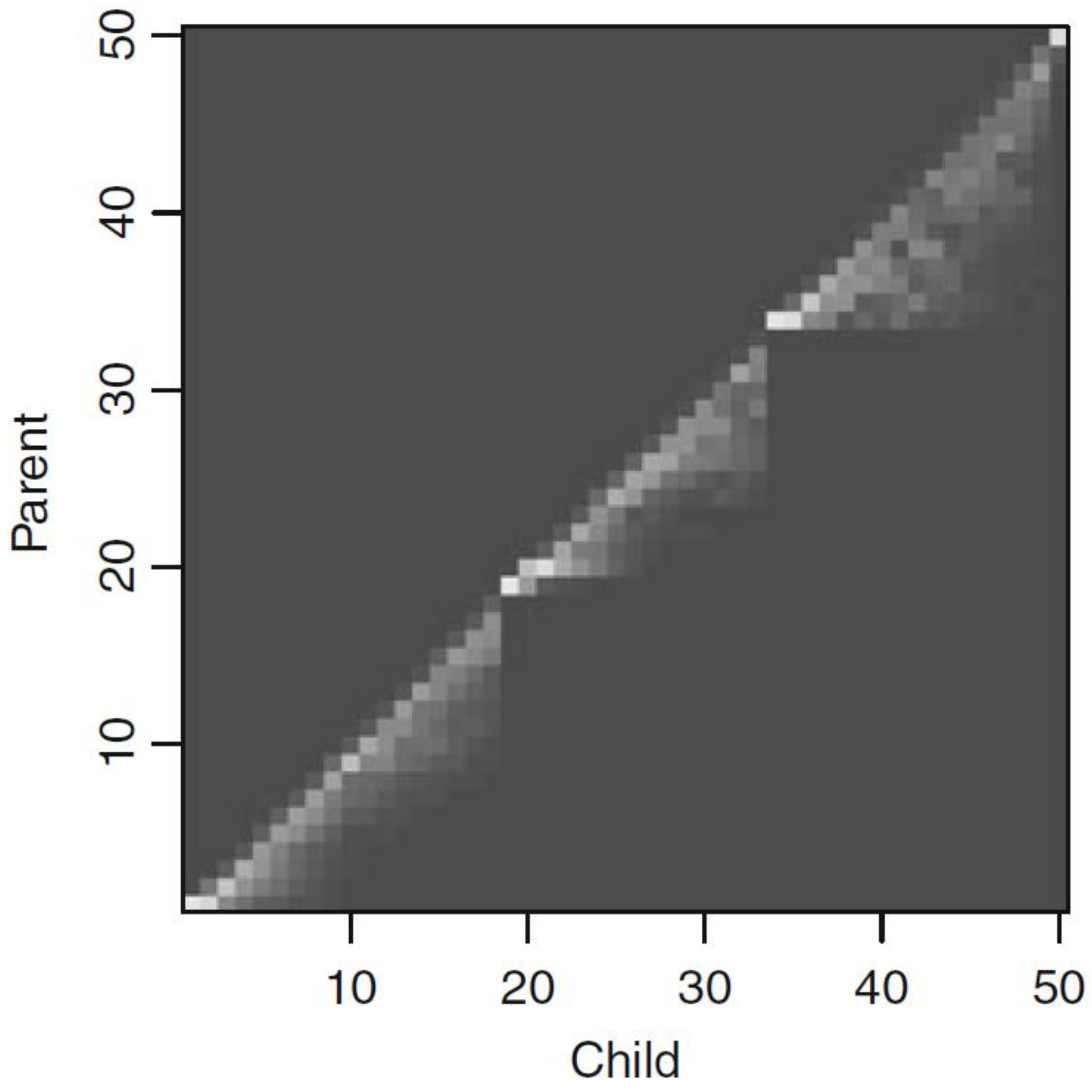
Example 1 - unmarked

Marginal posterior parent distributions



Example 1 - unmarked

All marginal posterior parent distributions



Example 2 – independent marks

- Formulation:

