Asymptotics for posterior hazards

Pierpaolo De Blasi

University of Turin

10th August 2007, BNR Workshop,
Isaac Newton Intitute, Cambridge, UK

Joint work with Giovanni Peccati (Université Paris VI)
and Igor Prünster (University of Turin)

Outline

Mixture hazard rate

Let Y be a positive absolutely continuous random variable (*lifetime*) with random hazard rate of the form

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

- k is a *kernel*, jointly measurable application $\mathbb{R}^+ \times \mathbb{X} \to \mathbb{R}^+$,
- $\tilde{\mu}$ is a random mixing measure on the space $\mathbb{X}$, modeled as a *completely random measure*.

The cumulative hazard is given by $\tilde{H}(t) = \int_0^t \tilde{h}(s)\mathrm{d}s$ and is required to satisfy $H\tilde{(}t) \to \infty$ for $t \to \infty$ a.s. Then

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

defines a random density function to be used in Bayesian nonparametric inference.

See Dykstra and Laud (1981) and Lo and Weng (1989).

Kernels

As examples,we consider the class of mixture hazard rates that arise under the following choice of 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*)

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

Completely random measures

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

Definition. A random element $\tilde{\mu}$ taking values in $\mathbb{M}$ such that

(i) $\tilde{\mu}(\emptyset) = 0$

(ii) for any disjoint sets $(A_i)_{i\geq 1}$:
$\tilde{\mu}(A_1), \tilde{\mu}(A_2), \ldots$ are mutually independent
$\tilde{\mu}(\cup_{j\geq 1} A_j) = \sum_{j\geq 1} \tilde{\mu}(A_j)$

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*

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

where ν stands for the intensity of the Poisson random measure underlying $\tilde{\mu}$.

- CRMs select almost surely discrete measures: hence, $\tilde{\mu}$ can always be represented as $\sum_{i\geq 1} J_i \delta_{X_i}$.
- Note that 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. In this case $\tilde{\mu}$ is termed *homogeneous*.
- A *non-homogeneous* CRM has $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v \,|\, x)\lambda(\mathrm{d}x)$. Non-homogeneous CRMs arise in the posterior.
- 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.
- We consider CRMs with Poisson intensity of the following form:

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

γ constant $\rightarrow$ *generalized gamma measure* (Brix, 1999)
$\sigma = 0 \rightarrow$ *weighted gamma measure* (Lo & Weng, 1989).

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.

Which conditions on f_0 *and* $\tilde{\mu}$ *are needed under the four kernels considered in order to achieve consistency?*

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

Let $\boldsymbol{Y} = (Y_1, ..., Y_n)$ be a set of observations. We are interested in establishing *Central Limit Theorems* of the type

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

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

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

Weak consistency

Denote by P_0 the probability distribution associated with f_0 and by P_0^∞ the infinite product measure. Also, let Π be the prior distribution of the random density function $\tilde{f}$ and Π_n the posterior distribution.

Definition. If, for any $\epsilon > 0$ and as $n \to \infty$,

$$\Pi_n(A_\epsilon(f_0)) \to 1 \qquad \text{a.s.} - P_0^\infty,$$

where $A_\epsilon(f_0)$ is a ϵ–neighborhood of f_0 in the weak topology, then Π is said to be weak consistent at f_0.

Remark. A *sufficient condition* for weak consistency requires a prior Π to assign positive probability to Kullback–Leibler (K-L) neighborhoods of f_0 (K-L condition):

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

where $\int \log(f_0/f) f_0$ denotes the K-L divergence between f_0 and f, whereas $\mathbb{F}$ is the space of density functions defined on $\mathbb{R}^+$.

General consistency result

Assume $\mathbb{X} = \mathbb{R}^+$ and postulate the existence of a boundedly finite measure μ_0 on $\mathbb{R}^+$ such that the true density f_0 has hazard

$$h_0(t) = \int_{\mathbb{R}^+} k(t,x)\mu_0(\mathrm{d}x)$$

where k is the kernel used for constructing $\tilde{f}$ via the mixture hazard $\tilde{h}$.

We translate the K-L condition w.r.t. f_0 into a condition of positive prior probability assigned to uniform neighborhoods of h_0 on $[0,T]$.

THEOREM *Weak consistency*

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 Π 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 δ.

Note that condition (ii) is related to the asymptotic characterization of functionals of the random hazard rate (see later).

Relaxing the condition $h_0(0) > 0$

For (DL) and (OU) kernels one may actually have $h_0(0) = 0$. Problems then arise in establish the K-L condition since $\log(h_0/\tilde{h})$ (hence $\log(f_0/\tilde{f})$) can become arbitrarily large around 0.

In order to cover also the case $h_0(0) = 0$, we resort to the *short time behaviour* of $\tilde{\mu}$, which determines the way $\tilde{h}$ vanishes in 0 for the (DL) and (OU) cases.

