# Asymptotics for posterior hazards

Igor Prünster

University of Turin, Collegio Carlo Alberto and ICER

Joint work with P. Di Biasi and G. Peccati

Workshop on Limit Theorems and Applications

Paris, 16th January 2008

# Outline

## Life-testing model with mixture hazard rate

- Mixture hazard rate models
- Completely random measures and kernels
- de Finetti's theorem and Bayesian inference
- Asymptotic issues

# Outline

## Life-testing model with mixture hazard rate

- Mixture hazard rate models
- Completely random measures and kernels
- de Finetti's theorem and Bayesian inference
- Asymptotic issues

## Posterior consistency

- Weak consistency
- Sufficient condition for Kullback-Leibler support
- Consistency for specific models

# Outline

## Life-testing model with mixture hazard rate
- Mixture hazard rate models
- Completely random measures and kernels
- de Finetti's theorem and Bayesian inference
- Asymptotic issues

## Posterior consistency
- Weak consistency
- Sufficient condition for Kullback-Leibler support
- Consistency for specific models

## CLTs for functionals of the hazard rate
- CLTs for the cumulative hazard and applications
- CLTs for the path–variance and applications
- CLTs for posterior functionals

# Mixture hazard rate

Let $Y$ be a positive absolutely continuous r.v. representing lifetime.
Assume that its random hazard rate is of the form

$$\tilde{h}(t) = \int_{\mathbb{X}} k(t, x)\tilde{\mu}(\mathrm{d}x)$$

- $k$ is a kernel
- the mixing measure $\tilde{\mu}$ is modeled as a completely random measure.

# Mixture hazard rate

Let $Y$ be a positive absolutely continuous r.v. representing lifetime.
Assume that its random hazard rate is of the form

$$\tilde{h}(t) = \int_{\mathbb{X}} k(t,x)\tilde{\mu}(\mathrm{d}x)$$

- $k$ is a kernel
- the mixing measure $\tilde{\mu}$ is modeled as a completely random measure.

*Remark.* Given $\tilde{\mu}$, $\tilde{h}$ represents the hazard rate of $Y$, that is

$$\tilde{h}(t)\,\mathrm{d}t = \mathbb{P}(t \le Y \le t + \mathrm{d}t | Y \ge t, \tilde{\mu}).$$

# Mixture hazard rate

Let $Y$ be a positive absolutely continuous r.v. representing lifetime. Assume that its random hazard rate is of the form

$$\tilde{h}(t) = \int_{\mathbb{X}} k(t, x)\tilde{\mu}(\mathrm{d}x)$$

- $k$ is a kernel
- the mixing measure $\tilde{\mu}$ is modeled as a completely random measure.

*Remark.* Given $\tilde{\mu}$, $\tilde{h}$ represents the hazard rate of $Y$, that is

$$\tilde{h}(t)\,\mathrm{d}t = \mathbb{P}(t \leq Y \leq t + \mathrm{d}t | Y \geq t, \tilde{\mu}).$$

The cumulative hazard is given by $\tilde{H}(t) = \int_0^t \tilde{h}(s)\mathrm{d}s$. Provided $\tilde{H}(t) \to \infty$ as $t \to \infty$ a.s., one can define a random density function as

$$\tilde{f}(t) = \tilde{h}(t)\ \exp(-\tilde{H}(t))$$

$\Longrightarrow$ Life testing model. See Dykstra and Laud (1981), Lo and Weng (1989), Ishwaran and James (2004) James (2005) for Bayesian nonparametric treatments, among others.

# Completely random measures

Let $\mathbb{M}$ be the space of boundedly finite measures on some Polish space $\mathbb{X}$.

**Definition.** (Kingman, 1967) A random element $\tilde{\mu}$ taking values in $\mathbb{M}$ such that, for any disjoint sets $(A_i)_{i\geq 1}$,

$$\tilde{\mu}(A_1), \tilde{\mu}(A_2), \ldots \qquad \text{are mutually independent}$$

is said to be a completely random measure (CRM) on $\mathbb{X}$.

# Completely random measures

Let $\mathbb{M}$ be the space of boundedly finite measures on some Polish space $\mathbb{X}$.

**Definition.** (Kingman, 1967) A random element $\tilde{\mu}$ taking values in $\mathbb{M}$ such that, for any disjoint sets $(A_i)_{i\geq 1}$,

$$\tilde{\mu}(A_1), \tilde{\mu}(A_2), \ldots \qquad \text{are mutually independent}$$

is said to be a completely random measure (CRM) on $\mathbb{X}$.

**Remark.** A CRM can always be represented as a linear functional of a Poisson random measure. In particular, $\tilde{\mu}$ is uniquely characterized by its *Laplace functional*, which is given by

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} g(x)\,\tilde{\mu}(\mathrm{d}x)}\right] = \mathrm{e}^{-\int_{\mathbb{R}^+\times\mathbb{X}}[1-\mathrm{e}^{-v\,g(x)}]\nu(\mathrm{d}v,\mathrm{d}x)} \qquad (\star)$$

for any $\mathbb{R}^+$–valued $g \in \mathscr{G}_\nu := \{g : \int_{\mathbb{R}^+\times\mathbb{X}}[1-\mathrm{e}^{-v\,g(x)}]\nu(\mathrm{d}v,\mathrm{d}x) < \infty\}$. In $(\star)$ $\nu$ stands for the intensity of the Poisson random measure and it will be used to characterize the corresponding CRM $\tilde{\mu}$.

- CRMs select almost surely discrete measures: hence, $\tilde{\mu}$ can always be represented as $\sum_{i \geq 1} J_i \delta_{X_i}$.

- CRMs select almost surely discrete measures: hence, $\tilde{\mu}$ can always be represented as $\sum_{i\geq 1} J_i \delta_{X_i}$.
- If $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\lambda(\mathrm{d}x)$ the law of the $J_i$'s and the $X_i$'s are independent and $\tilde{\mu}$ is termed *homogeneous CRM*.

  On the other hand, if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v \mid x)\lambda(\mathrm{d}x)$, we have a *non-homogeneous* CRM. Note that in the posterior we will always have non–homogeneous CRMs.
- If $\nu(\mathbb{R}^+, \mathrm{d}x) = \infty$ for any $x$, then the CRM jumps infinitely often on any bounded set $A$. However, recall that $\tilde{\mu}(A) < \infty$ a.s.

- CRMs select almost surely discrete measures: hence, $\tilde{\mu}$ can always be represented as $\sum_{i\geq 1} J_i \delta_{X_i}$.
- If $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\lambda(\mathrm{d}x)$ the law of the $J_i$'s and the $X_i$'s are independent and $\tilde{\mu}$ is termed *homogeneous CRM*.
  On the other hand, if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v \mid x)\lambda(\mathrm{d}x)$, we have a *non-homogeneous* CRM. Note that in the posterior we will always have non–homogeneous CRMs.
- If $\nu(\mathbb{R}^+, \mathrm{d}x) = \infty$ for any $x$, then the CRM jumps infinitely often on any bounded set $A$. However, recall that $\tilde{\mu}(A) < \infty$ a.s.
- Specific hazard rates will be obtained by considering CRMs with Poisson intensity measures of the following form:

$$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{1}{\Gamma(1-\sigma)} \frac{\mathrm{e}^{-\gamma(x)v}}{v^{1+\sigma}} \mathrm{d}v\, \lambda(\mathrm{d}x),$$

  where $\sigma \in [0, 1)$, $\gamma$ is a strictly positive function and $\lambda$ a $\sigma$–finite measure on $\mathbb{X}$. Such measures are termed non–homogeneous generalized gamma (NHGG) CRMs.
  If $\gamma$ is constant we obtain the *generalized gamma CRMs* (Brix, 1999), whereas if $\sigma = 0$ we have the *extended gamma measure* (Dykstra and Laud, 1981; Lo and Weng, 1989).