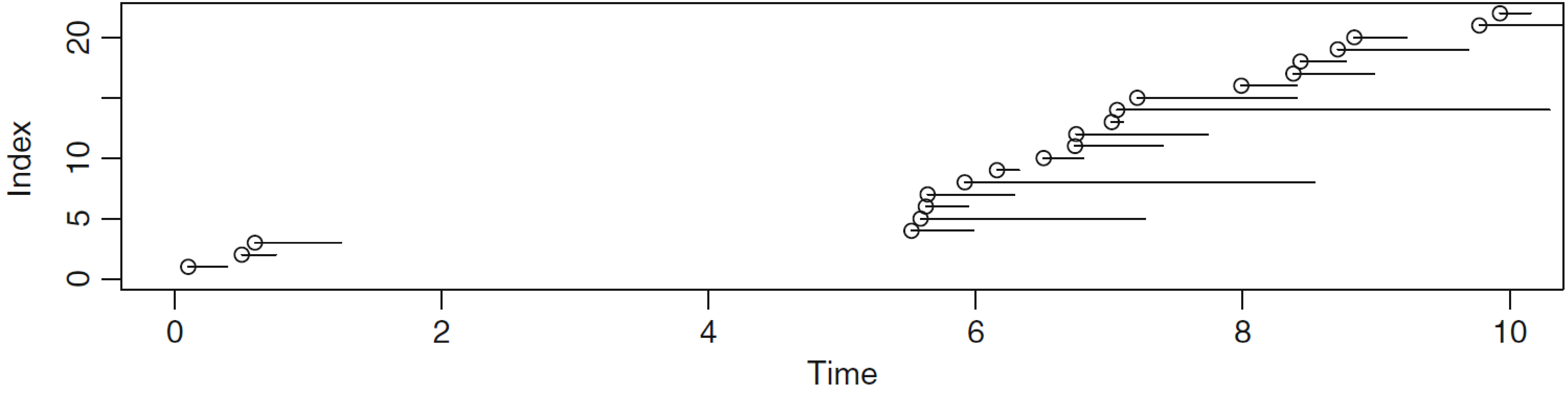
$$\mu(t) = \mu_1 \mathbf{1}(t \geq 0)$$

$$\alpha(\kappa) = \alpha_1 \kappa$$

$$\beta(t, \kappa) = \mathbf{1}\big(t \in (0, \kappa)\big)/\kappa$$

$$\gamma^*(\kappa \mid t) = \gamma_1 \exp(-\gamma_1 \kappa)$$

- Model population reproduction
  - γ_1 - inverse of mean survival time
  - α_1/γ_1 - mean offspring size