Proposition 1 *Short time behaviour*

Weak consistency holds for (DL) and (OU) kernels 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) $\liminf_{t\to 0} \tilde{\mu}((0,t])/t^r = \infty \quad \text{a.s.}$

In particular (b) holds if $\tilde{\mu}$ is a CRM belonging to the generalized gamma family with $\sigma \in (0,1)$ and $\lambda(\mathrm{d}x) = \mathrm{d}x$.

Note that the powers α and r do not need to satisfy any relation.

Uniform balls of the true hazard rate

In order to derive explicit consistency results for the mixture hazard models based on the four kernels considered, we need to establish:

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

To this aim, the following result is essential.

Denote by $\mathbb{G}$ the space of nondecreasing càdlàg functions on $\mathbb{R}^+$ such that $G(0)=0$ for any $G\in\mathbb{G}$. Note that any CRM induces a distribution on $\mathbb{G}$ and that, for $G_0(x)=\mu_0\big([0,x]\big)$, $G_0\in\mathbb{G}$.

LEMMA 1 *Support of CRMs*

Let $\tilde\mu$ be a CRM on $\mathbb{R}^+$, satisfying $\nu(\mathbb{R}^+,\mathrm{d}x)=\infty$ for any $x\in\mathbb{X}$, and denote by Q the distribution induced on $\mathbb{G}$. Then, for any $G_0\in\mathbb{G}$, any finite M and $\eta>0$,

$$Q\Big\{G\in\mathbb{G}:\ \sup_{x\le M}|G(x)-G_0(x)|<\eta\Big\}>0.$$

Different techniques are required for the (exp) kernel w.r.t. the other three kernels (different support of $k(t,x)$ seen as a function of x).

PROPOSITION 2 *Weak consistency for (DL), (rect) and (OU)*

Let $\tilde{\mu}$ be a generalized gamma CRM with $\sigma \in (0,1)$ and $\lambda(\mathrm{d}x) = \mathrm{d}x$. Then weak consistency holds for the (DL), (rect) and (OU) kernels provided that:

(i) $h_0(t) > 0$ for any $t > 0$;

(ii) $\int t^i f_0(t)\mathrm{d}t < \infty$ with $i = 1$ for (rect) and (OU), $i = 2$ for (DL).

Note that Condition (i) is automatically satisfied if μ_0 is absolutely continuous.

PROPOSITION 3 *Weak consistency for (exp)*

Weak consistency holds for the (exp) kernel provided that:

(i) $\tilde{h}(0) < \infty$ a.s. and $h_0(0) < \infty$;

(ii) $\int t\, f_0(t)\mathrm{d}t < \infty$.

The only relevant assumption consists in asking for both $\tilde{h}$ and h_0 not to explode in 0. Note that $h_0(t) > 0$ is automatically satisfied.

CLTs for linear and quadratic functionals

- CLTs provide a synthetic picture of the random hazard, like information about the trend, oscillation and overall variance for large time horizons.
- In Peccati and Pruenster (2006) functionals of the prior hazard rate are considered. Such results can serve as a guide for prior specification.
- Here we investigate the asymptotic behaviour of functionals of the posterior random hazard rate *given a fixed number of observations*.
- We found out that, in all the considered special cases, the CLTs associated with the posterior hazard rate are *the same as for the prior ones*, and this for any number of observations.
- Although consistency implies that a given model can be asymptotically directed towards any deterministic target, the overall structure of a posterior hazard rate is systematically determined by the prior choice.

Posterior distribution

We exploit an explicit posterior representation of the random mixture hazard model of James (2005).

- Let $\tilde{P}$ be the random probability measure associated with $\tilde{f}(t)$ and let $(Y_n)_{n\geq 1}$ be a sequence of exchangeable observations from $\tilde{P}$. Set $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$.
- Define a random probability measure on $\mathbb{X} \times \mathbb{R}^+$ as

$$\tilde{P}(\mathrm{d}x, \mathrm{d}t) = \mathrm{e}^{-\int_0^t \int_{\mathbb{X}} k(s,y)\tilde{\mu}(\mathrm{d}y)\,\mathrm{d}s} k(t,x)\tilde{\mu}(\mathrm{d}x)\,\mathrm{d}t.$$

  By introducing the latent variables $\boldsymbol{X} = (X_1, \ldots, X_n)$, we describe $\tilde{\mu} \mid \boldsymbol{Y}$ in terms of $\tilde{\mu} \mid \boldsymbol{Y}, \boldsymbol{X}$ mixed over $\boldsymbol{X} \mid \boldsymbol{Y}$.
- 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. Let n_j be the frequency of X_j^* and $C_j = \{r : x_r = x_j^*\}$. The likelihood function is then given by

$$\mathscr{L}(\mu; \boldsymbol{y}, \boldsymbol{x}) = \mathrm{e}^{-\int_{\mathbb{X}} \sum_{i=1}^n \int_0^{y_i} k(t,x)\mathrm{d}t\,\mu(\mathrm{d}x)} \prod_{j=1}^k \mu(\mathrm{d}x_j^*)^{n_j} \prod_{i\in C_j} k(y_i, x_j^*)$$

