Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 120 (2004) 65–84

journal of statistical planning and inference

www.elsevier.com/locate/jspi

# Bootstrap approximations for Bayesian analysis of Geo/G/1 discrete-time queueing models

Pier Luigi Conti*

*Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università di Roma "La Sapienza", P.le Aldo moro 5, Rome 00185, Italy*

Received 23 February 2001; accepted 2 October 2002

**Abstract**

In this paper we consider a Bayesian nonparametric approach to the analysis of discrete-time queueing models. The main motivation consists in applications to telecommunications, and in particular to asynchronous transfer mode (ATM) systems. Attention is focused on the posterior distribution of the overflow rate. Since the exact distribution of such a quantity is not available in a closed form, an approximation based on "proper" Bayesian bootstrap is proposed, and its properties are studied. Some possible alternatives to proper Bayesian bootstrap are also discussed. Finally, an application to real data is provided.


*MSC:* 62F15; 62G09; 62P30

*Keywords:* Nonparametrics; Overflow rate; Geo/G/1 models; Waiting time distribution; Bayesian bootstrap; Asymptotics

## 1. Introduction and preliminaries

The statistical analysis of queueing systems is an important interface between queueing theory and its applications to real cases. In fact, the parameters characterizing a queueing model are usually unknown, and must be estimated on the basis of the available real data. In the recent years there has been a considerable growth of interest in Bayesian inference for queueing models. See, among others, the papers by Mc Grath et al. (1987), Mc Grath and Singpurwalla (1987), and the series of papers by Armero

* Tel.: +39649940707; fax: +3964959241.
*E-mail address:* pierluigi.conti@uniromal.it (P.L. Conti).


doi:10.1016/S0378-3758(02)00499-8

(1994) and by Armero and Bayarri (1994a, b, 1996, 1998). All those papers deal with continuous-time parametric models, and the "unknown" part of the model is represented by a finite number of parameters. Furthermore, attention is focused on *Markovian* queues, where the arrivals are paced by a Poisson process, and the service time distribution is either exponential or Erlang (i.e. convolution of exponential distributions). These stringent assumptions allow one to write down explicitly the equilibrium waiting time distribution, which turns out to depend on a finite number of unknown parameters. In many cases, their posterior distribution is obtained in a closed form. To our knowledge, the only paper dealing with nonparametric statistics for continuous-time queueing models is Ruggeri et al. (1999). They consider a queueing model with Poisson input, and general service time. The results obtained in that paper, despite their interest, do not fit broadband teletraffic applications for several reasons. In fact, many real communication systems work in *discrete* time; continuous-time models are obtained by fluid-flow approximations, and could overestimate the performance of the systems (Roberts et al., 1996). Furthermore, in that paper no attention is paid to the overflow rate, which is the most important performance measure for many queueing systems.

Motivated by applications to teletraffic, and in particular to asynchronous transfer mode (ATM) broadband communication systems (Parker, 1995), nonparametric statistical problems for discrete-time queueing models have been recently considered by Conti (1999). In ATM, information is segmented into fixed-size transmission units called *cells*. The transmission time of a cell is the *cell-time*, and it is naturally assumed as a *time slot*. During a time slot, exactly one cell is transmitted, and $k \geqslant 0$ cells can simultaneously enter the system, from different connected sources, and have to be transmitted. The system works in discrete time, which is measured in terms of time slots. Cells that cannot be immediately transmitted are stored in a buffer, and form a queue. Their transmission is delayed according to a FIFO rule. In particular, in Conti (1999) the attention is focused on Bayesian inference for Geo/G/1 models. These models are considered realistic and useful for cell-level teletraffic applications; see Roberts et al., 1996; Brunel and Kim, 1993; Gravey et al. (1990) and references therein. Let $T_i$ be the $i$th inter-arrival time, (i.e. the number of time slots between the $(i-1)$th and the $i$th arrivals), and let $S_i$ be the r.v. $i$th service time (i.e. the number of cells corresponding to the $i$th arrival). From now on we assume that

(i) $T_i \sim \mathrm{Geo}(\lambda)$, i.e. $P(T_i = k|\lambda) = \lambda(1-\lambda)^{k-1},\ k \geqslant 1,\ 0 < \lambda < 1$.
(ii) $P(S_i = k|P_b) = b(k),\ k \geqslant 1\ (\sum_{k\geqslant 1} b(k) = 1)$, $P_b$ being the probability measure (concentrated over positive integers) corresponding to the service time distribution.
(iii) $\mu = E(S_i|P_b) < \infty$.
(iv) $(T_i;\ i \geqslant 1)$ and $(S_i;\ i \geqslant 1)$ are two sequences of i.i.d. r.v.'s, conditionally on $\lambda$ and $P_b$; the r.v.'s $T_i$'s are independent of $S_i$'s, conditionally on $\lambda$ and $P_b$.

The quantities $\lambda$ and $b(k)$'s have to be estimated on the basis of observed data that in our case are assumed to be $n$ inter-arrival times $\mathbf{T}_n = (T_1, \ldots, T_n)$, and the corresponding service times $\mathbf{S}_n = (S_1, \ldots, S_n)$. Note that no special parametric hypotheses are made on the service time distribution.

In Conti (1999), a nonparametric approach to the estimation of the probability generating function (p.g.f.) of the equilibrium waiting time distribution is taken. The p.g.f. of the waiting time distribution is given by

$$W(z) = \frac{(1-\rho)(1-z)}{1-\lambda-z+\lambda B(z)} I_{(0,1)}(\rho), \tag{1}$$

where

$$B(z) = \sum_{k=1}^{\infty} z^k b(k)$$

is the p.g.f. of the service time distribution, $I_A(\cdot)$ is the indicator function of the set $A$, and $\rho = \mu\lambda$.

Eq. (1) shows that the p.g.f. $W(\cdot)$ is a transform of both inter-arrival and service time distributions. In other words, it can be seen as a functional that maps the inter-arrival and service time distributions onto the equilibrium waiting time distribution.

Note that the p.g.f. (1) is proper if and only if (iff) $\rho < 1$, that is iff the model is stable. Roughly speaking, if $\rho \geqslant 1$, then we have $W(\cdot) \equiv 0$, and hence the probability distribution corresponding to $W(\cdot)$ takes finite values with probability zero. This corresponds to the well-known fact that the equilibrium waiting time is almost surely infinite whenever $\rho \geqslant 1$, that is whenever the model is unstable.

The approach based on the p.g.f. $W(\cdot)$ is motivated by two basic facts. First of all, the equilibrium waiting time distribution is known in a closed form only when the service time distribution is geometric (Hunter, 1983). Unfortunately, a geometric service time distribution cannot be reasonably assumed in any ATM application; see, for instance, Roberts et al. (1996). In the second place, all relevant queueing characteristics can be expressed as functionals of such a p.g.f. As a consequence, virtually all results in statistical analysis of queueing characteristics can be obtained as by-products. Finally, the consideration of the overflow rate (Section 3) naturally leads to use the p.g.f. $W(\cdot)$.

The definition of a prior law directly for $W(\cdot)$, and the computation of the corresponding posterior conditionally on the sample data, is practically "too difficult" (see the remarks in Conti, 1999). For this reason, it is preferable to resort to an "indirect" approach as in Conti (1999), based on the following steps.

(a) Constructing priors for inter-arrival and service time distributions.
(b) Updating them on the basis of sample data.
(c) Studying the induced posterior distribution of the p.g.f. of the waiting time distribution.

Since the posterior of the p.g.f. $W(\cdot)$ cannot be obtained in a closed form, after calculating the posterior laws of $\lambda$ and $b(k)$'s it is necessary to find out (at least) an approximation of the posterior law of $W(\cdot)$ conditionally on the sample data. In Conti (1999), the attention is focused on asymptotic approximations that hold when the sample size is "large enough". The most important result obtained in that paper is, at least from a practical point of view, an asymptotic approximation of the posterior mean of the overflow rate. No attempt is made to approximate its posterior distribution.

