

Probabilistic sampling of finite renewal processes

NELSON ANTUNES^{1,*} and VLADAS PIPIRAS^{2,**}

¹*University of Algarve/CEMAT, Faro, Portugal and INRIA Paris – Rocquencourt, Paris, France. E-mail: nantunes@ualg.pt*

²*Instituto Superior Técnico/CEMAT, Lisboa, Portugal and University of North Carolina, Chapel Hill, NC, USA. E-mail: pipiras@email.unc.edu*

Consider a finite renewal process in the sense that interrenewal times are positive i.i.d. variables and the total number of renewals is a random variable, independent of interrenewal times. A finite point process can be obtained by probabilistic sampling of the finite renewal process, where each renewal is sampled with a fixed probability and independently of other renewals. The problem addressed in this work concerns statistical inference of the original distributions of the total number of renewals and interrenewal times from a sample of i.i.d. finite point processes obtained by sampling finite renewal processes. This problem is motivated by traffic measurements in the Internet in order to characterize flows of packets (which can be seen as finite renewal processes) and where the use of packet sampling is becoming prevalent due to increasing link speeds and limited storage and processing capacities.

Keywords: IP flows; finite renewal process; interrenewal times; number of renewals; sampling; thinning; asymptotic normality; decompounding

1. Introduction

1.1. Motivation

The statistical and probabilistic problems considered in this work are motivated by questions arising from data traffic analysis in modern communication networks such as the Internet. Over these networks, information is sent in the form of packets, and packets are grouped into flows. (A flow corresponds to a group of packets sharing common characteristics. Ideally, a flow can be thought of as a set of packets that arises in the network through a remote terminal session or a web page download.) Each packet carries information about the flow it belongs to and also whether it is the first or the last packet in the flow. Examining each packet then allows all flows to be reconstructed. Knowing the structure of flows (flow level characteristics) helps network operators and networking researchers to understand and discover characteristics of data traffic.

A key difficulty with capturing each packet is that this rapidly leads to a huge amount of data to store and analyze. For instance, capturing all traffic for a few hours on a gigabit/sec. link at a medium load level yields several hundred gigabytes of data. A way to reduce the volume of

* Current address: Campus de Gambelas, 8900 Faro, Portugal.

** Current address: Av. Rovisco Pais, 1049-001 Lisboa, Portugal.

data is by sampling packets. One of the simplest approaches is probabilistic sampling, where each packet, independently of the others, is captured and analyzed with a fixed probability q . The basic problem, known as the *inversion problem* in the network traffic literature, is then to deduce the structure of the original flows from sampled packets.

This problem has recently attracted much attention in the networking community, with the focus almost exclusively on inference of the original distribution of flow sizes (total number of packets); see [11,17,24]. In this work, we take a closer look at statistical properties of the previously considered estimator of the original distribution of flow sizes and are also interested in inference of interarrival times between packets from sampled data. This distribution allows traffic burstiness to be characterized and leads to other major characteristics such as the duration of a flow.

1.2. Statement of problem and goals

From a mathematical standpoint, flows and their probabilistic sampling can be described as follows. As suggested by data traffic measurements (e.g., [18]), a flow can be modeled as a finite renewal process, where the total number of renewals W is random and the interrenewal times $D_i, i = 1, \dots, W - 1$, are positive i.i.d. random variables, independent of W . Suppose that the finite renewal process is sampled probabilistically with a fixed probability q to obtain a sampled finite point process. (As shown below, the resulting sampled point process is generally not a finite renewal process.) The sampled number of renewals (which could be zero) and sampled interrenewal times (which are available only when the number of sampled renewals is greater than 1) will be denoted by W_q and $D_{q,i}, i = 1, \dots, W_q - 1$, respectively. We illustrate this notation in Figure 1, where all circles (both empty and full) correspond to the finite renewal process, and full circles correspond to the sampled finite point process.

Given N i.i.d. copies of a sampled finite point process (some of which are empty when the number of sampled renewals is zero), one of our main goals is to infer the original distribution F_D of interrenewal times D_i . The focus is on nonparametric inference of F_D , statistical properties of the resulting estimator and its performance in simulations. The estimator of F_D involves a nonparametric estimator of the original distribution f_W of the total number of renewals W , which has previously been considered in the network traffic context [17]. Statistical properties of the estimator of f_W will also be studied here for the first time.

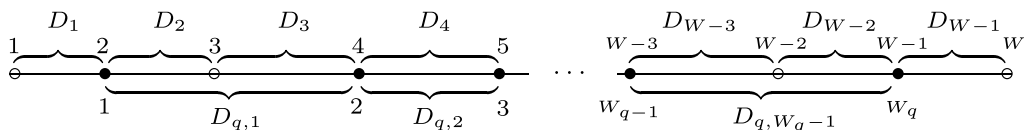


Figure 1. Finite renewal process and sampled finite point process.

1.3. Inference procedure

To be able to make inference about F_D , we first need to relate it to $F_{D_{q,i}|s}$, where the latter indicates the conditional distribution function of $D_{q,i}$ given $W_q = s$ (with $i < s$). As part of this work (see Theorem 3.1 in Section 3), we show that

$$F_{D_{q,i}|s} = \sum_{m=1}^{\infty} A_{s,m} F_D^{*m}, \tag{1.1}$$

where $*m$ denotes the m th convolution. In particular, the right-hand side of (1.1), and hence $F_{D_{q,i}|s}$, does not depend i . We also note for later reference that the definition of the sequence $A_s = \{A_{s,m}\}_{m \in \mathbb{N}}$ involves the distribution f_W . Somewhat independently of the main objectives above, we examine (1.1) for several underlying distributions f_W , such as geometric and heavy-tailed, that arise in network traffic studies (e.g., [6]). For example, in the geometric case, the distribution $F_{D_{q,i}|s}$ does not depend on s , although this seems to be an exception to the general rule. We also provide a result similar to (1.1) for a multidimensional vector $(D_{q,1}, \dots, D_{q,n})$. This allows (conditional) dependencies to be examined among sampled interrenewal times. For example, when the distribution f_W is geometric, the sampled interrenewal times turn out to be independent, in which case the sampled finite point process is a finite renewal process. We should also note that forms of conditioning other than on $W_q = s$ are possible, such as $W_q \geq s$; these will be discussed briefly.

