Bayesian nonparametric marked Hawkes processes

Athanasios Kottas

Department of Statistics, University of California, Santa Cruz

Joint work with Hyotae Kim (*Department of Statistical Science, Duke University*)

Bayesian Nonparametrics Networking Workshop 2023
Monash University, Caulfield campus, Melbourne, Australia
December 4-7, 2023

NPB methods

- Nonparametric Bayesian (NPB) methods avoid restrictive assumptions forced by specific parametric forms for the distributions/functions comprising a statistical model (density, regression, intensity functions ...)

- Bayesian nonparametrics
  - ▷ Requires priors for random distributions (or random functions).
  - ▷ About 50 years of history $\rightarrow$ advances in Bayesian computing have made their practical use feasible in the past 25 years or so.

- Powerful framework to build models that yield more general inferences and more reliable predictions than parametric models.

- Many success stories in the literature in terms of applications.

NPB methods

- Nonparametric Bayesian (NPB) methods avoid restrictive assumptions forced by specific parametric forms for the distributions/functions comprising a statistical model (density, regression, intensity functions ...)

- Bayesian nonparametrics
  - ▷ Requires priors for random distributions (or random functions).
  - ▷ About 50 years of history $\rightarrow$ advances in Bayesian computing have made their practical use feasible in the past 25 years or so.

- Powerful framework to build models that yield more general inferences and more reliable predictions than parametric models.

- Many success stories in the literature in terms of applications.

NPB methods

- Nonparametric Bayesian (NPB) methods avoid restrictive assumptions forced by specific parametric forms for the distributions/functions comprising a statistical model (density, regression, intensity functions ...)

- Bayesian nonparametrics
  - ▷ Requires priors for random distributions (or random functions).
  - ▷ About 50 years of history → advances in Bayesian computing have made their practical use feasible in the past 25 years or so.

- Powerful framework to build models that yield more general inferences and more reliable predictions than parametric models.

- Many success stories in the literature in terms of applications.

NPB methods

- Nonparametric Bayesian (NPB) methods avoid restrictive assumptions forced by specific parametric forms for the distributions/functions comprising a statistical model (density, regression, intensity functions ...)

- Bayesian nonparametrics
  - ▷ Requires priors for random distributions (or random functions).
  - ▷ About 50 years of history $\rightarrow$ advances in Bayesian computing have made their practical use feasible in the past 25 years or so.

- Powerful framework to build models that yield more general inferences and more reliable predictions than parametric models.

- Many success stories in the literature in terms of applications.

Earlier work with Poisson process models

Risk assessment of extreme values from environmental time series (Kottas, Wang & Rodriguez, 2012)

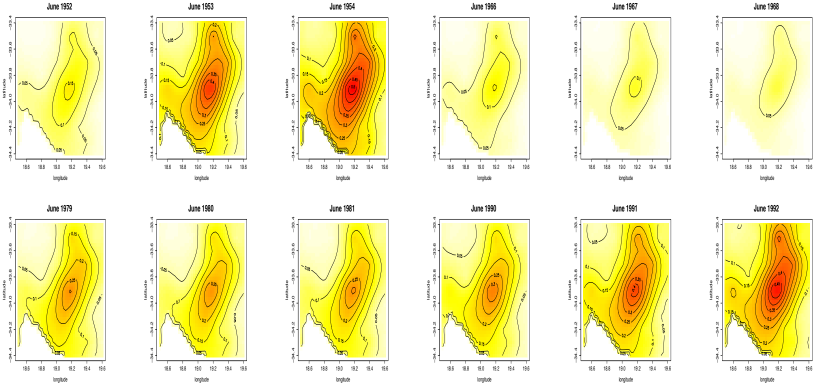

Figure: Cape Floristic Region rainfall data. Posterior mean estimate for the risk surface probability of at least one rainfall exceedance in the month of June for twelve years.

Earlier work with Poisson process models

Analysis of hurricane occurrences in the southeastern U.S. (Xiao, Kottas & Sansó, 2015)

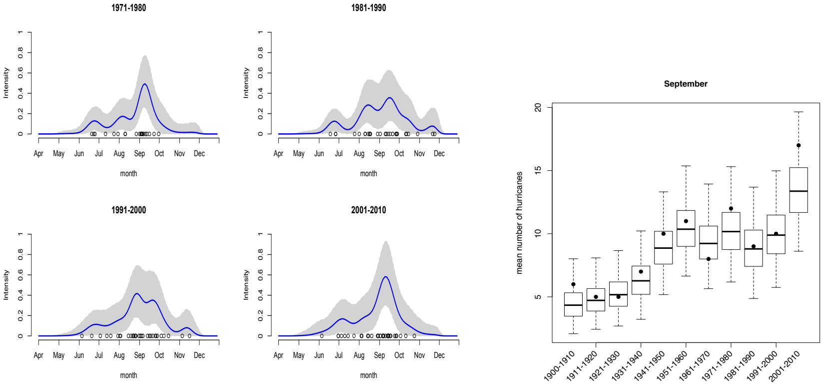

Figure: Left panel: posterior mean estimates and 95% interval bands of the hurricane intensity during 1971-2010. Right panel: boxplots of posterior samples of the mean number of hurricanes in September by decade.

Earlier work with Poisson process models

Assessing systematic risk in the S&P500 index (Rodriguez, Wang & Kottas, 2017)

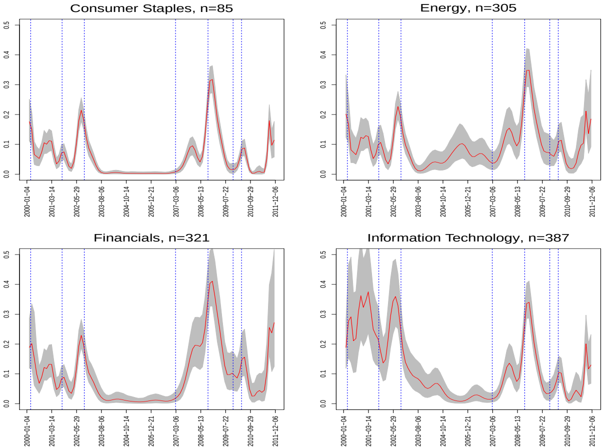

Figure: Posterior mean and 95% uncertainty bands for the overall intensity of extreme drops for four sectors of the S&P500 index.

Point pattern data

- A new story on NPB modeling & inference for Hawkes processes, a versatile class of point processes.
- Point processes: stochastic models to study the distribution and interactions in point patterns recorded over time and/or space.

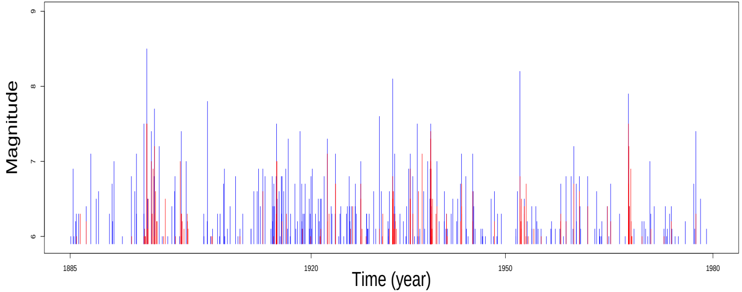

Figure: Earthquake occurrences (of magnitude $\geq$ 6) in northeastern Japan from 1885 through 1980; main shocks (blue), aftershocks (red) (Ogata, 1988).

Point pattern data

- A new story on NPB modeling & inference for Hawkes processes, a versatile class of point processes.
- Point processes: stochastic models to study the distribution and interactions in point patterns recorded over time and/or space.

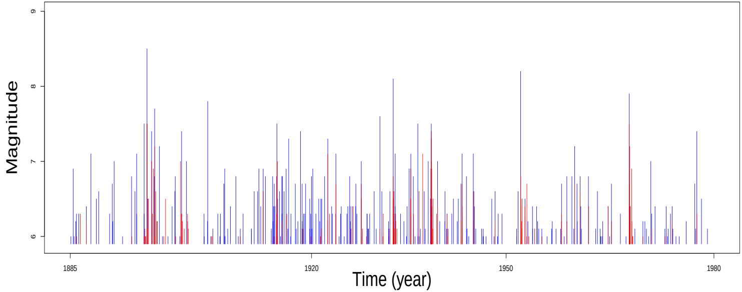

Figure: Earthquake occurrences (of magnitude $\geq$ 6) in northeastern Japan from 1885 through 1980; main shocks (blue), aftershocks (red) (Ogata, 1988).

Point processes

- Focus on temporal point processes.
  - Conditional intensity for a point process over time:

$$\lambda^*(t) \equiv \lambda(t \mid \mathcal{H}(t)) = \lim_{\Delta t \to 0} \frac{\Pr(N(t+\Delta t) - N(t) = 1 \mid \mathcal{H}(t))}{\Delta t}$$

    $N(t)$: number of events in $(0, t]$; $\mathcal{H}(t)$: process history up to time t.

  - Likelihood based on point pattern $0 < t_1 < \ldots < t_n < T$, observed in time window $(0, T)$:

$$\exp\left(-\int_0^T \lambda^*(u)\mathrm{d}u\right) \prod_{i=1}^{n} \lambda^*(t_i)$$

- Non-homogeneous Poisson process (NHPP) $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \lambda(t)$

Point processes

- Focus on temporal point processes.
  - ▷ Conditional intensity for a point process over time:

$$\lambda^*(t) \equiv \lambda(t \mid \mathcal{H}(t)) = \lim_{\Delta t \to 0} \frac{\Pr(N(t+\Delta t) - N(t) = 1 \mid \mathcal{H}(t))}{\Delta t}$$

    $N(t)$: number of events in $(0, t]$; $\mathcal{H}(t)$: process history up to time t.