The goal of the present paper is to obtain useful approximations of the posterior distributions of measures of the system performance. We focus our attention on the overflow rate. This considerably improves, from a theoretical point of view as well as from a practical one, the results already obtained in Conti (1999). For instance, it is now possible to construct (approximated) credibility regions.

The paper is organised as follows. Approximations of the posterior distribution of $W(\cdot)$, based on the "proper" Bayesian bootstrap (Muliere and Secchi, 1996) are considered in Section 2, and their limiting behaviour is studied. The main result (Proposition 1) essentially provides an infinite-dimensional version of Theorem 3.1 by Muliere and Secchi (1996).

As a by-product, an approximation of the posterior distribution of the overflow rate is obtained in Section 3. Attention is also focused on the behaviour of approximated credibility regions. The motivation to the study of Section 3 is that the overflow rate is of primary interest in the evaluation of the system performance. We stress also that no assumptions are made on the "true parameters" of the queueing model. In particular, the stability of the system is not assumed.

Bayesian bootstrap was originally proposed by Rubin (1981); a large sample study, in case of a Dirichlet prior, is in Lo (1987). Extensions named "Bayesian bootstrap clones" were subsequently proposed by Lo (1991). The rationale of all those Bayesian bootstrap schemes is that they provide approximations which are frequently better than the crude first-order asymptotic approximation based on Bernstein–Von Mises theorem. Experimental results for law school data are provided by Lo (1991). A theoretical study for the approximation of the posterior distribution of an unknown mean, when the prior is a Dirichlet process, can be found in Weng's (1989) paper.

The "proper" Bayesian bootstrap by Muliere and Secchi (1996) is different from all others Bayesian bootstrap schemes, since it possesses a special feature that makes it particularly attractive: it works for every *fixed* sample size $n$, and does not require any large sample argument. This is in fact the main motivation for its use. As it appears in the sequel, we will always consider the sample size $n$ as fixed. Our interest will be mainly in obtaining an approximation of the *actual* posterior distribution of special non-linear functionals of the Dirichlet process (see Section 3), without necessarily assuming $n$ "large". A discussion of the relative merits of proper Bayesian bootstrap will be given in Section 4.

## 2. Prior specifications and proper Bayesian bootstrap approximations for $W(\cdot)$

The assumptions on the prior distributions for $\lambda$ and $P_b$ are listed below.

(P1) The prior measure for $P_b$ is a Dirichlet process such that for every $k \geqslant 1$ the joint distribution of $(b(1), \ldots, b(k))$ is Dirichlet $\mathscr{D}(\beta(1), \ldots, \beta(k); \beta - \beta(1) - \cdots - \beta(k))$, where $\beta(\cdot)$ is a finite measure with support the set of all positive numbers. In the sequel, the sum of all $\beta(k)$'s is denoted by $\beta$, and it is further assumed that $\sum k\beta(k) < \infty$.

(P2) The prior for $\lambda$ is a natural conjugate Beta distribution $\mathrm{Be}(\alpha_1, \alpha_2)$.
(P3) The prior distributions of $P_b$ and $\lambda$ are independent.

Denote now by $\bar{\zeta}$ ($\geqslant 1$) the quantity

$$\bar{\zeta} = \sup\left\{ z \geqslant 1 : \sum_{k=1}^{\infty} E[b(k)] z^k < \infty \right\}$$

and let $\bar{z}$ be the greatest positive root of the equation $G(z) = 0$, where $G(z) = 1 - \lambda - z + \lambda B(z)$. Then, it is easily seen that $\bar{z} = 1$ ($\bar{z} > 1$, respectively) iff $\bar{\zeta} = 1$ ($\bar{\zeta} > 1$, respectively). Moreover, random function (1) is a.s. analytic in $(0, \bar{z})$, and continuous in $[0, \bar{z})$ ($[0, 1]$ if $\bar{z} = 1$).

In the sequel, we denote by $\hat{\lambda}$ and $\hat{b}(j)$ the quantities

$$\hat{\lambda} = n \left( \sum_{i=1}^{n} T_i \right)^{-1}$$

and

$$\hat{b}(j) = n^{-1} \sum_{i=1}^{n} I_{\{j\}}(S_i) \quad j \geqslant 1,$$

respectively.

Then the posterior means of $\lambda$, $\mu$, and $B(z)$ are given by

$$E[\lambda | \mathbf{T}_n, \mathbf{S}_n] = \frac{n + \alpha_1}{n\hat{\lambda}^{-1} + \alpha_1 + \alpha_2},$$

$$E[\mu | \mathbf{T}_n, \mathbf{S}_n] = \sum_{k=1}^{\infty} k E[b(k) | \mathbf{S}_n] = \sum_{k=1}^{\infty} k \left\{ \frac{\beta(k) + n\hat{b}(k)}{n + \beta} \right\},$$

$$E[B(z) | \mathbf{T}_n, \mathbf{S}_n] = \sum_{k=1}^{\infty} z^k \left\{ \frac{\beta(k) + n\hat{b}(k)}{n + \beta} \right\},$$

respectively.

As it will be clearer in Section 3, we are chiefly interested in the posterior distribution of some special (generally non-linear) functionals of $W(\cdot)$. Apart from some very special cases, it is not usually possible to obtain in a closed form the posterior distribution of such functionals. Hence, it is necessary to resort to some kind of approximation. Now, virtually every numerical approximation requires to simulate from the posterior of $(b(k);\ k \geqslant 1)$. Define

$$\pi_b(k) = E[b(k) | \mathbf{T}_n, \mathbf{S}_n] = \frac{1}{n + \beta} \beta(k) + \frac{n}{n + \beta} \hat{b}(k), \quad k \geqslant 1$$

and let $(n + \beta)\pi_b(\cdot)$ be the posterior driving measure of the Dirichlet process $(b(k);\ k \geqslant 1)$. The simulation problem is easily solved when $\pi_b(\cdot)$ possesses a finite support. Without loss of generality, it can be assumed equal to $\{1, \ldots, q\}$. Then, one has simply to simulate from a Dirichlet vector, and this can be done, for instance, by resorting to the well-known representation formula

$$b(k) \stackrel{\mathrm{d}}{=} \frac{Y_k}{\sum_{k=1}^{q} Y_k}, \qquad k = 1, \ldots, q,$$

where $Y_1, \ldots, Y_q$ are independent r.v.'s with Gamma $\mathrm{Ga}(\beta(k) + n\hat{b}(k), 1)$ distributions, and $\stackrel{\mathrm{d}}{=}$ denotes the equality in distribution. However, when $\pi_b(\cdot)$ has infinite support, it is no longer possible to simulate directly from the posterior law of $(b(k);\ k \geqslant 1)$.

In this section we use a Bayesian bootstrap scheme essentially due to Muliere and Secchi (1996). In Section 4, further possible choices will be discussed in Section 4. Let $\pi_b(\cdot)$ be the probability measure, concentrated over the non-negative integers, such that (conditionally on $\mathbf{S}_n$, $\mathbf{T}_n$)

$$\pi_b(k) = E[b(k)|\mathbf{T}_n, \mathbf{S}_n] = \frac{1}{n+\beta}\,\beta(k) + \frac{n}{n+\beta}\,\hat{b}(k), \qquad k \geqslant 1$$

and let $\mathbf{S}_m^* = (S_1^*, \ldots, S_m^*)$ be an i.i.d. sample of size $m$ from the probability measure $\pi_b$. Denote further by

$$\hat{b}_m^*(k) = \frac{1}{m} \sum_{i=1}^{m} I_{\{k\}}(S_i^*), \qquad k \geqslant 1 \tag{2}$$

the empirical counterparts of $\pi_b(k)$'s. Finally, let $(b_m^*(k);\ k \geqslant 1)$ be a Dirichlet process, independent of $\mathbf{T}_n$, with driving measure $((n+\beta)\hat{b}_m^*(k);\ k \geqslant 1)$, conditionally on $\mathbf{S}_n$, $\mathbf{S}_m^*$. The idea is to approximate the posterior distribution of $(b(k);\ k \geqslant 1)$ by the conditional distribution of $(b_m^*(k);\ k \geqslant 1)$, given $\mathbf{S}_n$.