Having the relation (1.1), we would naturally expect that it can be inverted, in the sense that

$$F_D = \sum_{m=1}^{\infty} a_{s,m} F_{D_{q,i}|s}^{*m}, \tag{1.2}$$

where $a_s = \{a_{s,m}\}_{m \in \mathbb{N}}$ is obtained by reversion of $A_s = \{A_{s,m}\}_{m \in \mathbb{N}}$, that is, their respective z -transformations (or formal power series) $G_{a_s}(z) = \sum_{m=1}^{\infty} a_{s,m} z^m$ and $G_{A_s}(z) = \sum_{m=1}^{\infty} A_{s,m} z^m$ satisfy $G_{a_s}(G_{A_s}(z)) = G_{A_s}(G_{a_s}(z)) = z$. The relation (1.2) suggests a natural estimator of F_D , defined as (see Section 4.2)

$$\widehat{F}_D = \sum_{m=1}^{\infty} \widehat{a}_{s,m} \widehat{F}_{D_{q,i}|s}^{*m}, \tag{1.3}$$

where $\widehat{F}_{D_{q,i}|s}$ is the empirical distribution function of $D_{q,i}$ given $W_q = s$, and \widehat{a}_s is the reversion of the sequence \widehat{A}_s , where the latter is defined as A_s by replacing f_W in its definition by the empirical distribution \widehat{f}_W . Note that the estimator \widehat{F}_D is defined for fixed s and i .

1.4. Contributions to the literature

Under technical assumptions, we will show that \widehat{F}_D is an asymptotically normal estimator of F_D , namely,

$$\sqrt{N}(\widehat{F}_D - F_D) \xrightarrow{d} X, \tag{1.4}$$

where X is a limiting Gaussian process and the convergence in distribution \xrightarrow{d} holds in a suitable space of functions (Theorem 4.3). The approach and some techniques behind this result are closely related to the works in the so-called decomposing framework by Buchmann and Grübel [5], Børgsted and Pitts [3] (see also [14,15]). These authors consider analogous estimators, but where the sequence A_s in (1.1) is known and hence so is its reversion sequence a_s . We thus deviate from these earlier works by assuming that a_s also needs to be estimated, which makes the analysis substantially more complex. More specifically, the limit X in (1.4) can be written as

$$X = - \sum_{n=1}^{\infty} (a_s^{(1)} * (\zeta \circ a_s))_n F_{D_{q,i}|s}^{*n} + \left(\sum_{n=1}^{\infty} a_{s,n} n F_{D_{q,i}|s}^{*(n-1)} \right) * Z, \tag{1.5}$$

where (ζ, Z) is a Gaussian process, \circ is a composition operation, $a_s^{(1)}$ is the “derivative” of a_s and $*$ is the convolution. The second term in (1.5) is the term found in the decomposing literature when A_s and a_s are supposed to be known. The first term in (1.5) is new and accounts for the variations in \widehat{a}_s here.

Since \widehat{a}_s involves the estimator \widehat{f}_W via \widehat{A}_s , we also need the asymptotic normality result for \widehat{f}_W . Although \widehat{f}_W already appears in the network traffic literature [17], its statistical properties have not been studied before to the best of our knowledge, and the asymptotic normality result is also derived here. We show that, under suitable assumptions,

$$\sqrt{N}(\widehat{f}_W - f_W) \xrightarrow{d} S(\xi), \tag{1.6}$$

where $S(\xi) = \{S(\xi)_w\}_{w \in \mathbb{N}}$ is a Gaussian process (Theorem 4.1 and Proposition 4.5). It is interesting to note that the result (1.6) is shown under technical assumptions which do not cover distributions with heavier tails (such as heavy-tailed distributions). It is proved, however, that the assumptions can be thought of as sharp (in the sense of Proposition 4.4 of Section 4.1). What the asymptotics of \widehat{f}_W are beyond these assumptions remains an interesting open question.

We would like to make several other related comments. First, it is well known from estimation of flow sizes in the network traffic context that the performance of the estimator \widehat{f}_W degrades rapidly as the sampling probability q decreases. For example, for $q = 0.1$, the inference becomes impractical for reasonable sample sizes N . We shall discuss this fact in light of the derived asymptotic normality result for \widehat{f}_W (Sections 4.1 and 5.1) by indicating two different regimes for the estimator performance in terms of the limiting asymptotic variance. The regimes are defined as:

- stable, if $\sup_{w \in \mathbb{N}} ES(\xi)_w^2 < \infty$;
 - explosive, if $\sup_{w \in \mathbb{N}} ES(\xi)_w^2 = \infty$,
- (1.7)

where $S(\xi)$ is the limiting Gaussian process in (1.6) and $ES(\xi)_w^2$ is expressed in terms of q , and \widehat{f}_W or f_{W_q} . The performance of the estimator \widehat{f}_W is satisfactory in the stable regime, but poor in the explosive regime. For small q , f_W typically belongs to the explosive regime and hence to the case of poor performance. Analogous difficulties for smaller q remain when using the

estimator \widehat{F}_D . Because of these practical considerations, in the network traffic context, alternative sampling schemes have been sought and considered, such as the sample-and-hold method [7,12], and the dual sampling technique [22]. We plan to study inference of interrenewal times in one of these frameworks and postpone any real data application to future work. This work will therefore be limited to a simulation study (Section 5).

In another direction, the results on characterizing the sampled number of renewals and sampled interrenewal times, such as the relation (1.1), contribute to a substantial literature on sampling of point processes. Sampling (also known as *thinning*) is discussed in several manuscripts such as [8,9,19]. To the best of our knowledge, sampling (thinning) of finite renewal processes has been largely unexplored. The results of our work (Section 3) show that sampling of finite renewal processes does not generally lead to finite renewal processes (in that, conditionally on the number of sampled renewals, sampled interrenewal times are generally dependent).

1.5. Structure of the paper

The rest of the paper is organized as follows. In Section 2 we collect the notation used throughout the work and include other preliminaries. Section 3 concerns properties of the finite point process obtained from sampling the finite renewal process. Inference of the original distributions of the total number of renewals and interrenewal times is studied in Section 4. A simulation study can be found in Section 5. For better readability, most of the proofs are deferred to Appendix A. The main body of the article contains the proofs of only those results which we consider key in individual sections. Finally, Appendix B contains a technical result used in Section 4.2, with bounds on remainder terms in Taylor expansions of compositions of formal power series.

2. Notation and other preliminaries

Here, we introduce notation and make a number of assumptions that will be used throughout the paper. As in Section 1, we consider a finite renewal process consisting of a finite but random number W (with $W \geq 1$) of renewals and positive i.i.d. interrenewal times $D_i, i = 1, \dots, W - 1$, independent of W . The terminology of finite renewal processes here follows that of [9], Example 5.3(b), page 125. We denote by D the variable with a common distribution of D_i . We also let V denote the total duration of the finite renewal process defined by $V = \sum_{i=1}^{W-1} D_i$ (if $W = 1$, then $V = 0$). Suppose now that each renewal, independently of the others, is sampled with a fixed probability $q \in (0, 1)$ to form a sampled finite point process. We let W_q denote the total number of sampled renewals, $D_{q,i}, i = 1, \dots, W_q - 1$, the sampled interrenewal times and $V_q = \sum_{i=1}^{W_q-1} D_{q,i}$ the total duration of the sampled finite point process.

