Predictive Bayesian nonparametric renewal function estimation

Pier Luigi Conti
Dipartimento di Statistica, Probabilità e Statistiche Applicate
Università di Roma "La Sapienza"
P.le A. Moro, 5
Roma, 00185, Italy
pierluigi.conti@uniroma1.it

Pietro Muliere
Istituto di Metodi Quantitativi
Università L. Bocconi
Viale Isonzo, 25
Milano, 20135, Italy
pietro.muliere@uni-bocconi.it

Abstract

In this paper we study the estimation of the renewal function in a Bayesian nonparametric approach, essentially in a Bayesian predictive perspective. After introducing the prior (a Dirichlet process), a proper bootstrap scheme that approximates the posterior law of the renewal function is studied, and its convergence properties are proved. Then, large-sample results (consistency and Bernstein-von Mises theorem) are obtained.

Key words and phrases. Renewal function, exchangeability, proper Bayesian bootstrap, asymptotics.

AMS 2000 subject classifications. 62G99, 60G09.

1 Introduction and motivation

Renewal processes play a key role in applications of stochastic processes, due to their widespread use in reliability, control processes, telecommunications networks, such as high-speed packed-switched networks. Generally speaking, renewal processes come into force when one is interested in modelling events that happen at random times. An important example is the Internet traffic. Measurement equipments usually record the number of events (for instance, opening of web pages, requested of sending IP packets, etc.) in time intervals, as well as times between consecutive events. Furthermore, often complex stochastic models can be decomposed into *regenerative cycles*, each of them being equivalent to a renewal process; see Resnick (1989).

A key tool in studying and using renewal processes is the *renewal function*, since it allows one to evaluate the expectation and variance (as well as many other characteristics) of the number of arrivals of the relevant events in time intervals; cfr. Gnedenko and Kovalenko (1992).

Let $(X_n;\ n \geq 1)$ be a sequence of non-negative, exchangeable r.v.s. From de Finetti's representation theorem (de Finetti (1937)), there exists a random distribution function F, conditionally on which the r.v.s X_n, $n \geq 1$, are independent and identically distributed (*i.i.d.*) for F. In symbols:

$$F(x) = Pr(X_i \leq x\,|F).$$

In the sequel, we assume that $F(0^-) = 0$, $F(0) < 1$, a.s.. In our setting, the r.v.s X_is can be though as random times between consecutive events.

The renewal counting process $N(t)$; $t \geq 0$ counts the number of events before time t, *i.e.*:

$$N(t) = \sum_{k=1}^{\infty} I_{[0,\,t]}(S_k)$$

where $S_k = X_1 + \cdots + X_k$.

The expectation of $N(t)$ (given F) is the renewal function (r.f.):

$$\begin{aligned} M(t) &= E\left[\sum_{k=1}^{\infty} I_{[0,\,t]}(S_k)\,\middle|\,F\right] \\ &= \sum_{k=1}^{\infty} Pr(S_k \leq t\,|F) \\ &= \sum_{k=1}^{\infty} F^{*k}(t). \end{aligned} \tag{1}$$

The symbol $*$ denotes here the convolution operator, defined as

$$F^{*0}(t) = I_{[0,\infty)}(t),\ F^{*1}(t) = F(t),\ F^{*2}(t) = (F * F)\,(t) = \int_0^t F(t-y)\,dF(y),$$

and, in general,

$$F^{*k}(t) = (F * F^{*k-1})\,(t),\quad k \geq 1.$$

The relationship $\{N(t) \leq k\} \Leftrightarrow \{S_k > t\}$ allow limit properties of $(N(t);\ t \geq 1)$ to be determined from the behaviour of $(X_i;\ i \geq 1)$. Note further that the event $N(t) \leq k$ depends only on $X_1, \ldots, X_k$.

The problem we deal with is the estimation of the r.f. (1), on the basis of sample data. In classical statistics, this problem has been studied by several authors, who mainly obtained asymptotic results; see Frees (1986), Harel *et al.* (1995), Grübel and Pitts (1993), and references therein.

In this paper we study the estimation of the r.f. in a Bayesian nonparametric approach. Although our approach is nonparametric, the use of a prior distribution allows one to convey in the estimation process all relevant non-sample information.

The paper is organized as follows. In Section 2 the prior and posterior distributions are discussed, and the relationship between renewal function and predictive approach is described. Section 3 is devoted to study a bootstrap scheme that approximates the posterior law of the renewal function. In Section 4, some large-sample results are obtained. Finally, a few suggestions on their use are provided in the conclusions.

2 Prior and posterior distributions

As already said, the exchangeability of r.v.s X_is implies that, *via* de Finetti's representation theorem, there exists a probability measure such that the joint law of $X_1, \ldots, X_n$, for any n, can be written as

$$Pr(X_1 \leq x_1, \ldots, X_n \leq x_n) = \int \left\{ \prod_{i=1}^{n} F(x_i) \right\} \mu(dF)$$

where μ is the de Finetti (or prior) measure. From a predictive point of view, A fundamental problem consists in computing the conditional probabilities:

$$Pr(X_{n+1} \leq x \,|X_1, \ldots, X_n)$$

that is, in our case, the probability law of the next random time, given the "past times" $X_1, \ldots, X_n$.