  - ▷ Likelihood based on point pattern $0 < t_1 < \ldots < t_n < T$, observed in time window $(0, T)$:

$$\exp\left(-\int_0^T \lambda^*(u)\mathrm{d}u\right) \prod_{i=1}^{n} \lambda^*(t_i)$$

- Non-homogeneous Poisson process (NHPP) $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \lambda(t)$

Hawkes process

- Example of a "self-exciting" process (Hawkes, 1971; Hawkes & Oakes, 1974)

$$\lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t-t_i) \;=\; \mu(t) + \gamma \sum_{t_i<t} f(t-t_i)$$

  → $\mu(t)$: immigrant (background) intensity function
  → $h(t)$: excitation function ($h(t) > 0$ s.t. $\gamma = \int_0^\infty h(u)\,\mathrm{d}u < \infty$)
  → γ: branching ratio; $f(t)$: offspring density.

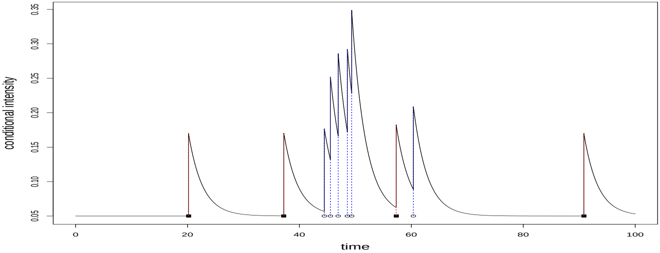

Figure: Conditional intensity realization based on constant immigrant intensity ($\mu(t) = 0.05$), $\gamma = 0.3$, and exponential offspring density.

Hawkes process

- Example of a "self-exciting" process (Hawkes, 1971; Hawkes & Oakes, 1974)

$$\lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t-t_i) \; = \; \mu(t) + \gamma \sum_{t_i<t} f(t-t_i)$$

→ $\mu(t)$: immigrant (background) intensity function
→ $h(t)$: excitation function ($h(t) > 0$ s.t. $\gamma = \int_0^\infty h(u)\,\mathrm{d}u < \infty$)
→ γ: branching ratio; $f(t)$: offspring density.

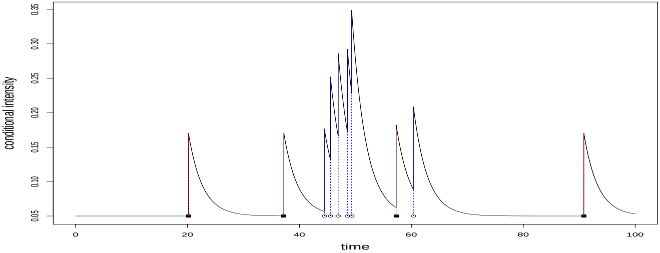

Figure: Conditional intensity realization based on constant immigrant intensity ($\mu(t) = 0.05$), $\gamma = 0.3$, and exponential offspring density.

Hawkes process branching structure

- Point pattern $\boldsymbol{t} = (t_1, \ldots, t_n)$ with (latent) branching variables $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$y_i = \begin{cases} 0 & \text{if } t_i \text{ is an immigrant} \\ j & \text{if } t_i \text{ has parent } t_j \end{cases}$$

- Given $\boldsymbol{y}$, $\boldsymbol{t}$ is partitioned into the set of immigrants, $I = \{t_j : y_j = 0\}$, and sets of offspring, $O_j = \{t_i : y_i = j\}$ (O_j collects all offspring of t_j)

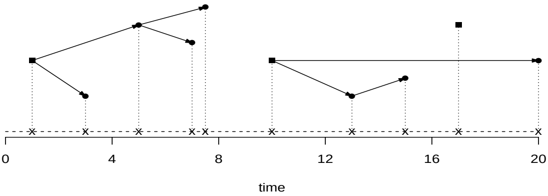

Figure: Branching structure: immigrants (squares), offspring (circles), full point pattern (x). $I = \{t_1, t_6, t_9\}$, $O_1 = \{t_2, t_3\}$, $O_3 = \{t_4, t_5\}$, $O_6 = \{t_7, t_{10}\}$, $O_7 = \{t_8\}$

Hawkes process branching structure

- Point pattern $\boldsymbol{t} = (t_1, \ldots, t_n)$ with (latent) branching variables $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$y_i = \begin{cases} 0 & \text{if } t_i \text{ is an immigrant} \\ j & \text{if } t_i \text{ has parent } t_j \end{cases}$$

- Given $\boldsymbol{y}$, $\boldsymbol{t}$ is partitioned into the set of immigrants, $I = \{t_j : y_j = 0\}$, and sets of offspring, $O_j = \{t_i : y_i = j\}$ (O_j collects all offspring of t_j)

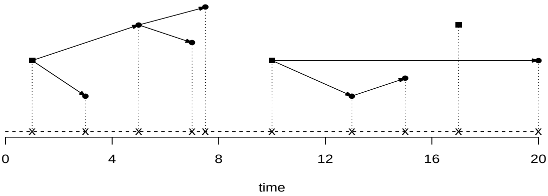

Figure: Branching structure: immigrants (squares), offspring (circles), full point pattern (x). $I = \{t_1, t_6, t_9\}$, $O_1 = \{t_2, t_3\}$, $O_3 = \{t_4, t_5\}$, $O_6 = \{t_7, t_{10}\}$, $O_7 = \{t_8\}$

Branching structure augmented likelihood

- Observed point pattern $\boldsymbol{t} = (t_1, ..., t_n)$ in time window $(0, T)$.
- Given $\boldsymbol{y}$, I and the O_j are realizations from independent Poisson processes
  - with intensity $\mu(t)$ for I
  - and intensity $h(t - t_j) = \gamma f(t - t_j)$ for O_j, for $j = 1, ..., n$ (when $O_j = \emptyset$, the likelihood contribution from O_j is an exponential term defining the probability that t_j has no offspring in $(0, T)$).
- Conditional likelihood (probability model for $\boldsymbol{t} \mid \boldsymbol{y}$)

$$\exp\left(-\int_0^T \mu(u)\mathrm{d}u\right)\left\{\prod_{t_i \in I} \mu(t_i)\right\} \exp\left(-\sum_{j=1}^{n}\int_0^T h(u - t_j)\mathrm{d}u\right)\left\{\prod_{t_i \in O} h(t_i - t_{p(i)})\right\}$$

where $O = \cup_{j=1}^{n} O_j$ is the set of all offspring, and $t_{p(i)} \equiv t_{y_i}$ is the parent of t_i

  - branching variables distribution: $\boldsymbol{y} \sim \delta_0(y_1) \prod_{j=2}^{n} \mathrm{Unif}(y_j \mid 0, 1, ..., j - 1)$

Branching structure augmented likelihood

- Observed point pattern $\boldsymbol{t} = (t_1, \ldots, t_n)$ in time window $(0, T)$.
- Given $\boldsymbol{y}$, I and the O_j are realizations from independent Poisson processes
  - ▷ with intensity $\mu(t)$ for I
  - ▷ and intensity $h(t - t_j) = \gamma f(t - t_j)$ for O_j, for $j = 1, \ldots, n$ (when $O_j = \emptyset$, the likelihood contribution from O_j is an exponential term defining the probability that t_j has no offspring in $(0, T)$).
- Conditional likelihood (probability model for $\boldsymbol{t} \mid \boldsymbol{y}$)

$$\exp\left(-\int_0^T \mu(u)\mathrm{d}u\right)\left\{\prod_{t_i \in I} \mu(t_i)\right\} \exp\left(-\sum_{j=1}^{n}\int_0^T h(u - t_j)\mathrm{d}u\right)\left\{\prod_{t_i \in O} h(t_i - t_{p(i)})\right\}$$

where $O = \cup_{j=1}^{n} O_j$ is the set of all offspring, and $t_{p(i)} \equiv t_{y_i}$ is the parent of t_i

  - ▷ branching variables distribution: $\boldsymbol{y} \sim \delta_0(y_1) \prod_{j=2}^{n} \mathrm{Unif}(y_j \mid 0, 1, \ldots, j - 1)$

Branching structure augmented likelihood

- Observed point pattern $\boldsymbol{t} = (t_1, \ldots, t_n)$ in time window $(0, T)$.
- Given $\boldsymbol{y}$, I and the O_j are realizations from independent Poisson processes
  - ▷ with intensity $\mu(t)$ for I
  - ▷ and intensity $h(t - t_j) = \gamma f(t - t_j)$ for O_j, for $j = 1, \ldots, n$ (when $O_j = \emptyset$, the likelihood contribution from O_j is an exponential term defining the probability that t_j has no offspring in $(0, T)$).
- Conditional likelihood (probability model for $\boldsymbol{t} \mid \boldsymbol{y}$)

$$\exp\left(-\int_0^T \mu(u)\mathrm{d}u\right)\left\{\prod_{t_i \in I} \mu(t_i)\right\} \exp\left(-\sum_{j=1}^{n}\int_0^T h(u - t_j)\mathrm{d}u\right)\left\{\prod_{t_i \in O} h(t_i - t_{p(i)})\right\}$$

where $O = \cup_{j=1}^{n} O_j$ is the set of all offspring, and $t_{p(i)} \equiv t_{y_i}$ is the parent of t_i

  - ▷ branching variables distribution: $\boldsymbol{y} \sim \delta_0(y_1)\prod_{j=2}^{n} \mathrm{Unif}(y_j \mid 0, 1, \ldots, j-1)$

Parametric models

- Conditional likelihood factorizes into terms for immigrants and offspring $\rightarrow$ more tractable than the (unconditional) likelihood for parametric Bayesian inference for the immigrant and offspring intensities (Rasmussen, 2013).

- The offspring intensity likelihood normalizing term still poses challenges w.r.t. flexible modeling for the excitation function/offspring density.

- Common choices for the immigrant intensity and offspring density:
  - typically, $\mu(t) \equiv \mu$ (along with $\gamma \in (0, 1)$, it yields a stationary HP).
  - Exponential offspring density; also, Lomax, $f(t) = \phi\,\theta^{\phi}(\theta + t)^{-(\phi+1)}$, associated with the ETAS (Epidemic Type Aftershock Sequence) model.

- Hawkes processes (including marked and space-time HPs) are widely used in various areas, including seismology, finance, and criminology.

Parametric models

- Conditional likelihood factorizes into terms for immigrants and offspring $\rightarrow$ more tractable than the (unconditional) likelihood for parametric Bayesian inference for the immigrant and offspring intensities (Rasmussen, 2013).