# Kernels

A kernel $k$ is a jointly measurable application from $\mathbb{R}^+ \times \mathbb{X}$ to $\mathbb{R}^+$, such that $\int_{\mathbb{X}} k(t,x)\lambda(\mathrm{d}x) < +\infty$ and $\int_{\cdot} k(t,x)\mathrm{d}t$ is a $\sigma$–finite measure on $\mathscr{B}(\mathbb{R}^+)$ for any $x$ in $\mathbb{X}$.

# Kernels

A kernel $k$ is a jointly measurable application from $\mathbb{R}^+ \times \mathbb{X}$ to $\mathbb{R}^+$, such that $\int_{\mathbb{X}} k(t,x)\lambda(\mathrm{d}x) < +\infty$ and $\int_{\cdot} k(t,x)\mathrm{d}t$ is a $\sigma$–finite measure on $\mathscr{B}(\mathbb{R}^+)$ for any $x$ in $\mathbb{X}$.

We will consider mixture hazard rates arising with the following specific kernels:

- Dykstra-Laud (DL) kernel (*monotone increasing hazard rates*)

$$k(t,x) = \mathbb{I}_{(0\le x\le t)}$$

- rectangular (rect) kernel with bandwidth $\tau > 0$

$$k(t,x) = \mathbb{I}_{(|t-x|\le\tau)}$$

- Ornstein–Uhlenbeck (OU) kernel with $\kappa > 0$

$$k(t,x) = \sqrt{2\kappa}\,\mathrm{e}^{-\kappa(t-x)}\mathbb{I}_{(0\le x\le t)}$$

- exponential (exp) kernel (*monotone decreasing hazard rates*)

$$k(t,x) = x^{-1}\mathrm{e}^{-\frac{t}{x}}.$$

# de Finetti's theorem and Bayesian inference

An (ideally) infinite sequence of absolutely continuous $\mathbb{R}^+$–valued observations $\boldsymbol{Y}^{(\infty)} = (Y_i)_{i\geq 1}$ is exchangeable if and only if there exists a probability measure $Q$ (de Finetti measure) on the space $\mathbb{F}$ of all density functions on $\mathbb{R}^+$ such that

$$\mathbb{P}\left[\boldsymbol{Y}^{(\infty)} \in A\right] = \int_{\mathbb{F}} \prod_{i=1}^{n} f(A_i)\, Q(\mathrm{d}f)$$

for any $n \geq 1$ and $A = A_1 \times \cdots \times A_n \times \mathbb{R}^+ \times \ldots$, where $f(A_i) = \int_{A_i} f(x)\mathrm{d}x$.

# de Finetti's theorem and Bayesian inference

An (ideally) infinite sequence of absolutely continuous $\mathbb{R}^+$–valued observations $\boldsymbol{Y}^{(\infty)} = (Y_i)_{i\geq 1}$ is exchangeable if and only if there exists a probability measure $Q$ (de Finetti measure) on the space $\mathbb{F}$ of all density functions on $\mathbb{R}^+$ such that

$$\mathbb{P}\left[\boldsymbol{Y}^{(\infty)} \in A\right] = \int_{\mathbb{F}} \prod_{i=1}^{n} f(A_i)\, Q(\mathrm{d}f)$$

for any $n \geq 1$ and $A = A_1 \times \cdots \times A_n \times \mathbb{R}^+ \times \ldots$, where $f(A_i) = \int_{A_i} f(x)\mathrm{d}x$.

Consider now the random density $\tilde{f} = \tilde{h}\,\mathrm{e}^{-\tilde{H}}$ and denote by $\Pi$ its law on $\mathbb{F}$ (which depends on the kernel $k$ and the CRM $\tilde{\mu}$). Thus, our inferential model amounts to assuming the lifetimes $Y_i$'s to be exchangeable with de Finetti measure $\Pi$.

# de Finetti's theorem and Bayesian inference

An (ideally) infinite sequence of absolutely continuous $\mathbb{R}^+$–valued observations $\boldsymbol{Y}^{(\infty)} = (Y_i)_{i\geq 1}$ is exchangeable if and only if there exists a probability measure $Q$ (de Finetti measure) on the space $\mathbb{F}$ of all density functions on $\mathbb{R}^+$ such that

$$\mathbb{P}\left[\boldsymbol{Y}^{(\infty)} \in A\right] = \int_{\mathbb{F}} \prod_{i=1}^{n} f(A_i)\, Q(\mathrm{d}f)$$

for any $n \geq 1$ and $A = A_1 \times \cdots \times A_n \times \mathbb{R}^+ \times \ldots$, where $f(A_i) = \int_{A_i} f(x)\mathrm{d}x$.

Consider now the random density $\tilde{f} = \tilde{h}\,\mathrm{e}^{-\tilde{H}}$ and denote by $\Pi$ its law on $\mathbb{F}$ (which depends on the kernel $k$ and the CRM $\tilde{\mu}$). Thus, our inferential model amounts to assuming the lifetimes $Y_i$'s to be exchangeable with de Finetti measure $\Pi$.

Given a set of observations $\boldsymbol{Y}^n = (Y_1, \ldots, Y_n)$, the posterior distribution is

$$\Pi(\mathrm{d}f|\boldsymbol{Y}^n) = \frac{\prod_{i=1}^n f(Y_i)\Pi(\mathrm{d}f)}{\int_{\mathbb{F}} \prod_{i=1}^n f(Y_i)\Pi(\mathrm{d}f)} \qquad (\bullet)$$

Then, the Bayes estimator of the density function of $Y$ is

$$\hat{f}_n(t) = \mathbb{E}[\tilde{f}(t)|\boldsymbol{Y}^n] = \int_{\mathbb{F}} f(t)\Pi(\mathrm{d}f|\boldsymbol{Y}^n)$$

$\Longrightarrow$ For concrete implementation, an explicit representation of $(\bullet)$ is needed.

# Asymptotic issues

Posterior consistency. First generate independent data from a "true" fixed density $f_0$, then check whether the sequence of posterior distributions of $\tilde{f}$ accumulates in some suitable neighborhood of $f_0$.

$\Longrightarrow$ *Which conditions on $\tilde{f}$ are needed in order to achieve consistency for large classes of $f_0$'s?*

# Asymptotic issues

Posterior consistency. First generate independent data from a “true” fixed density $f_0$, then check whether the sequence of posterior distributions of $\tilde{f}$ accumulates in some suitable neighborhood of $f_0$.

$\Longrightarrow$ *Which conditions on $\tilde{f}$ are needed in order to achieve consistency for large classes of $f_0$'s?*

CLTs for functionals of the random hazard. For a fixed $T > 0$, consider the functionals (i) $\tilde{H}(T)$ and (ii) $T^{-1}\int_0^T [\tilde{h}(t) - \tilde{H}(T)/T]^2 \mathrm{d}t$ (*path-variance*).