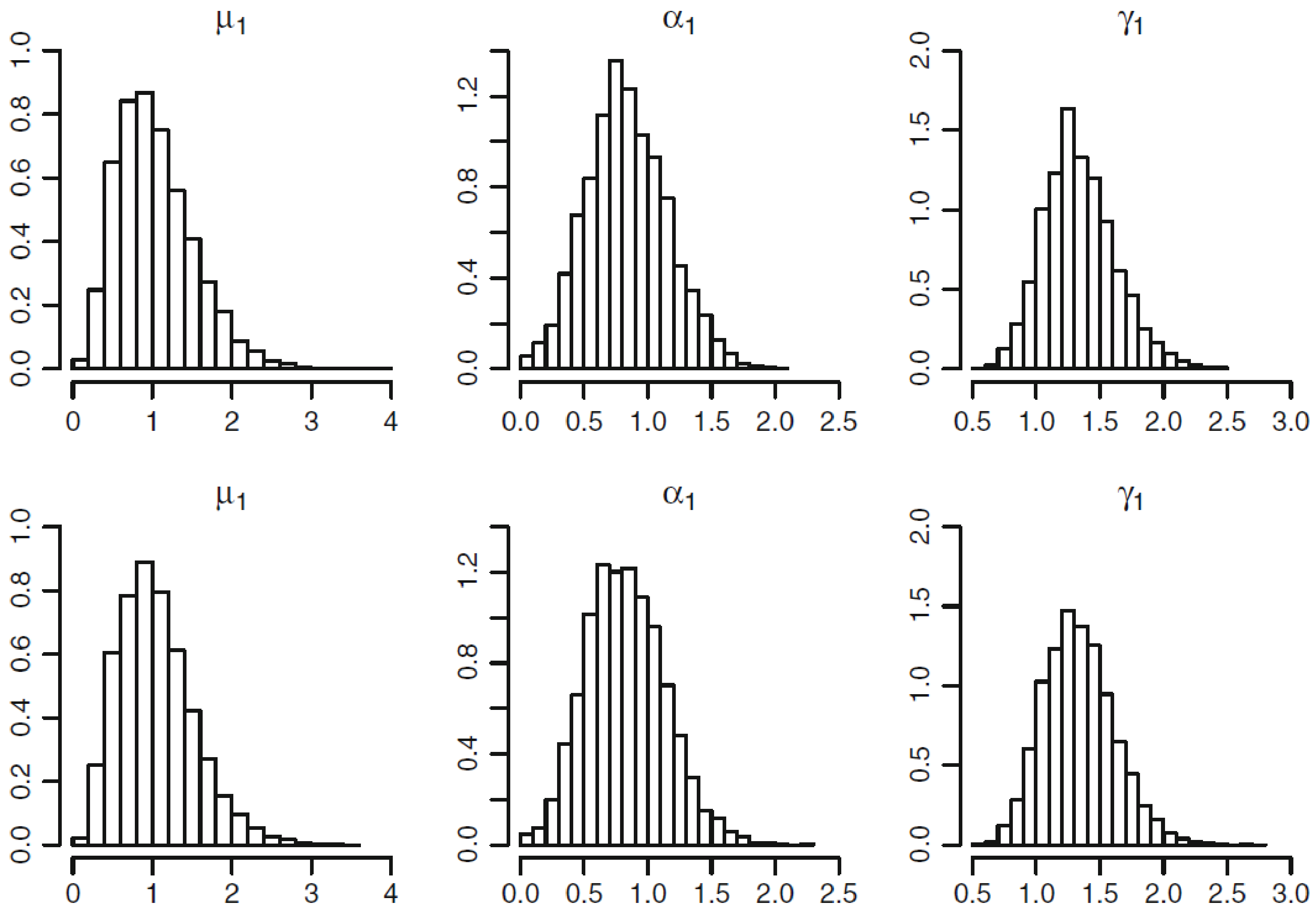


Example 2 - independent marks

- Dataset:
  - Points are events at instants in $[0, 10]$
  - Non-negative marks representing event duration