- Given $\boldsymbol{X}$ and $\boldsymbol{Y}$, the conditional distribution of $\tilde{\mu} \mid \boldsymbol{Y}, \boldsymbol{X}$ coincides with the distribution of the random measure

$$\tilde{\mu}_{\Delta^n} \stackrel{d}{=} \tilde{\mu}^n + \sum_{i=1}^k J_i \, \delta_{X_j^*} = \tilde{\mu}^n + \Delta^n$$

  - $\tilde{\mu}^n$ is a CRM with intensity measure

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

  - $\Delta^n := \sum_{i=1}^k J_i \, \delta_{X_j^*}$ with, for $i = 1, \ldots, k$, X_i^* a fixed points of discontinuity with corresponding jump J_i which are, conditionally on $\boldsymbol{X}$ and $\boldsymbol{Y}$, independent of $\tilde{\mu}^n$.

- The conditional distribution $\boldsymbol{X}|\boldsymbol{Y}$, say $\tilde{P}_{\boldsymbol{X}|\boldsymbol{Y}}$, can be expressed as the conditional distributions of $\boldsymbol{X}^* \mid \boldsymbol{p}, \boldsymbol{Y}$ and $\boldsymbol{p} \mid \boldsymbol{Y}$, where $\boldsymbol{p}$ is a *partitions* of the integers $\{1, \ldots, n\}$.
- We derive CLTs for the posterior random hazard given $\boldsymbol{X}$ and $\boldsymbol{Y}$:

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

and, then, average over the distribution of $\boldsymbol{X}|\boldsymbol{Y}$.

Limit theorems for shifted measures

Peccati and Pruenster (2006) provide sufficient conditions to have that linear and quadratic functionals associated with a random mixture hazard *without fixed atoms* verify a CLT.

Linear functional. For given $\boldsymbol{X}$ and $\boldsymbol{Y}$, define

$$\tilde{H}_{\Delta^n}(T) = \int_0^T \tilde{h}_{\Delta^n}(t)\mathrm{d}t = \tilde{H}^n(T) + \sum_{i=1}^k J_i \int_0^T k(t, X_i^*)\mathrm{d}t$$

Assume there exists a (deterministic) function $T \mapsto C_n(k, T)$ and $\sigma_n^2(k) > 0$ such that, as $T \to +\infty$,

$$C_n^2(k,T) \times \int_{\mathbb{R}^+ \times \mathbb{X}} \left[s \int_0^T k(t,x)\mathrm{d}t \right]^2 \nu^n(\mathrm{d}s, \mathrm{d}x) \to \sigma_n^2(k),$$

$$C_n^3(k,T) \times \int_{\mathbb{R}^+ \times \mathbb{X}} \left[s \int_0^T k(t,x)\mathrm{d}t \right]^3 \nu^n(\mathrm{d}s, \mathrm{d}x) \to 0,$$

Then

$$C_n(k,T) \times \left[\tilde{H}^n(T) - \mathbb{E}[\tilde{H}^n(T)] \right] \Big| \boldsymbol{Y} \xrightarrow{\text{law}} \mathrm{N}(0, \sigma_n^2(k)),$$

where $\sigma_n^2(k)$ may depend on $\boldsymbol{Y}$.

Assume moreover that

$$\lim_{T\to+\infty} C_n(k,T)\times \sum_{j=1}^{k} J_j \int_0^T k(t,X_j^*)\mathrm{d}t = m_n(\Delta^n,k), \qquad (1)$$

for some quantity $m_n(\Delta^n,k)\geq 0$ *that may depend on* $\boldsymbol{X}$. Then

$$C_n(k,T)\times\left[\tilde{H}_{\Delta^n}(T)-\mathbb{E}[\tilde{H}^n(T)]\right]\Big|\boldsymbol{Y},\boldsymbol{X} \xrightarrow{\text{law}} \mathrm{N}\big(m_n(\Delta^n,k),\sigma_n^2(k)\big).$$

If (1) holds for $\boldsymbol{X}=(x_1,\dots,x_n)\in\mathbb{X}^n$ in a set of $\tilde{P}_{\boldsymbol{X}|\boldsymbol{Y}}$–probability 1, then we can state the result as

$$\eta(T)\times\left[\tilde{H}(T)-\tau(T)\right]\Big|\boldsymbol{Y} \xrightarrow{\text{law}} \int_{\mathbb{X}^n} Z_{\boldsymbol{X},\boldsymbol{Y}}\,\mathrm{d}\tilde{P}_{\boldsymbol{X}|\boldsymbol{Y}},$$

where

- $\eta(T)=C_n(k,T)$;
- $\tau(T)=\mathbb{E}[\tilde{H}^n(T)|\boldsymbol{Y}]$;
- $Z_{\boldsymbol{X},\boldsymbol{Y}}$ is a Gaussian random variable with mean $m_n(\Delta^n,k)$ and variance $\sigma_n(k)$.