We are interested in establishing *Central Limit Theorems* of the type

$$\eta(T) \times [\tilde{H}(T) - \tau(T)] \xrightarrow{\text{law}} \mathcal{N}(0, \sigma^2) \quad \text{as } T \to +\infty$$

for appropriate positive functions $\tau(T)$ and $\eta(T)$ and variance $\sigma^2$.

$\Longrightarrow$ $\tau(\cdot)$, $\eta(\cdot)$ *and* $\sigma^2$ *provide an overall picture of the model.*

# Asymptotic issues

Posterior consistency. First generate independent data from a "true" fixed density $f_0$, then check whether the sequence of posterior distributions of $\tilde{f}$ accumulates in some suitable neighborhood of $f_0$.

$\Longrightarrow$ *Which conditions on $\tilde{f}$ are needed in order to achieve consistency for large classes of $f_0$'s?*

CLTs for functionals of the random hazard. For a fixed $T > 0$, consider the functionals (i) $\tilde{H}(T)$ and (ii) $T^{-1}\int_0^T [\tilde{h}(t) - \tilde{H}(T)/T]^2 \mathrm{d}t$ (*path-variance*).

We are interested in establishing *Central Limit Theorems* of the type

$$\eta(T) \times [\tilde{H}(T) - \tau(T)] \xrightarrow{\text{law}} \mathcal{N}(0, \sigma^2) \quad \text{as } T \to +\infty$$

for appropriate positive functions $\tau(T)$ and $\eta(T)$ and variance $\sigma^2$.

$\Longrightarrow$ *$\tau(\cdot)$, $\eta(\cdot)$ and $\sigma^2$ provide an overall picture of the model.*

Moreover, given the observations $\boldsymbol{Y}^n = (Y_1, \ldots, Y_n)$, we also want to derive CLTs for the functionals (i) and (ii) with respect to the posterior distribution.

$\Longrightarrow$ *How are $\tau(\cdot)$, $\eta(\cdot)$ and $\sigma^2$ influenced by the observed data?*

# Weak consistency

Denote by $P_0$ the probability distribution associated with $f_0$ and by $P_0^\infty$ the infinite product measure. Recall that $\Pi$ is the (prior) distribution of the random density function $\tilde{f} = \tilde{h}\mathrm{e}^{-\tilde{H}}$.

*Definition.* $\Pi$ is said to be weakly consistent at $f_0$, if, for any $\epsilon > 0$

$$\Pi\left(W_\epsilon \,\middle|\, \boldsymbol{Y}^n\right) \stackrel{n\to\infty}{\longrightarrow} 1 \qquad \text{a.s.} - P_0^\infty,$$

where $W_\epsilon$ is a $\epsilon$–neighbourhood of $P_0$ in the weak topology.

# Weak consistency

Denote by $P_0$ the probability distribution associated with $f_0$ and by $P_0^\infty$ the infinite product measure. Recall that Π is the (prior) distribution of the random density function $\tilde{f} = \tilde{h}\mathrm{e}^{-\tilde{H}}$.

*Definition.* Π is said to be weakly consistent at $f_0$, if, for any $\epsilon > 0$

$$\Pi(W_\epsilon \,|\, \boldsymbol{Y}^n) \stackrel{n\to\infty}{\longrightarrow} 1 \qquad \text{a.s.} - P_0^\infty,$$

where $W_\epsilon$ is a $\epsilon$–neighbourhood of $P_0$ in the weak topology.

$\Longrightarrow$ "What if" or frequentist approach to Bayesian consistency (Diaconis and Freedman, 1986): the Bayesian paradigm assumes the data to be exchangeable and updates the prior distribution accordingly (typically via Bayes theorem). But "what happens if" the data are not exchangeable but instead i.i.d. from some "true" distribution $P_0$? Does the posterior distribution concentrate in a neighbourhood of $P_0$?

# Weak consistency

Denote by $P_0$ the probability distribution associated with $f_0$ and by $P_0^\infty$ the infinite product measure. Recall that $\Pi$ is the (prior) distribution of the random density function $\tilde{f} = \tilde{h} e^{-\tilde{H}}$.

*Definition.* $\Pi$ is said to be weakly consistent at $f_0$, if, for any $\epsilon > 0$

$$\Pi(W_\epsilon \,|\, \boldsymbol{Y}^n) \stackrel{n\to\infty}{\longrightarrow} 1 \qquad \text{a.s.} - P_0^\infty,$$

where $W_\epsilon$ is a $\epsilon$–neighbourhood of $P_0$ in the weak topology.

$\Longrightarrow$ "What if" or frequentist approach to Bayesian consistency (Diaconis and Freedman, 1986): the Bayesian paradigm assumes the data to be exchangeable and updates the prior distribution accordingly (typically via Bayes theorem). But "what happens if" the data are not exchangeable but instead i.i.d. from some "true" distribution $P_0$? Does the posterior distribution concentrate in a neighbourhood of $P_0$?

*Remark.* A sufficient condition for weak consistency (Schwartz, 1965) requires a prior $\Pi$ to assign positive probability to Kullback–Leibler (K–L) neighborhoods of $f_0$ (K–L condition):

$$\Pi\Big(f \in \mathbb{F} : \int \log(f_0/f) f_0 < \epsilon\Big) > 0 \qquad \text{for any } \epsilon > 0,$$

where $\int \log(f_0/f) f_0$ is the K–L divergence between $f_0$ and $f$.

# General consistency result

The first result translates the K–L condition into a condition in terms of positive prior probability assigned to uniform neighbourhoods of $h_0$ on $[0, T]$, where $h_0$ is the hazard rate associated to $f_0$.

THEOREM

*Assume* (i) $h_0(t) > 0$ *for any* $t \geq 0$,
(ii) $\int_{\mathbb{R}^+} g(t)\, f_0(t)\mathrm{d}t < \infty$ , *where* $g(t) = \max\{\mathbb{E}[\tilde{H}(t)], t\}$.

*Then, a sufficient condition for* $\Pi$ *to be weakly consistent at* $f_0$ *is that*

$$\Pi\left\{h : \ \sup_{0<t\leq T} \left|h(t) - h_0(t)\right| < \delta\right\} > 0$$

*for any finite* $T$ *and positive* $\delta$.

# General consistency result

The first result translates the K–L condition into a condition in terms of positive prior probability assigned to uniform neighbourhoods of $h_0$ on $[0, T]$, where $h_0$ is the hazard rate associated to $f_0$.

THEOREM

*Assume* (i) $h_0(t) > 0$ *for any* $t \geq 0$,
(ii) $\int_{\mathbb{R}^+} g(t)\, f_0(t) \mathrm{d}t < \infty$ *, where* $g(t) = \max\{\mathbb{E}[\tilde{H}(t)], t\}$*.*

*Then, a sufficient condition for* $\Pi$ *to be weakly consistent at* $f_0$ *is that*

$$\Pi\Big\{h : \sup_{0<t\leq T} \big|h(t) - h_0(t)\big| < \delta\Big\} > 0$$

*for any finite* $T$ *and positive* $\delta$*.*

*Remark*.
(I) The theorem holds for any random hazard rate, not only for those with mixture structure we are focusing on here.
(II) For mixture hazards verifying positivity of uniform neighbourhoods of $h_0$ is much simpler than the K–L condition for $f_0$.
(III) The conditions on $f_0$ are not restrictive except for the requirement that $h_0(0) > 0$. Indeed, in many situations it is reasonable to have hazards which start from 0 and, thus, one would also like to check consistency with respect to "true" hazards $h_0$ for which $h_0(0) = 0$.