- The offspring intensity likelihood normalizing term still poses challenges w.r.t. flexible modeling for the excitation function/offspring density.

- Common choices for the immigrant intensity and offspring density:
  - typically, $\mu(t) \equiv \mu$ (along with $\gamma \in (0, 1)$, it yields a stationary HP).
  - Exponential offspring density; also, Lomax, $f(t) = \phi\,\theta^{\phi}(\theta + t)^{-(\phi+1)}$, associated with the ETAS (Epidemic Type Aftershock Sequence) model.

- Hawkes processes (including marked and space-time HPs) are widely used in various areas, including seismology, finance, and criminology.

Parametric models

- Conditional likelihood factorizes into terms for immigrants and offspring $\rightarrow$ more tractable than the (unconditional) likelihood for parametric Bayesian inference for the immigrant and offspring intensities (Rasmussen, 2013).

- The offspring intensity likelihood normalizing term still poses challenges w.r.t. flexible modeling for the excitation function/offspring density.

- Common choices for the immigrant intensity and offspring density:
  - typically, $\mu(t) \equiv \mu$ (along with $\gamma \in (0,1)$, it yields a stationary HP).
  - Exponential offspring density; also, Lomax, $f(t) = \phi\,\theta^{\phi}(\theta+t)^{-(\phi+1)}$, associated with the ETAS (Epidemic Type Aftershock Sequence) model.

- Hawkes processes (including marked and space-time HPs) are widely used in various areas, including seismology, finance, and criminology.

Parametric models

- Conditional likelihood factorizes into terms for immigrants and offspring $\rightarrow$ more tractable than the (unconditional) likelihood for parametric Bayesian inference for the immigrant and offspring intensities (Rasmussen, 2013).

- The offspring intensity likelihood normalizing term still poses challenges w.r.t. flexible modeling for the excitation function/offspring density.

- Common choices for the immigrant intensity and offspring density:
  - typically, $\mu(t) \equiv \mu$ (along with $\gamma \in (0,1)$, it yields a stationary HP).
  - Exponential offspring density; also, Lomax, $f(t) = \phi\,\theta^{\phi}(\theta + t)^{-(\phi+1)}$, associated with the ETAS (Epidemic Type Aftershock Sequence) model.

- Hawkes processes (including marked and space-time HPs) are widely used in various areas, including seismology, finance, and criminology.

Objectives

- Develop a nonparametric Bayesian modeling and inference framework for Hawkes processes.
- Increase the inferential scope for Hawkes processes by exploring flexible prior probability models for the immigrant intensity function and/or the excitation function (offspring density).
- Need to balance model flexibility with computational efficiency, especially in the prior models for the offspring intensity.
- We begin with the basic version of the temporal HP, and then extend to marked HPs in the context of earthquake modeling (mark $\rightarrow$ earthquake magnitude).

Objectives

- Develop a nonparametric Bayesian modeling and inference framework for Hawkes processes.
- Increase the inferential scope for Hawkes processes by exploring flexible prior probability models for the immigrant intensity function and/or the excitation function (offspring density).
- Need to balance model flexibility with computational efficiency, especially in the prior models for the offspring intensity.
- We begin with the basic version of the temporal HP, and then extend to marked HPs in the context of earthquake modeling (mark $\longrightarrow$ earthquake magnitude).

Objectives

- Develop a nonparametric Bayesian modeling and inference framework for Hawkes processes.
- Increase the inferential scope for Hawkes processes by exploring flexible prior probability models for the immigrant intensity function and/or the excitation function (offspring density).
- Need to balance model flexibility with computational efficiency, especially in the prior models for the offspring intensity.
- We begin with the basic version of the temporal HP, and then extend to marked HPs in the context of earthquake modeling (mark $\rightarrow$ earthquake magnitude).

Modeling the immigrant intensity

- Prior model for the immigrant intensity $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t-t_i)$
- Weighted combination of Erlang densities, with weights defined through increments of a random cumulative intensity function:

$$\mu(t) = \sum_{j=1}^{J} \omega_j \, \mathrm{ga}(t \mid j, \theta), \quad t \in \mathbb{R}^+$$
$$\omega_j = H(j\theta) - H((j-1)\theta), \quad H \sim \mathcal{G}(H_0, c_0)$$

  - $\mathcal{G}(H_0, c_0)$: gamma process specified through H_0, a centering cumulative intensity function, and precision parameter $c_0 > 0$ (Kalbfleisch, 1978)
  - Based on the independent increments of the gamma process,

$$\omega_j \mid \theta, c_0, H_0 \overset{ind.}{\sim} \mathrm{gamma}(c_0 \{H_0(j\theta) - H_0((j-1)\theta)\}, c_0), \;\; j = 1, ..., J$$

- Let h be a NHPP intensity, with cumulative intensity $H(t) = \int_0^t h(u)\mathrm{d}u$, such that $H(t) = O(t^m)$, as $t \rightarrow \infty$, for some $m > 0$. Then, as $\theta \rightarrow 0$ and $J \rightarrow \infty$, $\sum_{j=1}^{J} \{H(j\theta) - H((j-1)\theta)\} \, \mathrm{ga}(t \mid j, \theta)$ converges to $h(t)$ at every t s.t. $h(t) = \mathrm{d}H(t)/\mathrm{d}t$.

Modeling the immigrant intensity

- Prior model for the immigrant intensity $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i < t} h(t - t_i)$
- Weighted combination of Erlang densities, with weights defined through increments of a random cumulative intensity function:

$$\mu(t) = \sum_{j=1}^{J} \omega_j \, \mathrm{ga}(t \mid j, \theta), \quad t \in \mathbb{R}^+$$
$$\omega_j = H(j\theta) - H((j-1)\theta), \quad H \sim \mathcal{G}(H_0, c_0)$$

  - ▷ $\mathcal{G}(H_0, c_0)$: gamma process specified through H_0, a centering cumulative intensity function, and precision parameter $c_0 > 0$ (Kalbfleisch, 1978)
  - ▷ Based on the independent increments of the gamma process,

$$\omega_j \mid \theta, c_0, H_0 \overset{ind.}{\sim} \mathrm{gamma}(c_0 \{H_0(j\theta) - H_0((j-1)\theta)\}, c_0), \quad j = 1, \ldots, J$$

- Let h be a NHPP intensity, with cumulative intensity $H(t) = \int_0^t h(u) \mathrm{d}u$, such that $H(t) = O(t^m)$, as $t \rightarrow \infty$, for some $m > 0$. Then, as $\theta \rightarrow 0$ and $J \rightarrow \infty$, $\sum_{j=1}^{J} \{H(j\theta) - H((j-1)\theta)\} \, \mathrm{ga}(t \mid j, \theta)$ converges to $h(t)$ at every t s.t. $h(t) = \mathrm{d}H(t)/\mathrm{d}t$.

Modeling the immigrant intensity

- Prior model for the immigrant intensity $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t - t_i)$
- Weighted combination of Erlang densities, with weights defined through increments of a random cumulative intensity function:

$$\mu(t) = \sum_{j=1}^{J} \omega_j \, \mathrm{ga}(t \mid j, \theta), \quad t \in \mathbb{R}^+$$
$$\omega_j = H(j\theta) - H((j-1)\theta), \quad H \sim \mathcal{G}(H_0, c_0)$$

  - $\mathcal{G}(H_0, c_0)$: gamma process specified through H_0, a centering cumulative intensity function, and precision parameter $c_0 > 0$ (Kalbfleisch, 1978)
  - Based on the independent increments of the gamma process,

$$\omega_j \mid \theta, c_0, H_0 \overset{ind.}{\sim} \mathrm{gamma}(c_0 \{H_0(j\theta) - H_0((j-1)\theta)\}, c_0), \;\; j = 1, \ldots, J$$

- Let h be a NHPP intensity, with cumulative intensity $H(t) = \int_0^t h(u) \mathrm{d}u$, such that $H(t) = O(t^m)$, as $t \rightarrow \infty$, for some $m > 0$. Then, as $\theta \rightarrow 0$ and $J \rightarrow \infty$, $\sum_{j=1}^{J} \{H(j\theta) - H((j-1)\theta)\} \, \mathrm{ga}(t \mid j, \theta)$ converges to $h(t)$ at every t s.t. $h(t) = \mathrm{d}H(t)/\mathrm{d}t$.

Modeling the immigrant intensity

- Prior model for the immigrant intensity $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i < t} h(t - t_i)$
- Weighted combination of Erlang densities, with weights defined through increments of a random cumulative intensity function:

$$\mu(t) = \sum_{j=1}^{J} \omega_j \, \mathrm{ga}(t \mid j, \theta), \quad t \in \mathbb{R}^+$$
$$\omega_j = H(j\theta) - H((j-1)\theta), \quad H \sim \mathcal{G}(H_0, c_0)$$

  - $\mathcal{G}(H_0, c_0)$: gamma process specified through H_0, a centering cumulative intensity function, and precision parameter $c_0 > 0$ (Kalbfleisch, 1978)
  - Based on the independent increments of the gamma process,

$$\omega_j \mid \theta, c_0, H_0 \overset{ind.}{\sim} \mathrm{gamma}(c_0 \{H_0(j\theta) - H_0((j-1)\theta)\}, c_0), \; j = 1, \ldots, J$$

- Let h be a NHPP intensity, with cumulative intensity $H(t) = \int_0^t h(u)\mathrm{d}u$, such that $H(t) = O(t^m)$, as $t \to \infty$, for some $m > 0$. Then, as $\theta \to 0$ and $J \to \infty$, $\sum_{j=1}^{J} \{H(j\theta) - H((j-1)\theta)\} \, \mathrm{ga}(t \mid j, \theta)$ converges to $h(t)$ at every t s.t. $h(t) = \mathrm{d}H(t)/\mathrm{d}t$.

