**Università degli Studi di Modena e Reggio Emilia**

# On Bayesian Nonparametric Estimation of Smooth Hazard Rates with a View to Seismic Hazard Assessment

*Luca La Rocca*

R 38-05 Aprile 2005

# RESEARCH REPORT

**Dipartimento di Scienze Sociali, Cognitive e Quantitative**

Via G. Giglioli Valle, 9
42100 Reggio Emilia
Tel: +39 0522 522511
Fax: +39 0522 522512
Internet: http:://www.cei.unimore.it

Depositato ai sensi dell'articolo 1 del DLL 31/08/1945 n.660

# On Bayesian Nonparametric Estimation of Smooth Hazard Rates with a View to Seismic Hazard Assessment

Luca La Rocca *
University of Modena and Reggio Emilia, Italy
e-mail: larocca.luca@unimore.it

April 26, 2005

**Abstract**

Nonparametric inference on the hazard rate is an alternative to density estimation for positive variables which naturally deals with right censored observations. It is a classic topic of survival analysis which is here shown to be of interest in the applied context of seismic hazard assessment. This paper puts forth a new Bayesian approach to hazard rate estimation, based on building the prior hazard rate as a convolution mixture of a probability density with a compound Poisson process. The resulting new class of nonparametric priors is studied in view of its use for Bayesian inference: first, conditions are given for the prior to be well defined and to select smooth distributions; then, a procedure is developed to choose the hyperparameters so as to assign a constant expected prior hazard rate, while controlling prior variability; finally, an MCMC approximation of the posterior distribution is found. The proposed algorithm is implemented for the analysis of some Italian seismic event data and a possible adjustment to a well established class of prior hazard rates is discussed in some detail.

**Keywords:** Compound Poisson Process, Markov Chain Monte Carlo, Right Censoring, Survival Analysis.

*Correspondence address: Dipartimento di Scienze Sociali, Cognitive e Quantitative, Via Giglioli Valle, 9, 42100 Reggio Emilia, Italy.

## 1 Introduction

It is well known that the unknown distribution of positive variables can be described in terms of its hazard rate function, whenever it is safe to assume that such distribution be absolutely continuous with respect to the Borel-Lebesgue measure on the positive half-line. This is especially interesting in survival analysis, that is when variables are times to event and the hazard rate can be interpreted as the instantaneous conditional probability density of event. Here, the topic of statistical inference on the hazard rate is dealt with from a Bayesian nonparametric point of view and its relevance in the applied context of seismic hazard assessment is pointed out.

Interest in Bayesian nonparametric hazard rate estimation dates back to the early work by Dykstra and Laud (1981), who introduced the weighted gamma process—extended gamma process in their terminology—as a tool for modelling prior hazard rates. They obtained the Bayes estimator under quadratic loss, i.e. the pointwise posterior mean, both for the hazard rate itself and for the corresponding survival function. Padgett and Wei (1981) did the same, but building the prior hazard rate as a compound Poisson process with positive deterministic jump-sizes. The main drawback of both proposals is a lack of generality, in that they both assume an increasing hazard rate, as pointed out by the authors themselves. Furthermore, both processes are pure jump ones and therefore they cannot properly model smooth hazard rates.

The paper by Dykstra and Laud (1981) originated a narrow but fertile streamline of research, mainly aimed at achieving greater generality. First, Ammann (1984, 1985) obtained both "bath-tub" and "U-shaped" hazard rates, by considering differences of weighted gamma processes. Then, in the context of multiplicative intensity models, Lo and Weng (1989) were able to build an unrestricted—possibly smooth—hazard rate as the mixture of a kernel with a weighted gamma measure on an Euclidean space. Eventually, James (2003) extended the framework to deal with semiparametric models and also let the measure space be an arbitrary Polish one.

In principle, posterior computations for all of the above priors can be carried out exactly, the difficulty being essentially the same as for the basic priors by Dykstra and Laud (1981). However, these exact computations rely on enumerating all possible partitions of the observations and thus, in practice, simulation techniques are needed. Lo and Weng (1989) suggested to sample partitions by means of the so-called *chinese restaurant algorithm*, which was later improved by devising *weighted* versions of it; see in particular the recent paper by Ho and Lo (2001). For their part, Laud *et al.* (1996) discretised the time axis in order to find a posterior MCMC approximation for priors in the class by Dykstra and Laud (1981). More recently, without resorting to time discretisation, Ishwaran and James (2004) gave two Gibbs samplers for posteriors originated by priors in the wider class by James (2003).

On the other hand, to the best of the author's knowledge, the paper by Padgett and Wei (1981) remained isolated. This was probably due to the restrictive assumption of deterministic jump-sizes, which was crucial for the posterior computations to be feasible. Nowadays, thanks to the upsurge of MCMC methods, this restriction is no longer necessary. Indeed, building on the Ph.D. thesis by La Rocca (2003), this paper shows that a general compound Poisson process with positive jump-sizes can be effectively used in a convolution mixture with a probability density to build a flexible prior hazard rate admitting a straightforward Gibbs sampler. The suggested construction can be seen as generalising the forgotten one by Padgett and Wei (1981), in light of the above described contributions following the paper by Dykstra and Laud (1981).

The paper is organised as follows. Section 2 introduces the topic of Bayesian nonparametric inference on the hazard rate, showing its relevance in the applied context of seismic hazard assessment. Section 3 investigates some properties of the proposed class of prior distributions, eventually giving a time-scale equivariant procedure to elicitate a single prior in the class. Section 4 furnishes an MCMC approximation of the posterior distribution. Section 5 applies the suggested MCMC algorithm to the analysis of some Italian earthquake catalogue data. Section 6 discusses the opportunity of using a compound Poisson process as an approximation to a weighted gamma process.