The following notation will be used many times throughout the paper. For a discrete random variable X , its distribution or probability mass function (p.m.f.) will be denoted by $f_X(x) = P(X = x)$. For example, we shall write $f_W(w), f_{W_q}(w)$, etc. For a continuous random variable Y , its distribution function will be denoted by $F_Y(y) = P(Y \leq y)$ and its Laplace transform

(LT) will be denoted by $\tilde{F}_Y(v) = \int e^{-vy} F_Y(dy)$, $v \geq 0$ ($v \in \mathbb{R}_+$). We shall also use the conditional distribution functions $F_{Y|s}(y) = P(Y \leq y | W_q = s)$ and $F_{Y|s^+}(y) = P(Y \leq y | W_q \geq s)$. So, for example, we shall write $F_D, \tilde{F}_D, F_{D_{q,i}|s}$, etc. In several instances, we shall use analogous notation, but where Y is replaced by a multidimensional.

For a sequence $a = \{a_n\}_{n \in \mathbb{N}_0}$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and \mathbb{N} is the set of natural numbers, we denote its formal power series or its z -transformation by $G_a(z) = \sum_{n=0}^{\infty} a_n z^n$. Conversely, any such formal power series is associated with a sequence. When a stands for a p.m.f. f_X , we shall also write G_X in place of G_{f_X} . Since a sequence a will not necessarily be nonnegative, we shall also use the notation $a^+ = \{a_n^+\}_{n \in \mathbb{N}_0}$, defined as $a_n^+ = |a_n|$, and $\mathcal{G}_a(z) = G_{a^+}(z) = \sum_{n=0}^{\infty} |a_n| z^n$. In several instances, we shall use a multidimensional power series $G_a(z_1, \dots, z_m) = \sum_{n_1=0}^{\infty} \dots \sum_{n_m=0}^{\infty} a_{\mathbf{n}} z_1^{n_1} \dots z_m^{n_m}$ associated with $a = \{a_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{N}_0^m}$ and $\mathbf{n} = (n_1, \dots, n_m)$.

We shall also use the following common operations on sequences (or their formal power series) $a = \{a_n\}_{n \in \mathbb{N}_0}$ and $b = \{b_n\}_{n \in \mathbb{N}_0}$. The composition of a and b will be denoted $a \circ b$ and is defined by its formal power series as $G_{a \circ b}(z) = G_a(G_b(z))$. The k th derivative of $a = \{a_n\}_{n \in \mathbb{N}_0}$ will be denoted by $a^{(k)} = \{a_n^{(k)}\}_{n \in \mathbb{N}_0}$ and is defined by $G_{a^{(k)}}(z) = d^k G_a(z) / dz^k = \sum_{n=k}^{\infty} a_n n(n-1) \dots (n-k+1) z^{n-k}$. The reversion of a sequence $a = \{a_n\}_{n \in \mathbb{N}}$ will be defined as a sequence $b = \{b_n\}_{n \in \mathbb{N}}$ with its formal power series satisfying $G_b(G_a(z)) = G_a(G_b(z)) = z$. The reversion is defined for any sequence $a = \{a_n\}_{n \in \mathbb{N}}$ with $a_1 \neq 0$. As usual, the symbol $*$ will stand for convolution in either the discrete or continuous setting. With all these notions for sequences and their formal power series, we follow the nice monograph of [16].

3. From finite renewal process to sampled finite point process

In this section, we study several characteristics of the sampled finite point process in terms of the original finite renewal process. We first briefly consider the number of sampled renewals and then turn to sampled interrenewal times.

3.1. Number of sampled renewals

The relation between the probability mass functions of the number of sampled renewals and the number of original renewals is given by

$$\begin{aligned}
 f_{W_q}(s) &= \sum_{w=s}^{\infty} P(W_q = s | W = w) P(W = w) \\
 &= \sum_{w=s}^{\infty} \binom{w}{s} q^s (1-q)^{w-s} f_W(w), \quad s \geq 0.
 \end{aligned}
 \tag{3.1}$$

The function $G_{W_q}(z)$ is given by

$$\begin{aligned} G_{W_q}(z) &= \sum_{s=0}^{\infty} z^s \sum_{w=s}^{\infty} \binom{w}{s} q^s (1-q)^{w-s} f_W(w) \\ &= \sum_{w=1}^{\infty} f_W(w) (zq + (1-q))^w \\ &= G_W(zq + (1-q)), \quad |z| < 1. \end{aligned} \tag{3.2}$$

The same relations (3.1) and (3.2) also appear in [17] and will be used in Section 4.1 to obtain inversion results.

3.2. Sampled interrenewal times

Given that s renewals have been sampled, the next result characterizes the LT of the i th ($i < s$) sampled interrenewal time in terms of the LT of the original interrenewal times.

Theorem 3.1. *The LT of $D_{q,i}$ given $W_q = s$ can be expressed as*

$$\tilde{F}_{D_{q,i}|s}(v) = G_{A_s}(\tilde{F}_D(v)), \quad 1 \leq i < s, v \geq 0, \tag{3.3}$$

where $A_s = \{A_{s,m}\}_{m \in \mathbb{N}}$ and $A_{s,m}$ denotes the probability that the number of original renewals between the i th and $(i + 1)$ th sampled renewals is equal to $m - 1$ given $W_q = s$,

$$A_{s,m} = \frac{q^s}{f_{W_q}(s)} \sum_{w=s+m-1}^{\infty} f_W(w) \binom{w-m}{s-1} (1-q)^{w-s}. \tag{3.4}$$

Proof. For $1 \leq i < s$ and $t \geq 0$, we have

$$\begin{aligned} P(D_{q,i} \leq t, W_q = s) &= \sum_{w=s}^{\infty} f_W(w) P(D_{q,i} \leq t, W_q = s | W = w) \\ &= \sum_{w=s}^{\infty} f_W(w) P(W_q = s | W = w) P(D_{q,i} \leq t | W = w, W_q = s). \end{aligned} \tag{3.5}$$

Let M denote the number of original renewals not sampled between the i th and $(i + 1)$ th sampled renewals plus 1. If the number of original renewals is $W = w$ and the number of sampled renewals is $W_q = s$, then M can take the values $1, 2, \dots, w - s + 1$, and

$$P(D_{q,i} \leq t | W = w, W_q = s) = \sum_{m=1}^{w-s+1} F_D^{*m}(t) P(M = m | W = w, W_q = s). \tag{3.6}$$