Modeling the immigrant intensity

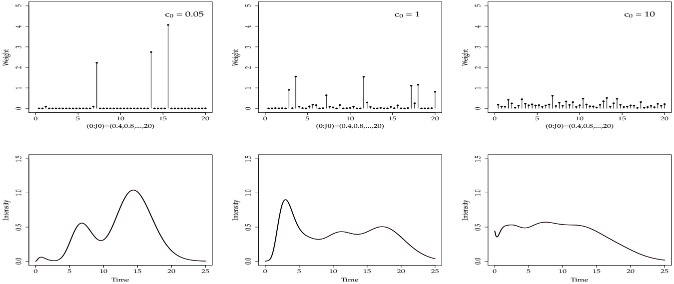

Figure: Prior realizations for the mixture weights and the corresponding intensity function for $c_0 = 0.05, 1, 10$. In all cases, $J = 50$, $\theta = 0.4$, and $H_0(t) = t/2$.

- $\exp(-\int_0^T \mu(u)\mathrm{d}u) = \prod_{j=1}^J \exp(-\omega_j \{\int_0^T \mathrm{ga}(u \mid j, \theta)\,\mathrm{d}u\})$
  - Latent variables that identify the basis densities to which the t_i are assigned.
  - Each ω_j can be updated from a gamma posterior conditional distribution that does not depend on the other ω_k, $k \neq j$.

Modeling the immigrant intensity

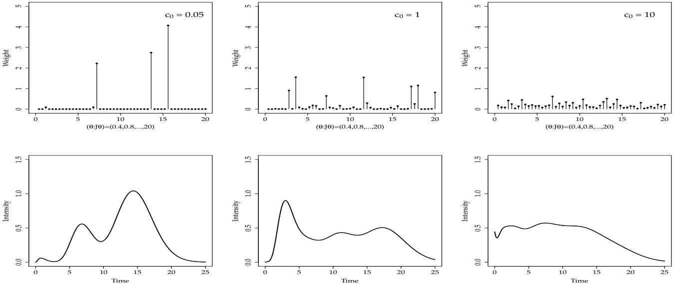

Figure: Prior realizations for the mixture weights and the corresponding intensity function for $c_0 = 0.05, 1, 10$. In all cases, $J = 50$, $\theta = 0.4$, and $H_0(t) = t/2$.

- $\exp(-\int_0^T \mu(u)\mathrm{d}u) = \prod_{j=1}^J \exp(-\omega_j \{\int_0^T \mathrm{ga}(u \mid j, \theta)\, \mathrm{d}u\})$
  - ▷ Latent variables that identify the basis densities to which the t_i are assigned.
  - ▷ Each ω_j can be updated from a gamma posterior conditional distribution that does not depend on the other ω_k, $k \neq j$.

Modeling the excitation function

- Prior model for the excitation function $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i < t} h(t - t_i)$
- Recall the condition $\int_0^\infty h(u)\,\mathrm{d}u < \infty$.
- Weighted combination of Erlang densities, with different prior for the weights:

$$h(t) = \sum_{\ell=1}^{L} V_\ell \, \mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$

$$V_\ell \mid \theta, F_0 \stackrel{ind.}{\sim} \mathrm{gamma}(\alpha_0 A_\ell, c), \quad \ell = 1, \ldots, L$$

  - $A_\ell = F_0(\ell\theta) - F_0((\ell-1)\theta)$, $\ell = 1, \ldots, L-1$; $A_L = 1 - F_0((L-1)\theta)$
  - $\alpha_0 > 0$, and F_0 is a distribution on $\mathbb{R}^+$
  - Branching ratio $\gamma = \int_0^\infty h(u)\mathrm{d}u = \sum_{\ell=1}^{L} V_\ell \sim \mathrm{gamma}(\alpha_0, c)$ prior

Modeling the excitation function

- Prior model for the excitation function $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t - t_i)$
- Recall the condition $\int_0^\infty h(u)\,\mathrm{d}u < \infty$.
- Weighted combination of Erlang densities, with different prior for the weights:

$$h(t) = \sum_{\ell=1}^{L} V_\ell \,\mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$

$$V_\ell \mid \theta, F_0 \overset{ind.}{\sim} \mathrm{gamma}(\alpha_0 A_\ell, c), \quad \ell = 1, \ldots, L$$

  - ▷ $A_\ell = F_0(\ell\theta) - F_0((\ell-1)\theta), \ell = 1, \ldots, L-1; A_L = 1 - F_0((L-1)\theta)$
  - ▷ $\alpha_0 > 0$, and F_0 is a distribution on $\mathbb{R}^+$
  - ▷ Branching ratio $\gamma = \int_0^\infty h(u)\mathrm{d}u = \sum_{\ell=1}^{L} V_\ell \sim \mathrm{gamma}(\alpha_0, c)$ prior

Modeling the excitation function

- Prior model for the excitation function $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \sum_{t_i<t} h(t - t_i)$
- Recall the condition $\int_0^\infty h(u)\,\mathrm{d}u < \infty$.
- Weighted combination of Erlang densities, with different prior for the weights:

$$h(t) = \sum_{\ell=1}^{L} V_\ell \,\mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$
$$V_\ell \mid \theta, F_0 \overset{ind.}{\sim} \mathrm{gamma}(\alpha_0 A_\ell, c), \quad \ell = 1, \ldots, L$$

  - $A_\ell = F_0(\ell\theta) - F_0((\ell-1)\theta)$, $\ell = 1, \ldots, L-1$; $A_L = 1 - F_0((L-1)\theta)$
  - $\alpha_0 > 0$, and F_0 is a distribution on $\mathbb{R}^+$
  - Branching ratio $\gamma = \int_0^\infty h(u)\mathrm{d}u = \sum_{\ell=1}^{L} V_\ell \sim \mathrm{gamma}(\alpha_0, c)$ prior

Modeling the excitation function

- Implied model for the offspring density

$$f(t) = \frac{h(t)}{\int_0^\infty h(u)\mathrm{d}u} = \sum_{\ell=1}^{L} \psi_\ell \, \mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$

  - ▷ $\psi_\ell = V_\ell / \{\sum_{r=1}^L V_r\} \rightarrow (\psi_1, ..., \psi_L) \mid \theta, F_0 \sim \mathrm{Dirichlet}(\alpha_0 A_1, ..., \alpha_0 A_L)$
  - ▷ Model corresponds to a DP-based Erlang mixture for the offspring density

$$\psi_\ell = F(\ell\theta) - F((\ell-1)\theta), \quad F \sim \mathrm{DP}(\alpha_0, F_0) \qquad (\psi_L = 1 - F((L-1)\theta))$$

- Offspring intensity likelihood normalizing term:

$$\exp\left(-\sum_{j=1}^{n} \int_0^T h(u - t_j)\mathrm{d}u\right) = \prod_{\ell=1}^{L} \exp(-V_\ell \, E_\ell(\theta))$$

  - ▷ $E_\ell(\theta) = \sum_{j=1}^n \int_0^{T-t_j} \mathrm{ga}(u \mid \ell, \theta)\mathrm{d}u$ (function of only one parameter, θ)

- MCMC: gamma full conditional for each V_ℓ (does not depend on V_k, $k \neq \ell$)!

Modeling the excitation function

- Implied model for the offspring density

$$f(t) = \frac{h(t)}{\int_0^\infty h(u)\mathrm{d}u} = \sum_{\ell=1}^{L} \psi_\ell \, \mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$

  - ▷ $\psi_\ell = V_\ell / \{\sum_{r=1}^L V_r\} \rightarrow (\psi_1, \ldots, \psi_L) \mid \theta, F_0 \sim \mathrm{Dirichlet}(\alpha_0 A_1, \ldots, \alpha_0 A_L)$
  - ▷ Model corresponds to a DP-based Erlang mixture for the offspring density

$$\psi_\ell = F(\ell\theta) - F((\ell-1)\theta), \quad F \sim \mathrm{DP}(\alpha_0, F_0) \qquad (\psi_L = 1 - F((L-1)\theta))$$

- Offspring intensity likelihood normalizing term:

$$\exp\left(-\sum_{j=1}^n \int_0^T h(u - t_j)\mathrm{d}u\right) = \prod_{\ell=1}^L \exp(-V_\ell \, E_\ell(\theta))$$

  - ▷ $E_\ell(\theta) = \sum_{j=1}^n \int_0^{T-t_j} \mathrm{ga}(u \mid \ell, \theta)\mathrm{d}u$ (function of only one parameter, θ)

- MCMC: gamma full conditional for each V_ℓ (does not depend on V_k, $k \neq \ell$)!

Modeling the excitation function

- Implied model for the offspring density

$$f(t) = \frac{h(t)}{\int_0^\infty h(u)\mathrm{d}u} = \sum_{\ell=1}^{L} \psi_\ell \, \mathrm{ga}(t \mid \ell, \theta), \quad t \in \mathbb{R}^+$$

  - $\psi_\ell = V_\ell / \{\sum_{r=1}^L V_r\} \rightarrow (\psi_1, \ldots, \psi_L) \mid \theta, F_0 \sim \mathrm{Dirichlet}(\alpha_0 A_1, \ldots, \alpha_0 A_L)$
  - Model corresponds to a DP-based Erlang mixture for the offspring density

$$\psi_\ell = F(\ell\theta) - F((\ell-1)\theta), \quad F \sim \mathrm{DP}(\alpha_0, F_0) \qquad (\psi_L = 1 - F((L-1)\theta))$$

- Offspring intensity likelihood normalizing term:

$$\exp\left(-\sum\nolimits_{j=1}^{n} \int_0^T h(u - t_j)\mathrm{d}u\right) = \prod\nolimits_{\ell=1}^{L} \exp(-V_\ell \, E_\ell(\theta))$$

  - $E_\ell(\theta) = \sum_{j=1}^n \int_0^{T-t_j} \mathrm{ga}(u \mid \ell, \theta)\mathrm{d}u$ (function of only one parameter, θ)

- MCMC: gamma full conditional for each V_ℓ (does not depend on V_k, $k \neq \ell$)!