The exchangeability assumption implies:

$$Pr(X_{n+1} \leq x \,|X_1, \ldots, X_n) = E\left[F(x)\,|X_1, \ldots, X_n\right] \tag{2}$$

In the predictive context, the renewal function does have a concrete, important meaning. Imagine to start counting time just after the nth event; in other words, the time axis is rescaled so that time 0 actually coincides with $X_1 + \cdots + X_n$. An important quantity is the expected number of events in $[0,\ t]$, conditionally on the knowledge of $X_1, \ldots, X_n$. In view of (2) and (1), this quantity is equal to

$$
\begin{aligned}
E[N(t)\,|X_1,\,\cdots,\,X_n] &= \sum_{k=1}^{\infty} E\left[I_{[0,\,t]}(X_{n+1}+\cdots+X_{n+k})\right] \\
&= \sum_{k=1}^{\infty} Pr(X_{n+1}+\cdots+X_{n+k} \le t\,|X_1,\,\cdots,\,X_n) \\
&= E[M(t)\,|X_1,\,\cdots,\,X_n] \\
&= E\left[\sum_{k=1}^{\infty} F^{*k}(t)\,\middle|\, X_1,\,\cdots,\,X_n\right]
\end{aligned}
\tag{3}
$$

Relationship (3) shows that the renewal function $M(t)$ plays a central role in predicting the number of events in the (rescaled) time interval $[0,\,t]$. The renewal function plays a central role also in predicting the number of events in different intervals. Denote by $N(t,\,t+h) = N(t+h)+N(t)$ the number of events in $(t,\,t+h]$. From (3) it follows that:

$$
E[N(t,\,t+h)\,||X_1,\,\cdots,\,X_n] = E[M(t+h) - M(t)\,|X_1,\,\cdots,\,X_n].
$$

The problem is how to select the prior. The Bayesian approach to the evaluation of the conditional probability (2) requires to elicit a prior distribution for F on the space of distribution function, and then to use the posterior distribution for F for the computation of the expected value appearing in (2).

An interesting prior for F was introduced by Ferguson (1973) in a fundamental paper an Bayesian approach to nonparametric statistics. We will indicate this prior, called the Dirichlet process, by $\mathcal{D}(a,\,\pi)$, where $a > 0$ is a positive number and π is a proper d.f. on $[0,\,\infty)$. The d.f. π can be interpreted as a prior guess at F, whereas a is the "strength" of this guess. For the definition of the Dirichlet process and a review of its main features, we refer to Ferguson (1973), Ferguson (1974).

Denote by

$$
F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(X_i \le x)}
$$

the empirical distribution function (e.d.f.) based on $X_1,\,\cdots,\,X_n$. It is possible to show (cfr. Regazzini (1978), Lo (1991)) that for every $n \ge 1$

$$
\widetilde{F}_n(x) = Pr(X_{n+1} \le x\,|X_1,\,\cdots,\,X_n) = \frac{a}{n+a}\pi(x) + \frac{n}{n+a}F_n(x) \tag{4}
$$

with $a > 0$ and $\pi(\cdot)$ a distribution function, if and only if F is a Dirichlet process $\mathcal{D}(a,\,\pi)$.

Relationship (3) requires the knowledge of the posterior law of $M(t)$, given $X_1,\,\ldots,\,X_n$. However, it cannot be explicitly obtained. In fact, as well known, the posterior law of F, given $X_1,\,\cdots,\,X_n$, is still a Dirichlet process $\mathcal{D}((n+a),\,\widetilde{F}_n)$. However, neither the posterior law of F^{*k},

$k \geq 2$, nor the posterior law of $M(t) = \sum F^{*k}(t)$ can be written in a closed form. This suggests to resort to some approximation of the posterior law of $M(t)$.

In principle, it is possible to define a prior directly on $M(t)$; $t \geq 0$. However, this is a very difficult task, because of the properties of $M(t)$, that of course should be incorporated into such a prior. First of all, the elementary renewal theorem (cfr. Resnick (1989)) implies that

$$M(t) \sim \frac{t}{E[X_1\,|F]} \quad as\ t \to \infty \tag{5}$$

i.e. $M(t)$ is "asymptotically linear". Furthermore, from the Blackwell's theorem (cfr. Resnick (1989)) it follows that

$$\frac{M(t+h) - M(t)}{t} \sim \frac{h}{E[X_1\,|F]} \quad as\ t \to \infty \tag{6}$$

for every positive h. Features such as (5) and (6) are clearly difficult to be incorporated into a prior. Hence, it is considerably more realistic to construct a prior for F, the d.f. of times between consecutive events. This provides rather a strong motivation to resort to some approximation scheme for the posterior distribution of $M(t)$. In the sequel, we suggest some ideas based on a proper bootstrap scheme.

3 Proper bootstrap approximation

Rubin's bootstrap (cfr. Rubin (1981)) provides a simple way for approximating the posterior law of functional of the Dirichlet process, when a is going to zero. In other words, it essentially ignores the prior law. Of course, it is useful (see Lo (1987)) when the sample size n is large. However, it is of not so great importance if either a is not so small comparatively to n, or π is a proper distribution function.

We present here a technique for approximating the posterior law of $M(t)$. It is based on a proposal by Muliere and Secchi (1996), that we consider here in slightly more general setting. In more detail, the basic ideas of the so-called "proper Bayesian bootstrap" are developed in the renewal processes setting.

The starting point consists in drawing a sample $\overline{\boldsymbol{X}}_m = (\overline{X}_1, \ldots, \overline{X}_m)$ of size m from $\widetilde{F}_n$ (4), and in taking the corresponding e.d.f.

$$\overline{F}_m(x) = \frac{1}{m}\sum_{j=1}^{m} I_{(\overline{X}_j \leq x)} \tag{7}$$

This corresponds to elicit a prior opinion about F by a proper distribution π, and a positive number a measuring our faith on such a "guess".