## 2 Motivation

Let $t_1, \ldots, t_n$ be $n$ observed survival times which are either exact or right censored, according to some binary variables $o_1, \ldots, o_n$: if $o_i = 1$, the event $\{T_i = t_i\}$ has been observed and $t_i$ is *exact*; if $o_i = 0$, the event $\{T_i > t_i\}$ has been observed and $t_i$ is *right censored.* In the above, $T_1, \ldots, T_n$ are positive random variables modelling the observations, which are supposed to be *i.i.d.* with probability density $\rho(t) \exp\{-\int_0^t \rho(s)ds\}$, $t \geq 0$, given the unknown hazard rate $\rho$. Assuming *non-informative censoring*, that is a censoring mechanism independent of the observations, the model likelihood is given by

$$L(\rho) = \prod_{i=1}^{n} \rho(t_i)^{o_i} \exp\left\{-\int_0^{t_i} \rho(s)ds\right\}. \tag{1}$$

In order to carry out Bayesian inference on $\rho$, first $\rho$ itself needs to be built as a stochastic process such that

$$\rho \geq 0, \qquad \exists t > 0 : \int_0^t \rho(s)ds < \infty, \qquad \int_0^\infty \rho(s)ds = \infty. \tag{2}$$

Conditions (2) are easily shown to be necessary and sufficient for $\rho$ to be a valid hazard rate. Then, for each observation $T_i$ as above, it holds that

$$\rho(s) = \lim_{h \downarrow 0} \frac{1}{h} \mathcal{P}(s \leq T_i \leq s + h \mid T_i \geq s), \quad s \geq 0, \tag{3}$$

where $\mathcal{P}$ is conditional on $\rho$ itself, as the model is a Bayesian one. Therefore, the distribution of $\rho$ will express prior beliefs on the unknown instantaneous conditional probability density of event and this will guide its elicitation.

Once $\rho$ has been built, its posterior distribution can be found via the Bayes formula. For example, if the Bayes estimator of $\rho$ under quadratic loss is the posterior quantity of interest, as will be the case for seismic hazard assessment, one needs to compute

$$\mathbb{E}\left[\rho(s) \mid (t_i, o_i)_{i=1}^n\right] = \frac{\mathbb{E}\left[\rho(s)L(\rho)\right]}{\mathbb{E}\left[L(\rho)\right]}, \qquad s \geq 0. \tag{4}$$

A proof of Bayes formula for general dominated models can be found, for example, in the book by Schervish (1995, pages 16–17). From a practical point of view, an MCMC algorithm may well be needed in order to exploit formula (4); see the paper by Tierney (1994) for a nice introduction to MCMC methods.

Interest for the hazard rate in survival analysis is mainly due to the interpretation given by formula (3). However, for positive variables, it is always possible to consider the hazard rate in place of the probability density, as the two exist precisely for the same distributions. In particular, this will be convenient if there are right censored observations, as it is often the case in survival analysis, because in this case Bayes formula becomes awkward to deal with, once the likelihood (1) has been rewritten in terms of the unknown probability density. Moreover, hazard rate estimation can be of interest in the applied context of seismic hazard assessment, as the next subsection shows.

## 2.1 Seismic hazard assessment

According to Vere-Jones (1995), when the object is statistical analysis, an earthquake is essentially described by five coordinates: latitude, longitude and depth of its first motion, together with its origin time and the so-called magnitude, which is a measure of the event size on a logarithmic scale. Then, a suitable framework for statistical modelling is offered by the theory of point processes, which reduces to the theory of counting processes, if the analysis concentrates on the distribution of origin times. This is commonly done by fixing a suitable space-magnitude window, i.e. by only considering strong events in a given seismogenic region.

Let $0 = S_0 < S_1 < \cdots < S_n < \cdots$ be an increasing sequence of random variables modelling the event times at issue. An equivalent representation is given by the counting process

$$N(t) = \sum_{i \geq 1} \mathbb{I}_{\{S_i \leq t\}}, \qquad t \geq 0$$

or, alternatively, by the sequence of inter-event times $T_i = S_i - S_{i-1}$, $i \geq 1$. A nice way to specify the distribution of $N$ is by assuming exchangeability of the inter-event times, that is by letting $N$ be a renewal process, conditionally on the unknown distribution of $T_i$. This is to be considered a reasonable assumption, if the strongest earthquakes only are at issue; see for example the reflections by Wu *et al.* (1995). Indeed, there are many examples in the literature where a parametric renewal model is used to analyse a seismic event sequence. Here, however, the emphasis will be on the nonparametric point of view, which avoids imposing any functional form upon the inter-event time distribution.

An important goal in seismic hazard assessment is the evaluation of what Vere-Jones (1995) calls the *geophysical risk*, that is the instantaneous conditional expected number of events per time unit. Indeed, this is nothing else than the stochastic intensity

$$\lambda(t) = \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}[N(t+h) - N(t) | \mathcal{O}_t], \qquad t \geq 0 \tag{5}$$

of the counting process $N$ with respect to the observed history $\mathcal{O}$, where $\mathcal{O}_t$ is the $\sigma$-algebra of events generated by $N(s)$, $0 \leq s \leq t$; refer to the book by Brémaud (1981) for a rigourous treatment of stochastic intensity. The importance of geophysical risk evaluation is made clear by a simple yet penetrating decision model due to Ellis (1985) in which the optimal strategy to call an earthquake alert is to wait for $\lambda$ to exceed a suitable threshold.

Letting the inter-event times $T_i$, $i \geq 1$ be *i.i.d.* $\sim \rho(t) \exp\{-\int_0^t \rho(s) ds\} dt$, conditionally on the unknown hazard rate $\rho$, as discussed above, it is possible to compute the geophysical risk (5) in what will be called the *nonparametric renewal model* as

$$\begin{aligned} \lambda(t) &= \hat{\rho}_t \left( t - S_{N(t)} \right) \\ \hat{\rho}_t(s) &= h_{N(t)} \left( s; T_1, \ldots, T_{N(t)}, t - S_{N(t)} \right) \\ h_n(s; t_1, \ldots, t_n, t_{n+1}) &= \mathbb{E}[\rho(s) \mid T_1 = t_1, \ldots, T_n = t_n, T_{n+1} > t_{n+1}] \end{aligned}$$

that is through the Bayes estimator of $\rho$ under quadratic loss, where it is worth noting that the last observation is right censored. This result can be proven by first conditioning on $\rho$ and $\mathcal{O}_t$ toghether, thus finding the well known renewal intensity $\rho(t - S_{N(t)})$, $t \geq 0$, then noting that the trace of $\mathcal{O}_t$ on $\{N(t) = n\}$ is the same as the trace on $\{T_{n+1} > t - S_n\}$ of the $\sigma$-algebra of events generated by $T_1, \ldots, T_n$. In this way, Bayesian nonparametric hazard rate estimation, carried out on the inter-event times, becomes a tool for nonparametric geophysical risk evaluation.