# Relaxing the condition $h_0(0) > 0$

Allowing $h_0(0) = 0$ can create problems with the K–L condition since $\log(h_0/\tilde{h})$ (hence $\log(f_0/\tilde{f})$) can become arbitrarily large around 0. However, by taking a mixture hazard model and studying the short time behaviour of $\tilde{\mu}$, which determines the way $\tilde{h}$ vanishes in 0, we can remove the condition.

# Relaxing the condition $h_0(0) > 0$

Allowing $h_0(0) = 0$ can create problems with the K–L condition since $\log(h_0/\tilde{h})$ (hence $\log(f_0/\tilde{f})$) can become arbitrarily large around 0. However, by taking a mixture hazard model and studying the short time behaviour of $\tilde{\mu}$, which determines the way $\tilde{h}$ vanishes in 0, we can remove the condition.

PROPOSITION

*Weak consistency holds also with $h_0(0) = 0$ provided there exist $\alpha, r > 0$ such that:*

(a) $\lim_{t\to 0} h_0(t)/t^{\alpha} = 0$;

(b) $\liminf_{t\to 0} \tilde{\mu}((0,t])/t^{r} = \infty \quad \text{a.s.}$ *(short time behaviour condition)*

*and a mild condition on the kernel is satisfied.*
*In particular,* (b) *holds if $\tilde{\mu}$ is a NHGG CRM with $\sigma \in (0,1)$ and $\lambda(\mathrm{d}x) = \mathrm{d}x$.*

Condition (a) requires the "true" hazard to leave the origin not faster than an arbitrary power.
Condition (b) requires $\tilde{\mu}$ to leave the origin at least as fast as a power.
Note that the $\alpha$ and $r$ in (a) and (b) do not need to satisfy any relation.

# Dykstra–Laud mixture hazards

Consider DL mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} \mathbb{I}_{(0 \leq x \leq t)} \tilde{\mu}(\mathrm{d}x) = \tilde{\mu}((0, t])$, which give rise to non–decreasing hazard rates.

THEOREM

*Let $\tilde{h}$ be a mixture hazard with DL kernel with $\tilde{\mu}$ satisfying the short time behaviour condition.*

*Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_1$, where $\mathscr{F}_1$ is defined as the set of densities which satisfy the following conditions:*

(i) $\int_{\mathbb{R}^+} t^2 f_0(t)\mathrm{d}t < \infty$;

(ii) $h_0(0) = 0$ *and* $h_0(t) > 0$ *for any* $t \geq 0$;

(iii) $h_0$ *is non–decreasing*.

$\Longrightarrow$ Recall that the short time behaviour condition holds for NHGG CRMs.

# Dykstra–Laud mixture hazards

Consider DL mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} \mathbb{I}_{(0 \le x \le t)} \tilde{\mu}(\mathrm{d}x) = \tilde{\mu}((0, t])$, which give rise to non–decreasing hazard rates.

THEOREM

*Let $\tilde{h}$ be a mixture hazard with DL kernel with $\tilde{\mu}$ satisfying the short time behaviour condition.*

*Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_1$, where $\mathscr{F}_1$ is defined as the set of densities which satisfy the following conditions:*

(i) $\int_{\mathbb{R}^+} t^2 f_0(t)\mathrm{d}t < \infty$;

(ii) $h_0(0) = 0$ *and* $h_0(t) > 0$ *for any* $t \ge 0$;

(iii) $h_0$ *is non–decreasing*.

$\Longrightarrow$ Recall that the short time behaviour condition holds for NHGG CRMs.

*Remark*. Steps for proving this and the following results (techniques different but the strategy is the same):

(1) prove positivity of uniform neighbourhoods of $h_0$ (and, hence, weak consistency) for "true" $h_0$ of the type $h_0(t) = \int_{\mathbb{R}^+} k(t, x)\mu_0(\mathrm{d}x)$;

(2) show that "true" $h_0$ of the type $h_0(t) = \int_{\mathbb{R}^+} k(t, x)\mu_0(\mathrm{d}x)$ are arbitrarily close in the uniform metric to any $h_0$ belonging to a class of hazards having a suitable qualitative feature.

# Rectangular mixture hazards

Consider rectangular mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} \mathbb{I}_{(|t-x| \leq \tilde{\tau})} \tilde{\mu}(\mathrm{d}x)$, where the bandwidth $\tau$ is treated as a hyper–parameter and an independent prior $\pi$ is assigned to it. So we have two sources of randomness $\tilde{\tau}$, with distribution $\pi$, and $\tilde{\mu}$, with distribution $\mathscr{L}$: hence, the prior distribution $\Pi$ on $\tilde{h}$ is induced by $\pi \times \mathscr{L}$ via the map $(\tau, \mu) \to h(\,\cdot\,|\tau, \mu) := \int \mathbb{I}_{(|\,\cdot\,-x| \leq \tau)} \mu(\mathrm{d}x)$.

# Rectangular mixture hazards

Consider rectangular mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} \mathbb{I}_{(|t-x| \leq \tilde{\tau})} \tilde{\mu}(\mathrm{d}x)$, where the bandwidth $\tau$ is treated as a hyper–parameter and an independent prior $\pi$ is assigned to it. So we have two sources of randomness $\tilde{\tau}$, with distribution $\pi$, and $\tilde{\mu}$, with distribution $\mathscr{L}$: hence, the prior distribution $\Pi$ on $\tilde{h}$ is induced by $\pi \times \mathscr{L}$ via the map $(\tau, \mu) \to h(\,\cdot\,|\tau, \mu) := \int \mathbb{I}_{(|\,\cdot\,-x| \leq \tau)} \mu(\mathrm{d}x)$.

THEOREM

*Let $\tilde{h}$ be a mixture hazard with rectangular kernel and random bandwidth. Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_2$, where $\mathscr{F}_2$ is defined as the set of densities which satisfy the following conditions:*

(i) $\int_{\mathbb{R}^+} t\, f_0(t)\mathrm{d}t < \infty$;

(ii) $h_0(t) > 0$ *for any* $t > 0$;

(iii) *$h_0$ is bounded and Lipschitz*.

# Ornstein–Uhlenbeck mixture hazards

Consider OU mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}+} \sqrt{2\kappa}\, \mathrm{e}^{-\kappa(t-x)} \mathbb{I}_{(0\le x\le t)} \tilde{\mu}(\mathrm{d}x)$.
For any differentiable decreasing function $g$ define the local exponential decay rate as $-g'(y)/g(y)$.

THEOREM

*Let $\tilde{h}$ be a mixture hazard with OU kernel with $\tilde{\mu}$ satisfying the short time behaviour condition.*
*Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_3$, where $\mathscr{F}_3$ is defined as the set of densities which satisfy the following conditions:*
(i) $\int_{\mathbb{R}+} t\, f_0(t)\mathrm{d}t < \infty$;
(ii) $h_0(0) = 0$ *and* $h_0(t) > 0$ *for any* $t \ge 0$;
(iii) $h_0$ *is differentiable and for any* $t > 0$ *such that* $h_0'(t) < 0$ *the corresponding local exponential decay rate is smaller than* $\kappa\sqrt{2\kappa}$.