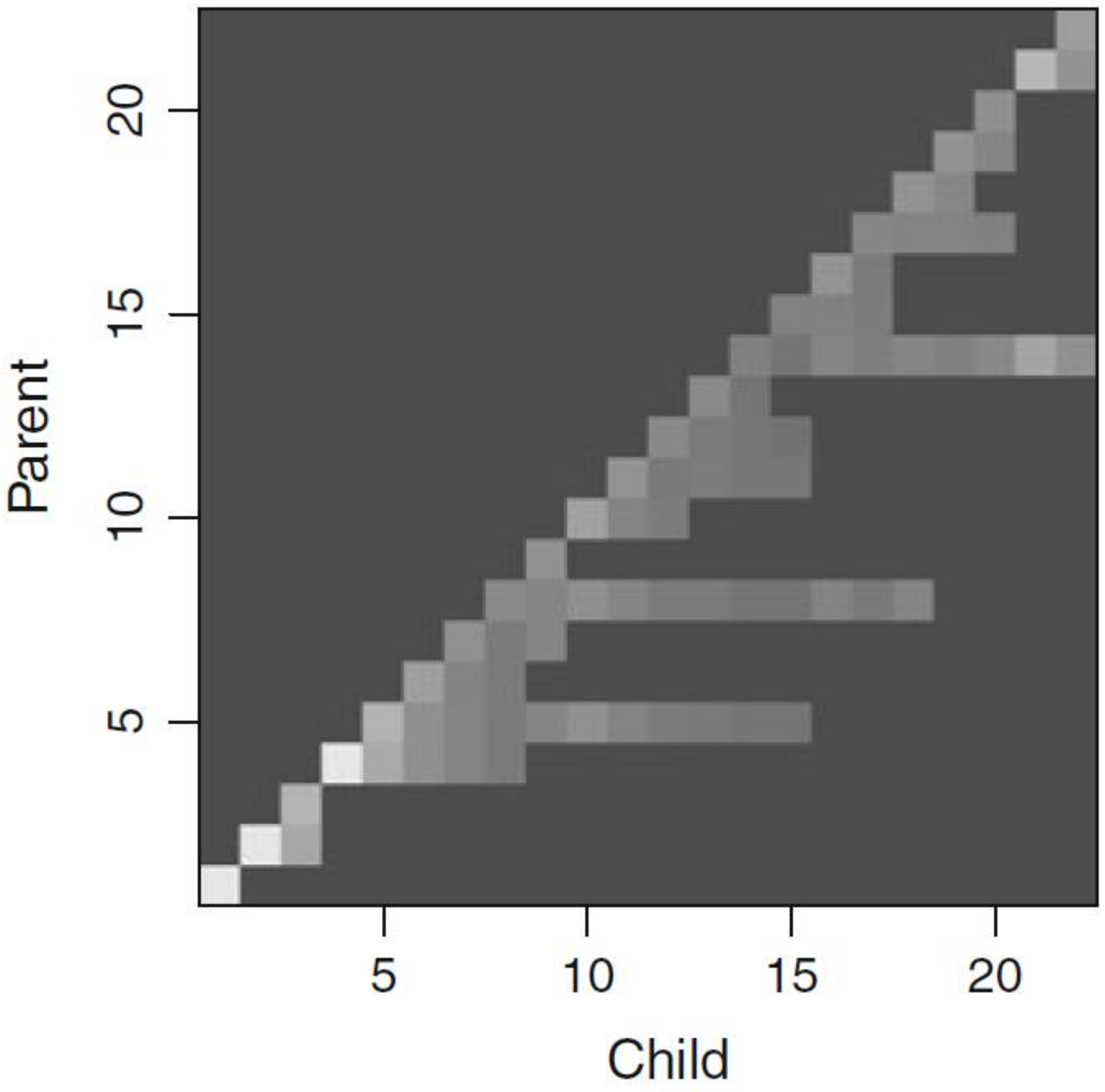
Example 2 – independent marks

- Posteriors
  - Top - CIF
  - Bottom - cluster process
- MRE of 50% in parameter means



Example 2 – independent marks

All marginal posterior parent distributions



Example 3 – dependent marks

- Model seismic events
- Magnitudes m are exponential i.i.d.
- Positions $(x, y) \in [0, 10] \times [0, 10]$:
  - Uniform for primary earthquake
  - Normal around epicenter for aftershocks
- Mean number of aftershocks spawned by primary earthquake:

$$\frac{\alpha_1 \gamma_1}{\gamma_1 - \alpha_2}$$

Example 3 – dependent marks

- Formulation:

$$\mu(t) = \mu_1 \mathbf{1}(t \geq 0)$$

$$\alpha(\kappa) = \alpha_1 \exp(\alpha_2 m)$$

$$\beta(t,\kappa) = \frac{\beta_2}{\beta_1}\left(1 + \frac{t}{\beta_1}\right)^{\beta_2 - 1}$$