Next, let

$$\overline{M}_m(t) = \sum_{k=1}^{\infty} \overline{F}_m^{*k}(t) \tag{8}$$

be the r.f. corresponding to (7). The key idea in Muliere and Secchi (1996) is that, conditionally on $X_1, \ldots, X_n$, the probability law of the random function $\overline{F}_m(\cdot)$ (which is a mixture of Dirichlet processes with finite support) approximates the (posterior) probability law of the random function $F(\cdot)$. As a consequence, the probability law of $\overline{M}_m(t)$ (8) should approximate the posterior law of $M(t)$.

The main result of the present section is Proposition 1, where the above claimed convergence is proved. We begin by a couple of lemmas, that used to prove Proposition 1.

Lemma 1 *Assume π has support $[0, \infty)$. Conditionally on $X_1, \ldots, X_n$, as m increases:*

(i) $\overline{F}_m^{*k}(x)$ converges weakly to $F^{*k}(x)$ for every $k \geq 1$ and $x \geq 0$;

(ii) $K \geq 1$, $\sum_{k \leq K} \overline{F}_m^{*k}(x)$ converges weakly to $\sum_{k \leq K} F^{*k}(x)$, for every $K \geq 1$ and $x \geq 0$.

Proof From our assumption on π, conditionally on $X_1, \ldots, X_n$, $F(t)$ is continuous at the point t with probability 1. For this reason, in sequel of the proof we will tacitely admit that t is a continuity point for F.

Let

$$\phi(u) = \int_0^\infty e^{iux}\, dF(x), \quad \overline{\phi}_m(u) = \int_0^\infty e^{iux}\, d\overline{F}_m(x), \ m \geq 1$$

be the characteristic functions of $F(\cdot)$, $\overline{F}_m(\cdot)$, respectively. In Muliere and Secchi (2003) it is shown that, conditionally on $X_1, \ldots, X_n$, the sequence of random functions $(\overline{F}_m(\cdot);\ m \geq 1)$ converges weakly to $F(\cdot)$ in the Skorokhod topology on $D[0, \infty]$. From the continuous mapping theorem (cfr. Billingsley (1968)), and using the uniform continuity of characteristic functions, the sequence of random functions $(\overline{\phi}_m(\cdot);\ m \geq 1)$ converges weakly to $\phi(\cdot)$ as $m \to \infty$. Weak convergence takes place in $C[-\infty, +\infty]$ equipped by the sup-norm. In symbols: $\overline{\phi}_m(\cdot) \overset{w}{\to} \phi(\cdot)$ as $m \to \infty$. In the same way, it can be shown that:

$$\begin{aligned} \overline{\phi}_m(\cdot)^k &\overset{w}{\to} \phi(\cdot)^k \ \textit{as}\ m \to \infty,\ k \geq 1 \\ \sum_{k=1}^{K} \overline{\phi}_m(\cdot)^k &\overset{w}{\to} \sum_{k=1}^{K} \phi(\cdot)^k \ \textit{as}\ m \to \infty,\ K \geq 1 \end{aligned}$$

in $C[-\infty, +\infty]$ endowed by the sup-norm.

From the Skorokhod representation theorem (cfr. Pollard (1984)) there exist, on an appropriate probability space, random functions $\widehat{\phi}_m^k(\cdot)$, $\widehat{\phi}^k(\cdot)$ such that:

(a) $\widehat{\phi}^k(\cdot) \overset{d}{=} \phi^k(\cdot)$, $\widehat{\phi}_m^k(\cdot) \overset{d}{=} \overline{\phi}_m^k(\cdot)$, $m \geq 1$, where $\overset{d}{=}$ denotes equality in distribution;

(b) $sup_{-\infty<u<+\infty} \left| \widehat{\phi}_m(u)^k - \widehat{\phi}(u)^k \right| \overset{a.s.}{\to} 0$ as $m \to \infty$.

Denote now by $\widehat{F}^{*k}(t)$, $\widehat{F}_m^{*k}(t)$ the inverse Fourier-Stieltjes transforms of $\widehat{\phi}^k$, $\widehat{\phi}_m^k$, respectively. By well known properties of the characteristic functions (cfr. Lukacs (1960), pp. 35-38), and recalling that F may be assumed continuous at x, it is possible to argue that

(c) $\widehat{F}^{*k}(t) \stackrel{d}{=} F^{*k}(t)$, $\widehat{F}_m^{*k}(t) \stackrel{d}{=} \overline{F}_m^{*k}(t)$, $m \geq 1$, $k \geq 1$;

(d) $\widehat{F}_m^{*k}(t) \stackrel{a.s.}{\to} \widehat{F}^{*k}(t)$ as $m \to \infty$, for every $k \geq 1$.

From (d) it is possible to conclude that $\widehat{F}_m^{*k}(t)$ tends also in distribution to $\widehat{F}(t)$ as $m \to \infty$, and using (c) conclusion (i) follows. Statement (ii) can be shown in a similar way.□

Lemma 2 *The inequality*

$$F^{*k}(t) \leq F(t)^k \tag{9}$$

holds for every $k \geq 1$.