$\Longrightarrow$ Choosing a large $\kappa$ leads to less smooth trajectories of $\tilde{h}$, but, on the other hand, ensures also consistency w.r.t. to $h_0$'s which have abrupt decays in certain regions.

# Exponential mixture hazards

Finally, consider exponential mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} x^{-1} \mathrm{e}^{-\frac{t}{x}} \tilde{\mu}(\mathrm{d}x)$, which produces decreasing hazards.
Recall that a function $g$ on $\mathbb{R}^+$ is completely monotone if it possesses derivatives $g^{(n)}$ of all orders and $(-1)^n g^{(n)}(y) \geq 0, \qquad y > 0.$

# Exponential mixture hazards

Finally, consider exponential mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} x^{-1} \mathrm{e}^{-\frac{t}{x}} \tilde{\mu}(\mathrm{d}x)$, which produces decreasing hazards.
Recall that a function $g$ on $\mathbb{R}^+$ is completely monotone if it possesses derivatives $g^{(n)}$ of all orders and $(-1)^n g^{(n)}(y) \geq 0, \qquad y > 0$.

**THEOREM**

*Let $\tilde{h}$ be a mixture hazard with exponential kernel such that $\tilde{h}(0) < \infty$ a.s. Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_4$, where $\mathscr{F}_4$ is defined as the set of densities which satisfy the following conditions:*
(i) $\int_{\mathbb{R}^+} t\, f_0(t)\mathrm{d}t < \infty$;
(ii) $h_0(0) < \infty$;
(iii) *$h_0$ is completely monotone*.

*Remark.* By taking a homogeneous CRM with base–measure $\lambda(\mathrm{d}x) = x^{-1/2}\mathrm{e}^{-1/x}(2\sqrt{\pi})^{-1}\mathrm{d}x$, we have $\tilde{h}(0) < \infty$ a.s. and, interestingly, the prior mean of $\tilde{h}$ is centered on a quasi–Weibull hazard i.e. $\mathbb{E}[\tilde{h}(t)] = c(t+1)^{-1/2}$ (nonparametric envelope of a parametric model).

# Exponential mixture hazards

Finally, consider exponential mixture hazards $\tilde{h}(t) = \int_{\mathbb{R}^+} x^{-1} \mathrm{e}^{-\frac{t}{x}} \tilde{\mu}(\mathrm{d}x)$, which produces decreasing hazards.
Recall that a function $g$ on $\mathbb{R}^+$ is completely monotone if it possesses derivatives $g^{(n)}$ of all orders and $(-1)^n g^{(n)}(y) \geq 0, \qquad y > 0$.

THEOREM
*Let $\tilde{h}$ be a mixture hazard with exponential kernel such that $\tilde{h}(0) < \infty$ a.s. Then $\Pi$ is weakly consistent for any $f_0 \in \mathscr{F}_4$, where $\mathscr{F}_4$ is defined as the set of densities which satisfy the following conditions:*
(i) $\int_{\mathbb{R}^+} t\, f_0(t)\mathrm{d}t < \infty$;
(ii) $h_0(0) < \infty$;
(iii) $h_0$ *is completely monotone*.

*Remark.* By taking a homogeneous CRM with base–measure $\lambda(\mathrm{d}x) = x^{-1/2}\mathrm{e}^{-1/x}(2\sqrt{\pi})^{-1}\mathrm{d}x$, we have $\tilde{h}(0) < \infty$ a.s. and, interestingly, the prior mean of $\tilde{h}$ is centered on a quasi–Weibull hazard i.e. $\mathbb{E}[\tilde{h}(t)] = c(t+1)^{-1/2}$ (nonparametric envelope of a parametric model).

*Remark.* All the previous results hold also for data subject to right–censoring.

# CLTs for linear and quadratic functionals

CLTs provide a synthetic picture of the random hazard, which is very useful for understanding the behaviour of the model and the influence of the various parameters $\Longrightarrow$ prior specification

# CLTs for linear and quadratic functionals

CLTs provide a synthetic picture of the random hazard, which is very useful for understanding the behaviour of the model and the influence of the various parameters $\Longrightarrow$ prior specification

We are going to establish CLTs, as $T \to \infty$, of the type

$$\eta_1(T) \times \left[\tilde{H}(T) - \tau_1(T)\right] \xrightarrow{\text{law}} X_1 \sim \mathcal{N}(0, \sigma_1) \quad \text{(Linear functional)}$$

$\Longrightarrow$ How fast does the cumulative hazard diverge from its long–term trend?

$$\eta_2(T) \times \left[\frac{1}{T}\int_0^T \left[\tilde{h}(t) - \frac{\tilde{H}(T)}{T}\right]^2 \mathrm{d}t - \tau_2(T)\right] \xrightarrow{\text{law}} X_2 \sim \mathcal{N}(0, \sigma_2) \quad \text{(Path–variance)}$$

$\Longrightarrow$ How big are the oscillations of $\tilde{h}(t)$ around its average value?

# CLTs for linear and quadratic functionals

CLTs provide a synthetic picture of the random hazard, which is very useful for understanding the behaviour of the model and the influence of the various parameters $\Longrightarrow$ prior specification

We are going to establish CLTs, as $T \to \infty$, of the type

$$\eta_1(T) \times \left[\tilde{H}(T) - \tau_1(T)\right] \xrightarrow{\text{law}} X_1 \sim \mathcal{N}(0, \sigma_1) \quad \text{(Linear functional)}$$

$\Longrightarrow$ How fast does the cumulative hazard diverge from its long–term trend?

$$\eta_2(T) \times \left[\frac{1}{T}\int_0^T \left[\tilde{h}(t) - \frac{\tilde{H}(T)}{T}\right]^2 \mathrm{d}t - \tau_2(T)\right] \xrightarrow{\text{law}} X_2 \sim \mathcal{N}(0, \sigma_2) \quad \text{(Path–variance)}$$

$\Longrightarrow$ How big are the oscillations of $\tilde{h}(t)$ around its average value?

The results are obtained by suitably adapting and extending the techniques introduced in Peccati and Taqqu (2007) for deriving CLTs for double Poisson integrals.

# CLTs for the cumulative hazard

Define $k_T^{(0)}(s,x) = s\int_0^T k(t,x)\,\mathrm{d}t$. In the following, we suppose a few technical integrability conditions involving the kernel and Poisson intensity measure to be satisfied.

THEOREM

*If there exists a strictly positive function $T \mapsto C_0(k,T)$, such that, as $T \to +\infty$,*

$$\begin{aligned} C_0^2(k,T) \times \int_{\mathbb{R}^+\times\mathbb{X}} \left[k_T^{(0)}(s,x)\right]^2 \nu(\mathrm{d}s,\mathrm{d}x) &\rightarrow \sigma_0^2(k) > 0, \\ C_0^3(k,T) \times \int_{\mathbb{R}^+\times\mathbb{X}} \left[k_T^{(0)}(s,x)\right]^3 \nu(\mathrm{d}s,\mathrm{d}x) &\rightarrow 0. \end{aligned}$$

*Then,*

$$C_0(k,T) \times \left[\tilde{H}(T) - \mathbb{E}[\tilde{H}(T)]\right] \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, \sigma_0^2(k)\right),$$

*Remark.*
(I) The asymptotic variance depends on the specific structure of the CRM $\tilde{\mu}$ and of the kernel $k$.
(II) The two conditions only involve the analytic form of the kernel $k$ and do not make use of the asymptotic properties of the law of the process $\tilde{h}(t)$.