By considering the location of the i th sampled renewal in the total number of renewals w along with the possible distinct locations of the sampled renewals before the i th and after the $(i + 1)$ th sampled renewals, we obtain that

$$\begin{aligned}
 P(M = m | W = w, W_q = s) &= \sum_{l=i}^{w-m-(s-i-1)} \binom{l-1}{i-1} \binom{w-m-l}{s-i-1} / \binom{w}{s} \\
 &= \binom{w-m}{s-1} / \binom{w}{s},
 \end{aligned}
 \tag{3.7}$$

where l is the location of the i th sampled renewal and the last equality follows from the identity

$$\sum_{l=a}^{x-b} \binom{l-1}{a-1} \binom{x-l}{b} = \binom{x}{a+b}, \quad a \geq 1, b \geq 0, a + b \leq x.
 \tag{3.8}$$

By using (3.5)–(3.7), we deduce that

$$\begin{aligned}
 F_{D_{q,i}|s}(t) &= \frac{P(D_{q,i} \leq t, W_q = s)}{f_{W_q}(s)} \\
 &= \frac{1}{f_{W_q}(s)} \sum_{w=s}^{\infty} f_W(w) q^s (1-q)^{w-s} \sum_{m=1}^{w-s+1} \binom{w-m}{s-1} F_D^{*m}(t) \\
 &= \sum_{m=1}^{\infty} A_{s,m} F_D^{*m}(t),
 \end{aligned}
 \tag{3.9}$$

where $A_{s,m}$ is given by (3.4). The relation (3.3) follows by taking the LT in (3.9). □

Theorem 3.1 implies that, when conditioning on $W_q = s$, the distribution of the i th ($i < s$) sampled interrenewal time depends, in general, only on s and not on i , and hence that the times between consecutive sampled renewals are identically distributed conditionally on W_q . In contrast to the original finite renewal process which assumes independence of W and D_i , the sampled quantities W_q and $D_{q,i}$ are dependent.

Other forms of conditioning on the number of sampled renewals are possible. For example, from (3.3), the LT of $D_{q,i}$ given $W_q \geq s$ can be expressed as

$$\begin{aligned}
 \tilde{F}_{D_{q,i}|s^+}(v) &= \sum_{s'=s}^{\infty} \tilde{F}_{D_{q,i}|s'}(v) P(W_q = s' | W_q \geq s) \\
 &= \frac{1}{P(W_q \geq s)} \sum_{s'=s}^{\infty} \sum_{m=1}^{\infty} A_{s',m} \tilde{F}_D(v)^m f_{W_q}(s') \\
 &= G_{A_{s^+}}(\tilde{F}_D(v)),
 \end{aligned}
 \tag{3.10}$$

where $A_{s^+} = \{A_{s^+,m}\}_{m \in \mathbb{N}}$ and $A_{s^+,m}$ is the probability that the number of original renewals between the i th and $(i + 1)$ th sampled renewals is equal to $m - 1$ given $W_q \geq s$,

$$\begin{aligned}
 A_{s^+,m} &= \frac{1}{P(W_q \geq s)} \sum_{s'=s}^{\infty} \sum_{w=s'+m-1}^{\infty} f_W(w) \binom{w-m}{s'-1} q^{s'} (1-q)^{w-s'} \\
 &= \frac{q(1-q)^{m-1}}{P(W_q \geq s)} \sum_{w=m+s-1}^{\infty} f_W(w) \left(1 - \sum_{s'=0}^{s-2} \binom{w-m}{s'} q^{s'} (1-q)^{w-m-s'}\right).
 \end{aligned}
 \tag{3.11}$$

In the rest of the paper, we shall focus only on the conditioning $W_q = s$, but similar results can be derived for other forms of conditioning such as $W_q \geq s$.

The next result gives the joint LT for a finite number of sampled interrenewal times. We shall use this general result below to investigate dependence between sampled interrenewal times (see Section 3.2.1); see Appendix A for the proof.

Theorem 3.2. *The joint LT of $\mathbf{D}_{q,n} = (D_{q,i_1}, \dots, D_{q,i_n})$ with $1 \leq i_1 < \dots < i_n$ given $W_q = s$, $s > i_n$, can be expressed as*

$$\tilde{F}_{\mathbf{D}_{q,n}|s}(\mathbf{v}) = G_{B_s}(\tilde{F}_D(v_1), \dots, \tilde{F}_D(v_n)), \quad \mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}_+^n,
 \tag{3.12}$$

where $B_s = \{B_{s,\mathbf{m}}\}_{\mathbf{m} \in \mathbb{N}^n}$ with $\mathbf{m} = (m_1, \dots, m_n)$ and $B_{s,\mathbf{m}}$ denotes the probability that the number of renewals between the i_j th and $(i_j + 1)$ th sampled renewals is equal to $m_j - 1$, $1 \leq j \leq n$, given $W_q = s$:

$$B_{s,\mathbf{m}} = \frac{q^s}{f_{W_q}(s)} \sum_{w=s+m-n}^{\infty} f_W(w) \binom{w-m}{s-n} (1-q)^{w-s}
 \tag{3.13}$$

with $m = m_1 + \dots + m_n$.

By Theorem 3.2, the joint distribution of a subset of sampled interrenewal times given $W_q = s$ depends only on s and not on the indices of the sampled times considered. Another relation of interest is between the duration of the sampled finite point process and the original interrenewal times; see Appendix A for the proof.

Proposition 3.1. *The LT of V_q given that $W_q \geq 2$ can be expressed as*

$$\tilde{F}_{V_q|2^+}(v) = G_C(\tilde{F}_D(v)), \quad v \geq 0,
 \tag{3.14}$$

where $C = \{C_m\}_{m \in \mathbb{N}}$ and C_m is the probability that the number of renewals between the first and last sampled renewals is equal to $m - 1$, given $W_q \geq 2$:

$$C_m = \frac{q^2}{P(W_q \geq 2)} \sum_{w=m+1}^{\infty} f_W(w)(w-m)(1-q)^{w-m-1}.
 \tag{3.15}$$

Since the LT of the duration of the original finite renewal process $V = \sum_{i=1}^{W-1} D_i$, given $W \geq 2$, can be expressed as

$$\tilde{F}_{V|W \geq 2}(v) = \frac{1}{P(W \geq 2)} \sum_{w=1}^{\infty} f_W(w+1) \tilde{F}_D(v)^w,$$

we expect that

$$\tilde{F}_D(v) = \sum_{n=1}^{\infty} D_n \tilde{F}_{V|W \geq 2}(v)^n, \tag{3.16}$$

where $D = \{D_n\}_{n \in \mathbb{N}}$ is the reversion of $\{f_W(w+1)/P(W \geq 2)\}_{w \in \mathbb{N}}$ (defined in Section 2). By plugging (3.16) into (3.14), we can obtain an expression for the duration V_q of the sampled finite point process in terms of the duration V of the finite renewal process.