Proof In view of the monotonicity of F, we have

$$\begin{aligned} F^{*k}(t) &= \int_0^t F(t-y)\, dF^{*k-1}(y) \\ &\leq F(t)F^{*k-1}(t) \end{aligned}$$

from which (9) easily follows.□

Proposition 1 *Assume that the support of* π *is* $[0, \infty)$. *Then, conditionally on* $X_1, \ldots, X_n$, $\overline{M}_m(t)$ *converges in distribution to* $M(t)$, *as* m *goes to infinity.*

Proof Let K be a positive integer. From Lemma 2 we have

$$\begin{aligned} \sum_{k=K+1}^{\infty} F^{*k}(t) &\leq \sum_{k=K+1}^{\infty} F(t)^k \\ &= \frac{F(t)^K}{1-F(t)} \end{aligned}$$

and since $F(t) < 1$ with probability 1, we obtain:

$$\sum_{k=K+1}^{\infty} F^{*k}(t) \stackrel{p}{\to} 0 \ \ as\ K \to \infty. \tag{10}$$

Let x be a continuity point of the (posterior) d.f. of $M(t)$, so that also the d.f. of $F^{*k}(t)$ is continuous at x.

Conditionally on $\boldsymbol{X}_n = (X_1, \ldots, X_n)$, from Lemma 1 we obtain:

$$\begin{aligned} Pr\left(\sum_{k=1}^{\infty} \overline{F}_m^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) &\leq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) \\ &\to Pr\left(\sum_{k=1}^{K} \overline{F}^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) \quad as\ m \to \infty \end{aligned} \tag{11}$$

for any $K \geq 1$.

Take now positive δ, ϵ. By (10), there exists $K = K(\delta, \epsilon)$ such that

$$Pr\left(\frac{F^{*K}(t)}{1-F(t)} \geq \delta \,\middle|\, \boldsymbol{X}_n\right) \leq \frac{\epsilon}{2} \tag{12}$$

Hence, from Lemma 1, for every m "large enough", the inequality

$$Pr\left(\frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)} \geq \delta \,\middle|\, \boldsymbol{X}_n\right) \leq \epsilon \tag{13}$$

is obtained. Using now inequality (13), we have

$$\begin{aligned}
Pr\left(\sum_{k=1}^{\infty} \overline{F}_m^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) &= Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \sum_{k=K+1}^{\infty} \overline{F}_m^{*k}(t) \,\middle|\, \boldsymbol{X}_n\right) \\
&\geq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)} \,\middle|\, \boldsymbol{X}_n\right) \\
&\geq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)},\, \frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)} < \delta \,\middle|\, \boldsymbol{X}_n\right) \\
&\geq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \delta,\, \frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)} < \delta \,\middle|\, \boldsymbol{X}_n\right) \\
&\geq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \delta \,\middle|\, \boldsymbol{X}_n\right) + Pr\left(\frac{\overline{F}_m^{*K}(t)}{1-\overline{F}_m(t)} < \delta \,\middle|\, \boldsymbol{X}_n\right) - 1 \\
&\geq Pr\left(\sum_{k=1}^{K} \overline{F}_m^{*k}(t) \leq x - \delta \,\middle|\, \boldsymbol{X}_n\right) - \epsilon.
\end{aligned} \tag{14}$$

From (11) and (14) the couple of inequalities

$$\begin{aligned}
\liminf_{m\to\infty} Pr\left(\sum_{k=1}^{\infty} \overline{F}_m^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) &\geq Pr\left(\sum_{k=1}^{K} F^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) \\
\limsup_{m\to\infty} Pr\left(\sum_{k=1}^{\infty} \overline{F}_m^{*k}(t) \leq x \,\middle|\, \boldsymbol{X}_n\right) &\leq Pr\left(\sum_{k=1}^{K} F^{*k}(t) \leq x - \delta \,\middle|\, \boldsymbol{X}_n\right) - \epsilon
\end{aligned}$$

and by letting $K \to \infty$, ϵ, $\delta \to 0$, the proposition in proved.□

Proposition 1 essentially tells that (*i*) the approximation to the posterior law of $M(t)$ provided by the proper bootstrap works for every *fixed* n, and (*ii*) the larger m, the better the approximation.

From a practical point of view, the actual distribution of $\overline{M}_m(t)$ is difficult to be obtained exactly. For this reason, we describe a simple algorithm providing a useful numerical approximation.

1. Generate a sample $\overline{\boldsymbol{X}}_m = (\overline{X}_1, \ldots, \overline{X}_m)$ of size m from $\widetilde{F}_n$ (4);
2. Compute the e.d.f. $\overline{F}_m$ (7);
3. Compute the renewal function $\overline{M}_m(t)$ (8);
4. Repeat steps 1-3 r times, in order to obtain r values:

$$\overline{M}_{m,1}(t), \ldots, \overline{M}_{m,r}(t) \tag{15}$$

The e.d.f. computed on the basis of (15):

$$\widehat{H}^t_{m,r}(x) = \frac{1}{r}\sum_{j=1}^{r} I_{(\overline{M}_{m,j}(t) \leq x)}$$

provides an approximation of $Pr(M(t) \leq x \,|\boldsymbol{X}_n)$. The algorithm is essentially a bootstrap procedure, since it samples from a mixture of the e.d.f generated by $\boldsymbol{X}_n$ and π, that, together with the weight a, elicits the prior opinion on F. It is substantially different from Rubin's bootstrap scheme, because in Rubin's bootstrap π and a do not play any role.