# 3 Prior distribution

It is here suggested that the prior hazard rate $\rho$ be built as

$$\rho(t) = \xi_0 k_0(t) + \sum_{j=1}^{\infty} \xi_j k(t - \sigma_j), \qquad t \geq 0, \tag{6}$$

where $\xi_0, \xi_1, \xi_2, \ldots$ are positive independent random variables, with $\xi_1, \xi_2, \ldots$ identically distributed, $\sigma_0 = 0$ and $\sigma_j = \tau_1 + \cdots + \tau_j$, $j \geq 1$, with $\tau_1, \tau_2, \ldots$ independent of $\xi_0, \xi_1, \xi_2, \ldots$ and *i.i.d.* $\sim \mathcal{E}(q)$, while $k_0$ is a positive real function defined on $\mathbb{R}_+$ and integrable on a neighborhood of 0, and $k$ is a probability density on $\mathbb{R}$. Note that $\mathcal{E}(q)$ is the exponential distribution having expected value $q^{-1}$, where $q > 0$. It will been shown in the following that formula (6) defines, under mild conditions, a valid and possibly smooth hazard rate function.

Neglecting for a moment the first term, whose role will be clarified later, the proposed construction turns out to be a convolution mixture of the probability density $k$ with the compound Poisson process $\mu$ having jump-times $\sigma_j$, $j \geq 1$ and jump-sizes $\xi_j$, $j \geq 1$. In fact, formula (6) can be rewritten as $\rho(t) = \xi_0 k_0(t) + \int_{\mathbb{R}_+} k(t-s)\mu(ds)$, $t \geq 0$, thus proving itself a special case of the general mixture by Lo and Weng (1989). The novelty here is the use of a compound Poisson process in place of the weighted gamma process, which allows to replace the integral with a series and thus leads, first of all, to a useful interpretation of the adopted construction.

Introduce a family $\theta_{ij}$, $j \geq 0$, $i \geq 1$ of positive random variables such that, conditionally on $\sigma$ and $\xi$, they are independent and $\theta_{ij}$ follows the distribution determined by the hazard rate $\xi_j k_j(t - \sigma_j)$, $t \geq 0$, where $k_j = k$ for $j \geq 1$. Then, define the positive random variables $T'_i = \min_j \theta_{ij}$, $i \geq 1$ and observe that, conditionally on $\sigma$ and $\xi$, they have the same distribution as the observations and can therefore replace them in the statistical model. This allows to interpret each observation $T_i$ as being originated by countably many competing *hazard sources*, in that it is the minimum of a sequence of latent survival times. Note that, in this interpretation, each hazard source is characterized by its *location* $\sigma_j$ together with its *size* $\xi_j$ and contributes to the hazard shape via $k$, for $j \geq 1$, or via $k_0$, for $j = 0$.

Before investigating further the suggested class of prior distributions, it is appropriate to check that its definition be well posed. To this end, it is sufficient to show that $\rho$ defined by (6) almost surely satisfies conditions (2). In fact, this is done below in the proof of Theorem 1. Preliminarily, it is worth considering a couple of formulas concerning the integration of a Borel function $f \geq 0$ with respect to a compound Poisson process $\mu$:

$$\begin{aligned} \mathbb{E} \int_{\mathbb{R}_+} f d\mu &= q\mathbb{E}[\xi_1] \int_{\mathbb{R}_+} f(s) ds, & (7) \\ \mathbb{V}\mathrm{ar} \int_{\mathbb{R}_+} f d\mu &= q\mathbb{E}\left[\xi_1^2\right] \int_{\mathbb{R}_+} f^2(s) ds, & (8) \end{aligned}$$

where it is assumed that all quantities of interest are finite. About the proof, both (7) and (8) can be proven via a standard monotone argument.

**Theorem 1** *The trajectories of the stochastic process $\rho$ defined by (6) are almost surely valid hazard rates, if $\mathbb{E}[\xi_0] < \infty$, $\mathbb{E}[\xi_1] < \infty$ and $\mathbb{P}\{\xi_1 = 0\} < 1$.*

*Proof.* First, it holds that $\rho \geq 0$ by construction. Then, thanks to Fubini-Tonelli theorem and the Strong Law of Large Numbers, it also holds that

$$\int_0^t \rho(s) ds = \xi_0 K_0(t) + \sum_{j=1}^{\infty} \xi_j [K(t - \sigma_j) - K(-\sigma_j)] \to \infty, \qquad \text{as} \ \ t \to \infty,$$

where $K_0(t) = \int_0^t k_0(s) ds$, $K(t) = \int_{-\infty}^t k(s) ds$ and the convergence is almost sure. Finally, it follows from formula (7) with $f(s) = K(t - s) - K(-s)$, $s \in \mathbb{R}_+$, that

$$\mathbb{E} \int_0^t \rho(s) ds = \mathbb{E}[\xi_0] K_0(t) + q\mathbb{E}[\xi_1] \int_{-\infty}^t [t - \max\{s, 0\}] k(s) ds,$$

which is finite for $t$ small enough, implying $\int_0^t \rho(s) ds < \infty$ almost surely. □

*Remark.* Under the hypotheses of Theorem 1, which can be thought of as in force from now on, each latent variable $\theta_{ij}$ with $j \neq 0$ is defective, as almost surely $\mathcal{P}\{\theta_{ij} = \infty\} = \exp\{-\int_0^\infty \xi_j k(s - \sigma_j) ds\} \geq \exp\{-\xi_j\} > 0$. As for $j = 0$, each $\theta_{ij}$ is defective if and only if $K_0(\infty) = \int_0^\infty k_0(s) ds < \infty$.

A nice feature ot the construction given by formula (6) is that it gives rise to smooth hazard rates, if such are $k_0$ and $k$. More precisely, the following result holds, where $k^{(i)}$ denotes the $i^{\text{th}}$ derivative of $k$.