# Applications

We first consider general homogeneous CRM i.e. with intensity $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\lambda(\mathrm{d}x)$ and set, unless otherwise specified, $\lambda(\mathrm{d}x) = \mathrm{d}x$. Define $K_\rho^{(i)} = \int_0^\infty s^i \rho(\mathrm{d}s)$ for $i = 1, 2$.

# Applications

We first consider general homogeneous CRM i.e. with intensity $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\lambda(\mathrm{d}x)$ and set, unless otherwise specified, $\lambda(\mathrm{d}x) = \mathrm{d}x$. Define $K_\rho^{(i)} = \int_0^\infty s^i \rho(\mathrm{d}s)$ for $i = 1, 2$.

(i) Rectangular mixture hazards

$$\frac{1}{\sqrt{T}}\left[\tilde{H}(T) - 2\tau K_\rho^{(1)} T\right] \xrightarrow{\text{law}} X \sim \mathscr{N}\left(0, 4K_\rho^{(2)}\tau^2\right)$$

$\Longrightarrow$ Except for constants, exactly the same result holds for Ornstein–Uhlenbeck mixture hazards.

# Applications

We first consider general homogeneous CRM i.e. with intensity $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\lambda(\mathrm{d}x)$ and set, unless otherwise specified, $\lambda(\mathrm{d}x) = \mathrm{d}x$. Define $K_\rho^{(i)} = \int_0^\infty s^i \rho(\mathrm{d}s)$ for $i = 1, 2$.

(i) Rectangular mixture hazards

$$\frac{1}{\sqrt{T}}\left[\tilde{H}(T) - 2\tau K_\rho^{(1)} T\right] \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, 4K_\rho^{(2)}\tau^2\right)$$

$\Longrightarrow$ Except for constants, exactly the same result holds for Ornstein–Uhlenbeck mixture hazards.

(ii) Exponential mixture hazards with base–measure $\lambda(\mathrm{d}x) = x^{-1/2}\mathrm{e}^{-1/x}(2\sqrt{\pi})^{-1}$

$$\frac{1}{T^{\frac{1}{4}}}\left[\tilde{H}(T) - K_\rho^{(1)}\sqrt{T}\right] \xrightarrow{\text{law}} X, \sim \mathcal{N}\left(0, (2-\sqrt{2})K_\rho^{(2)}\right)$$

$\Longrightarrow$ Long–term trend is Weibull $\sqrt{T}$

## (iii) Dykstra–Laud mixture hazards

$$\frac{1}{T^{\frac{3}{2}}}\left[\tilde{H}(T)-\frac{K_\rho^{(1)}}{2}T^2\right]\xrightarrow{\text{law}} X, \sim \mathcal{N}\left(0,\frac{K_\rho^{(2)}}{3}\right)$$

(iii) Dykstra–Laud mixture hazards

$$\frac{1}{T^{\frac{3}{2}}}\left[\tilde{H}(T)-\frac{K_\rho^{(1)}}{2}T^2\right]\xrightarrow{\text{law}} X, \sim \mathcal{N}\left(0,\frac{K_\rho^{(2)}}{3}\right)$$

*Remark.* If $\tilde{\mu}$ is the generalized gamma CRM i.e.

$$\nu(\mathrm{d}v,\mathrm{d}x)=\frac{1}{\Gamma(1-\sigma)}\frac{\mathrm{e}^{-\gamma v}}{v^{1+\sigma}}\,\mathrm{d}v\,\mathrm{d}x \qquad \sigma\in[0,1),\gamma>0,$$

then $K_\rho^{(2)}$ which enters the asymptotic variance is $\frac{1-\sigma}{\gamma^{2-\sigma}}$. This confirms the empirical finding that a small $\gamma$ induces a large variance i.e. a non–informative prior.

Consider now a non–homogeneous CRM: specifically we take the extended gamma CRM i.e. $\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-\gamma(x)v}}{v}\,\mathrm{d}v\,\mathrm{d}x$ with $\gamma$ a strictly positive function.

Case I. If $\gamma$ is bounded, then CLTs with the same trends and rates $C_0(k, T)$ are obtained.

Consider now a non–homogeneous CRM: specifically we take the extended gamma CRM i.e. $\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-\gamma(x)v}}{v}\,\mathrm{d}v\,\mathrm{d}x$ with $\gamma$ a strictly positive function.

Case I. If $\gamma$ is bounded, then CLTs with the same trends and rates $C_0(k, T)$ are obtained.

Case II. If $\gamma$ diverges, interesting phenomena appear.
(i) Rectangular mixture hazard with $\gamma(x) = 1 + \sqrt{x}$

$$\frac{1}{\sqrt{\log(T)}}\left[\tilde{H}(T) - 4T^{1/2} - 2\log(T)\right] \xrightarrow{\text{law}} X \sim \mathscr{N}(0, 4)$$

Compared to the homogeneous case:
(a) the trend is reduced from $T$ to $\sqrt{T}$
(b) the rate of divergence from the trend is reduced from $\sqrt{T}$ to $\sqrt{\log(T)}$.

Consider now a non–homogeneous CRM: specifically we take the extended gamma CRM i.e. $\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-\gamma(x)v}}{v}\,\mathrm{d}v\,\mathrm{d}x$ with $\gamma$ a strictly positive function.

Case I. If $\gamma$ is bounded, then CLTs with the same trends and rates $C_0(k, T)$ are obtained.

Case II. If $\gamma$ diverges, interesting phenomena appear.
(i) Rectangular mixture hazard with $\gamma(x) = 1 + \sqrt{x}$

$$\frac{1}{\sqrt{\log(T)}}\left[\tilde{H}(T) - 4T^{1/2} - 2\log(T)\right] \xrightarrow{\text{law}} X \sim \mathcal{N}(0, 4)$$

Compared to the homogeneous case:
(a) the trend is reduced from $T$ to $\sqrt{T}$
(b) the rate of divergence from the trend is reduced from $\sqrt{T}$ to $\sqrt{\log(T)}$.

By suitably choosing $\gamma$ the trend and the rate of divergence from the trend can be tuned at any order less than $T$ and $\sqrt{T}$, respectively.

(ii) DL mixture hazard with $\gamma(x) = 1 + \sqrt{x}$

$$\frac{1}{\sqrt{\log(T)}\,T}\left[\tilde{H}(T) - 4/3\,T^{3/2} - \log(T)T\right] \xrightarrow{\text{law}} X \sim \mathcal{N}(0,1)$$

Compared to the homogeneous case:
(a) the trend is reduced from $T^2$ to $T^{3/2}$
(b) the rate of divergence from the trend is reduced from $T^{3/2}$ to $\sqrt{\log(T)}T$.

(ii) DL mixture hazard with $\gamma(x) = 1 + \sqrt{x}$

$$\frac{1}{\sqrt{\log(T)}\,T}\left[\tilde{H}(T) - 4/3\,T^{3/2} - \log(T)T\right] \xrightarrow{\text{law}} X \sim \mathcal{N}(0,1)$$

Compared to the homogeneous case:
(a) the trend is reduced from $T^2$ to $T^{3/2}$
(b) the rate of divergence from the trend is reduced from $T^{3/2}$ to $\sqrt{\log(T)}T$.