This approximation naturally leads to approximate the posterior distribution of $B(z)$ by the conditional distribution of

$$B_m^*(z) = \sum_{k=1}^{\infty} b_m^*(k) z^k.$$

Hence, the posterior distribution of $W(z)$ is approximated by the conditional distribution of

$$W_m^*(z) = \frac{(1-\rho_m^*)(1-z)}{1-\lambda-z+\lambda B_m^*(z)}\, I_{(0,1)}(\rho_m^*),$$

where $\rho_m^* = \lambda\mu_m^*$, $\mu_m^* = \sum kb_m^*(k)$.

As a consequence of the Glivenko–Cantelli theorem, conditionally on $\mathbf{S}_n$ and $\mathbf{T}_n$, $\hat{b}_m^*(k)$'s tend a.s. to $\pi_b(k)$'s as $m$ goes to infinity. As proved in Proposition 1, this implies that the distribution of $W^*(z)$ weakly converges to the posterior of $W(z)$.

**Lemma 1.** *Under the assumption* P1, *the sequence of stochastic processes* $(B_m^*(\cdot);\ m \geqslant 1)$ *converges weakly to* $B(\cdot)$ *in the space* $C[0,r]$ *endowed by the sup-norm*, *for every* $1 \leqslant r < \bar{\zeta}$.

**Proof.** See Appendix A.

**Proposition 1.** *Under the assumptions of Lemma* 1, *the sequence of stochastic processes* $(W_m^*(\cdot))$ *converges weakly in* $C[0,r]$, *equipped with the sup-norm, to* $W(\cdot)$.

**Proof.** See Appendix A.

## 3. Applications to the approximation of the posterior distribution of the overflow rate

The results obtained in Section 2 are mainly interesting because they provide an easy way to obtain approximations for the posterior distribution of quantities of interest in queueing analysis. As specified in Section 1, the goal of statistical analysis of queueing systems is usually the evaluation of their performance. Since measures of performance are essentially functionals of the (equilibrium) waiting time distribution, they can be also expressed as functionals of its p.g.f. More formally, virtually every measure of performance $\theta$ of the systems can be represented as $\theta = \Theta(W)$, $\Theta(\cdot)$ being an appropriate functional of $W$. As a by-product of the approximation of the posterior of $W$, one can easily obtain an approximation of the posterior of $\theta$ by replacing the posterior of $\Theta(W)$ by the distribution of $\Theta(W_m^*)$, provided that $\Theta(\cdot)$ is continuous w.r.t the topology of the weak convergence. In fact, in this case the probability law of $\Theta(W_m^*)$ converges to the probability law of $\Theta(W)$ whenever $W_m^*$ converges weakly to $W$.

In practice, the approximation is based on Monte Carlo method. The basic idea, as described in Muliere and Secchi (1996) consists of the following steps.

*Step* 1: Generate a sample $S_m^*$ of size $m$ from the probability measure $\pi_b$, and compute the $\hat{b}_m^*(k)$'s as in (2).

*Step* 2: Generate two independent realizations of the variates $\lambda$ and $(b_m^*(k);\ k \geqslant 1)$ from their posterior distributions.

*Step* 3: Compute the corresponding p.g.f. $W^*(\cdot)$.

*Step* 4: Compute the value of $\theta^* = \Theta(W^*)$.

*Step* 5: Repeat Steps 1–4, a large number of times, say $r$ times, to obtain the values $\theta_1^*, \dots, \theta_r^*$.

The empirical distribution function of $\theta_1^*, \dots, \theta_r^*$, putting mass $r^{-1}$ on $\theta_j^*$'s provides the Bayesian bootstrap approximation for the posterior of $\theta$ given the data.

We consider here the approximation of the posterior distribution of the probability of a "long delay", i.e. the probability $Q(M)$ that the waiting time is greater than a given constant $M$. Such a quantity is of primary importance in assessing the performance of the system under examination. In fact, in terms of teletraffic applications, the waiting time can be viewed as the "buffer content". Since in concrete applications the buffer size is finite, $M$, say, the probability $Q(M)$ turns out to be the *overflow rate*, i.e. the probability that a cell cannot be neither transmitted nor stored in the buffer. Further discussion on the importance of $Q(M)$ is in Roberts et al. (1996).

As a slight generalization of Conti (1999), the probability $Q(M)$ is "well approximated" by

$$\theta = \frac{1-\rho}{\lambda B'(\bar{z})-1}\,\bar{z}^{-(M+1)} I_{(\rho<1)} + I_{(\rho\geqslant 1)}. \tag{3}$$

Note that $\theta = 1$ whenever $\rho \geqslant 1$, i.e. in case of an unstable queueing model.

The relative error of approximation tends to zero as $M$ increases. Since in real cases the value of $M$ (buffer size) is large (usually $M \geqslant 100$), inference based on the posterior of $Q(M)$ is virtually identical to inference based on the posterior of $\theta$.

Let $G_m^*(z)$ be equal to $1-\lambda-z+\lambda B_m^*(z)$, and let $\bar{z}_m^*$ be the largest real root of the equation $G_m^*(z)=0$. Finally, let $\theta^*$ be defined exactly as $\theta$, except that $\bar{z}$, $\rho$, $G'(\cdot)$, are replaced by $\bar{z}_m^*$, $\rho_m^*$, $G_m^{*'}(\cdot)$, respectively.

For the sake of simplicity, let $H(x)$ and $H_m^*(x)$ be equal to $P(\theta \leqslant x|\mathbf{S}_n,\mathbf{T}_n)$ and $P(\theta_m^* \leqslant x|\mathbf{S}_n,\mathbf{T}_n)$, respectively. Furthermore, if $\theta_{m1}^*,\ldots,\theta_{mr}^*$ are the $\theta_{mj}^*$'s values obtained from steps 1–5, denote by

$$\hat{H}_{m,r}^*(x) = \frac{1}{r}\sum_{j=1}^{r} I_{(-\infty,x]}(\theta_{mj}^*)$$

their empirical distribution function (e.d.f., for short).

The next proposition shows that the bootstrap distribution of $\theta_m^*$ gives an effective approximation of the posterior distribution of $\theta$. It is essentially in the spirit of Theorem 3.1 in Muliere and Secchi (1996). Part (ii) is the corresponding $\hat{H}_{m,r}^*$-version. It shows that the approximation still holds when the genuine bootstrap distribution of $\theta_m^*$ is replaced by its simulated version. To be precise, conditionally on $\mathbf{T}_n$, $\mathbf{S}_n$, the generated sequences $(\theta_{m1}^*,\ldots,\theta_{mr}^*)$ are assumed to live in an appropriate probability space $(\Omega^*,\mathscr{F}^*,P^*)$. In Proposition 2, $r$ and $m$ are both assumed to tend to infinity, and $r$ is implicitly taken dependent on $m$: $r=r_m$.

**Proposition 2.** *Under assumptions* P1–P3, *the following two statements hold true*:

(i) $\lim_{m\to\infty} H_m^*(x) = H(x)$ *for every continuity point* $x$ *of* $H(\cdot)$.
(ii) $\lim_{m\to\infty} \hat{H}_{m,r}^*(x) = H(x)$ *for every continuity point* $x$ *of* $H(\cdot)$, *a.s.*-$P^*$.

**Proof.** See Appendix A.

Let now $u(q)=\inf\{x: H(x)\geqslant q\}$ be the $q$th quantile of the posterior distribution of $\theta$, and let $\hat{u}_{m,r}^*(q)=\inf\{x:\hat{H}_{m,r}^*(x)\geqslant q\}$ be the corresponding $q$th quantile of $\hat{H}_{m,r}^*(\cdot)$. Proposition 1 suggests to take the interval

$$[\hat{u}_{m,r}^*(\alpha/2), \hat{u}_{m,r}^*(1-\alpha/2)]$$

as an approximated credibility region of probability $1-\alpha$ for the posterior of $\theta$. The sequel of the present section is devoted to prove that this procedure is correct, i.e. that the quantiles $\hat{u}_{m,r}^*(q)$ converge to $u(q)$ as $m$ and $r$ tend to infinity. To prove this result, we need the following preliminary lemma.