**Theorem 2** *Let $k_0$ and $k$ be $r$ times continuously differentiable on their respective domains. Furthermore, for all $i = 0 \ldots r$, let $k^{(i)}$ be integrable on $\mathbb{R}$ and such that $|k^{(i)}(x)| \downarrow 0$, as $x \to -\infty$. Then, the trajectories of $\rho$ defined by (6) are almost surely $r$ times continuously differentiable on $\mathbb{R}_+$.*

*Proof.* First of all, it is clear that the first term $\xi_0 k_0$ in the right hand side of formula (6) is $r$ times continuously differentiable on $\mathbb{R}_+$ if and only if such is $k_0$. As for the series $\sum_{j=1}^{\infty} \xi_j k(t - \sigma_j)$, $t \geq 0$, take for now $r = 0$ and consider the issue of its continuity.

It will be enough, due to a well known result on real functions defined by series, to prove almost sure uniform convergence on $[0, t]$, for all $t \geq 0$. To this aim, having fixed $t \geq 0$, note that

$$\sum_{j=\eta}^{\infty} \xi_j k(s - \sigma_j) \leq \sum_{j=\eta}^{\infty} \xi_j k(t - \sigma_j), \quad \forall s \leq t$$

if $\eta$ is taken big enough for $k(x), x \leq t - \sigma_\eta$ to be decreasing; recall that by hypothesis $k(x) \downarrow 0$, as $x \to -\infty$, and by construction $\sigma_j \uparrow \infty$, as $j \to \infty$. Therefore, uniform convergence on $[0, t]$ follows from simple convegence in $t$.

Now, let $r = 1$ and consider the first derivative of $\rho$. The "series of the derivatives" is almost surely absolutely convergent, because it holds that

$$\mathbb{E}\left[\sum_{j=1}^{\infty} \xi_j |k'(t - \sigma_j)|\right] = q\mathbb{E}[\xi_1] \int_{-\infty}^{t} |k'(x)| dx < \infty$$

thanks to formula (7), a change of variable and the integrability of $k'$ on $\mathbb{R}$. Furthermore, following the same reasoning as above, it can be shown that convergence is uniform. Therefore, almost surely, the trajectories of $\rho$ are continuously differentiable on $\mathbb{R}_+$ and it holds that

$$\rho'(t) = \xi_0 k_0'(t) + \sum_{j=1}^{\infty} \xi_j k'(t - \sigma_j), \qquad t \geq 0. \tag{9}$$

The general case $r > 1$ can be dealt with similarly, thus finding an expression analogous to (9) for higher order derivatives of $\rho$. □

As an example, take $k(x) = (2\pi v)^{-\frac{1}{2}} \exp\{-x^2/2v\}$, $x \in \mathbb{R}$, that is let $k$ be a zero mean normal probability density with variance $v$. With this choice, the hypotheses of Theorem 2 are satisfied for all positive $r$, so that the hazard rate $\rho$ is almost surely infinitely smooth. For the sake of simplicity and concreteness, all what follows will be based on this particular choice; the next subsection shows that the resulting family of prior distributions for smooth hazard rates is, at least, flexible enough to express weak opinions.

### 3.1 Elicitation

In order to devise an elicitation procedure for priors in the family defined by equation (6) it is worth considering the pointwise prior mean hazard rate

$$\mathbb{E}[\rho(t)] = \mathbb{E}[\xi_0]k_0(t) + q\mathbb{E}[\xi_1]K(t), \quad t \geq 0, \tag{10}$$

where $K(t) = \int_{-\infty}^{t} k(s)ds$ and equation (10) can be proven via formula (7). Equation (10) shows that, given a valid hazard rate function $r$, it is possible to let $\mathbb{E}[\rho(t)] = r(t)$, for all $t \geq 0$, whenever $r$ is bounded away from zero, by taking $k_0 = \mathbb{E}[\xi_0]^{-1}(r - q\mathbb{E}[\xi_1]K)$. In particular, the special case of interest $r \equiv r_0$, expressing weak prior opinions, will be considered in the following; in this case, the constraint on $r$ can be written as $q\mathbb{E}[\xi_1] \leq r_0$ and can be conveniently satisfied by letting $q\mathbb{E}[\xi_1] = r_0$, so that $k_0(t) \to 0$ as $t \to \infty$ and the hazard source in the origin is "local" as the other ones. Indeed, once $k_0$ has been chosen after $k$ so as to make the prior hazard rate flat, there are no apparent reasons to further distinguish the zeroth hazard source from the other ones and the simplifying assumption $\mathbb{E}[\xi_0] = \mathbb{E}[\xi_1]$, implying $k_0 = q(1 - K)$, can also be made.

Furthermore, consider the pointwise prior variance of $\rho$

$$\mathbb{V}\mathrm{ar}[\rho(t)] = \mathbb{V}\mathrm{ar}[\xi_0]k_0^2(t) + q\mathbb{E}\left[\xi_1^2\right] \int_{-\infty}^{t} k^2(u)\, du, \quad t \geq 0 \tag{11}$$

and take its limit as $t \to \infty$, that is $q\mathbb{E}[\xi_1^2]\|k\|_2^2$, as a single number measure of prior variability. Equation (11) can be obtained via formula (8) and thus, since $\|k\|_2^2 = (4\pi v)^{-\frac{1}{2}}$ for the chosen zero mean normal probability density, the overall prior variability will be measured by $\mathcal{V} = q(4\pi v)^{-\frac{1}{2}}\mathbb{E}\left[\xi_1^2\right]$.

Finally, consider the pointwise prior mean square slope of $\rho$

$$\mathbb{E}\left[\rho'(t)^2\right] = \mathbb{V}\mathrm{ar}[\xi_0]k_0'(t)^2 + q\mathbb{E}\left[\xi_1^2\right] \int_{-\infty}^{t} k'(u)^2\, du, \quad t \geq 0 \tag{12}$$

and take its limit as $t \to \infty$, that is $q\mathbb{E}[\xi_1^2]\|k'\|_2^2$, as an average measure of prior instantaneous rate of change. Equation (12) can be derived by first noting that $\mathbb{E}[\rho'(t)] = 0$, for all $t \geq 0$, then applying formula (8) as above. Thus, the average prior instantaneous rate of change will be measured by $\mathcal{Z} = q(16\pi v^3)^{-\frac{1}{2}}\mathbb{E}\left[\xi_1^2\right]$, since for $k'(x) = -x(2\pi v^3)^{-\frac{1}{2}}\exp\{-x^2/2v\}$, $x \in \mathbb{R}$, it holds that $\|k'\|_2^2 = (16\pi v^3)^{-\frac{1}{2}}$.