From the Glivenko-Cantelli theorem, as r increases we obtain:

$$\sup_x \left| \widehat{H}^t_{m,r}(x) - Pr(\overline{M}_m(t) \leq x \,|\boldsymbol{X}_n) \right| \overset{a.s.}{\to} 0 \quad as\ r \to \infty \tag{16}$$

for every fixed m. By combining (16) and Proposition 1, it easy to obtain the following further result.

Proposition 2 *Assume that r goes to ∞ as m does. Then:*

$$\widehat{H}^t_{m,r}(x) \overset{a.s.}{\to} Pr(M(t) \leq x \,|\boldsymbol{X}_n)$$

as $m \to \infty$, for every point x at which $Pr(M(t) \leq x \,|\boldsymbol{X}_n)$ is continuous.□

4 Consistency and Bernstein-von Mises theorem for the posterior distribution

The goal of the present section is to study the large sample behaviour of the posterior law of $M(\cdot)$. From a theoretical viewpoint, this kind of study is of interest because in Bayesian nonparametrics there are many cases where the asymptotic normality of the posterior (*i.e.* Bernstein-von Mises theorem) does not hold. See, for instance, the paper Freedman (1999) for negative results, and by Kim and Lee (2004) for positive results; seen also Ghosh and Ramamoorthi (2003).

From a practical point of view, on the other hand, Bernstein-von Mises theorem provides a rationale basis to approximate the posterior law of quantities of interest, and to construct Bayesian confidence regions.

As a by-product of our results, we will also prove the consistency of the posterior distribution of $M(\cdot)$. The importance of consistency as a "validation" of Bayesian nonparametric procedures is stressed, for instance in Wasserman (1998), Ghosh and Ramamoorthi (2003).

The regularity conditions needed are listed below.

*A*1. There exists a "true" (population) d.f. F_0, such that $(X_n;\ n \geq 1)$ are independent and identically distributed (*i.i.d.*) with common d.f. F_0. This essentially means that the sequence of observables $(X_n;\ n \geq 1)$ lives on the probability space $(I\!R^\infty,\ \mathcal{B}(I\!R)^\infty,\ P_0^\infty)$, with

$$P_0(B) = \int_B dF_0(x), \quad B \in \mathcal{B}(I\!R)$$

and P_0^∞ is the product measure generated by P_0.

*A*2. The d.f. F_0 is continuous.

In the sequel, we will denote by

$$M_0(t) = \sum_{k=1}^{\infty} F_0^{*k}(t)$$

the "true" population renewal function, and by $\widetilde{M}_n(t)$ the quantity

$$\widetilde{M}_n(t) = \sum_{k=1}^{\infty} \widetilde{F}_n^{*k}(t)$$

where $\widetilde{F}_n(x) = E[F(x)\,|X_1,\ \cdots,\ X_n]$. Note that $\widetilde{M}_n(t)$ is a proper renewal function, and that $\widetilde{M}_n(t) \approx E[M(t)\,|X_1,\ \cdots,\ X_n]$.

In the sequel, we will essentially find the large sample posterior law of $\sqrt{n}(M(t) - \widetilde{M}_n(y))$. Proposition 5 is essentially the "Bayesian counterpart" of asymptotic results obtained by other authors in classical statistics: Grübel and Pitts (1993), Harel *et al.* (1995). We consider here weak convergence in the space $D[0,\ b]$ of *cadlag* functions equipped by the Skorokhod metric, with finite b. A different metric, that allows one to study weak convergence in the whole space $D[0,\ \infty]$, is in Grübel and Pitts (1993).

The approach we use is based on the so-called "Hungarian constructions" (Komlós *et al.* (1975)), that provide an almost sure representation of the posterior law of the stochastic process $M(\cdot)$. The idea of using Hungarian constructions in Bayesian asymptotics goes back to Lo (1987), who obtained the large sample distribution of the "Bayesian bootstrapped" version of $F(\cdot)$. Hungarian constructions are also used in Conti (2004) to study the limiting distribution of the quantile process in a Bayesian setting. The reason why the Hungarian construction approach is useful in our problem is that it allows one to choose in a very convenient way the probability space where the quantile process lives.

A *Kiefer process* $(K(s,\ t);\ 0 \leq s \leq 1;\ t \geq 0)$ is a two-parameter Gaussian process with mean function and covariance kernel given by

$$E[K(s,\ t)] = 0, \quad E[K(s_1,\ t_1)\,K(s_2,\ t_2)] = (\min(s_1,\ s_2) - s_1 s_2)\ \min(t_1,\ t_2),$$

respectively. Clearly, for every fixed $t > 0$ the process $(t^{-1/2}K(s,\, t);\ 0 \leq s \leq 1)$ is a Brownian bridge.

Proposition 3 *Assume that the support of* π *is* $[0,\, \infty)$. *Under A1, A2, there exist a a process* $\widehat{F}(\cdot)$ *and a Kiefer process* $K(\cdot, \cdot)$, *defined on an appropriate probability space, such that*

(i) $\widehat{F}(\cdot) \stackrel{d}{=} F(\cdot)$, *conditionally to* $X_1, \ldots, X_n$, *for every* $n \geq 1$.

(ii) $\sup_x \left| \sqrt{n}(\widehat{F}(x) - \widetilde{F}_n(x)) - \frac{1}{\sqrt{n}} K(F_0(x),\, n) \right| = O\left(\frac{(\log\log n)^{1/4} (\log n)^{1/2}}{n^{1/4}} \right)$ *a.s.-*P_0^{∞}.

Proof See Lo (1987).□

The interest of Proposition 4 is twofold. On one hand, it establishes the consistency of the posterior for $M(\cdot)$. On the other hand, it will be used in proving Proposition 5.