**Lemma 2.** *The distribution function $H(x)$ possesses the following properties*:

(i) $H(x)=0$ *for every* $x<0$, $H(1^-)=P(\rho<1|\mathbf{S}_n,\mathbf{T}_n)$, $H(x)=1$ *for every* $x \geqslant 1$.
(ii) $H(x)$ *is strictly increasing in the interval* $(0,1)$.

**Proof.** See Appendix A.

**Proposition 3.** *Under hypotheses* P1–P3, *the approximated quantiles* $\hat{u}^*_{m,r}(q)$ *converge*, *a.s.-*$P^*$, *to* $u(q)$, *as* $m$ *and* $r$ *both tend to infinity*, *for every* $0<q<1-P(\rho \geqslant 1|\mathbf{S}_n,\mathbf{T}_n)$.

**Proof.** See Appendix A.

## 4. Discussion and remarks on different approximations

In Sections 2 and 3, a crucial role is played by the approximation of the posterior law of $(b(k);\ k \geqslant 1)$. Motivated by the remarks made by both the referees, in the sequel we shortly discuss some alternatives to proper Bayesian bootstrap.

As remarked by a referee, the proper Bayesian bootstrap scheme considered in this paper is (slightly) more complex than Rubin's Bayesian bootstrap (1981) from a computational point of view, since it requires preliminary generations of samples $S^*_m$'s. Furthermore, the asymptotic arguments on which Rubin's Bayesian bootstrap rests (Lo, 1987) essentially work as long as the posterior has a Brownian bridge limiting distribution, for whatever prior (not only a Dirichlet process). In this case, due to the discreteness of the problem, the posterior of the cumulative d.f. of $b(k)$'s possesses, for large $n$, a limiting distribution which is not a Brownian bridge. It is actually a Gaussian process with discontinuous covariance function. However, it is easy to show that Rubin's Bayesian bootstrap still approximates the posterior law of $(b(k);\ k \geqslant 1)$. As already stressed, the disadvantage of Rubin's Bayesian bootstrap is that it only provides a first-order approximation based on Bernstein–Von Mises theorem. Our main concern is mainly in studying approximations that hold for *every* sample size.

A possible alternative to proper Bayesian bootstrap, as suggested by a referee, could consist in truncating the driving measure of the posterior law of $(b(k);\ k \geqslant 1)$ at some point $q$, and then to simulate from a Dirichlet process driven by such a truncated measure. Formally, take $m \geqslant 1$, and denote by $q_m$ the smallest positive integer such that

$$\sum_{k=1}^{q_m} \pi_{b,m}(k) \geqslant 1-m^{-1}, \tag{4}$$

and let $\pi_{b,m}(k)$ be equal to $\pi_b(k)$ if $k=1,\dots,q_m-1$, and $\pi_{b,m}(q_m)=\pi_b(q_m)+\sum_{k \geqslant q_m} \pi_b(k)$. Then, one could approximate the posterior law of $(b(k);\ k \geqslant 1)$ by a Dirichlet process with driving measure $(n+\beta)\pi_{b,m}(\cdot)$. Since this measure possesses a finite support, the problem reduces to simulate from a (finite) Dirichlet vector.

This clearly opens the road to further approximation schemes, that are shortly discussed in the sequel. We begin by a simple result, which is a minor variation of Proposition 1. Consider a (countable) family $\{(\tilde{b}_m(k);\ k \geqslant 1);\ m \geqslant 1\}$ of random probability measures concentrated over the set of all positive integer numbers. Let further $\tilde{B}_m(z)$ be equal to $\sum_{k\geqslant 1} \tilde{b}_m(k)z^k$, $\tilde{\rho}_m = \lambda\tilde{B}_m'(1)$, and

$$\tilde{W}_m(z) = \frac{(1-\tilde{\rho}_m)(1-z)}{1-\lambda-z+\lambda\tilde{B}_m(z)} I_{(0,\,1)}(\tilde{\rho}_m).$$

Suppose now that $\{(\tilde{b}_m(k);\ k \geqslant 1);\ m \geqslant 1\}$ converges weakly to $(b(k);\ k \geqslant 1)$ as $m$ goes to infinity, in the space $[0,1]^\infty$, equipped by the sup-norm. The corresponding Borel $\sigma$-field coincides with the usual product $\sigma$-field $\otimes\mathscr{B}([0,1])$, $\mathscr{B}([0,1])$ being the Borel class over $[0,1]$. Furthermore, in this case the weak convergence is equivalent to the convergence of the finite dimensional distributions. Since the r.v.'s $\tilde{b}_m(k)$, $b(k)$ are bounded, the Helly–Bray theorem (see, e.g., Breiman, 1992, p. 160) implies that $E[\tilde{b}_m(k)|\mathbf{S}_n,\mathbf{T}_n]$ tends to $E[b(k)|\mathbf{S}_n,\mathbf{T}_n]$ as $m$ tends to infinity, for every positive integer $k$. More definitely, the following proposition holds.

**Proposition 4.** *Assume that, conditionally on* $\mathbf{S}_n$, $\mathbf{T}_n$, $\{(\tilde{b}_m(k);\ k \geqslant 1);\ m \geqslant 1\}$ *converges weakly to* $(b(k);\ k \geqslant 1)$ *as m increases, and that the series* $\sum_{k\geqslant 1} E[\tilde{b}_m(k)|\mathbf{S}_n,\mathbf{T}_n]z^k$ *possesses (at least) the same radius of convergence of* $\sum_{k\geqslant 1}\beta(k)z^k$, *for every* $m \geqslant 1$. *Then*:

(i) *the sequence of stochastic processes* $(\tilde{B}_m(\cdot);\ m \geqslant 1)$ *converges weakly to* $B(\cdot)$ *in the space* $C[0,r]$ *endowed by the sup-norm, for every* $1 \leqslant r < \bar{\zeta}$;
(ii) *the sequence of stochastic processes* $(\tilde{W}_m(\cdot);\ m \geqslant 1)$ *converges weakly to* $W(\cdot)$ *in the space* $C[0,r]$ *endowed by the sup-norm, for every* $1 \leqslant r < \bar{\zeta}$.

**Proof.** Similar to those of Lemma 1 and Proposition 1. □

The essence of Proposition 4 is that if the conditional law of $(\tilde{b}_m(k);\ k \geqslant 1)$, as $m$ is "large", approximates the posterior law of $(b(k);\ k \geqslant 1)$, then also the (conditional) law of $\tilde{W}_m(\cdot)$ approximates the posterior distribution of $W(\cdot)$. Of course, the approximation is useful when the realizations of the random probability measure $(\tilde{b}_m(k);\ k \geqslant 1)$ possess a.s. a *finite* support, since it is usually not particularly difficult to simulate from them.

The Bayesian bootstrap is obtained as a special case of the scheme described above (apart from a change of symbols, $(b_m^*(k);\ k \geqslant 1)$ is a mixture of Dirichlet processes). The same holds for the procedure suggested by the referee: it is enough to assume that $(\tilde{b}_m(k);\ k \geqslant 1)$ is a Dirichlet process driven by the measure $(n+\beta)\pi_{b,m}(\cdot)$. The conditions of Proposition 4 are fulfilled, since $\pi_{b,m}(\cdot)$ converges in total variation to $\pi_b(\cdot)$, and this obviously implies that $(\tilde{b}_m(k);\ k \geqslant 1)$ converges weakly to $(b(k);\ k \geqslant 1)$.