Also, note that, equivalently, the relations (3.3), (3.12) and (3.14) can be expressed in terms of the distributions functions (see (3.9), and relations (A.4) and (A.7) in Appendix A, respectively) or via the characteristic functions.

3.2.1. Examples

In the following, we examine characteristics of the sampled finite point process for particular distributions of W .

Geometric distribution

Suppose that W is a geometric random variable with parameter c , that is, $f_W(w) = c^{w-1}(1-c)$, $w \geq 1$, $c \in (0, 1)$. Substituting this $f_W(w)$ into (3.4), we obtain, after straightforward calculations, that

$$A_{s,m} = (c(1-q))^{m-1} (1-c(1-q)), \quad m \geq 1.$$

Therefore, A_s is the p.m.f. of a geometric distribution with parameter $c(1-q)$ which does not depend on s . From (3.3) and using the generating function of the geometric distribution, we have

$$\tilde{F}_{D_{q,i}|s}(v) = \frac{(1-c(1-q))\tilde{F}_D(v)}{1-c(1-q)\tilde{F}_D(v)}.$$

Now, substituting $f_W(w)$ into (3.13), we conclude, after some algebra, that

$$B_{s,\mathbf{m}} = A_{s,m_1} \cdots A_{s,m_n}, \quad \mathbf{m} = (m_1, \dots, m_n) \in \mathbb{N}^n,$$

and hence, from (3.12),

$$\tilde{F}_{\mathbf{D}_{q,n}|s}(\mathbf{v}) = \prod_{i=1}^n \frac{(1-c(1-q))\tilde{F}_D(v_i)}{1-c(1-q)\tilde{F}_D(v_i)}.$$

The latter expression shows that when W is geometrically distributed, the sampled interrenewal times are independent and identically distributed (i.e., the sampled finite point process is a finite

renewal process). The LT of V_q in (3.14), where C is now the p.m.f. of a geometric distribution with parameter $1 - c$, simplifies to

$$\tilde{F}_{V_q|2^+}(v) = \frac{c\tilde{F}_D(v)}{1 - (1 - c)\tilde{F}_D(v)}.$$

Pareto distribution

Suppose now that W is a discrete Pareto distribution, that is, $f_W(w) = w^{-\alpha-1}/\zeta(\alpha + 1)$, $w \geq 1$, $\alpha > 0$ and $\zeta(z) = \sum_{w=1}^{\infty} w^{-z}$ is the Riemann zeta function. Unlike in the previous example, closed forms for LT's of the sampled interrenewal times are not available. Nevertheless, we can show that for a particular example, the sampled interrenewal times are not conditionally independent. Taking $s = 3$, $i_1 = 1$ and $i_2 = 2$, the random variables $D_{q,1}$ and $D_{q,2}$ are independent given $W_q = 3$ if and only if

$$G_{B_3}(\tilde{F}_D(v_1), \tilde{F}_D(v_2)) = G_{A_3}(\tilde{F}_D(v_1))G_{A_3}(\tilde{F}_D(v_2)) \tag{3.17}$$

for all $(y_1, y_2) \in \mathbb{R}_+^2$. Since $\tilde{F}_D(y)$ is continuous in y , it takes all the values in $(0, 1)$. Then, (3.17) implies that $G_{B_3}(w_1, w_2) = G_{A_3}(w_1)G_{A_3}(w_2)$ for all $w_1, w_2 \in (0, 1)$ and hence $B_{3,\mathbf{m}} = A_{3,m_1}A_{3,m_2}$ for all $\mathbf{m} = (m_1, m_2) \in \mathbb{N}^2$. The coefficients of the term $\tilde{F}_D(v_1)\tilde{F}_D(v_2)$ on the right- and left-hand sides of (3.17) are, respectively,

$$A_{3,1}^2 = \frac{9(\text{Li}_{\alpha-1}(1 - q) - 3\text{Li}_{\alpha}(1 - q) + 2\text{Li}_{\alpha+1}(1 - q))^2}{\text{Li}_{\alpha-2}(1 - q) - 3\text{Li}_{\alpha-1}(1 - q) + 2\text{Li}_{\alpha}(1 - q)}$$

and

$$B_{3,(1,1)} = \frac{6(1 - q + \text{Li}_{\alpha}(1 - q) - 2\text{Li}_{\alpha+1}(1 - q))}{\text{Li}_{\alpha-2}(1 - q) - 3\text{Li}_{\alpha-1}(1 - q) + 2\text{Li}_{\alpha}(1 - q)},$$

where $\text{Li}_n(a) = \sum_{w=1}^{\infty} a^w/w^n$, $|a| < 1$, is the polylogarithm function. Since $A_{3,1}^2 \neq B_{3,(1,1)}$, the random variables $D_{q,1}$ and $D_{q,2}$ are not conditionally independent. This is also why we use the term ‘‘sampled finite point process’’ throughout the paper instead of ‘‘sampled finite renewal process’’.

Heavy-tailed distributions

Consider the case of a heavy-tailed distribution for W , in the sense that

$$f_W(w) \sim c\alpha w^{-\alpha-1} \tag{3.18}$$

as $w \rightarrow \infty$, where $\alpha \in (1, 2)$ and $c > 0$. Such distributions are common models for long flow sizes in network traffic studies (e.g., [6]). As in the example with Pareto distribution above, closed forms for LT's of the sampled interrenewal times are not available under (3.18). We can nevertheless provide a number of interesting qualitative results concerning the case (3.18).

The relation (3.18) implies that

$$P(W > w) \sim cw^{-\alpha} \tag{3.19}$$

as $w \rightarrow \infty$. As indicated in [17], page 70, by the results of [2], page 333, this implies that

$$P(W_q > w) \sim q^\alpha c w^{-\alpha}, \tag{3.20}$$

that is, W_q is also heavy-tailed with the same parameter α . (Note that (3.20) does not, in general, imply that $f_{W_q}(w) \sim q^\alpha c \alpha w^{-\alpha-1}$.) Also, note that in the case (3.19), the total duration V is expected to be heavy-tailed, even for light-tailed interrenewal times D_i . Indeed, by Robert and Segers [21], Theorem 3.2, if (3.19) holds and

$$P(D_1 > t) = o(t^{-\alpha}), \tag{3.21}$$

then

$$P(V > t) \sim P\left(W > \frac{t}{ED_1}\right) \sim c(ED_1)^\alpha t^{-\alpha}, \tag{3.22}$$

that is, the tail of total duration is dominated by that of W .