Assume now that $T_\infty$ is a finite time-horizon of interest, meaning that the goal of the analysis is to find $\hat{\rho}(t)$, $0 \leq t \leq T_\infty$, and let $M_\infty$ be a "typical" number of extreme points (maxima and minima) to be found within such time horizon in a prior hazard rate trajectory. Then, as each extreme roughly corresponds to travelling away from and back to the mean for a round trip distance of approximately two standard deviations, it is sensible to impose $\mathcal{Z}^{\frac{1}{2}}T_\infty = 2M_\infty\mathcal{V}^{\frac{1}{2}}$ and this allows to set $v = 8^{-1}T_\infty^2 M_\infty^{-2}$.

Eventually, let $\xi_0, \xi_1, \xi_2, \ldots$ be *i.i.d.* $\sim \mathcal{G}(a,b)$ where $\mathcal{G}(a,b)$ is the gamma distribution having expected value $ab^{-1}$ and variance $ab^{-2}$. This choice is a conjugate one, aimed at making Gibbs sampling easy, as the next section will show. Since $\mathbb{E}[\xi_1] = q^{-1}r_0$, it holds that $b = aqr_0^{-1}$ and it is enough to fix $a$ together with $q$. To this aim, let $\mathcal{V}^{\frac{1}{2}} = Hr_0$, where $H$ plays the role of an overall coefficient of variation; then $\mathbb{E}[\xi_1^2] = (1+a^{-1})\mathbb{E}[\xi_1]^2$ implies $(1+a^{-1}) = qH^2\|k\|_2^{-2}$, where $\|k\|_2^2 = 2^{\frac{1}{2}}\pi^{-\frac{1}{2}}M_\infty T_\infty^{-1}$, and $q$ follows from $a$. Since large values of $a$ correspond to small values of $q$, that is few large-sized hazard sources, while small values of $a$ correspond to large values of $q$, that is many small-sized hazard sources, the hyperparameter $a^{-1}$ can be seen as a measure of "nonparametricity" with a critical value of $a = 1$ corresponding to a change in the shape of the gamma distribution. Notice that the expected number of hazard sources in $[0, T_\infty]$ will be $qT_\infty = 2^{\frac{1}{2}}\pi^{-\frac{1}{2}}M_\infty H^{-2}(1+a^{-1})$.

In conclusion, a single prior from the family defined by equation (6) is selected once $r_0$, $H$, $T_\infty$, $M_\infty$ and $a$ have been chosen. It is worth noting that the elicitation procedure is time-scale equivariant: indeed, for $T_i^* = T_i/T_0$, $i \geq 1$, the obvious choices are $r_0^* = T_0 r_0$, $H^* = H$, $T_\infty^* = T_\infty/T_0$, $M_\infty^* = M_\infty$ and $a^* = a$, so that $\rho_*(u)$, $u \geq 0$, is distributed as $T_0\rho(T_0 u)$, $u \geq 0$, that is as the unknown hazard rate of $T_i^*$ when $\rho$ is the unknown hazard rate of $T_i$.

# 4 Posterior distribution

It is here shown that, when the prior hazard rate is defined by equation (6), it is straightforward to find an MCMC approximation of the corresponding posterior distribution. Indeed, the interpretation of equation (6) in terms of competing hazard sources allows to devise a sort of Gibbs sampler which admits a direct implementation in any programming language. To this aim, let $\gamma_i$ be the hazard source originating $t_i$, for all $i = 1 \ldots n$, so that $t_i = \theta_{i\gamma_i}$; the complete likelihood is then given by

$$L(\xi, \sigma) = \prod_{i=1}^{n} [\xi_{\gamma_i} k_{\gamma_i}(t_i - \sigma_{\gamma_i})]^{o_i} e^{-\xi_0 K_0(t_i) - \sum_{j=1}^{\infty} \xi_j [K(t_i - \sigma_j) - K(-\sigma_j)]},$$

where it is worth noting that $\gamma_i$ plays a role for exact observations only, that is when $o_i = 1$. Considering the prior distribution for $(\sigma, \xi)$, the following *full-conditionals* for $(\gamma, \sigma, \xi)$ are found:

$$\begin{aligned}
\wp(\gamma_i | \gamma_{-i}, \sigma, \xi) &\propto \xi_{\gamma_i} k_{\gamma_i}(t_i - \sigma_{\gamma_i}); && (13)\\
\wp(\xi_0 | \gamma, \sigma, \xi_{-0}) &\propto \xi_0^{a_0 - 1 + \sum_{i=1}^{n} \mathbb{I}_{\{\gamma_i = 0\}}} \\
&\quad \cdot\ e^{-\xi_0 \left\{ b_0 + \sum_{i=1}^{n} K_0(t_i) \right\}}; && (14)\\
\wp(\xi_j | \gamma, \sigma, \xi_{-j}) &\propto \xi_j^{a_1 - 1 + \sum_{i=1}^{n} \mathbb{I}_{\{\gamma_i = j\}}} \\
&\quad \cdot\ e^{-\xi_j \left\{ b_1 + \sum_{i=1}^{n} [K(t_i - \sigma_j) - K(-\sigma_j)] \right\}}; && (15)\\
\wp(\sigma_j | \gamma, \sigma_{-j}, \xi) &\propto e^{-\xi_j \sum_{i=1}^{n} [K(t_i - \sigma_j) - K(-\sigma_j)]} \\
&\quad \cdot\ \mathbb{I}_{(\sigma_{j-1}, \sigma_{j+1})}(\sigma_j) \prod_{i:\gamma_i = j} k(t_i - \sigma_j). && (16)
\end{aligned}$$

Note that, in light of the above remark, equation (13) is to be sampled for $i$ such that $o_i = 1$ only, while equations (15) and (16) matter for all $j \geq 1$.