Denote now by $\widehat{M}(t)$ the renewal function

$$\widehat{M}(t) = \sum_{k=1}^{\infty} \widehat{F}^{*k}(t).$$

In view of (i) in Proposition 3, we clearly have $\widehat{M}(t) \stackrel{d}{=} M(t)$ conditionally to $X_1, \ldots, X_n$, for every $n \geq 1$. Hence, to study the large sample behaviour of $\sqrt{n}(M(\cdot) - \widetilde{M}_n(\cdot))$, it is enough to study the limit law of $\sqrt{n}(\widehat{M}(\cdot) - \widetilde{M}_n(\cdot))$ as n increases.

Proposition 4 *Let* $b > 0$ *such that* $F_0(b) < 1$. *If the support of* π *is* $[0,\, \infty)$ *and assumptions A1, A2 are fulfilled, then we have:*

$$\lim_{n \to \infty} P\left(\sup_{0 \leq t \leq b} |M(t) - M_0(t)| > \epsilon \,|\, \boldsymbol{X}_n \right) = 1 \ \ \forall \epsilon > 0. \tag{17}$$

$$\lim_{n \to \infty} P\left(\sup_{0 \leq t \leq b} \left| M(t) - \widetilde{M}_n(t) \right| > \epsilon \,\middle|\, \boldsymbol{X}_n \right) = 1 \ \ \forall \epsilon > 0. \tag{18}$$

Proof First of all, after some algebra it is not difficult to see that the inequality

$$\sup_{0 \leq x \leq b} \left| F^{*k}(x) - F_0^{*k}(x) \right| \leq \left\{ \sum_{j=0}^{k-1} F_0^j(b)\, F^{k-j-1}(b) \right\} \sup_{0 \leq x \leq b} |F(x) - F_0(x)| \tag{19}$$

holds for every $k \geq 1$. From (19) it is then possible to obtain the following relationship:

$$\begin{aligned}
\sup_{0 \leq x \leq b} |M(x) - M_0(x)| &= \sup_{0 \leq x \leq b} \left| \sum_{k=1}^{\infty} (F^{*k}(x) - F_0^{*k}(x)) \right| \\
&\leq \left\{ \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} F_0(b)^{*j} F(b)^{k-j-1} \right\} \sup_{0 \leq x \leq b} |F(x) - F_0(x)| \\
&= \left\{ \sum_{j \geq 0} \sum_{k \geq j+1} F_0(b)^{*j}\, F(b)^{k-j-1} \right\} \sup_{0 \leq x \leq b} |F(x) - F_0(x)|
\end{aligned}$$

$$
\begin{aligned}
&= \left\{ \sum_{j=0}^{\infty} F_0^{*j}(b) \left(\sum_{k-j-1=0}^{\infty} F(b)^{k-j-1} \right) \right\} \sup_{0 \le x \le b} |F(x) - F_0(x)| \\
&= M_0(b) \frac{1}{1 - F(b)} \sup_{0 \le x \le b} |F(x) - F_0(x)| . \qquad (20)
\end{aligned}
$$

From well-known results on the consistency of the Dirichlet process (see, for instance, Ghosh and Ramamoorthi (2003), p. 123) we have further

$$
\lim_{n \to \infty} P \left(\sup_{0 \le x < \infty} |F(x) - F_0(x)| > \epsilon \,|\, \boldsymbol{X}_n \right) = 1 \;\; \forall \epsilon > 0 \qquad (21)
$$

and from (21) and (20) result (17) follows. Relationship (18) is shown in a similar way.□

Proposition 5 *Let $b > 0$ such that $F_0(b) < 1$. Under assumptions A1, A2, there exists a set of sequences $(X_n;\ n \ge 1)$ of observables having P_0^{∞}-probability 1, such that the sequence of stochastic processes $(\sqrt{n}(M(t) - \widetilde{M}_n(t));\ 0 \le t \le b)$, $n \ge 1$, conditionally on $X_1, \ldots, X_n$, converges weakly in $D[0, b]$ to a Gaussian random process having the form*

$$
G(t) = \sum_{k=1}^{\infty} k \int_0^t B(F_0(t-y))\, dF_0^{*k-1}(y) \qquad (22)
$$

where $B(\cdot)$ is a standard Brownian bridge.

Proof As already said, it is enough to show that $\sqrt{n}(\widehat{M}(\cdot) - \widetilde{M}_n(\cdot))$ converges weakly to the Gaussian process (22). For the sake of simplicity, in the sequel we will use the following notation:

$$
\begin{aligned}
W_n(x) &= \sqrt{n}(\widehat{F}(x) - \widetilde{F}_n(x)) \\
B_n(x) &= \frac{1}{\sqrt{n}} K(F_0(x), n)
\end{aligned}
$$

In the first place, it is not difficult to see that

$$
\widehat{F}^{*k}(x) - \widetilde{F}_n^{*k}(x) = \sum_{j=1}^{k} \left\{ (\widehat{F}^{*j} * \widetilde{F}_n^{*k-j})(x) - (\widehat{F}^{*j-1} * \widetilde{F}_n^{*k-j+1})(x) \right\} . \qquad (23)
$$