Synthetic data examples (excitation function)

- Data generated from HPs with constant immigrant intensity, and:
  - top row: $\gamma = 0.8$, (unimodal) Weibull offspring density; $n_I = 145$, $n_O = 553$
  - bottom row: $\gamma = 0.9$, Weibull mixture offspring density; $n_I = 117$, $n_O = 1123$

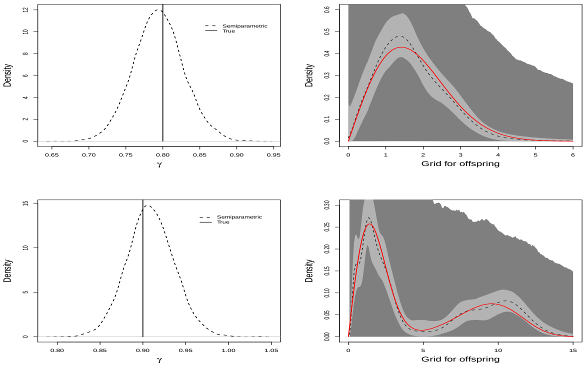

Figure: Posterior density for the branching ratio (left column), and posterior mean and 95% prior and posterior interval estimates for the offspring density (right column).

Modeling decreasing offspring densities

- Prior model for the offspring density $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \gamma \sum_{t_i < t} f(t - t_i)$ for problems where a decreasing offspring density is desirable.

- Uniform mixture representation for decreasing f on $\mathbb{R}^+$ (Brunner & Lo, 1989)

$$f(t) = \int_{\mathbb{R}^+} \theta^{-1} 1_{[0,\theta)}(t) \, \mathrm{d}G(\theta) = \sum_{\ell=1}^{\infty} \eta_\ell \, \mathrm{Unif}(t \mid 0, \theta_\ell), \quad t \in \mathbb{R}^+$$

  - DP(α, G_0) prior for $G \rightarrow \eta_\ell = \zeta_\ell \prod_{r=1}^{\ell-1}(1-\zeta_r)$, with $\zeta_\ell \overset{i.i.d.}{\sim} \mathrm{Beta}(1, \alpha)$
  - Geometric weights prior for $G \rightarrow \eta_\ell = (1-\zeta)\,\zeta^{\ell-1}$, with ζ following a distribution on $(0, 1)$.
  - θ_ℓ arise independently from an inverse-gamma distribution.

- MCMC: truncate to L mixture components; posterior full conditional for each θ_ℓ is a piecewise density with truncated inverse-gamma components; slice sampling steps to update the variable(s) ($\{\zeta_\ell\}$ or ζ) that define the mixture weights.

Modeling decreasing offspring densities

- Prior model for the offspring density $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \gamma \sum_{t_i < t} f(t - t_i)$ for problems where a decreasing offspring density is desirable.

- Uniform mixture representation for decreasing f on $\mathbb{R}^+$ (Brunner & Lo, 1989)

$$f(t) = \int_{\mathbb{R}+} \theta^{-1} 1_{[0,\theta)}(t)\, \mathrm{d}G(\theta) = \sum_{\ell=1}^{\infty} \eta_\ell \,\mathrm{Unif}(t \mid 0, \theta_\ell), \quad t \in \mathbb{R}^+$$

  - $\mathrm{DP}(\alpha, G_0)$ prior for $G \rightarrow \eta_\ell = \zeta_\ell \prod_{r=1}^{\ell-1}(1 - \zeta_r)$, with $\zeta_\ell \overset{i.i.d.}{\sim} \mathrm{Beta}(1, \alpha)$
  - Geometric weights prior for $G \rightarrow \eta_\ell = (1 - \zeta)\, \zeta^{\ell-1}$, with ζ following a distribution on $(0, 1)$.
  - θ_ℓ arise independently from an inverse-gamma distribution.

- MCMC: truncate to L mixture components; posterior full conditional for each θ_ℓ is a piecewise density with truncated inverse-gamma components; slice sampling steps to update the variable(s) ($\{\zeta_\ell\}$ or ζ) that define the mixture weights.

Modeling decreasing offspring densities

- Prior model for the offspring density $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \gamma \sum_{t_i < t} f(t - t_i)$ for problems where a decreasing offspring density is desirable.

- Uniform mixture representation for decreasing f on $\mathbb{R}^+$ (Brunner & Lo, 1989)

$$f(t) = \int_{\mathbb{R}+} \theta^{-1} 1_{[0,\theta)}(t) \, \mathrm{d}G(\theta) = \sum_{\ell=1}^{\infty} \eta_\ell \, \mathrm{Unif}(t \mid 0, \theta_\ell), \quad t \in \mathbb{R}^+$$

  - ▷ $\mathrm{DP}(\alpha, G_0)$ prior for $G \rightarrow \eta_\ell = \zeta_\ell \prod_{r=1}^{\ell-1}(1 - \zeta_r)$, with $\zeta_\ell \overset{i.i.d.}{\sim} \mathrm{Beta}(1, \alpha)$
  - ▷ Geometric weights prior for $G \rightarrow \eta_\ell = (1 - \zeta)\, \zeta^{\ell-1}$, with ζ following a distribution on $(0, 1)$.
  - ▷ θ_ℓ arise independently from an inverse-gamma distribution.

- MCMC: truncate to L mixture components; posterior full conditional for each θ_ℓ is a piecewise density with truncated inverse-gamma components; slice sampling steps to update the variable(s) ($\{\zeta_\ell\}$ or ζ) that define the mixture weights.

Modeling decreasing offspring densities

- Prior model for the offspring density $\rightarrow \lambda(t \mid \mathcal{H}(t)) = \mu(t) + \gamma \sum_{t_i<t} f(t - t_i)$ for problems where a decreasing offspring density is desirable.

- Uniform mixture representation for decreasing f on $\mathbb{R}^+$ (Brunner & Lo, 1989)

$$f(t) = \int_{\mathbb{R}^+} \theta^{-1} 1_{[0,\theta)}(t)\, \mathrm{d}G(\theta) = \sum_{\ell=1}^{\infty} \eta_\ell \, \mathrm{Unif}(t \mid 0, \theta_\ell), \quad t \in \mathbb{R}^+$$

  - $\mathrm{DP}(\alpha, G_0)$ prior for $G \rightarrow \eta_\ell = \zeta_\ell \prod_{r=1}^{\ell-1}(1 - \zeta_r)$, with $\zeta_\ell \stackrel{i.i.d.}{\sim} \mathrm{Beta}(1, \alpha)$
  - Geometric weights prior for $G \rightarrow \eta_\ell = (1 - \zeta)\, \zeta^{\ell-1}$, with ζ following a distribution on $(0, 1)$.
  - θ_ℓ arise independently from an inverse-gamma distribution.

- MCMC: truncate to L mixture components; posterior full conditional for each θ_ℓ is a piecewise density with truncated inverse-gamma components; slice sampling steps to update the variable(s) ($\{\zeta_\ell\}$ or ζ) that define the mixture weights.

Synthetic data example (decreasing offspring density)

- Data generated from a HP with constant immigrant intensity, $\gamma = 0.8$, Lomax offspring density (shape parameter 1.1, infinite variance); $n_I = 156$, $n_O = 627$.

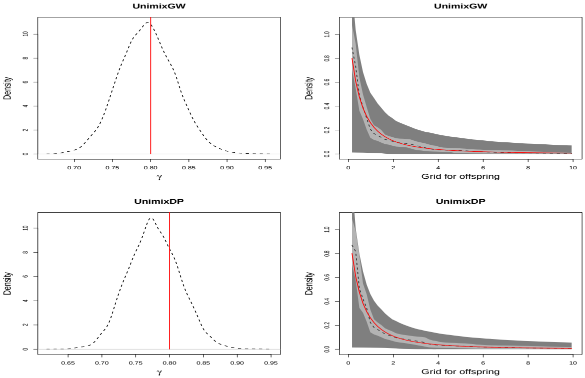

Figure: Posterior density for the branching ratio (left column), and posterior mean and 95% prior and posterior interval estimates for the offspring density (right column).

Earthquake data analysis

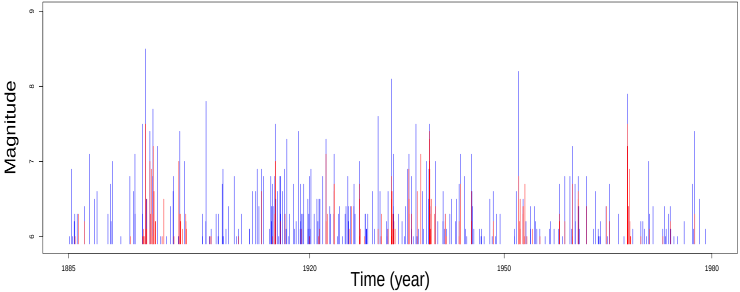

Figure: Japan earthquakes point pattern: main shocks (blue) and aftershocks (red).

- Occurrence times and magnitudes for 483 earthquakes (of depth < 100 kilometers, and with magnitude ≥ 6) in northeastern Japan from 1885 through 1980 (time window of about 35000 days) (Ogata, 1988).
- Each shock is classified as a main shock (258), aftershock (200) or foreshock (25) $\rightarrow$ we work with 458 time points, excluding the foreshocks.

Estimates for immigrant intensity and offspring density

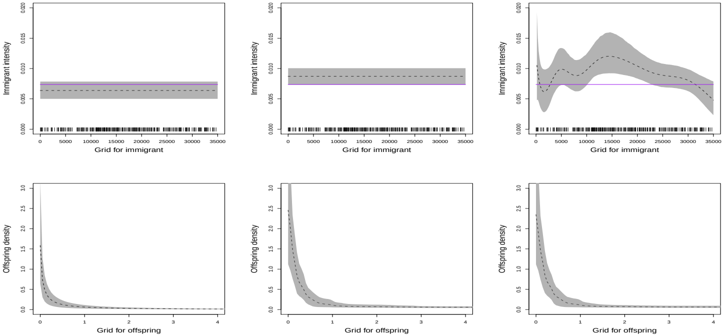

Figure: Inference results under models $\mathcal{M}_1$ (left), $\mathcal{M}_2$ (middle), and $\mathcal{M}_3$ (right). The purple solid line corresponds to $\tilde{\mu}$ such that $\int_0^T \tilde{\mu} dt = 258$, the number of immigrant points.