By suitably selecting $\gamma$ the trend and the rate of divergence from the trend can be tuned at any order in the range $(T, T^2]$ and $(T, T^{3/2}]$, respectively.

# CLTs for the path–variance

We now provide a CLT for the path–variance $\int_0^T \left[\tilde{h}(t) - \frac{\tilde{H}(T)}{T}\right]^2 \mathrm{d}t$ of a mixture hazard rate.

**THEOREM**

Under a series of technical assumptions, we have

$$C_1(k,T) \times \left\{ \frac{1}{T}\int_0^T \left[\tilde{h}(t) - \frac{\tilde{H}(T)}{T}\right]^2 \mathrm{d}t - \frac{1}{T}\int_0^T \mathbb{E}\left[\tilde{h}(t) - \frac{\mathbb{E}(\tilde{H}(T))}{T}\right]^2 \mathrm{d}t \right\}$$

$$\xrightarrow{\text{law}} X, \sim \mathcal{N}\left(0, \sigma_1^2\right)$$

$\Longrightarrow$ Some of the conditions are quite delicate: for instance, they do not hold for DL and exponential mixture hazards. This leads to conjecture that our conditions do not hold for hazards with un–regular oscillations over time. A DL mixture hazard accumulates all the randomness of $\tilde{\mu}$ as $T$ increases and, hence, the path–variance increases over time. An exponential mixture hazard dampens the influence of $\tilde{\mu}$ as $T$ increases and, hence, the path–variance decreases over time.

# Applications

## (i) Ornstein–Uhlenbeck mixture hazard

$$\sqrt{T} \times \left\{ \frac{1}{T} \int_0^T [\tilde{h}(t) - \frac{1}{T}\tilde{H}(T)]^2 \mathrm{d}t - K_\rho^{(2)} \right\} \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, \frac{2\,(K_\rho^{(2)})^2}{\kappa} + K_\rho^{(4)}\right)$$

If $\tilde{\mu}$ is the generalized gamma CRM, the variance of the limiting normal r.v. is

$$\sigma_1^2 = \frac{(1-\sigma)\ (2(1-\sigma)\gamma^\sigma + \kappa(2-\sigma)_2)}{\kappa\gamma^{4-\sigma}}$$

# Applications

## (i) Ornstein–Uhlenbeck mixture hazard

$$\sqrt{T} \times \left\{ \frac{1}{T} \int_0^T [\tilde{h}(t) - \frac{1}{T}\tilde{H}(T)]^2 \mathrm{d}t - K_\rho^{(2)} \right\} \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, \frac{2\,(K_\rho^{(2)})^2}{\kappa} + K_\rho^{(4)}\right)$$

If $\tilde{\mu}$ is the generalized gamma CRM, the variance of the limiting normal r.v. is

$$\sigma_1^2 = \frac{(1-\sigma)\ (2(1-\sigma)\gamma^\sigma + \kappa(2-\sigma)_2)}{\kappa\gamma^{4-\sigma}}$$

$\Longrightarrow$ Hints for prior specification:
(a) $\sigma_1^2 \downarrow$ decreases as $\kappa \uparrow$ and $\gamma \uparrow$
(b) $\sigma_1^2$ is maximized by $\sigma = 0$ for low values of $\kappa$ and $\gamma$, whereas, for moderately large values of $\kappa$ and $\gamma$, the maximizing $\sigma$ increases as $\kappa$ and $\gamma$ increase. E.g., $\sigma_1^2$ is maximized by $\sigma = 0$ if $\kappa = 0.5$ and $\gamma = 2$, whereas it is maximized by $\sigma \approx 0.1$, if $\kappa = 1$ and $\gamma = 5$.

# Applications

## (i) Ornstein–Uhlenbeck mixture hazard

$$\sqrt{T} \times \left\{ \frac{1}{T} \int_0^T [\tilde{h}(t) - \frac{1}{T}\tilde{H}(T)]^2 \mathrm{d}t - K_\rho^{(2)} \right\} \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, \frac{2\,(K_\rho^{(2)})^2}{\kappa} + K_\rho^{(4)}\right)$$

If $\tilde{\mu}$ is the generalized gamma CRM, the variance of the limiting normal r.v. is

$$\sigma_1^2 = \frac{(1-\sigma)\ (2(1-\sigma)\gamma^\sigma + \kappa(2-\sigma)_2)}{\kappa\gamma^{4-\sigma}}$$

$\Longrightarrow$ Hints for prior specification:
(a) $\sigma_1^2 \downarrow$ decreases as $\kappa \uparrow$ and $\gamma \uparrow$
(b) $\sigma_1^2$ is maximized by $\sigma = 0$ for low values of $\kappa$ and $\gamma$, whereas, for moderately large values of $\kappa$ and $\gamma$, the maximizing $\sigma$ increases as $\kappa$ and $\gamma$ increase. E.g., $\sigma_1^2$ is maximized by $\sigma = 0$ if $\kappa = 0.5$ and $\gamma = 2$, whereas it is maximized by $\sigma \approx 0.1$, if $\kappa = 1$ and $\gamma = 5$.

## (ii) Rectangular mixture hazard

$$\sqrt{T} \times \left\{ \frac{1}{T} \int_0^T [\tilde{h}(t) - \frac{1}{T}\tilde{H}(T)]^2 \mathrm{d}t - 2\tau K_\rho^{(2)} \right\} \xrightarrow{\text{law}} X \sim \mathcal{N}\left(0, 4\tau^2 \left[\frac{8\tau(K_\rho^{(2)})^2}{3} + K_\rho^{(4)}\right]\right)$$

# The posterior mixture hazard (James, 2005)

As previously, denote by $\boldsymbol{Y}^n = (Y_1, \ldots, Y_n)$ a set of $n$ observations.

The key to the derivation of the posterior hazard is the introduction of a suitable set of latent variables $\boldsymbol{X}^n = (X_1, \ldots, X_n)$. Since the $X_i$'s may feature ties, we denote by $\boldsymbol{X}^* = (X_1^*, \ldots, X_k^*)$ the $k \leq n$ distinct latent variables. Then, one can describe the conditional distribution of the CRM given $\boldsymbol{Y}^n$, $\tilde{\mu} \mid \boldsymbol{Y}^n$, in terms of $\tilde{\mu} \mid \boldsymbol{Y}^n, \boldsymbol{X}^n$ mixed over $\boldsymbol{X}^n \mid \boldsymbol{Y}^n$.

The law of a mixture hazard rate, conditionally on $\boldsymbol{Y}^n$ and $\boldsymbol{X}^n$, coincides with the distribution of the random object

$$\tilde{h}^n(t) = \int_{\mathbb{X}} k(t,x) \tilde{\mu}^n(\mathrm{d}x) + \sum_{i=1}^{k} J_i k(t, X_i^*),$$

where $\mu^n$ is a CRM with updated intensity measure

$$\nu^n(\mathrm{d}v, \mathrm{d}x) := \mathrm{e}^{-v \sum_{i=1}^{n} \int_0^{Y_i} k(t,x)\mathrm{d}t} \rho(\mathrm{d}v|x)\, \lambda(\mathrm{d}x) \qquad (\bullet)$$

and the latent variables $X_i^*$'s are the location of the random jumps $J_i$'s (for $i = 1, \ldots, k$), which are, conditionally on $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$, independent of $\tilde{\mu}^n$.