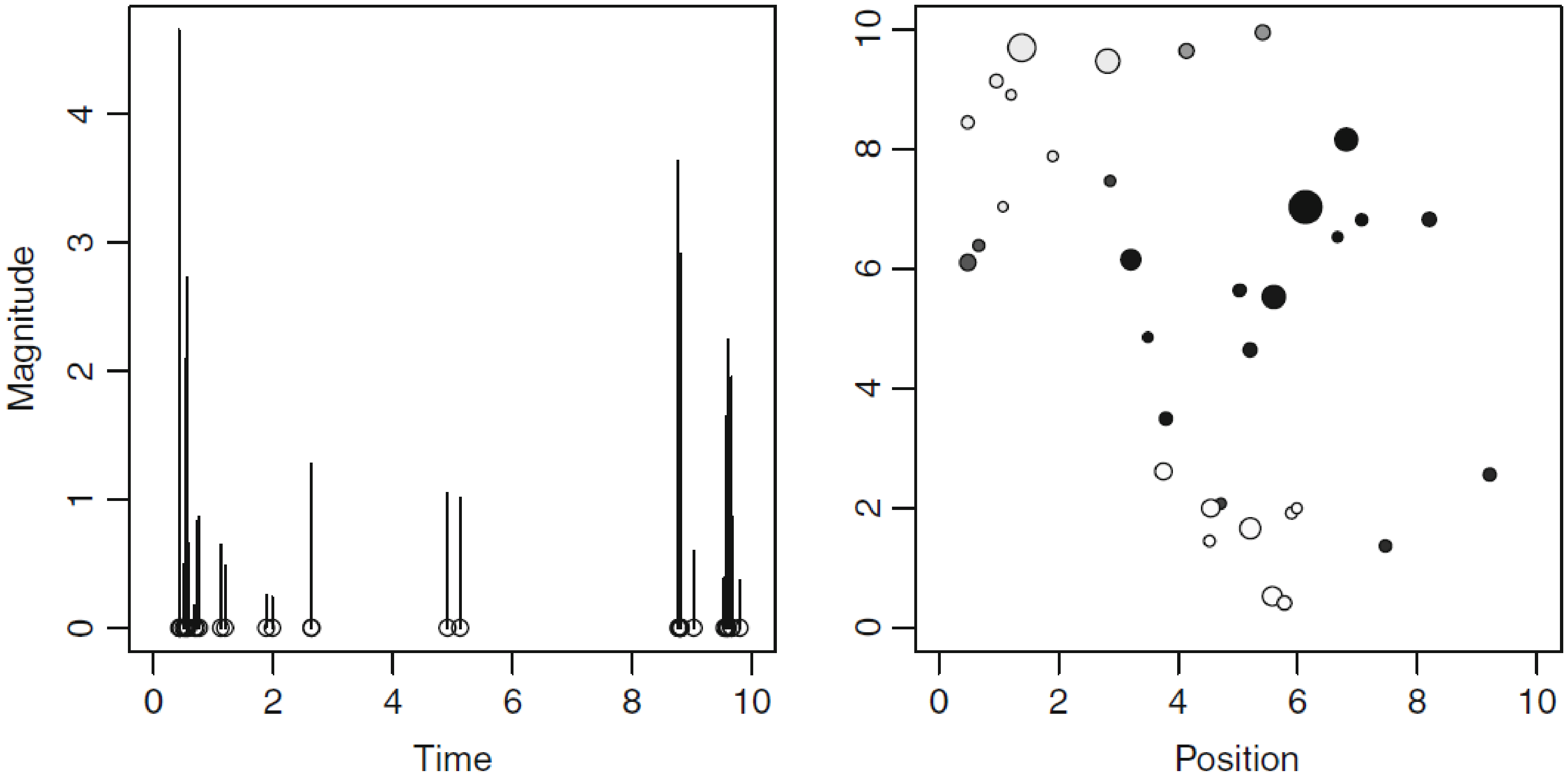
- Immigrant mark density:

$$\gamma(\kappa \mid t) = \gamma_1 \exp(-\gamma_1 m)\,\frac{\mathbf{1}\big((x,y) \in W\big)}{|W|}$$

- Offspring mark density:

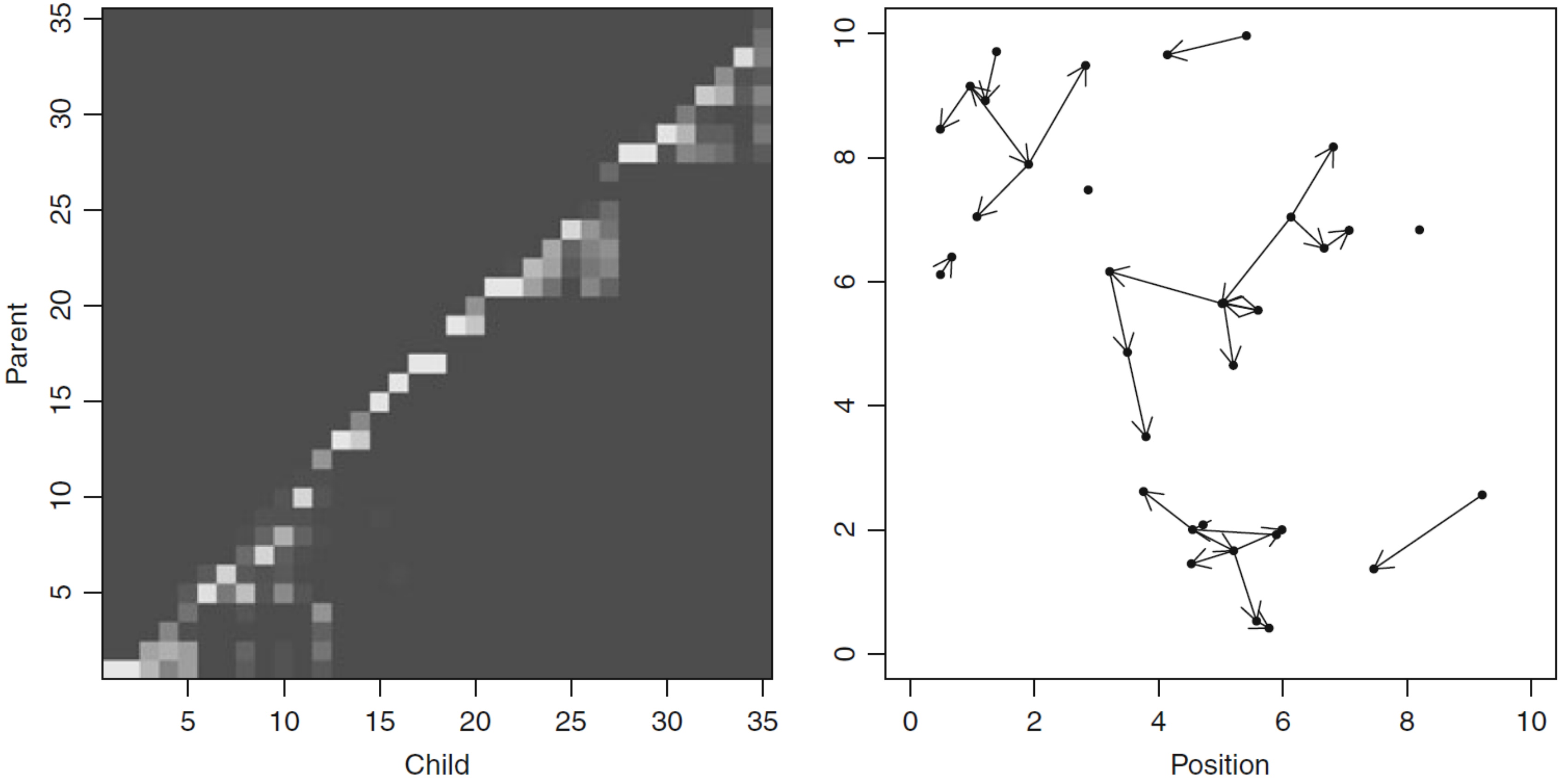
$$\gamma\big(\kappa \big| t, (t_{pa}, \kappa_{pa})\big) = \gamma_1 e^{-\gamma_1 m}\,\frac{1}{2\pi\gamma_2^2}\exp\left(-\frac{\left\|(x,y) - (x_{pa}, y_{pa})\right\|^2}{2\gamma_2^2}\right)$$

Example 3 – dependent marks



Example 3 – dependent marks

- Close match between methods
- 54% MRE in parameter means
- Good estimate of mean offspring size
- Less aftershocks for earthquakes of high magnitude



Example 3 – dependent marks

- Left: influence of position
- Right: mode of parents points to element

Conclusion

- Method comparison
  - Matching posteriors
  - High variation of simulation times
  - Cluster method is less complex
    - Appropriate for large datasets
- Simplifications
  - No events in $t < 0$
    - Offspring ignored
    - Small dataset and heavy tailed β
    - Affects both methods
  - Metropolis-within-Gibbs MCMC

Thank you