Sampling from the full-conditional (13) is trivial, once the attention has been restricted to a finite number of hazard sources, say $F$, as it is in practice necessary. This should be done so as to obtain a good approximation of $\rho$ on $[0, T_\infty]$, that is letting $\sigma_{F+1} > T_\infty + 3v^{\frac{1}{2}}$. The simplest possibility is to take $F$ such that *a priori* $\mathbb{P}\{\Pi_0 \leq F\} \geq 0.95$, where $\Pi_0$ is a Poisson variable having expected value $qT_\infty + 3qv^{\frac{1}{2}}$; the same condition should then be checked *a posteriori*, via the MCMC output, possibly coming to a larger value of $F$. Another possibility would be to let $F$ vary at each iteration, so that $\sigma_{F+1}$ is the location of the first hazard source to the right of $T_\infty + 3v^{\frac{1}{2}}$; this was done by Arjas and Gasbarra (1994) in a similar context and avoids checking *a posteriori*. In both cases, the full-conditional (16) for $j = F$ should be changed by replacing $\mathbb{I}_{(\sigma_{j-1},\sigma_{j+1})}(\sigma_j)$ with $\mathbb{I}_{(\sigma_{F-1},\infty)}(\sigma_F)e^{-q\sigma_F}$ to account for the lack of conditioning on $\sigma_{F+1}$.

The full-conditionals (14) and (15) are gamma distributions and thus simulation from them is standard, once it has been noted that $K$ is well known and $K_0$ can be rewritten as

$$K_0(t) = q\left\{t[1-K(t)] + (2\pi)^{-\frac{1}{2}}v^{\frac{1}{2}}[1-\exp\left(-t^2/2v\right)]\right\}, \quad t \geq 0$$

thanks to Fubini-Tonelli theorem and some direct computations.

Finally, sampling from the full-conditional (16) is not immediate, but it is simplified by the fact that the support is compact. A convenient possibility is the *slice sampling* technique described by Neal (2003) which can be used as a step within any random scan Gibbs sampler, preserving reversibility and target distribution. See, for example, the recent book by Madras (2002, pages 72–73) for a discussion on reversibility and Gibbs sampling.

## 5 Data analysis

Historical exposition of Italy to seismic hazard gave rise to an outstanding tradition of earthquake data recording and studying. A very recent outcome of this tradition is the catalogue by Gruppo di Lavoro CPTI (2004), which can be considered as the state of the art for the Italian region. The CPTI04, as it is named for short, records 2550 events starting from the Ancient World (first event dated June 253 BC) up to the end of year 2002. Every event is described via detailed information, including its magnitude, origin time and location; furthermore, most events are associated with one of the 36 seismogenic zones described by Gruppo di Lavoro MPS (2004, Appendix 2).

In what follows, attention will be focussed on seismogenic zone 923 (namely *Appennino Abruzzese*) where the terrific Avezzano earthquake of 1915 was originated. Indeed, this was the zone originating the greatest number of events in CPTI04 and the catalogue is deemed to be reasonably complete for this zone, so that it is a good test-bed for the nonparametric

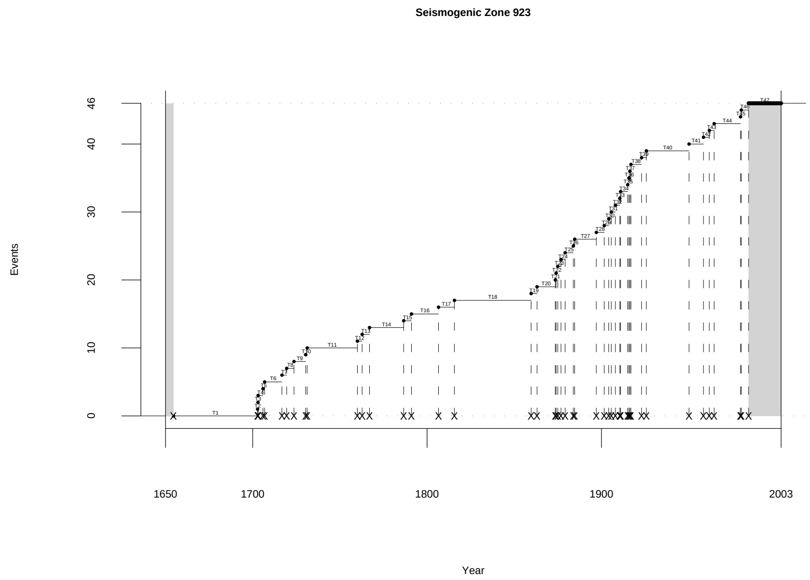


Figure 1: Observed sample path of the process counting seismic events with moment magnitude greater than 5.1 in zone 923 from 1650 to 2002.

renewal model. In particular, according to Gruppo di Lavoro MPS (2004, page 34), mainly on the basis of historical evidence, the event sequence of zone 923 can be considered complete starting from year 1650 for earthquakes with *moment magnitude* greater than 5.1. There are 47 such earthquakes in CPTI04 associated with zone 923, up to the end of year 2002, and the observed sample path of the process counting them is reported in Figure 1.

The 46 exact inter-event times of Figure 1 are considered for the analysis, together with the right censored one spanning from the origin of the last event to the end of the catalogue and corresponding to the grey area on the right of Figure 1. Note that there is also a grey area on the left of Figure 1, corresponding to a left censored inter-event time which unfortunately cannot be handled by the nonparametric renewal model. Roughly speaking, inter-event times (in years) are exact up to the second decimal digit, as the origin day is available but for a couple of events and for these it can be set to the 15$^{\text{th}}$ of the origin month. Furthemore, note that the Gregorian calendar was already in force in year 1650, so that it is not necessary to adjust for it; see the book by Stewart (1997) for a nice discussion of this issue.