*Remark.* The intensity in $(\bullet)$ is always non–homogeneous and the jump sizes of the CRM become smaller as $n$ increases.

Finally, denote the distribution of $\boldsymbol{X}^n | \boldsymbol{Y}^n$ by $\tilde{P}_{\boldsymbol{X}^n | \boldsymbol{Y}^n}$.

# CLTs for posterior functionals

Relying upon the previous posterior characterization, we derive CLTs for functionals *given a fixed number of observations*: the idea consists in first conditioning on $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$ and then in obtaining CLTs conditionally on $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$. Finally, the desired CLTs are obtained by averaging over $\tilde{P}_{\boldsymbol{X}^n|\boldsymbol{Y}^n}$.

# CLTs for posterior functionals

Relying upon the previous posterior characterization, we derive CLTs for functionals *given a fixed number of observations*: the idea consists in first conditioning on $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$ and then in obtaining CLTs conditionally on $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$. Finally, the desired CLTs are obtained by averaging over $\tilde{P}_{\boldsymbol{X}^n|\boldsymbol{Y}^n}$.

For given $\boldsymbol{X}^n$ and $\boldsymbol{Y}^n$, define $\tilde{H}^n(T)$ to be the cumulative hazard associated to the posterior hazard without fixed points of discontinuity. Assume there exists a (deterministic) function $T \mapsto C_n(k,T)$ and $\sigma_n^2(k) > 0$ such that

$$\lim_{T\to+\infty} C_n^2(k,T) \times \int_{\mathbb{R}^+\times\mathbb{X}} \left[k_T^{(0)}(s,x)\right]^2 \nu^n(\mathrm{d}s,\mathrm{d}x) = \sigma_n^2(\boldsymbol{Y}^n)$$

$$\lim_{T\to+\infty} C_n(k,T) \times \sum_{j=1}^k J_j \int_0^T k(t,X_j^*)\mathrm{d}t = m_n(\boldsymbol{X}^n) \geq 0 \quad a.s.$$

Note that $\sigma_n^2(\boldsymbol{Y}^n)$ may depend on $\boldsymbol{Y}^n$ and $m_n(\boldsymbol{X}^n)$ may depend on $\boldsymbol{X}^n$. Then, under suitable conditions we can state that

$$C_n(k,T) \times \left[\tilde{H}(T) - \mathbb{E}[\tilde{H}^n(T)]\right] \Big| \, \boldsymbol{Y}$$

converges, as $T \to \infty$ to a location mixture of Gaussian random variables with mean $m_n(\boldsymbol{X}^n)$ and variance $\sigma_n^2(\boldsymbol{Y}^n)$ and mixing distribution $\tilde{P}_{\boldsymbol{X}|\boldsymbol{Y}}$.

$\Longrightarrow$ The same strategy allows to obtain a CLT for the path–variance.

# Applications

How do the CLTs look like for the specific cases?

(I) For all specific kernels $m_n(\boldsymbol{X}^n) = 0$ a.s. This implies that the limiting distribution is a simple Gaussian random variable with mean 0 and variance $\sigma_n^2(\boldsymbol{Y}^n)$.

This behaviour is somehow expected since it seems reasonable that the part of the posterior *with fixed points of discontinuity* affects the asymptotic behaviour just once an infinite number of observations is collected.

# Applications

How do the CLTs look like for the specific cases?

(I) For all specific kernels $m_n(\boldsymbol{X}^n) = 0$ a.s. This implies that the limiting distribution is a simple Gaussian random variable with mean 0 and variance $\sigma_n^2(\boldsymbol{Y}^n)$.

This behaviour is somehow expected since it seems reasonable that the part of the posterior *with fixed points of discontinuity* affects the asymptotic behaviour just once an infinite number of observations is collected.

(II) The surprising fact is that the asymptotic variance $\sigma_n^2(\boldsymbol{Y}^n)$ does not depend on the data $\boldsymbol{Y}^n$. This implies that the CLTs associated with the posterior hazard rate are exactly the same as for the prior hazard for any fixed number of observations.

Indeed, one would expect $\sigma_n^2(\boldsymbol{Y}^n)$ to depend negatively on the number of observations increases: recall that the intensity measure of the posterior is $\mathrm{e}^{-v\sum_{i=1}^n \int_0^{Y_i} k(t,x)\mathrm{d}t}\,\rho(\mathrm{d}v|x)\,\lambda(\mathrm{d}x)$. Hence, the CRM is really just "killed" in the limit and the choice of the model matters.

# Concluding remarks

- We provided a comprehensive investigation of weak consistency of the mixture hazard model. $L_1$–consistency and rates of convergence of the posterior distribution are the natural following targets.

# Concluding remarks

- We provided a comprehensive investigation of weak consistency of the mixture hazard model. $L_1$–consistency and rates of convergence of the posterior distribution are the natural following targets.
- CLTs provide a rigorous guide for prior specification: they show e.g. how the choice of the kernel and of the CRM determine the trend of random cumulative hazard and, moreover, dictate the divergence this trend. The parameters of the kernel and of the CRM enter the asymptotic variance and so can be used to fine–tune the model.

# Concluding remarks

- We provided a comprehensive investigation of weak consistency of the mixture hazard model. $L_1$–consistency and rates of convergence of the posterior distribution are the natural following targets.
- CLTs provide a rigorous guide for prior specification: they show e.g. how the choice of the kernel and of the CRM determine the trend of random cumulative hazard and, moreover, dictate the divergence this trend. The parameters of the kernel and of the CRM enter the asymptotic variance and so can be used to fine–tune the model.
- The coincidence of the asymptotic behaviour of the posterior and the prior hazard means that the data do not influence the behaviour of the model for times larger than the largest observed lifetime $Y_{(n)}$. The fact that the overall variance is not influenced by the data is somehow counterintuitive: since the contribution of the CRM vanishes in the limit, one would expect the variance to become smaller and smaller as more data come in.

# References


- Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Prob.* **31**, 929–953.
- De Blasi, P., Peccati, G. and Prünster, I. (2007). Asymptotics for posterior hazards. *Preprint*.
- Diaconis, P. and Freedman, D. (1986). On the consistency of Bayes estimates. *Ann. Statist.* **14**, 1–67.
- Dykstra, R.L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367.
- Ishwaran, H. and James, L. F. (2004). Computational methods for multiplicative intensity models using weighted gamma processes: Proportional hazards, marked point processes, and panel count data. *J. Amer. Stat. Assoc.* **99**, 175–190.
- James, L.F. (2005). Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33**, 1771–1799.
- Kingman, J.F.C. (1967). Completely random measures. *Pacific J. Math.* **21**, 59–78.
- Lo, A.Y. and Weng, C.S. (1989). On a class of Bayesian nonparametric estimates. II. Hazard rate estimates. *Ann. Inst. Statist. Math.* **41**, 227–245.
- Peccati, G. and Prünster, I. (2007). Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Ann. Appl. Probab.*, in press.
- Peccati, G. and Taqqu, M.S. (2007). Central limit theorems for double Poisson integrals. *Bernoulli*, in press.
- Schwartz, L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete* **4**, 10–26.