From (23) the relationship

$$
\begin{aligned}
\sqrt{n}(\widehat{M}(t) - \widetilde{M}_n(t)) &= \sum_{k=1}^{\infty} \sqrt{n} \sum_{j=1}^{k} \left\{ (\widehat{F}^{*j} * \widetilde{F}_n^{*k-j})(x) - (\widehat{F}^{*j-1} * \widetilde{F}_n^{*k-j+1})(x) \right\} \\
&= \sum_{k=1}^{\infty} \sum_{j=1}^{k} \left(\widetilde{F}_n^{*j-1} * \widetilde{F}_n^{*k-j} * W_n \right)(x) \\
&\quad + \sum_{k=1}^{\infty} \sum_{j=1}^{k} \left((\widehat{F}^{*j-1} - \widetilde{F}_n^{*j-1}) * \widetilde{F}_n^{*k-j} * W_n \right)(x) \\
&= A_{1n}(x) + A_{2n}(x) \qquad (24)
\end{aligned}
$$

follows, where

$$A_{1n}(x) = \sum_{k=1}^{\infty}\sum_{j=1}^{k}\left(\widetilde{F}_n^{*j-1} * \widetilde{F}_n^{*k-j} * W_n\right)(x)$$

$$A_{2n}(x) = \sum_{k=1}^{\infty}\sum_{j=1}^{k}\left((\widehat{F}^{*j-1} - \widetilde{F}_n^{*j-1}) * \widetilde{F}_n^{*k-j} * W_n\right)(x)$$

We now show that A_{2n} is asymptotically negligible. To this purpose, we observe that

$$\begin{aligned}
\sup_{0\le x\le b} |A_{2n}(x)| &= \sup_{0\le x\le b}\left|\sum_{k=1}^{\infty}\sum_{j=1}^{k}\left((\widehat{F}^{*j-1} - \widetilde{F}_n^{*j-1}) * \widetilde{F}_n^{*k-j} * W_n\right)(x)\right| \\
&= \sup_{0\le x\le b}\left|\left(\sum_{j=1}^{\infty}(\widehat{F}^{*j-1} - \widetilde{F}_n^{*j-1}) * \left(\sum_{k=j}^{\infty}\widetilde{F}_n^{*k-j}\right) * W_n\right)(x)\right| \\
&= \sup_{0\le x\le b}\left|((\widehat{M} - \widetilde{M}_n) * \widetilde{M}_n * W_n)(x)\right| \\
&\le \widetilde{M}_n(b) \sup_{0\le x\le b}\left|\widehat{M}(x) - \widetilde{M}_n(x)\right| \sup_{0\le x<\infty}|W_n(x)| \qquad (25)
\end{aligned}$$

A result by Lo (1983) shows that the posterior law of $\sup_x |W_n(x)|$ tends to distribution of the supremum of a Brownian bridge. Furthermore, (i) Proposition 4 implies that the posterior of $\sup_{0\le x\le b} |\widehat{M}(x) - \widetilde{M}_n(x)|$ tends, as n increases, to a distribution degenerate at zero, and (ii) $\widetilde{M}_n(b)$ tends to $M_0(b)$ a.s.-P_0^{∞}, as a consequence of Frees (1986) and (4). As a consequence, from (25) the relationship

$$\lim_{n\to\infty} P\left(\sup_{0\le x\le b} |A_{2n}(x)| > \epsilon \,|\boldsymbol{X}_n\right) = 1 \;\; \forall \epsilon > 0. \qquad (26)$$

The same approach can be used in order to prove that $A_{1n}(x)$ is asymptotically equivalent to

$$\begin{aligned}
\sum_{k=1}^{\infty} k\,(F_0^{*k-1} * W_n)(x) &= \sum_{k=1}^{\infty} k \int_0^x W_n(x-y)\,dF^{*k-1}(y) \\
&= \sum_{k=1}^{\infty} k \int_0^x B_n(x-y)\,dF^{*k-1}(y) \\
&\quad + \sum_{k=1}^{\infty} k \int_0^x (W_n(x-y) - B_n(x-y))\,dF^{*k-1}(y) \\
&= \sum_{k=1}^{\infty} k \int_0^x B_n(x-y)\,dF^{*k-1}(y) + O\left(\frac{(\log\log n)^{1/4}(\log n)^{1/2}}{n^{1/4}}\right) \qquad (27)
\end{aligned}$$

Taking into account that B_n is a Brownian bridge on the scale of F_0, result (22) is obtained as n goes to infinity.□

The Gaussian process (22) possesses a mean function and a covariance kernel equal to:

$$\begin{aligned} E[G(t)] &= 0 \\ E[G(t)\,G(y)] &= E\left[\sum_{k=1}^{\infty}\sum_{k=1}^{\infty} k\,h \int_0^t\int_0^y B(F_0(t-u))\,B(F_0(y-v))\,dF_0^{*k-1}(u)\,dF_0^{*h-1}(v)\right] \qquad (28) \\ &= \sum_{k=1}^{\infty}\sum_{k=1}^{\infty} k\,h \int_0^t\int_0^y \{\min(F_0(t-u),\,F_0(y-v)) \\ &\quad -F_0(t-u)\,F_0(y-v)\}\,dF_0^{*k-1}(u)\,dF_0^{*h-1}(v). \qquad (29) \end{aligned}$$