On the sampling side, the total duration is also characterized by heavy tails. Indeed, by Proposition 3.1 above and [21], Theorem 3.2,

$$P(V_q > t) \sim P\left(C > \frac{t}{ED_1}\right) \sim c_0(ED_1)^\alpha t^{-\alpha}, \tag{3.23}$$

where a random variable C is such that $P(C = m) = C_m$, where C_m is defined in (3.15), as long as (3.21) holds and

$$C_m \sim c_0 \alpha m^{-\alpha-1}. \tag{3.24}$$

To show (3.24), observe that for large m , using (3.18),

$$\begin{aligned} C_m &\sim \frac{q^2 c \alpha}{P(W_q \geq 2)} \sum_{w=m+1}^{\infty} w^{-\alpha-1} (w-m)(1-q)^{w-m-1} \\ &= m^{-\alpha-1} \frac{q^2 c \alpha}{P(W_q \geq 2)} \sum_{k=1}^{\infty} \left(1 + \frac{k}{m}\right)^{-\alpha-1} k(1-q)^{k-1} \\ &\sim m^{-\alpha-1} \frac{q^2 c \alpha}{P(W_q \geq 2)} \sum_{k=1}^{\infty} k(1-q)^{k-1} \\ &= m^{-\alpha-1} \frac{q^2 c \alpha}{P(W_q \geq 2)} \left(-\frac{d}{dq} \sum_{k=0}^{\infty} (1-q)^k\right) \\ &= m^{-\alpha-1} \frac{q^2 c \alpha}{P(W_q \geq 2)} \left(-\frac{d}{dq} \frac{1}{q}\right) = m^{-\alpha-1} \frac{c \alpha}{P(W_q \geq 2)} =: m^{-\alpha-1} c_0 \alpha. \end{aligned} \tag{3.25}$$

Note also that the argument (3.25) above does not apply to the sampled interrenewal times. For example, for the first sampled interrenewal time $D_{q,1}$, the analogous argument would show

that, for coefficients in (3.4),

$$A_{s,m} \sim c(1 - q)^m m^{-\alpha-1} \tag{3.26}$$

as $m \rightarrow \infty$, which is not heavy-tailed, in contrast to (3.24).

4. From sampled finite point process to finite renewal process

In this section, we study inference of the original distributions of the number of renewals and interrenewal times.

4.1. Distribution of number of renewals

We are interested here in estimating the p.m.f. $f_W(w)$ from i.i.d. observations of W_q . We revisit a nonparametric estimator of $f_W(w)$ introduced in [17] and clarify several issues surrounding its use and properties. A number of open questions will also be raised.

Estimation of $f_W(w)$ is based on a theoretical inversion of the relation (3.1) which will be discussed first. The relation (3.2) can be written as $G_W(z) = G_{W_q}(q^{-1}z + (1 - q^{-1}))$, which has the same form as the original relation (3.2) when W and W_q are exchanged, and q is replaced by q^{-1} . In view of (3.1), we would then expect that

$$\begin{aligned} f_W(w) &= \sum_{s=w}^{\infty} \binom{s}{w} (q^{-1})^w (1 - q^{-1})^{s-w} f_{W_q}(s) \\ &= \sum_{s=w}^{\infty} \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1 - q)^{s-w} f_{W_q}(s), \quad w \geq 1. \end{aligned} \tag{4.1}$$

Hohn and Veitch [17] claim that (4.1) holds when $q \in (0.5, 1)$ (with this choice, note that $q^{-s}(1 - q)^s \in (0, 1)$ in (4.1)). As the following elementary result shows, this is not a necessary condition. The relation (4.1) also holds for $q \in (0, 0.5]$ as long as $f_{W_q}(w)$ or $f_W(w)$ decays to zero fast enough; see Appendix A for the proof.

Proposition 4.1. *If*

$$\sum_{s=n}^{\infty} \binom{s}{n} \frac{(1 - q)^{s-n}}{q^s} f_{W_q}(s) = \sum_{w=n}^{\infty} \binom{w}{n} 2^{w-n} (1 - q)^{w-n} f_W(w) < \infty, \quad n \geq 1, \tag{4.2}$$

then the relation (4.1) holds.

For example, for geometric W satisfying $f_W(w) = c^{w-1}(1 - c)$, $w \geq 1$, $c \in (0, 1)$, the condition (4.2) holds when $2 - (2 - \varepsilon)q < c^{-1}$, where $\varepsilon > 0$ is arbitrarily small and where we have used the fact that $\binom{w}{n}$ in (4.2) can be bounded by w^n up to a multiplicative constant. Note also that (4.2) always holds for $q \in (0.5, 1)$. When $q \in (0, 0.5]$, on the other hand, a number of distributions $f_W(w)$ of interest, such as heavy-tailed distributions (Section 3.2.1), do not satisfy (4.2).

In these cases, and generally for $q \in (0, 0.5]$, the relation (4.1) needs to be modified by a procedure used and referred to as an analytic continuation in [17], page 71. The procedure is defined below. We should note that this procedure is viewed below as a convenient algebraic trick that makes series converge (see, e.g., the proof of Proposition 4.3) rather than as a suitable analytic continuation, which is a complementary viewpoint followed in [17].

Let $z_0 = 1 - q$ and pick an arbitrary sequence $z_k, k = 1, \dots, l$, such that $1 > z_0 > z_1 > \dots > z_{l-1} > z_l = 0$ and $z_k \in C_{k-1} = \{z : |z - z_{k-1}| < 1 - z_{k-1}\}, k = 1, \dots, l$. For a sequence $x = \{x_n\}_{n \in \mathbb{N}}$, define formally and recursively sequences $T^{(k)}(x) = \{T^{(k)}(x)_n\}_{n \in \mathbb{N}}, k = 1, \dots, l$, as

$$T^{(k)}(x)_n = \sum_{i=n}^{\infty} \binom{i}{n} T^{(k-1)}(x)_i (z_k - z_{k-1})^{i-n}, \quad n \geq 1, \tag{4.3}$$

where $T^{(0)}(x)_i = x_i/q^i$. It is also convenient to define the mapping underlying (4.1), that is, $S(x) = \{S(x)_n\}_{n \in \mathbb{N}}$, where

$$S(x)_n = \sum_{i=n}^{\infty} \binom{i}{n} \frac{(-1)^{i-n}}{q^i} (1 - q)^{i-n} x_i, \quad n \geq 1. \tag{4.4}$$

The following elementary result relates $S(x)$ and $T^{(l)}(x)$ when x satisfies a natural condition.