- $\mathcal{M}_1$: ETAS model (constant immigrant intensity, Lomax offspring density)
- $\mathcal{M}_2$: constant immigrant intensity, uniform DP mixture offspring density
- $\mathcal{M}_3$: Erlang mixture immigrant intensity, uniform DP mixture offspring density

Posterior predictive inference

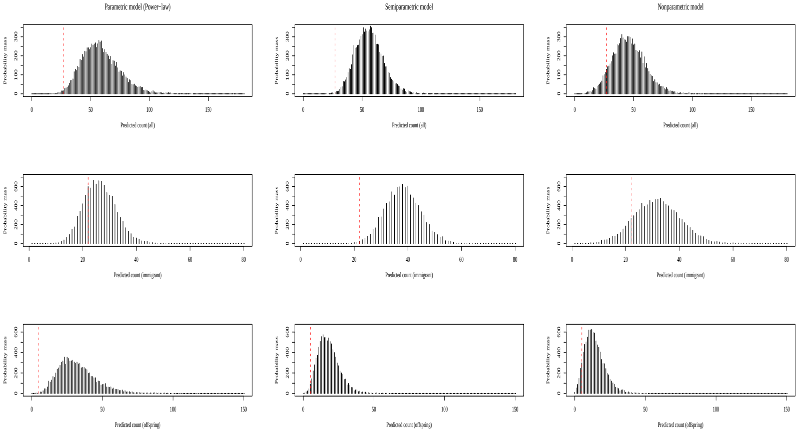

Figure: Posterior predictive distributions, under models $\mathcal{M}_1$ (left), $\mathcal{M}_2$ (middle), and $\mathcal{M}_3$ (right), for the total count (first row), immigrant count (second row), and offspring count (third row) of earthquakes that occurred in (1970, 1980). The red dashed line in each panel indicates the corresponding observed count (22 main shocks, and 5 aftershocks).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) = \mu(t) + \sum_{t_i < t} h(t - t_i, \kappa_i) = \mu(t) + \sum_{t_i < t} \alpha(\kappa_i)\, g_{\kappa_i}(t - t_i)$$

where $h(t, \kappa)$, for $(t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

- ▷ total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u, \kappa)\, \mathrm{d}u < \infty$, for all κ;
- ▷ branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\, \mathrm{d}\kappa < \infty$ ($\rho \in (0, 1)$ for stability);
- ▷ mark-dependent offspring density:

$$g_\kappa(t) = h(t, \kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

(density on $\mathbb{R}^+$ for any κ).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t-t_i,\kappa_i) = \mu(t) + \sum_{t_i<t} \alpha(\kappa_i)\, g_{\kappa_i}(t-t_i)$$

where $h(t,\kappa)$, for $(t,\kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

  - total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u,\kappa)\,\mathrm{d}u < \infty$, for all κ;
  - branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\,\mathrm{d}\kappa < \infty$ ($\rho \in (0,1)$ for stability);
  - mark-dependent offspring density:

$$g_\kappa(t) = h(t,\kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

(density on $\mathbb{R}^+$ for any κ).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) \;=\; \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i) \;=\; \mu(t) + \sum_{t_i<t} \alpha(\kappa_i)\, g_{\kappa_i}(t - t_i)$$

  where $h(t, \kappa)$, for $(t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

  - total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u, \kappa)\, \mathrm{d}u < \infty$, for all κ;
  - branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\, \mathrm{d}\kappa < \infty$ ($\rho \in (0, 1)$ for stability);
  - mark-dependent offspring density:

$$g_\kappa(t) \;=\; h(t, \kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

    (density on $\mathbb{R}^+$ for any κ).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i) = \mu(t) + \sum_{t_i<t} \alpha(\kappa_i)\, g_{\kappa_i}(t - t_i)$$

where $h(t, \kappa)$, for $(t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

- ▷ total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u, \kappa)\, \mathrm{d}u < \infty$, for all κ;
- ▷ branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\, \mathrm{d}\kappa < \infty$ ($\rho \in (0, 1)$ for stability);
- ▷ mark-dependent offspring density:

$$g_\kappa(t) = h(t, \kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

(density on $\mathbb{R}^+$ for any κ).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) \;=\; \mu(t) + \sum_{t_i<t} h(t-t_i, \kappa_i) \;=\; \mu(t) + \sum_{t_i<t} \alpha(\kappa_i)\, g_{\kappa_i}(t-t_i)$$

where $h(t, \kappa)$, for $(t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

  - total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u, \kappa)\, \mathrm{d}u < \infty$, for all κ;
  - branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\, \mathrm{d}\kappa < \infty$ ($\rho \in (0, 1)$ for stability);
  - mark-dependent offspring density:

$$g_\kappa(t) \;=\; h(t, \kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

(density on $\mathbb{R}^+$ for any κ).

Marked Hawkes processes

- Methodology in the context of earthquake modeling; focus on earthquake magnitude as the mark variable, κ, with values in mark space $\mathcal{K}$.
- Marked HP stochastic model built from: ground process intensity; and, mark density (*unpredictable marks* $\rightarrow f^*(\kappa \mid t)$, or, simpler, $f^*(\kappa)$).
- Ground process intensity function

$$\lambda_g^*(t) \;=\; \mu(t) + \sum_{t_i < t} h(t - t_i, \kappa_i) \;=\; \mu(t) + \sum_{t_i < t} \alpha(\kappa_i)\, g_{\kappa_i}(t - t_i)$$

where $h(t, \kappa)$, for $(t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$, is a non-negative function such that:

  - total offspring intensity: $\alpha(\kappa) = \int_0^\infty h(u, \kappa)\, \mathrm{d}u < \infty$, for all κ;
  - branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\, \mathrm{d}\kappa < \infty$ ($\rho \in (0, 1)$ for stability);
  - mark-dependent offspring density:

$$g_\kappa(t) \;=\; h(t, \kappa)/\alpha(\kappa), \quad t \in \mathbb{R}^+$$

(density on $\mathbb{R}^+$ for any κ).

Marked ETAS model

- ETAS model for earthquake occurrences:
  - ▷ exponential mark density $f^*(\kappa)$ (not dependent on current time t);
  - ▷ $\mu(t) \equiv \mu$, constant immigrant intensity;
  - ▷ $\alpha(\kappa) = \alpha_1 \exp(\alpha_2 \kappa)$, with $\alpha_2 > 0$, total offspring intensity exponentially increasing with magnitude;
  - ▷ $h(t, \kappa) = \alpha(\kappa)\, g(t)$, with

$$g(t) = \phi\, \theta^{\phi} (\theta + t)^{-(\phi+1)}, \quad t \in \mathbb{R}^+$$

    i.e., Lomax offspring density that does not change with magnitude.

- Key motivation: modeling that allows for mark-dependent offspring densities, i.e., enables estimation of aftershock densities that vary with the magnitude of the main shock.

Marked ETAS model

- ETAS model for earthquake occurrences:
  - exponential mark density $f^*(\kappa)$ (not dependent on current time t);
  - $\mu(t) \equiv \mu$, constant immigrant intensity;
  - $\alpha(\kappa) = \alpha_1 \exp(\alpha_2 \kappa)$, with $\alpha_2 > 0$, total offspring intensity exponentially increasing with magnitude;
  - $h(t, \kappa) = \alpha(\kappa)\, g(t)$, with

$$g(t) = \phi\, \theta^{\phi} (\theta + t)^{-(\phi+1)}, \quad t \in \mathbb{R}^+$$

    i.e., Lomax offspring density that does not change with magnitude.

- Key motivation: modeling that allows for mark-dependent offspring densities, i.e., enables estimation of aftershock densities that vary with the magnitude of the main shock.

Marked ETAS model

- ETAS model for earthquake occurrences:
  - exponential mark density $f^*(\kappa)$ (not dependent on current time t);
  - $\mu(t) \equiv \mu$, constant immigrant intensity;
  - $\alpha(\kappa) = \alpha_1 \exp(\alpha_2 \kappa)$, with $\alpha_2 > 0$, total offspring intensity exponentially increasing with magnitude;
  - $h(t, \kappa) = \alpha(\kappa)\, g(t)$, with

    $$g(t) \;=\; \phi\, \theta^{\phi} (\theta + t)^{-(\phi+1)}, \quad t \in \mathbb{R}^+$$

    i.e., Lomax offspring density that does not change with magnitude.

- Key motivation: modeling that allows for mark-dependent offspring densities, i.e., enables estimation of aftershock densities that vary with the magnitude of the main shock.

Marked ETAS model

- ETAS model for earthquake occurrences:
  - exponential mark density $f^*(\kappa)$ (not dependent on current time t);
  - $\mu(t) \equiv \mu$, constant immigrant intensity;
  - $\alpha(\kappa) = \alpha_1 \exp(\alpha_2 \kappa)$, with $\alpha_2 > 0$, total offspring intensity exponentially increasing with magnitude;
  - $h(t, \kappa) = \alpha(\kappa)\, g(t)$, with

$$g(t) \;=\; \phi\, \theta^{\phi} (\theta + t)^{-(\phi+1)}, \quad t \in \mathbb{R}^+$$

    i.e., Lomax offspring density that does not change with magnitude.

- Key motivation: modeling that allows for mark-dependent offspring densities, i.e., enables estimation of aftershock densities that vary with the magnitude of the main shock.

Modeling the mark-dependent excitation function

- Prior model for the excitation function $\rightarrow \lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i)$
- Take $\mathcal{K} = [k_0, k_{\max}]$ as the support for the earthquake magnitude distribution.
- Weighted combination of basis functions, constructed by the product of Erlang densities and polynomial functions:

$$h(t,\kappa) = \sum_{\ell=1}^{L}\sum_{m=1}^{M} \omega_{\ell m}\, \mathrm{ga}(t \mid \ell, \theta)\, b_m(\kappa), \qquad (t,\kappa) \in \mathbb{R}^+ \times \mathcal{K}$$

  - $b_m(\kappa) \propto \{z(\kappa)\}^{m-1}$, where $z(\kappa) = (\kappa - k_0)/(k_{\max} - k_0) \in (0,1)$.
  - $\omega_{\ell m} = H(A_{\ell m})$, where $A_{\ell m} = [(\ell-1)\theta, \ell\theta) \times [(m-1)/M, m/M)$, for $\ell = 1, \ldots, L$ and $m = 1, \ldots, M$.
  - H is a random measure on $\mathbb{R}^+ \times (0,1)$ assigned a gamma process prior, $H \sim \mathcal{G}(H_0, c_0)$.