The mixture in $\int_{\mathbb{X}^n} Z_{\boldsymbol{X},\boldsymbol{Y}}\mathrm{d}\tilde{P}_{\boldsymbol{X}|\boldsymbol{Y}}$ is over the quantity $m_n(\Delta^n,k)$ that may change according with the realization of $\boldsymbol{X}$.

Applications

- We derive CLTs for functionals of posterior hazards based on the four kernels combined with generalized gamma CRMs.
- In general $m_n(\Delta^n, k) = 0$, not depending on $\boldsymbol{Y}, \boldsymbol{X}$, that is the part of the posterior *with fixed points of discontinuity* affects the asymptotic behaviour just in the limit. Then, the convergence is simply expressed in terms of a Gaussian random variable.

LINEAR CLT 1 *(OU) kernel*

$$T^{-\frac{1}{2}}\left[\tilde{H}(T) - \sqrt{\tfrac{2}{\kappa}\gamma^{-(1-\sigma)}}T\right]\bigg|\boldsymbol{Y} \xrightarrow{\text{law}} \mathrm{N}(0, 2\gamma^{-(1-\sigma)})$$

LINEAR CLT 2 *(DL) kernel*

$$T^{-\frac{3}{2}}\left[\tilde{H}(T) - (2\gamma^{-(1-\sigma)})^{-1}T^2\right]\bigg|\boldsymbol{Y} \xrightarrow{\text{law}} \mathrm{N}\left(0, \frac{1-\sigma}{3\gamma^{2-\sigma}}\right)$$

- **Most important**: $C_n(k, T) = C_0(k, T)$, $\mathbb{E}[\tilde{H}(T)] \sim \mathbb{E}[\tilde{H}(T)|\boldsymbol{Y}]$ and $\sigma_n(k) = \sigma_0(k)$, that is we get the same trend, oscillation around the trend and overall variance we get with the prior cumulative hazard.

- One might think that the lack of dependence on the data of posterior functionals may be due to the boundedness of the support of the (DL), (OU) and (rect) kernels, as it happens that

$$\nu^n = \nu \qquad \text{in } \mathbb{R}^+ \times [Y_{(n)}, \infty),$$

  where $Y_{(n)}$ is the largest observed lifetime.
- Consider then the (exp) kernel.
  In particular, we set $\lambda(\mathrm{d}x) = x^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{x}} (2\sqrt{\pi})^{-1}$ in the Poisson intensity $\nu(\mathrm{d}v, \mathrm{d}x)$ of generalized gamma CRM, which gives us $\tilde{h}(0) < \infty$ a.s. and a prior mean centered on a quasi Weibull hazard, $\mathbb{E}[\tilde{h}(t)] = \gamma^{-1+\sigma} / \sqrt{t+1}$.

LINEAR CLT 3 *(exp) kernel*

$$T^{-\frac{1}{2}} \left[\tilde{H}(T) - (\gamma^{-1+\sigma})^{-1} T^{1/2} \right] \Big| \boldsymbol{Y} \xrightarrow{\text{law}} \mathrm{N}\left(0, \frac{(2-\sqrt{2})(1-\sigma)}{\gamma^{2-\sigma}}\right)$$

- **Again**, the oscillation rate, the trend and the overall variance coincides with the prior ones, that is they do not depend on the observation $\boldsymbol{Y}$.

Concluding remarks

- We provide a comprehensive investigation of posterior weak consistency of the mixture hazard model. L_1–consistency would be the natural successive target.
- The investigation of CLTs for functionals represents a completely new line of research and has interest in its own.
- The coincidence of 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 $Y_{(n)}$; and, this, despite the fact that these models are consistent.
- 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.
- Since this does not happen, our findings show a clear evidence that the choice of the CRM really matters whatever the size of the dataset and provide a guideline in incorporating prior knowledge appropriately into the model.

For Further Reading

Brix, A. (1999)
Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Prob.* **31**, 929–953.

Dykstra, R.L. and Laud, P. (1981)
A Bayesian nonparametric approach to reliability. *Ann. Statist.* **9**, 356–367.

James, L.F. (2005)
Bayesian Poisson process partition calculus with an application to Bayesian Lévy moving averages. *Ann. Statist.* **33**, 1771–1799.

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. (2006)
Linear and quadratic functionals of random hazard rates: an asymptotic analysis. *Preprint*.