It is of some interest to notice that also the approximation procedure developed by Muliere and Tardella (1998) fits the scheme discussed in this section. When particularised to our case, Muliere and Tardella procedure consists in approximating the posterior law of the Dirichlet process $(b(k);\ k \geqslant 1)$ by the law of a random measure defined as

$$\tilde{b}_m(k) = h_m I_{\{k\}}(Y_0) + \sum_{i=1}^{n_m} p_i I_{\{k\}}(Y_i), \quad k \geqslant 1,$$

where

$$p_1 = \xi_1 \quad p_i = \xi_i(1-\xi_{i-1})\cdots(1-\xi_1) \quad \forall i \geqslant 2,$$

$$n_m = \inf\left\{ j \geqslant 1 : \sum_{i=1}^{j} p_i \geqslant 1 - m^{-1} \right\}, \quad h_m = 1 - \sum_{i=1}^{n_m} p_i,$$

and $\xi_i$'s are i.i.d. r.v. with Beta distribution Be$(1, \beta+n)$, $Y_i$'s are i.i.d. r.v.'s, independent of $\xi_i$'s and with $P(Y_i = k) = \pi_b(k)$ for every $k \geqslant 1$, $i \geqslant 0$. Again, it is easy to check that the conditions of Proposition 4 are fulfilled.

The arguments developed so far naturally lead to a simple variation of the procedure outlined in the previous section. In fact, it is enough to replace the special $(b_m^*(k);\ k \geqslant 1)$ considered in Section 3 by the more general $(\tilde{b}_m(k);\ k \geqslant 1)$, and $W_m^*(\cdot)$ by $\tilde{W}_m(\cdot)$, respectively, and apply steps 2–4. In this way, $r$ values $\tilde{\theta}_{m1}, \ldots, \tilde{\theta}_{mr}$ are obtained. Again, the posterior d.f. $H(x)$ can be then approximated by its "empirical counterpart"

$$\tilde{H}_{m,r}(x) = \frac{1}{r} \sum_{j=1}^{r} I_{(-\infty,x]}(\tilde{\theta}_{mj})$$

and the quantiles $u(q) = \inf\{x : H(x) \geqslant q\}$ by the corresponding "empirical quantiles" $\tilde{u}_{m,r}(q) = \inf\{x : \tilde{H}_{m,r}(x) \geqslant q\}$. If $r = r_m$ goes to infinity as $m$ does, as a consequence of Proposition 4 and Lemma 2 we have that $\tilde{H}_{m,r}(x)$ converges in law to $H(x)$, and that $\tilde{u}_{m,r}(q)$ converges to $u(q)$. These two facts are summarised in Proposition 5.

**Proposition 5.** *Suppose that the assumptions* P1–P3 *and the conditions of Propositions* 4 *are fulfilled. Assume further that $r = r_m$ tends to infinity as $m$ increases, and that the generated sequences $(\tilde{\theta}_{m1}, \ldots, \tilde{\theta}_{mr})$ live in a probability space $(\Omega^*, \mathscr{F}^*, P^*)$. Then, the following two statements hold.*

(i) $\lim_{m\to\infty} \tilde{H}_{m,r}(x) = H(x)$ *for every continuity point of* $H(\cdot)$ *a.s.-*$P^*$.
(ii) $\lim_{m\to\infty} \tilde{u}_{m,r}(q) = u(q)$ *for every* $0 < q < 1 - P(\rho \geqslant 1 | \mathbf{S}_n, \mathbf{T}_n)$ a.s.-$P^*$.

**Proof.** Similar to those of Propositions 2, 3. □

Since several possible choices are possible, a short comparison is necessary. In particular, we examine three different approximations of (the posterior law of) $(b(k);\ k \geqslant 1)$, namely: (i) the proper Bayesian bootstrap by Muliere and Secchi; (ii) a Dirichlet process driven by the truncated measure $(n+\beta)\pi_{b,m}(\cdot)$ defined in (4) (denote by $(\tilde{b}_m^R(k);\ k \geqslant 1)$ the corresponding approximating random measure); the Muliere and

Tardella procedure (denote by $(\tilde{b}_m^{\mathrm{MT}}(k);\ k \geqslant 1)$ the corresponding approximating random measure).

The most important aspect of $(\tilde{b}_m^{\mathrm{MT}}(k);\ k \geqslant 1)$ is that its Prohorov distance from $(b(k);\ k \geqslant 1)$ is smaller than $m^{-1}$ (Muliere and Tardella, 1998). This allows one to fix $\varepsilon$, and to take $m = \lceil \varepsilon^{-1} \rceil$. However, this *does not* imply that the same holds for the non-linear functional $\theta = \theta(W)$ (3). Since $\theta$ is a continuous functional, we can only say that the approximation becomes better when $\varepsilon$ increases. However, it is not possible to quantify (with obvious symbols) how close the law of $\theta(\tilde{W}_m^{\mathrm{MT}})$ is close to the law of $\theta(W)$.

In the second place, it is easy to see that

$$E[b_m^*(k)|\mathbf{S}_n, \mathbf{T}_n] = E[\tilde{b}_m^{\mathrm{MT}}(k)|\mathbf{S}_n, \mathbf{T}_n] = \pi_b(k) \quad \forall k \geqslant 1,\ m \geqslant 1.$$

In a sense, the approximations provided by the proper Bayesian bootstrap and by the Muliere and Tardella scheme are "unbiased". The same does not hold for $(\tilde{b}_m^R(k);\ k \geqslant 1)$. $E[\tilde{b}_m^R(k)|\mathbf{S}_n, \mathbf{T}_n]$ is equal to zero for every $k > q_m$. Since we are mainly interested in the posterior law of the overflow rate, this could be a negative feature. In fact, the higher the service times, the higher the overflow rate. Since $(\tilde{b}_m^R(k);\ k \geqslant 1)$ systematically deletes the highest service times, the approximated posterior of $\theta$ tends to be concentrated on small values, and hence the performance of the system could be (at least potentially) overestimated.

An important issue is the computational burden of different approximation schemes. To generate a realization of $(b_m^*(k);\ k \geqslant 1)$ one has first to generate $m$ i.i.d. r.v.'s $S_1^*, \dots, S_m^*$ with common probability mass function $(\pi_b(k);\ k \geqslant 1)$. Then, conditionally on $S_j^*$'s, one has to generate $m$ (at most) independent Gamma r.v.'s. Efficient simulations algorithms for the Gamma distribution are described, for instance, in Devroye (1986).

As far as the Muliere and Tardella scheme is concerned, fix $0 < \varepsilon < 1$, and let $m_0$ be equal to $\lceil -\log \varepsilon \rceil$. To implement the scheme, one has to first generate $n_{m_0}$ independent Beta r.v.'s (again, for fast simulation algorithms, see Devroye, 1986), and then $n_{m_0} + 1$ i.i.d. discrete r.v.'s with probability mass function $(\pi_b(k);\ k \geqslant 1)$. It is possible to show (Muliere and Tardella, 1998) that $n_{m_0} - 1$ has Poisson distribution with mean $(n+\beta)m_0$. To get a good approximation $\varepsilon$ should be "small", i.e. $m_0$ should be "large". Since the average number of r.v. to be generated is $(n+\beta)m_0$, the computational burden could be heavy, at least when $n$ is large (in our application, $n$ is 5328).

For all the above-mentioned reasons, in the application considered in Section 5 we have used the proper Bayesian bootstrap.