Modeling the mark-dependent excitation function

- Prior model for the excitation function $\rightarrow \lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i)$
- Take $\mathcal{K} = [k_0, k_{\max}]$ as the support for the earthquake magnitude distribution.
- Weighted combination of basis functions, constructed by the product of Erlang densities and polynomial functions:

$$h(t, \kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m} \, \mathrm{ga}(t \mid \ell, \theta) \, b_m(\kappa), \quad (t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$$

  - $b_m(\kappa) \propto \{z(\kappa)\}^{m-1}$, where $z(\kappa) = (\kappa - k_0)/(k_{\max} - k_0) \in (0, 1)$.
  - $\omega_{\ell m} = H(A_{\ell m})$, where $A_{\ell m} = [(\ell - 1)\theta, \ell\theta) \times [(m - 1)/M, m/M)$, for $\ell = 1, \ldots, L$ and $m = 1, \ldots, M$.
  - H is a random measure on $\mathbb{R}^+ \times (0, 1)$ assigned a gamma process prior, $H \sim \mathcal{G}(H_0, c_0)$.

Modeling the mark-dependent excitation function

- Prior model for the excitation function $\rightarrow \lambda_g^*(t) = \mu(t) + \sum_{t_i < t} h(t - t_i, \kappa_i)$
- Take $\mathcal{K} = [k_0, k_{\max}]$ as the support for the earthquake magnitude distribution.
- Weighted combination of basis functions, constructed by the product of Erlang densities and polynomial functions:

$$h(t, \kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m} \, \mathrm{ga}(t \mid \ell, \theta) \, b_m(\kappa), \quad (t, \kappa) \in \mathbb{R}^+ \times \mathcal{K}$$

  - $b_m(\kappa) \propto \{z(\kappa)\}^{m-1}$, where $z(\kappa) = (\kappa - k_0)/(k_{\max} - k_0) \in (0, 1)$.
  - $\omega_{\ell m} = H(A_{\ell m})$, where $A_{\ell m} = [(\ell - 1)\theta, \ell\theta) \times [(m-1)/M, m/M)$, for $\ell = 1, \ldots, L$ and $m = 1, \ldots, M$.
  - H is a random measure on $\mathbb{R}^+ \times (0, 1)$ assigned a gamma process prior, $H \sim \mathcal{G}(H_0, c_0)$.

Modeling the mark-dependent excitation function

- Prior model for the excitation function $\rightarrow \lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i)$
- Take $\mathcal{K} = [k_0, k_{\max}]$ as the support for the earthquake magnitude distribution.
- Weighted combination of basis functions, constructed by the product of Erlang densities and polynomial functions:

$$h(t,\kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m} \, \mathrm{ga}(t \mid \ell, \theta) \, b_m(\kappa), \qquad (t,\kappa) \in \mathbb{R}^+ \times \mathcal{K}$$

  - $b_m(\kappa) \propto \{z(\kappa)\}^{m-1}$, where $z(\kappa) = (\kappa - k_0)/(k_{\max} - k_0) \in (0,1)$.
  - $\omega_{\ell m} = H(A_{\ell m})$, where $A_{\ell m} = [(\ell-1)\theta, \ell\theta) \times [(m-1)/M, m/M)$, for $\ell = 1, \ldots, L$ and $m = 1, \ldots, M$.
  - H is a random measure on $\mathbb{R}^+ \times (0,1)$ assigned a gamma process prior, $H \sim \mathcal{G}(H_0, c_0)$.

Modeling the mark-dependent excitation function

- Prior model for the excitation function $\rightarrow \lambda_g^*(t) = \mu(t) + \sum_{t_i<t} h(t - t_i, \kappa_i)$
- Take $\mathcal{K} = [k_0, k_{\max}]$ as the support for the earthquake magnitude distribution.
- Weighted combination of basis functions, constructed by the product of Erlang densities and polynomial functions:

$$h(t,\kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m} \, \mathrm{ga}(t \mid \ell, \theta) \, b_m(\kappa), \quad (t,\kappa) \in \mathbb{R}^+ \times \mathcal{K}$$

  - ▷ $b_m(\kappa) \propto \{z(\kappa)\}^{m-1}$, where $z(\kappa) = (\kappa - k_0)/(k_{\max} - k_0) \in (0,1)$.
  - ▷ $\omega_{\ell m} = H(A_{\ell m})$, where $A_{\ell m} = [(\ell-1)\theta, \ell\theta) \times [(m-1)/M, m/M)$, for $\ell = 1, \ldots, L$ and $m = 1, \ldots, M$.
  - ▷ H is a random measure on $\mathbb{R}^+ \times (0,1)$ assigned a gamma process prior, $H \sim \mathcal{G}(H_0, c_0)$.

Marked HP functionals

- Total offspring intensity:

$$\alpha(\kappa) = \int_0^\infty h(u,\kappa)\,\mathrm{d}u = \sum_{m=1}^{M} \left\{ \sum\nolimits_{\ell=1}^{L} \omega_{\ell m} \right\} b_m(\kappa) = \sum_{m=1}^{M} q_m\, b_m(\kappa)$$

an increasing function of earthquake magnitude.

- Mark-dependent offspring density:

$$g_\kappa(t) = \sum_{\ell=1}^{L} \left\{ \frac{\sum_{m=1}^{M} \omega_{\ell m}\, b_m(\kappa)}{\sum_{m=1}^{M} q_m\, b_m(\kappa)} \right\} \mathrm{ga}(t \mid \ell, \theta) = \sum_{\ell=1}^{L} \psi_\ell(\kappa)\, \mathrm{ga}(t \mid \ell, \theta)$$

a weighted combination of Erlang densities with magnitude-dependent weights.

- Branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\,\mathrm{d}\kappa = \sum_{m=1}^{M} q_m \left\{ \int_{k_0}^{k_{\max}} b_m(\kappa) f^*(\kappa)\,\mathrm{d}\kappa \right\}$,

with the integral available if f^* is a (rescaled) Beta, or Beta mixture, density.

  - By construction, $\alpha(\kappa) < \infty$ and $\rho < \infty$, and $\rho \in (0,1)$ can be achieved with large prior probability.

Marked HP functionals

- Total offspring intensity:

$$\alpha(\kappa) = \int_0^\infty h(u,\kappa)\,\mathrm{d}u = \sum_{m=1}^{M} \left\{ \sum\nolimits_{\ell=1}^{L} \omega_{\ell m} \right\} b_m(\kappa) = \sum_{m=1}^{M} q_m\, b_m(\kappa)$$

an increasing function of earthquake magnitude.

- Mark-dependent offspring density:

$$g_\kappa(t) = \sum_{\ell=1}^{L} \left\{ \frac{\sum_{m=1}^{M} \omega_{\ell m}\, b_m(\kappa)}{\sum_{m=1}^{M} q_m\, b_m(\kappa)} \right\} \mathrm{ga}(t \mid \ell, \theta) = \sum_{\ell=1}^{L} \psi_\ell(\kappa)\, \mathrm{ga}(t \mid \ell, \theta)$$

a weighted combination of Erlang densities with magnitude-dependent weights.

- Branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\,\mathrm{d}\kappa = \sum_{m=1}^{M} q_m \left\{ \int_{k_0}^{k_{\max}} b_m(\kappa) f^*(\kappa)\,\mathrm{d}\kappa \right\}$,

  with the integral available if f^* is a (rescaled) Beta, or Beta mixture, density.

  - By construction, $\alpha(\kappa) < \infty$ and $\rho < \infty$, and $\rho \in (0,1)$ can be achieved with large prior probability.

Marked HP functionals

- Total offspring intensity:

$$\alpha(\kappa) = \int_0^\infty h(u,\kappa)\,\mathrm{d}u = \sum_{m=1}^{M} \left\{ \sum\nolimits_{\ell=1}^{L} \omega_{\ell m} \right\} b_m(\kappa) = \sum_{m=1}^{M} q_m\, b_m(\kappa)$$

an increasing function of earthquake magnitude.

- Mark-dependent offspring density:

$$g_\kappa(t) = \sum_{\ell=1}^{L} \left\{ \frac{\sum_{m=1}^{M} \omega_{\ell m}\, b_m(\kappa)}{\sum_{m=1}^{M} q_m\, b_m(\kappa)} \right\} \mathrm{ga}(t \mid \ell, \theta) = \sum_{\ell=1}^{L} \psi_\ell(\kappa)\, \mathrm{ga}(t \mid \ell, \theta)$$

a weighted combination of Erlang densities with magnitude-dependent weights.

- Branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\,\mathrm{d}\kappa = \sum_{m=1}^{M} q_m \left\{ \int_{k_0}^{k_{\max}} b_m(\kappa) f^*(\kappa)\,\mathrm{d}\kappa \right\}$,
with the integral available if f^* is a (rescaled) Beta, or Beta mixture, density.
  - ▷ By construction, $\alpha(\kappa) < \infty$ and $\rho < \infty$, and $\rho \in (0,1)$ can be achieved with large prior probability.

Marked HP functionals

- Total offspring intensity:

$$\alpha(\kappa) = \int_0^\infty h(u,\kappa)\,\mathrm{d}u = \sum_{m=1}^{M} \left\{ \sum\nolimits_{\ell=1}^{L} \omega_{\ell m} \right\} b_m(\kappa) = \sum_{m=1}^{M} q_m\, b_m(\kappa)$$

an increasing function of earthquake magnitude.

- Mark-dependent offspring density:

$$g_\kappa(t) = \sum_{\ell=1}^{L} \left\{ \frac{\sum_{m=1}^{M} \omega_{\ell m}\, b_m(\kappa)}{\sum_{m=1}^{M} q_m\, b_m(\kappa)} \right\} \mathrm{ga}(t \mid \ell, \theta) = \sum_{\ell=1}^{L} \psi_\ell(\kappa)\, \mathrm{ga}(t \mid \ell, \theta)$$

a weighted combination of Erlang densities with magnitude-dependent weights.