The analysis is carried out by means of the R language and environment for statistical computing and graphics. First, a couple of classical techniques are used to assess the relevance of the nonparametric renewal model: a plot of the inter-event time series auto-correlation function shows no evidence against the renewal hypothesis, while the two-sided Kolmogorov-Smirnov test rejects at 5% significance the hypothesis of (exponential inter-event times with) constant hazard rate $\sum_{i=1}^{n} o_i / \sum_{i=1}^{n} t_i = 0.132$, thus raising interest in the nonparametric point of view. Then, the methodology put forth in this paper is applied: a prior in the proposed family is selected by letting $r_0 = 0.1$, $T_\infty = 50$, $H = 1$, $M_\infty = 1$, $a = 1$, and the corresponding posterior is approximated via the suggested MCMC algorithm. Note that $r_0$ is set to the "correct" order of magnitude for the data, while $T_\infty$ roughly corresponds to their range, $H$ is intended to be a "big" coefficient of variation for positive quantities, $M_\infty$ espresses the belief that seismic hazard is slowly varying and $a$ takes its critical value. The posterior hazard rate is plotted in Figure 2, both as pointwise expected value and as pointwise 2.5% and 97.5% quantiles, computed on a chain of length 8000 obtained by *thinning* the 80000 samples following a *burn-in* of 40000 iterations. Finally, the simple conjugate gamma-exponential model is considered for comparison: assuming $\rho(t) = \rho_0$, for all $t \geq 0$, and letting $\rho_0 \sim \mathcal{G}(a, b)$, with $a > 0$ and $b > 0$, it is well known that $\rho_0 \mid t, o \sim \mathcal{G}\left(a + \sum_{i=1}^{n} o_i, b + \sum_{i=1}^{n} t_i\right)$ and thus, for the data at issue, and passing to the non-informative limit $a = b = 0$, the posterior distribution $\rho_0 \mid t, o \sim \mathcal{G}(46, 348)$ is found.

According to Figure 2, the estimated hazard rate for the nonparametric renewal model is "bath-tub" shaped: there is an increase in seismic hazard immediately after an event occurs, then the hazard goes down to a sort of quiescence level and eventually it goes up again, possibly due to stress accumulation. Note that the after-event increase has nothing to do with aftershocks, as neither these nor foreshocks are recorded in the CPTI04. Indeed, the most prominent feature of the "bath-tub" is quiescence, as the mean hazard rate can be seen going well below the lower bound of the credible band for the conjugate gamma-exponential model.

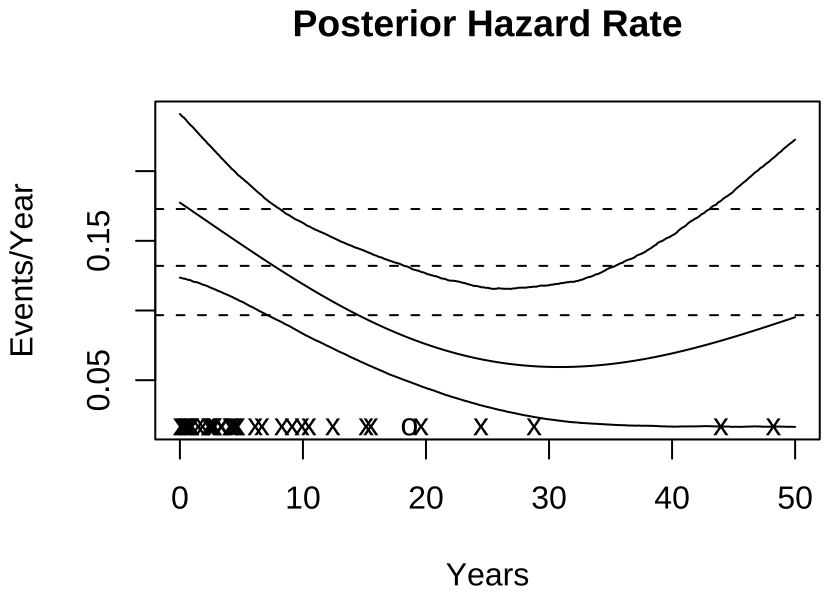


Figure 2: Posterior hazard rate for the inter-event times of zone 923. Solid lines refer to the nonparametric renewal model, dashed lines to the conjugate gamma-exponential one. Pointwise 2.5% quantile, mean and 97.5% quantile are plotted for both models. The 46 exact observations are marked with X, the right censored one with O.

# 6 Discussion

The opportunity of using prior distributions in the family put forth by this paper as an approximation of prior distributions in the family by Lo and Weng (1989) is here discussed. This is intended to be a viable alternative to the exhisting approximation techniques mentioned in the Introduction, with no need for time discretisation and very general applicability.

Let $\mu$ be a completely random measure on the positive half-line, having no deterministic component. It is well known that the distribution of $\mu$ is characterised by the pointwise Laplace transform of its distribution function, that is by

$$\mathbb{E}\left[e^{-z\mu(t)}\right] = \exp\left\{-\int_0^\infty (1-e^{-zu})\, L_t(du)\right\}, \quad z \geq 0,$$

where $L_t$ is the Lévy measure in the Lévy-Khinchine representation of $\mu(t)$, for $t \geq 0$; see the monograph by Kingman (1993, pages 79–82) for details. In particular, if $\mu$ is a weighted gamma process, then

$$L_t(du) = u^{-1} \int_0^t e^{-b(s)u}\, a(ds)\, du \tag{17}$$

for suitable $a$ and $b$; see the paper by Laud *et al.* (1996). Moreover, simple computations show that, if $\mu$ is a compound Poisson process, then

$$L_t(du) = qtg(u)\, du, \tag{18}$$

where $g(s)ds$ denotes the law of its jump-sizes. In the following, it will be shown that (17) can be effectively approximated by (18).