Remark When $y = t$, formula (29) considerably simplifies. In fact, by observing that

$$\begin{aligned} &\int_0^t\int_0^t \min(F_0(t-u),\,F_0(t-v))\,dF_0^{*k-1}(u)\,dF_0^{*h-1}(v) \\ = &\int_0^t\int_0^v F_0(t-v)\,dF_0^{*k-1}(u)\,dF_0^{*h-1}(v) + \int_0^t\int_0^u F_0(t-u)\,dF_0^{*k-1}(u)\,dF_0^{*h-1}(v) \\ = &\int_0^t F_0(t-v)\,F_0^{*k-1}(v)\,dF_0^{*h-1}(v) + \int_0^t F_0(t-u)\,F_0^{*h-1}(u)\,dF_0^{*k-1}(u) \\ = &\int_0^t F_0(t-u)\,d(F_0^{*k-1}(u)\,F_0^{*h-1}(u)) \\ = &\int_0^t F_0^{*k-1}(t-u)\,F_0^{*h-1}(t-u)\,dF_0(u) \\ = &(F_0 * (F_0^{*k-1}\,F_0^{*h-1}))(t); \qquad (30) \\ &\int_0^t F_0(t-u)\,dF_0^{*k-1}(u) = F^{*k}(t); \qquad (31) \\ &\int_0^t F_0(t-v)\,dF_0^{*h-1}(v) = F^{*h}(t); \qquad (32) \end{aligned}$$

from (29) and (30)-(32) it is easy to deduce the following relationship

$$E[G(t)^2] = \sum_{k=1}^{\infty}\sum_{k=1}^{\infty} k\,h\left\{(F_0 * (F_0^{*k-1}\,F_0^{*h-1}) - F_0^{*k}\,F_0^{*h})(t)\right\}. \qquad (33)$$

Proposition 5 provides a simple idea to approximate the posterior of $M(\cdot)$, since it suggests to approximate the posterior of $(\sqrt{n}(M(t) - \widetilde{M}_n(t));\ 0 \le t \le b)$ with a Gaussian process having null mean function and covariance kernel (29). The representation formula (22) provide a simple idea on how to perform such an approximation. In view of (18), it is enough to take an integer K sufficiently large, and to approximate the posterior law of $\sqrt{n}(M(\cdot) - \widetilde{M}_n(\cdot))$ with the probability distribution of

$$\sum_{k=1}^{K} k\int_0^t B(\widetilde{F}_n(t-y))\,d\widetilde{F}_n^{*k-1}(y). \qquad (34)$$

Because of the special structure of (34), the problem of simulating trajectories of the stochastic process (34) essentially reduces to simulate trajectories from a Brownian bridge.

5 Conclusions

In this paper we have provided a simple (and convenient, as well) way to approximate the posterior distribution of the renewal function, based on Bayesian bootstrap. Such an approximation is particularly remarkable because of its simplicity, and also because, in view of Proposition 1, it provides an approximation for the whole random function $M(\cdot)$.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Conti, P. L. (2004). Approximated inference for the quantile function via Dirichlet process. *Metron*, **LXII**, 201–222.

de Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives. *Annales de l'Institut H. Poicaré*, **7**, 1–68.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, **1**, 209–230.

Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *The Annals of Statistics*, **2**, 615–629.

Freedman, D. A. (1999). A Bernstein-von Mises Theorem in the Nonparametric Right-censoring Model. *The Annals of Statistics*, **27**, 1119–1140.

Frees, E. W. (1986). Nonparametric Renewal Function Estimation. *The Annals of Statistics*, **14**, 1366–1378.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics.* Springer Verlag, New York.

Gnedenko, B. V. and Kovalenko, I. V. (1992). *Introduction to queueing theory.* Birkhäuser, Boston.

Grübel, R. and Pitts, S. M. (1993). Nonparametric estimation in renewal theory i: the empirical renewal function. *The Annals of Statistics*, **21**, 1431–1451.

Harel, M., O'Cinneide, C. A., and Schneider, H. (1995). Asymptotics of the Sample Renewal Function. *Journal of Mathematical Analysis and Applications*, **189**, 240–255.

Kim, J. and Lee, J. (2004). The Bernstein-von Mises Theorem of Survival Models. *The Annals of Statistics*, **32**, 1492–1512.

Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. (part i). *Z. Wahrsch. verw. Gebiete*, **32**, 111–131.

Lo, A. Y. (1983). Weak convergence for Dirichlet processes. *Sankhya A*, **45**, 105–111.

Lo, A. Y. (1987). A large sample study of the Bayesian bootstrap. *The Annals of Statistics*, **15**, 360–375.

Lo, A. Y. (1991). A characterization of the Dirichlet process. *Statistics and Probability Letters*, **12**, 185–187.

Lukacs, E. (1960). *Characteristic Functions.* Griffin, London.

Muliere, P. and Secchi, P. (1996). Bayesian nonparametric predictive inference and bootstrap techniques. *Annals of the Institute of Statistical Mathematics*, **48**, 663–673.

Muliere, P. and Secchi, P. (2003). Weak convergence of a Dirichlet-multinomial process. *Georgian Mathematical Journal*, **10**, 319–324.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer Verlag, New York.

Regazzini, E. (1978). Intorno ad alcune questioni relative alla definizione del premio secondo la teoria della credibilità. *Giornale dell'Istituto Italiano degli Attuari*, **1-2**, 77–89.

Resnick, S. I. (1989). *Adventures in Stochastic Processes.* Birkhäuser, Boston.

Rubin, D. B. (1981). The Bayesian bootstrap. *The Annals of Statistics*, **9**, 130–134.

Wasserman, B. (1998). Asymptotic properties of nonparametric Bayesian procedures. In: *Practical Nonparametric and Semiparametric Bayesian Statistics* (D. Dey, P. Müller, D. Sinha Eds.). Springer Verlag, New York.