## 5. Application to real data

In the present section, the techniques developed so far are applied to telecommunication data coming from the experimental European ATM network (Parker, 1995). The data are obtained by the superposition of three different kinds of applications: videoconference (20 sources simultaneously connected), teleteaching (30 sources simultaneously

Table 1
Values of $\hat{b}(k)$'s

| $k$ | $\hat{b}(k)$ | $k$ | $\hat{b}(k)$ | $k$ | $\hat{b}(k)$ | $k$ | $\hat{b}(k)$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.653528500 | 7 | 0.011448950 | 13 | 0.003753754 | 19 | 0.000000000 |
| 2 | 0.168918900 | 8 | 0.011261269 | 14 | 0.002627628 | 20 | 0.000187688 |
| 3 | 0.056118620 | 9 | 0.007695195 | 15 | 0.001126126 | 21 | 0.000000000 |
| 4 | 0.030968470 | 10 | 0.006381381 | 16 | 0.001689189 | 22 | 0.000000000 |
| 5 | 0.017079580 | 11 | 0.007319820 | 17 | 0.000938438 | 23 | 0.000375375 |
| 6 | 0.012950450 | 12 | 0.005630631 | 18 | 0.000000000 | | |

connected), and transport of routing information (20 sources simultaneously connected). All these applications use the Internet Protocol over ATM.

The transmission capacity requirement Peak Cell Rate (PCR) is 8320 cells/s for videoconferences, 13104 cells/s for teleteaching, and 9434 cells/s for transport of routing information. The resulting Sustainable Cell Rates (SCR) are 510, 226, and 43 cells/s, respectively. See ITU-T Recommendation I.371 (1996) for the meaning of PCR, SCR and other traffic indicators. The ATM link bandwidth or the service rate of the server is 80,000 cells/s (34 Mbit/s). The sample size is equal to 5328. The value of $\hat{\lambda}$ is 0.243; the values of $\hat{b}(k)$'s are given in Table 1.

Due to the large value of the sample size, the prior of $\lambda$ does not play a very important role. At any rate, in order to be as realistic as possible, we have based our choice of such a prior on a "training sample" of experimental measurements made on more or less similar traffic. As a prior of $\lambda$, a Beta distribution with parameters $\alpha_1$, $\alpha_2$ is adopted. Since the first two (sample) moments of the inter-arrival times in the training sample are equal to 3.19 and 15.64, respectively, the prior parameters $\alpha_1$ and $\alpha_2$ have been selected in such a way that the two relationships

$$E[\lambda^{-1}] = \frac{\alpha_1 + \alpha_2 - 1}{\alpha_1 - 1} = 3.19, \tag{5}$$

$$E[\lambda^{-2}] = \frac{(\alpha_1 + \alpha_2 - 1)(\alpha_1 + \alpha_2 - 2)}{(\alpha_1 - 1)(\alpha_1 - 2)} = 15.64 \tag{6}$$

hold true. From (5) and (6) the equalities $\alpha_1 = 3.28$, $\alpha_2 = 4.99$ follow.

The choice of the prior measure $\beta(\cdot)$ is essentially based on similar considerations. First of all, for mathematical reasons the measure $\beta(\cdot)$ is taken proportional to a Poisson measure: $\beta(k) = c\mathrm{e}^{-\eta}\eta^k/k!$, $k \geqslant 1$.

Since the mean service time in the training sample is 2.39, the value of the parameter $\eta$ is selected in order to ensure the following equality:

$$\sum_{k=1}^{\infty} k\,\frac{\beta(k)}{\beta} = \frac{\eta}{1 - \exp\{-\eta\}} = 2.39$$

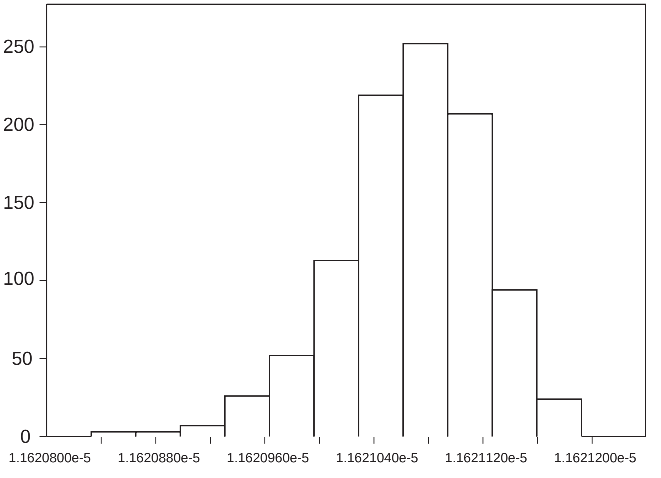


Fig. 1. Approximated posterior distribution of $\theta$; $M = 50$.

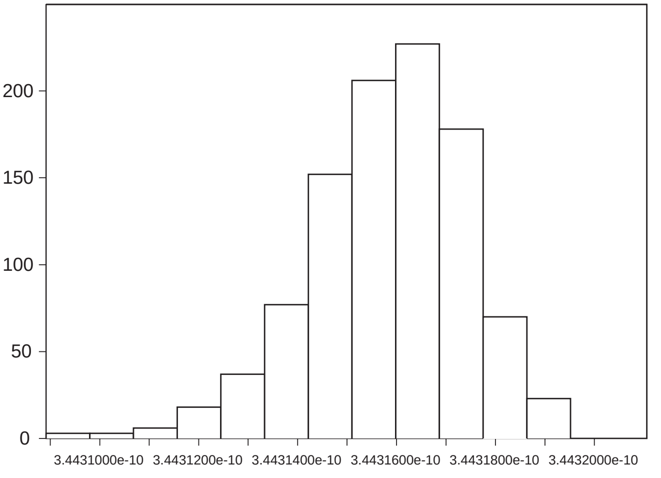


Fig. 2. Approximated posterior distribution of $\theta$; $M = 100$.

from which the relationship $\eta = 2.10$ follows. The constant $c$ has been taken equal to the size of the training sample: $c = 115$.

To approximate the posterior of $\theta$, the values $m = 25$, $r = 5000$ have been adopted. The histograms of the approximated posterior of $\theta$, corresponding to $M = 50$, $M = 100$, $M = 200$, are displayed in Figs. 1–3.

Approximated posterior regions of level 0.95 and 0.99 are reported in Table 2.

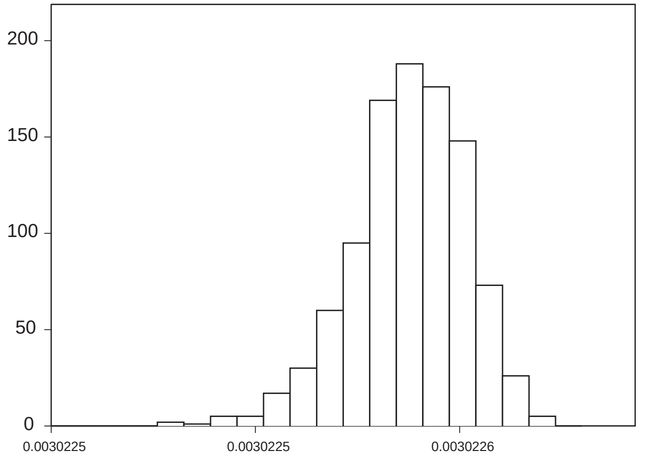


Fig. 3. Approximated posterior distribution of $\theta * 10^{16}$; $M = 200$.

Table 2
Approximated posterior regions for $\theta$

| $M$ | Level | |
|---|---|---|
| | 0.95 | 0.99 |
| 50 | $3.62 \times 10^{-5}$ $1.18 \times 10^{-4}$ | $6.22 \times 10^{-5}$ $1.04 \times 10^{-4}$ |
| 100 | $9.44 \times 10^{-9}$ $1.19 \times 10^{-8}$ | $1.08 \times 10^{-10}$ $5.58 \times 10^{-8}$ |
| 200 | $3.02 \times 10^{-18}$ $2.86 \times 10^{-17}$ | $1.93 \times 10^{-18}$ $5.12 \times 10^{-17}$ |

## Acknowledgements

Thanks are due to two anonymous referees for their valuable comments.

## Appendix A

**Proof of Lemma 1.** The proof uses arguments similar to those of Lemma 3 in Conti (1999), with some changes. We have to show that (i) the finite-dimensional distributions of $(B_m^*(\cdot);\ m \geqslant 1)$ converge to those of $(B(\cdot))$, and (ii) the sequence of random functions $(B_m^*(\cdot);\ m \geqslant 1)$ is tight.