Consider for simplicity $b \equiv b_0$ (namely a *homogeneuos* weighted gamma process) and let $a(t) = a_0 t$, for all $t \geq 0$. Then, equation (17) becomes $L_t(du) = a_0 t\, u^{-1} e^{-b_0 u} du$ and it would be tempting to take $g(u) = u^{-1}e^{-b_0 u}$ and $q = a_0$ in (18). Unfortunately, this would give $\int_0^\infty g(u)du = \infty$, which is clearly unacceptable. The point is that weighted gamma processes jump infinitely many times in finite time intervals, differently from compound Poisson processes. A somewhat natural solution is to ignore the smallest jumps, that is to consider $L_t^d(du) = a_0 t u^{-1} e^{-b_0 u} \mathbb{I}_{[d,\infty)}(u) du$, $t \in \mathbb{R}_+$, for some suitable threshold $d > 0$. It is immediate to check that $L_t^d$ corresponds to a compound Poisson process with intensity $q_d = a_0 \int_d^\infty u^{-1}e^{-b_0 u}\, du$ and jump-size density $g_d(u) = a_0 q_d^{-1} u^{-1} e^{-b_0 u} \mathbb{I}_{[d,\infty)}(u)$, $u \geq 0$. The effectiveness of this approximation relies on the fact that, if $(d_n)_{n\geq 0}$ is an infinitesimal sequence of thresholds, then

$$\int_0^\infty (1 - e^{-zu})\, L_t^{d_n}(du) \to \int_0^\infty (1 - e^{-zu})\, L_t^{d}(du), \quad \text{as } n \to \infty$$

for all positive $t$ and $z$, and thus the approximating sequence of compound Poisson process converges in law to the given wieghted gamma one; see the book by Kallenberg (1975, page 22).

With regards to the posterior MCMC algorithm, it has to be noted that $g_d$ is not a gamma density, because it is zero on $(0, d)$ and the exponent of $u$ is $-1$. Therefore, the approximating compound Poisson process, having Lévy measure $L_t^d$, slightly differs from the ones considered in the previous sections. The only change, however, concerns the jump-size full-conditional

$$\wp(\xi_j|\gamma, \sigma, \xi_{-j}) \propto \xi_j^{\sum_{i=1}^n \mathbb{I}_{\{\gamma_i=j\}}-1} e^{-\xi_j \left\{b_0 + \sum_{i=1}^n [K(t_i-\sigma_j)-K(\sigma_j)]\right\}} \mathbb{I}_{\{\xi_j \geq d\}}$$

which is a truncated gamma, if only $\sum_{i=1}^n \mathbb{I}_{\{\gamma_i=j\}} \geq 1$, while *slice-sampling*, for example, can deal with the case $\sum_{i=1}^n \mathbb{I}_{\{\gamma_i=j\}} = 0$.

# Acknowledgements

The author is indebted to his Ph.D. thesis advisor, Renata Rotondi, and to his Ph.D. program head, Eugenio Regazzini, for suggesting and making possible the research activity originating this paper. The author would also like to thank Gianpaolo Scalia Tomba, Fabrizio Ruggeri and Guido Consonni, for their support and advice, as well as Lancelot James, for some useful comments.

# References

Ammann L.P. (1984) Bayesian nonparametric inference for quantal response data, *The Annals of Statistics*, 12, 636–645.

Ammann L.P. (1985) Conditional Laplace transforms for Bayesian nonparametric inference in reliability theory, *Stochastic Processes and their Applications*, 20, 197–212.

Arjas E. and Gasbarra D. (1994) Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler, *Statistica Sinica*, 4, 505–524.

Brémaud P. (1981) *Point Processes and Queues. Martingale Dynamics*, Springer-Verlag.

Dykstra R.L. and Laud P. (1981) A Bayesian nonparametric approach to reliability, *The Annals of Statistics*, 9, 356–367.

Ellis S.P. (1985) An optimal statistical decision rule for calling earthquake alerts, *Earthquake Prediction Research*, 3, 1–10.

Gruppo di Lavoro CPTI (2004) *Catalogo Parametrico dei Terremoti Italiani, versione 2004 (CPTI04)*, INGV, Bologna, `http://emidius.mi.ingv.it/CPTI/`, in Italian.

Gruppo di Lavoro MPS (2004) *Redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 del 20 marzo 2003. Rapporto Conclusivo per il Dipartimento della Protezione Civile*, INGV, Milano-Roma, aprile 2004, 65pp. + 5 appendici, `http://zonesismiche.mi.ingv.it/`, in Italian.

Ho M.W. and Lo A.Y. (2001) Bayesian nonparametric estimation of a monotone hazard rate, in: *System and Bayesian Reliability*, World Scientific, 301–314.

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, *Journal of the American Statistical Association*, 99, 175–190.

James L.F. (2003) Bayesian calculus for gamma processes with applications to semiparametric intensity models, *Sankhyā, Series A*, 65, 196–223.

Kallenberg O. (1975) *Random Measures*, Akademie Verlag.

Kingman J.F.C. (1993) *Poisson Processes*, Oxford University Press.

La Rocca L. (2003) *Stima bayesiana non parametrica della funzione di intensità con applicazione alla valutazione della pericolosità sismica di alcune zone sismogenetiche italiane*, Ph.D. thesis, Dottorato in Statistica Matematica, Università degli Studi di Pavia, XV Ciclo, in Italian.

Laud P.W., Smith A.F.M. and Damien P. (1996) Monte Carlo methods for approximating a posterior hazard rate process, *Statistics and Computing*, 6, 77–83.

Lo A.Y. and Weng C.S. (1989) On a class of Bayesian nonparametric estimates: II. Hazard rate estimates, *Annals of the Institute of Statistical Mathematics*, 41, 227–245.

Madras N. (2002) *Lectures on Monte Carlo Methods*, American Mathematical Society, Providence, Rhode Island.

Neal R.M. (2003) Slice sampling, *The Annals of Statistics*, 31, 705–741, with discussion, pages 742–767.

Padgett W.J. and Wei L.J. (1981) A Bayesian nonparametric estimator of survival probability assuming increasing failure rate, *Communications in Statistics, Part A—Theory and Methods*, 10, 49–63.

Schervish M.J. (1995) *Theory of Statistics*, Springer-Verlag, corrected second printing, 1997.

Stewart I. (1997) *The Magical Maze*, Phoenix.

Tierney L. (1994) Markov chains for exploring posterior distributions, *The Annals of Statistics*, 22, 1701–1728, with discussion, pages 1728–1762.

Vere-Jones D. (1995) Forecasting earthquakes and earthquake risk, *International Journal of Forecasting*, 11, 503–538.

Wu S.C., Cornell C.A. and Winterstein S.R. (1995) A hybrid recurrence model and its implication on seismic hazard results, *Bulletin of the Seismological Society of America*, 85, 1–16.