Proposition 4.2. *If a sequence $x = \{x_n\}_{n \in \mathbb{N}}$ satisfies*

$$\sum_{i=n}^{\infty} \binom{i}{n} \frac{(1 - q)^{i-n}}{q^i} |x_i| < \infty, \quad n \geq 1, \tag{4.5}$$

then

$$T^{(l)}(x) = S(x), \tag{4.6}$$

where $T^{(l)}(x)$ and $S(x)$ are defined in (4.3) and (4.4), respectively.

The next result formalizes the fact that f_W can be obtained as $T^{(l)}(f_{W_q})$. Note that this is true without any assumptions on q and f_W .

Proposition 4.3. *For any $q \in (0, 1)$ and p.m.f. f_W , we have*

$$f_W = T^{(l)}(f_{W_q}), \tag{4.7}$$

where $T^{(l)}$ is defined in (4.3).

With theoretical inversion formulas (4.1) and (4.7), we can now turn to estimation. Let $W_{q,k}, k = 1, \dots, N$, be i.i.d. copies of the variable W_q and

$$\widehat{f}_{W_q}(s) = \frac{1}{N} \sum_{k=1}^N 1_{\{W_{q,k}=s\}}, \quad s \geq 0, \tag{4.8}$$

be the empirical p.m.f. of f_{W_q} , where 1_A denotes an indicator function of an event A . Note that we assume, in particular, that the event $W_{q,k} = 0$ can be observed in practice (or, equivalently, N and $W_{q,k} \geq 1$ can be observed in practice). In the network traffic context, we naturally observe only $W_{q,k} \geq 1$. The total number of flows N is deduced from the additional information in the sampled packet headers (e.g., [10]). (To be more precise, N is actually estimated, but we suppose it to be known for the sake of simplicity.)

In view of (4.7), it is natural to introduce the following nonparametric estimator of f_W :

$$\widehat{f}_W(w) = T^{(l)}(\widehat{f}_{W_q})_w \tag{4.9}$$

$$= S(\widehat{f}_{W_q})_w = \sum_{s=w}^{\infty} \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1-q)^{s-w} \widehat{f}_{W_q}(s), \quad w \geq 1, \tag{4.10}$$

where $T^{(l)}(x)$ and $S(x)$ are defined in (4.3) and (4.4), respectively. The first equality in (4.10) follows from Proposition 4.2 since only a finite number of the $\widehat{f}_{W_q}(s)$'s are nonzero.

We will next show that the estimator (4.9) is asymptotically normal under suitable assumptions. The suitable assumptions are quite strong, but we do not expect that they can be weakened much, as explained in Proposition 4.4 below and a discussion surrounding it. For $w \geq 1$, we also let

$$R_{q,w} = \sum_{s=w}^{\infty} \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} f_{W_q}(s) \tag{4.11}$$

$$= \sum_{k=w}^{\infty} f_W(k) (1-q)^{k-2w} \binom{k}{w} \sum_{s=w}^k \binom{s}{w} \binom{k-w}{s-w} (q^{-1}-1)^s, \tag{4.12}$$

where the second equality follows from (3.1).

Theorem 4.1. *Suppose that*

$$R_{q,w} < \infty, \quad w \geq 1, \tag{4.13}$$

where $R_{q,w}$ is defined in (4.11). Then, as $N \rightarrow \infty$,

$$\{\sqrt{N}(\widehat{f}_W(w) - f_W(w))\}_{w \in \mathbb{N}} \rightarrow \{S(\xi)_w\}_{w \in \mathbb{N}}, \tag{4.14}$$

where the convergence is in the sense of finite-dimensional distributions, $S(x)$ is defined in (4.4) and $\xi = \{\xi_s\}_{s \in \mathbb{N}}$ is a zero-mean Gaussian process with the covariance structure

$$E\xi_{s_1}\xi_{s_2} = f_{W_q}(s_1)\delta_{s_1,s_2} - f_{W_q}(s_1)f_{W_q}(s_2) \tag{4.15}$$

(and, as usual, $\delta_{s_1,s_2} = 1$ if $s_1 = s_2$, and $= 0$ if $s_1 \neq s_2$, this being the Kronecker symbol). In particular, the limiting variables $S(\xi)_w$ are zero-mean, Gaussian and have the variance

$$ES(\xi)_w^2 = R_{q,w} - f_W(w)^2. \tag{4.16}$$

Proof. We only consider the convergence (4.14) at fixed $w \geq 1$. Note that, using Proposition 4.2, $\sqrt{N}(\widehat{f}_W(w) - f_W(w)) = S(\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}))_w$. For fixed $j \geq w$ and $x = \{x_n\}_{n \in \mathbb{N}}$, define

$$S(x)_{j,w} = \sum_{i=w}^j \binom{i}{w} \frac{(-1)^{i-w}}{q^i} (1-q)^{i-w} x_i.$$

Using [1], Theorem 3.2, page 28, it is enough to show that:

- (i) $S(\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}))_{j,w} \xrightarrow{d} S(\xi)_{j,w}$ as $N \rightarrow \infty$;
- (ii) $S(\xi)_{j,w} \xrightarrow{d} S(\xi)_w$ as $j \rightarrow \infty$;
- (iii) for any $\delta > 0$,

$$\limsup_{j \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|S(\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}))_{j,w} - S(\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}))_w| > \delta) = 0.$$

The convergence in (i) is elementary since $\{\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))\}_{s \in \mathbb{N}}$ converges to $\xi = \{\xi_s\}_{s \in \mathbb{N}}$ in the sense of finite-dimensional distributions and since, for fixed j , $S(x)_{j,w}$ involves only a finite number of elements of $x = \{x_n\}_{n \in \mathbb{N}}$.