As far as (i) is concerned, we can confine ourselves to one-dimensional distributions (the proof is similar for higher dimensional distributions). Take positive $\varepsilon$ and $\gamma$ such that $\tau = r + \gamma < \bar{\zeta}$, and take $K$ large enough such that $\sum_{k>K} (r/\tau)^k \leqslant 1$. Then it is easy to see that

$$
\begin{aligned}
P(B_m^*(z) \leqslant t|\mathbf{S}_n) &\leqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t+\varepsilon \,\middle|\, \mathbf{S}_n\right) + P\left(\sum_{k>K} b_m^*(k)z^k \geqslant \varepsilon \,\middle|\, \mathbf{S}_n\right) \\
&\leqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t+\varepsilon \,\middle|\, \mathbf{S}_n\right) \\
&\quad + P\left(\bigcup_{k>K} \left\{ b_m^*(k)z^k \geqslant \varepsilon(z/\tau)^k \right\} \,\middle|\, \mathbf{S}_n\right) \\
&\leqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t+\varepsilon \,\middle|\, \mathbf{S}_n\right) + \frac{1}{\varepsilon}\sum_{k>K} E[b_m^*(k)|\mathbf{S}_n]\tau^k \\
&\leqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t+\varepsilon \,\middle|\, \mathbf{S}_n\right) + \frac{1}{\varepsilon}\sum_{k>K} \frac{\beta(k)+n\hat{b}(k)}{n+\beta}\,\tau^k.
\end{aligned}
$$

Since the finite-dimensional distribution of $(b_m^*(k);\ k \geqslant 1)$ tend to those of $((b(k);$ $k \geqslant 1)$ as $m$ goes to infinity (Freedman, 1963), it is easy to see that

$$
\begin{aligned}
\lim_{m\to\infty} P\left(\sum_{k\geqslant 1} b_m^*(k)z^k \leqslant t \,\middle|\, \mathbf{S}_n\right) &\leqslant P\left(\sum_{k\leqslant K} b(k)z^k \leqslant t+\varepsilon \,\middle|\, \mathbf{S}_n\right) \\
&\quad + \frac{1}{\varepsilon}\sum_{k>K} \frac{\beta(k)+n\hat{b}(k)}{n+\beta}\tau^k.
\end{aligned}
$$

Letting $K$ tend to infinity and $\varepsilon$ tend to zero in such a way that

$$
\frac{1}{\varepsilon}\sum_{k>K} \{\beta(k)+n\hat{b}(k)\}\tau^k
$$

tends to zero, we conclude that

$$
\lim_{m\to\infty} P\left(\sum_{k\geqslant 1} b_m^*(k)z^k \leqslant t \,\middle|\, \mathbf{S}_n\right) \leqslant P\left(\sum_{k\geqslant 1} b(k)z^k \leqslant t^+ \,\middle|\, \mathbf{S}_n\right).
$$

To prove the reverse inequality, it is enough to observe that

$$
P\left(\sum_{k=1}^{\infty} b_m^*(k)z^k \leqslant t \,\middle|\, \mathbf{S}_n\right) \geqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t-\varepsilon \,\middle|\, \mathbf{S}_n\right)
$$

$$+P\left(\sum_{k>K} b_m^*(k)z^k \leqslant \varepsilon \,\middle|\, \mathbf{S}_n\right) - 1$$

$$\geqslant P\left(\sum_{k\leqslant K} b_m^*(k)z^k \leqslant t-\varepsilon \,\middle|\, \mathbf{S}_n\right)$$

$$-P\left(\sum_{k>K} b_m^*(k)z^k > \varepsilon \,\middle|\, \mathbf{S}_n\right)$$

and apply the same technique as before.

We have now to prove the tightness of the sequence of random functions $(B_m^*(\cdot);\ m\geqslant 1)$. Observing that $B(0)=0$ a.s. for every $m\geqslant 1$, and using Theorem 8.3 in Billingsley (1968), we only have to show that for each positive $\varepsilon$ and $\eta$, there exist a positive $\delta$ and an integer $m_0$ such that

$$\frac{1}{\delta} P\left(\sup_{z\leqslant s\leqslant z+\delta} |B_m^*(s)-B_m^*(z)| \geqslant \varepsilon \,\middle|\, \mathbf{S}_n\right) \leqslant \eta \quad \forall m\geqslant m_0. \tag{7}$$

Take again a positive $\gamma$ such that $\tau = r+\gamma \in [1,\bar{\zeta})$, and a positive $\alpha$ smaller than 1. We have first, by Markov inequality and taking into account that $b_m^*(k)$'s are smaller than 1,

$$\frac{1}{\delta} P\left(\sup_{z\leqslant s\leqslant z+\delta} |B_m^*(s)-B_m^*(z)| \geqslant \varepsilon \,\middle|\, \mathbf{S}_n\right)$$

$$\leqslant \frac{1}{\delta} P\left(\sum_{k=1}^{\infty} b_m^*(k) \sup_{z\leqslant s\leqslant z+\delta} |s^k - z^k| \geqslant \varepsilon \,\middle|\, \mathbf{S}_n\right)$$

$$\leqslant \frac{1}{\delta} P\left(\delta \sum_{k=1}^{\infty} b_m^*(k) k (z+\delta)^{k-1} \geqslant \varepsilon \left(1-\frac{z+\delta}{\tau}\right) \sum_{k=1}^{\infty} \left(\frac{z+\delta}{\tau}\right)^{k-1} \,\middle|\, \mathbf{S}_n\right)$$

$$\leqslant \frac{1}{\delta} \sum_{k=1}^{\infty} P\left(b_m^*(k)\tau^{k-1} \geqslant \frac{\varepsilon}{k\delta}(1-r/\tau) \,\middle|\, \mathbf{S}_n\right)$$

$$\leqslant \frac{\delta^{\alpha}\tau^{1+\alpha}}{(\varepsilon\gamma)^{1+\alpha}} \sum_{k=1}^{\infty} k^2 \tau^{(1+\alpha)(k-1)} E[b_m^*(k)^{1+\alpha} | \mathbf{S}_n].$$

Observing that $E[b_m^*(k)^{1+\alpha}|\mathbf{S}_n] = (\beta(k)+n\hat{b}(k))/(n+\beta)$, and that

$$C = \frac{\tau^{1+\alpha}}{(\varepsilon\gamma)^{1+\alpha}} \sum_{k=1}^{\infty} k^2 \tau^{2(k-1)} \left\{\frac{\beta(k)+n\hat{b}(k)}{n+\beta}\right\} < \infty,$$

relationship (7) follows by taking $\delta$ smaller than $(\eta/C)^{1/\alpha}$. □

**Proof of Proposition 1.** It is sufficient to observe that, because of the properties of $B_m^*(\cdot)$, the mapping

$$f(B_m^*(\cdot)) = \frac{(1-\rho_m^*)(1-z)}{1-\lambda-z+\lambda B_m^*(z)}$$

is continuous in $C[0,r]$ w.r.t. the sup-norm, and then to apply the continuous mapping theorem (Billingsley, 1968, p. 30). □

**Proof of Proposition 2.** To prove part (i), observe first that, as a consequence of Lemma 2 in Conti (1999), the functional $\Theta(W)=\bar{z}$ is continuous w.r.t. the sup-norm. Since the radii of convergence of $B'(\cdot)$ and $B(\cdot)$ coincide, it is immediate to verify that $\bar{z}$ is in the circle of convergence of $B(\cdot)$. Using again the same arguments as in Lemma 2 in Conti (1999), it is immediate to conclude that the functional $\theta$ is continuous w.r.t. the sup-norm. By applying Corollary 1 and the continuous mapping theorem (Billingsley, 1968, p. 30), statement (i) follows.