- Branching ratio: $\rho = \int_{\mathcal{K}} \alpha(\kappa) f^*(\kappa)\,\mathrm{d}\kappa = \sum_{m=1}^{M} q_m \left\{ \int_{k_0}^{k_{\max}} b_m(\kappa) f^*(\kappa)\,\mathrm{d}\kappa \right\}$,
with the integral available if f^* is a (rescaled) Beta, or Beta mixture, density.
  - ▷ By construction, $\alpha(\kappa) < \infty$ and $\rho < \infty$, and $\rho \in (0,1)$ can be achieved with large prior probability.

Earthquake data analysis (revisited)

- Models are fitted to the subset of the data from year 1885 to 1950 (340 earthquakes: 192 main shocks, 148 aftershocks), using the remaining portion of the catalog (118 earthquakes: 66 main shocks, 52 aftershocks) to test predictions.

- First set of results contrasting three models, all based on constant immigrant intensity: $\lambda_g^*(t) = \mu + \sum_{t_i<t} h(t - t_i, \kappa_i)$
  - ETAS model: $h(t,\kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ Lomax offspring density
  - Semiparametric model: $h(t,\kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ uniform DP mixture offspring density
  - Nonparametric model: $h(t,\kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m}\, \mathrm{ga}(t \mid \ell, \theta)\, b_m(\kappa)$

- Improved predictive results based on the more general nonparametric model: Erlang mixture for the immigrant intensity $\mu(t)$, combined with the model for the magnitude-dependent excitation function.

Earthquake data analysis (revisited)

- Models are fitted to the subset of the data from year 1885 to 1950 (340 earthquakes: 192 main shocks, 148 aftershocks), using the remaining portion of the catalog (118 earthquakes: 66 main shocks, 52 aftershocks) to test predictions.

- First set of results contrasting three models, all based on constant immigrant intensity: $\lambda_g^*(t) = \mu + \sum_{t_i < t} h(t - t_i, \kappa_i)$
  - ▷ ETAS model: $h(t, \kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ Lomax offspring density
  - ▷ Semiparametric model: $h(t, \kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ uniform DP mixture offspring density
  - ▷ Nonparametric model: $h(t, \kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m}\, \mathrm{ga}(t \mid \ell, \theta)\, b_m(\kappa)$

- Improved predictive results based on the more general nonparametric model: Erlang mixture for the immigrant intensity $\mu(t)$, combined with the model for the magnitude-dependent excitation function.

Earthquake data analysis (revisited)

- Models are fitted to the subset of the data from year 1885 to 1950 (340 earthquakes: 192 main shocks, 148 aftershocks), using the remaining portion of the catalog (118 earthquakes: 66 main shocks, 52 aftershocks) to test predictions.

- First set of results contrasting three models, all based on constant immigrant intensity: $\lambda_g^*(t) = \mu + \sum_{t_i < t} h(t - t_i, \kappa_i)$
  - ETAS model: $h(t, \kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ Lomax offspring density
  - Semiparametric model: $h(t, \kappa) = \alpha(\kappa)\, g(t)$; $\alpha(\kappa) \rightarrow$ exponentially increasing function; $g(t) \rightarrow$ uniform DP mixture offspring density
  - Nonparametric model: $h(t, \kappa) = \sum_{\ell=1}^{L} \sum_{m=1}^{M} \omega_{\ell m}\, \mathrm{ga}(t \mid \ell, \theta)\, b_m(\kappa)$

- Improved predictive results based on the more general nonparametric model: Erlang mixture for the immigrant intensity $\mu(t)$, combined with the model for the magnitude-dependent excitation function.

Comparison with ETAS and semiparametric models

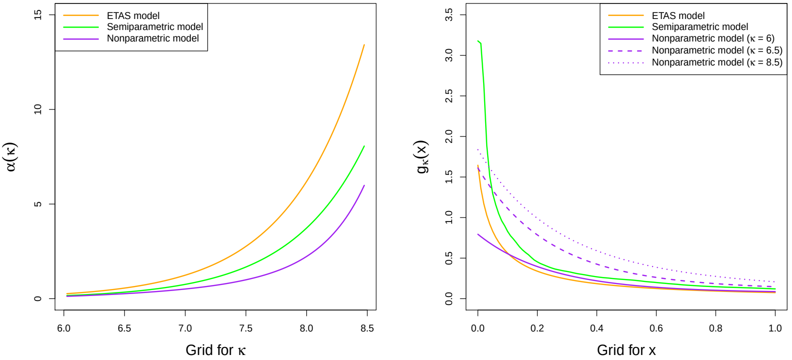

Figure: Posterior mean estimates for the total offspring intensity (left), and the offspring density (right), under the ETAS model, the semiparametric model, and the nonparametric model.

Magnitude-dependent estimates (general model)

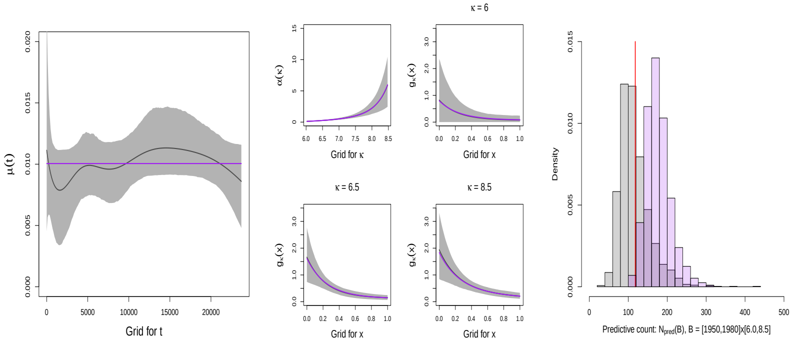

Figure: Posterior mean and 95% interval estimates for: the immigrant intensity (left); and, the total offspring intensity, and offspring densities at three magnitude values (middle). The right panel shows the posterior predictive distribution for the number of earthquakes in (1950, 1980). The red line indicates the observed count. For comparison, the purple histogram plots the posterior predictive distribution under the nonparametric model with constant immigrant intensity.

Summary

- Modeling framework for marked Hawkes processes based on nonparametric priors for the immigrant intensity and the mark-dependent excitation function.

- Papers/manuscripts in progress:
  - Kim, H. and Kottas, A. (2022). "Erlang mixture modeling for Poisson process intensities." *Statistics and Computing*, 32:3.
  - Kim, H. and Kottas, A. "Nonparametric Bayesian modeling and inference for Hawkes processes."
  - Kim, H. and Kottas, A. "Bayesian nonparametric marked Hawkes processes for earthquake modeling."

- Acknowledgment: funding from the NSF under award SES 1950902, and from the UCSC Committee on Research.

Summary

- Modeling framework for marked Hawkes processes based on nonparametric priors for the immigrant intensity and the mark-dependent excitation function.

- Papers/manuscripts in progress:
  - Kim, H. and Kottas, A. (2022). "Erlang mixture modeling for Poisson process intensities." *Statistics and Computing*, 32:3.
  - Kim, H. and Kottas, A. "Nonparametric Bayesian modeling and inference for Hawkes processes."
  - Kim, H. and Kottas, A. "Bayesian nonparametric marked Hawkes processes for earthquake modeling."

- Acknowledgment: funding from the NSF under award SES 1950902, and from the UCSC Committee on Research.

Summary

- Modeling framework for marked Hawkes processes based on nonparametric priors for the immigrant intensity and the mark-dependent excitation function.

- Papers/manuscripts in progress:
  - ▷ Kim, H. and Kottas, A. (2022). "Erlang mixture modeling for Poisson process intensities." *Statistics and Computing*, 32:3.
  - ▷ Kim, H. and Kottas, A. "Nonparametric Bayesian modeling and inference for Hawkes processes."
  - ▷ Kim, H. and Kottas, A. "Bayesian nonparametric marked Hawkes processes for earthquake modeling."

- Acknowledgment: funding from the NSF under award SES 1950902, and from the UCSC Committee on Research.

MANY THANKS !

Fully nonparametric models

- Combine the Erlang mixture for the immigrant intensity with either the Erlang mixture for the excitation function or the uniform mixture for the (decreasing) offspring density.
- The structure of the conditional likelihood (given the branching latent variables) allows combining MCMC updates for the parameters of the prior models for the two functions that define the conditional intensity.
- But, is it possible to effectively estimate both the immigrant and offspring intensity based on nonparametric prior models?
- Results for the model based on the Erlang mixture for the immigrant intensity and for the excitation function, using synthetic data from different scenarios:
  - ▷ constant immigrant intensity, exponential offspring density ($n_I = 156$, $n_O = 627$)
  - ▷ unimodal immigrant intensity, bimodal offspring density ($n_I = 504$, $n_O = 1037$)
  - ▷ bimodal immigrant intensity and offspring density ($n_I = 629$, $n_O = 1060$).

Synthetic data example (fully NPB model)

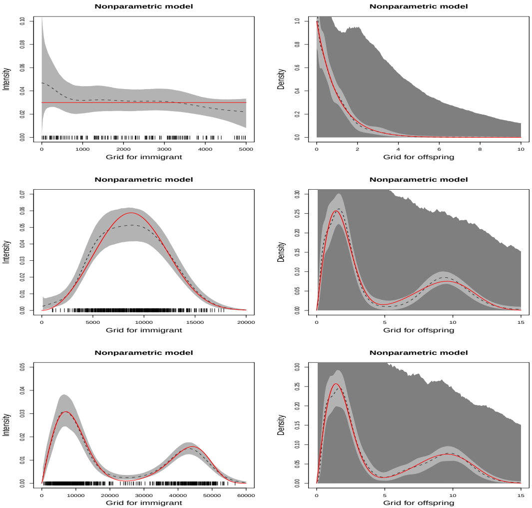

Figure: Posterior mean and interval estimates for the immigrant intensity (left column) and the offspring density (right column), under three scenarios for the true HP functions (red lines).