To prove part (ii), it is enough to show that $\hat{H}_{m,r}^*(x) - H_m^*(x)$ tends to zero as $m$ and $r$ go to infinity. Let

$$A_{m,r} = \sup_x |\hat{H}_{m,r}^*(x) - H_m^*(x)|.$$

By Dvoretzky–Kiefer–Wolfowitz (1956) inequality, we have first

$$P^*(A_{m,r} > \varepsilon|\mathbf{S}_n, \mathbf{T}_n) \leqslant C\exp\{-2r_m\varepsilon^2\},$$

where $C$ being an absolute, positive constant, and $\varepsilon$ a positive real number. Hence

$$\begin{aligned}\sum_{m=1}^{\infty} P^*(A_{m,r} > \varepsilon|\mathbf{S}_n, \mathbf{T}_n) &\leqslant C\sum_{m=1}^{\infty}\exp\{-2r_m\varepsilon^2\}\\ &\leqslant C\sum_{m=1}^{\infty}\exp\{-2m\varepsilon^2\}\\ &< C(1-\mathrm{e}^{-2\varepsilon^2})^{-1}\end{aligned}$$

for every $\varepsilon > 0$. By the first Borel–Cantelli lemma, the conclusion $A_{m,r} \to 0$ with $P^*$-probability 1 follows, which proves statement (ii). □

**Proof of Lemma 2.** Part (i) is obvious. It is enough to take into account that $\theta$ lies in the interval $[0,1]$ with probability 1, and that $\theta=1$ iff $\rho \geqslant 1$.

To prove part (ii), it is sufficient to show that

$$P(x < \theta < x+h|\mathbf{S}_n, \mathbf{T}_n) > 0 \tag{8}$$

for every positive $x$, $h$, with $x+h<1$.

The support (w.r.t. the topology of the weak convergence) of the posterior law of $P_b$ is the set of all probability measures whose support is contained in that of $(\beta(k);\ k \geqslant 1)$. Such a set of probability measures can be equipped by the Levy distance

$L(\cdot,\cdot)$ (Dudley, 1989) that makes the space separable and complete. Furthermore, in view of the assumptions (P1), the set $\mathscr{P}$ of all probability measures (over the positive integers) that possess finite p.g.f. $B(z)$ for every real $z$ does have posterior probability 1. From Theorem 2.1 in Parthasarathy (1967), it follows that for every $v_b \in \mathscr{P}$ and for every $r > 0$ all Levy neighbours $S(v_b, r) = \{v'_b \in \mathscr{P} : L(v_b, v'_b) < r\}$ do have positive posterior probability. Consider now $v_b \in \mathscr{P}$ and $\lambda_0$ such that the corresponding $\theta$ is in the interval $(x, x+h)$. Since the mapping $z \mapsto B(z)$ is continuous w.r.t. the topology of the weak convergence, it is easy to conclude that there exist a Levy neighbour $S(v_b, r)$ and an open interval $(\lambda_0 - u, \lambda_0 + u)$ such that $P(P_b \in S(v_b, r)|\mathbf{S}_n)$, $P(\lambda_0 - u < \lambda < \lambda_0 + u|\mathbf{T}_n)$ are both positive, and such that the $\theta$'s values corresponding to all $v'_b$'s in $S(v_b, r)$ and $\lambda$'s in $(\lambda_0 - u, \lambda_0 + u)$ are in $(x, x+h)$. This proves relationship (8). □

**Proof of Proposition 3.** The set of continuity points of $H(\cdot)$ is dense in $\mathbb{R}$. Hence, for every $\varepsilon' > 0$ there is $0 < \varepsilon < \varepsilon'$ such that $H(\cdot)$ is continuous in both $u(q) - \varepsilon$ and $u(q) + \varepsilon$. By Lemma 2(ii), the inequalities

$$H(u(q) - \varepsilon) < q < H(u(q) + \varepsilon)$$

hold. Then by Lemma 1(ii), the inequalities

$$\hat{H}^*_{m,r}(u(q) - \varepsilon) < q < \hat{H}^*_{m,r}(u(q) + \varepsilon) \tag{9}$$

hold for all but finitely many $m$'s, a.s.-$P^*$. Inequalities (9) are equivalent to (see, e.g., Serfling, 1980, Lemma 1.1.4)

$$u(q) - \varepsilon < \hat{u}^*_{m,r}(q) < u(q) + \varepsilon$$

for all but finitely many $m$'s. By repeating the same argument for every positive $\varepsilon'$, the proposition follows. □

## References


Armero, C., 1994. Bayesian inference in Markovian queues. Queueing Systems 15, 419–426.

Armero, C., Bayarri, M.J., 1994a. Bayesian prediction in M/M/1 queues. Queueing Systems 15, 401–417.

Armero, C., Bayarri, M.J., 1994b. Prior assessments for prediction in queues. The Statistician 43, 139–153.

Armero, C., Bayarri, M.J., 1996. Bayesian questions and answers in queues. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, Vol. 5. North-Holland, Amsterdam, pp. 3–23.

Armero, C., Bayarri, M.J., 1998. Dealing with uncertainties in queues and networks: a Bayesian approach. In: Ghosh, S. (Ed.), Mutivariate, Design and Sampling. Marcel Dekker, New York.

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

Breiman, L., 1992. Probability. SIAM, Philadelphia.

Brunel, H., Kim, B.G., 1993. Discrete-Time Models for Communication Systems Including ATM. Kluwer Academic Publishers, Boston.

Conti, P.L., 1999. Large sample Bayesian analysis for Geo/G/1 discrete time queueing models. Ann. Statist. 27 (6), 1785–1807.

Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer, New York.

Dudley, R.M., 1989. Trial Analysis and Probability. Wadsworth, Pacific Grove.

Dvoretzky, A., Kiefer, J.K., Wolfowitz, J., 1956. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. Ann. Math. Statist. 27, 642–669.

Freedman, D., 1963. On the asymptotic behavior of Bayes' estimates in the discrete case. Ann. Math. Statist. 34, 1386–1403.

Gravey, A., Louvion, J.R., Boyer, P., 1990. On the Geo/D/1 and Geo/D/n queues. Performance Eval. 11, 117–125.

Hunter, J.J., 1983. Mathematical Techniques of Applied Probability, Vol. 2. Academic Press, New York.

ITU-T Recommendation I.371, 1996. Traffic control and congestion in B-ISDN. International Telecommunication Union—Telecommunication.

Lo, A.Y., 1987. A large sample study of the Bayesian bootstrap. Ann. Statist. 15, 360–375.

Lo, A.Y., 1991. Bayesian bootstrap clones and a biometry function. Sankhyā-A 53, 320–333.

McGrath, M.F., Singpurwalla, N.D., 1987. A subjective Bayesian approach to the theory of queues II—inference and information. Queueing Systems 1, 335–353.

Mc Grath, M.F., Gross, D., Singpurwalla, N.D., 1987. A subjective Bayesian approach to the theory of queues: I—modelling. Queueing Systems 1, 317–333.

Muliere, P., Secchi, P., 1996. Bayesian nonparametric predictive inference and bootstrap techniques. Ann. Inst. Statist. Math. 48 (4), 663–673.

Muliere, P., Tardella, L., 1998. Approximating distributions of random functionals of Ferguson–Dirichlet priors. Canad. J. Statist. 26 (2), 283–297.

Parker, M., 1995. The European ATM pilot. In: Proceedings of XV International Switching Symposium ISS'95.

Parthasarathy, K.R., 1967. Probability Measures on Metric Spaces. Academic Press, New York.

Roberts, J., Mocci, U., Virtamo, J., 1996. Broadband Network Teletraffic. Springer, Berlin.

Rubin, D.B., 1981. The Bayesian bootstrap. Ann. Statist. 9, 130–134.

Ruggeri, F., Wiper, M., Rios-Insua, D., 1999. Bayesian Model Selection for M/G/1 queues. In: Racugno, W. (Ed.), Proceedings of the Workshop on Model Selection (Cagliari, 1997). Pitagora, Bologna.

Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Weng, C.-S., 1989. On a second-order asymptotic property of the Bayesian bootstrap mean. Ann. Statist. 17, 705–710.