The convergence in (ii) can be proven in a stronger sense, that of almost sure convergence. For this, observe that in the sense of finite-dimensional distributions,

$$\begin{aligned} \{\xi_s\}_{s \in \mathbb{N}} &\stackrel{d}{=} \{B(F_{W_q}(s)) - B(F_{W_q}(s-1))\}_{s \in \mathbb{N}} + \{f_{W_q}(s)B(1)\}_{s \in \mathbb{N}} \\ &=: \{\xi_s^{(1)}\}_{s \in \mathbb{N}} + \{\xi_s^{(2)}\}_{s \in \mathbb{N}}, \end{aligned} \tag{4.17}$$

where $B = \{B(t)\}_{t \in [0,1]}$ is a standard Brownian motion. It is then enough to show almost sure convergence of the corresponding terms for $\xi^{(1)}$ and $\xi^{(2)}$. Doing this for $\xi^{(2)}$ is elementary and thus we do so only for $\xi^{(1)}$. Note that

$$\{\xi_s^{(1)}\}_{s \in \mathbb{N}} \stackrel{d}{=} \{f_{W_q}(s)^{1/2} \eta_s\}_{s \in \mathbb{N}}, \tag{4.18}$$

where η_s are i.i.d. $N(0, 1)$ random variables. By the three series theorem, $S(\xi^{(1)})_{j,w} \rightarrow S(\xi^{(1)})_w$ a.s. as long as the condition (4.13) holds.

For the convergence in (iii), note that the probability in (iii) can be expressed and bounded as

$$\begin{aligned} &P\left(\left|\sum_{s=j+1}^{\infty} \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1-q)^{s-w} \sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))\right| > \delta\right) \\ &\leq \delta^{-2} \sum_{s=j+1}^{\infty} \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} E(\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s)))^2 \\ &\quad + 2\delta^{-2} \sum_{j+1 \leq s_1 < s_2} \binom{s_1}{w} \binom{s_2}{w} \frac{(1-q)^{s_1+s_2-2w}}{q^{s_1+s_2}} \end{aligned}$$

$$\begin{aligned} & \times |EN(\widehat{f}_{W_q}(s_1) - f_{W_q}(s_1))(\widehat{f}_{W_q}(s_2) - f_{W_q}(s_2))| \\ &= \delta^{-2} \sum_{s=j+1}^{\infty} \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} (f_{W_q}(s) - f_{W_q}(s)^2) \\ & \quad + 2\delta^{-2} \sum_{j+1 \leq s_1 < s_2} \binom{s_1}{w} \binom{s_2}{w} \frac{(1-q)^{s_1+s_2-2w}}{q^{s_1+s_2}} f_{W_q}(s_1) f_{W_q}(s_2) \leq \delta^{-2}(Q_j + 2Q_j^2), \end{aligned}$$

where $Q_j = \sum_{s=j+1}^{\infty} \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} f_{W_q}(s)$. It remains to observe that $Q_j \rightarrow 0$ as $j \rightarrow \infty$, by the assumption (4.13). The proof of (4.16) is now elementary. \square

Note that the condition (4.13) is quite strong and, for example, does not allow for heavy-tailed distributions when $q \in (0, 0.5]$. When this condition does not hold, since $\sqrt{N}(\widehat{f}_W(w) - f_W(w)) = T^{(l)}(\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}))_w$, we may think that the limit of the left-hand side of (4.14) should be

$$T^{(l)}(\xi) = \{T^{(l)}(\xi)_w\}_{w \in \mathbb{N}}, \tag{4.19}$$

where the process ξ is as in Theorem 4.1. This, however, is not expected to hold. In fact, without the condition (4.13), the expected limit (4.19) is not even well defined, as the result below shows. Consider the case of $l = 2$, $z_2 = 0$ and $T^{(2)}(\xi)$ for simplicity (the argument can also be extended to general l). With the representation (4.17), the term $T^{(2)}(\xi^{(2)})$ is well defined by arguing as in the proof of Proposition 4.3. We will show that without the condition (4.13), $T^{(2)}(\xi^{(1)})$ is not well defined, in the following natural sense. Note that

$$T^{(1)}(\xi^{(1)})_n = \sum_{s=n}^{\infty} \binom{s}{n} \frac{(z_1 - (1-q))^{s-n}}{q^s} f_{W_q}(s)^{1/2} \eta_s$$

is well defined, where we take $\xi_s^{(1)} = f_{W_q}(s)^{1/2} \eta_s$ with η_s as in (4.18) and let

$$T^{(2)}(\xi^{(1)})_{j,w} = \sum_{k=w}^j \binom{k}{w} (-z_1)^{k-w} T^{(1)}(\xi^{(1)})_k.$$

The proof of the following result can be found in Appendix A.

Proposition 4.4. *The condition (4.13) is necessary for $T^{(2)}(\xi^{(1)})_{j,w}$ to have a limit in distribution as $j \rightarrow \infty$.*

We shall conclude this section by considering the convergence (4.14) in a suitable space of sequences. We let $l_{\infty,a} = \{x = \{x_n\}_{n \in \mathbb{N}_0} : \|x\|_{\infty,a} := \sup_{n \geq 0} a^n |x_n| < \infty\}$. See Appendix A for the proof.

Proposition 4.5. *If, for $b' > 0$,*

$$\sum_{s=0}^{\infty} f_{W_q}(s)^{1/2} \left(\frac{b' + 1 - q}{q} \right)^s < \infty, \tag{4.20}$$

then the convergence (4.14) also holds in the space $l_{\infty,b}$ for any $0 < b < b'$.

Note that the condition (4.20) is stronger than (4.13).

Theorem 4.1 and Proposition 4.5 suggest several possibilities for confidence intervals (CI's) of f_W . For example, following Theorem 4.1, we can define the $100\alpha\%$ CI of $f_W(w)$ for fixed w as

$$\left(\widehat{f}_W(w) - z_\alpha N^{-1/2} (\widehat{ES(\xi)_w^2})^{1/2}, \widehat{f}_W(w) + z_\alpha N^{-1/2} (\widehat{ES(\xi)_w^2})^{1/2} \right), \tag{4.21}$$

where $\widehat{ES(\xi)_w^2}$ is defined as in (4.16) by replacing f_{W_q} in $R_{q,w}$ and $f_W(w)$ by \widehat{f}_{W_q} , and z_α is the 100α th percentile of $|\mathcal{N}(0, 1)|$. Another possibility is to use Proposition 4.5 and set the $100\alpha\%$ confidence ‘‘interval’’ (set) across all w simultaneously as

$$\{f : \|\widehat{f}_W - f\|_{\infty,b} \leq N^{-1/2} \widehat{q}(\alpha)\}. \tag{4.22}$$

Following [13] and [3], the quantity $\widehat{q}(\alpha)$ should be taken as the 100α th percentile of

$$\widehat{R}_N(z) = \frac{1}{N^N} \sum_{(i_1, \dots, i_N) \in \{1, \dots, N\}^N} 1_{[0,z]} \left(\sqrt{N} \left\| S \left(N^{-1} \sum_{k=1}^N 1_{\{W_{q,i_k} = \cdot\}} \right) - S(\widehat{f}_{W_q}) \right\|_{\infty,b} \right), \tag{4.23}$$

which can be viewed as a bootstrapped version of the probability

$$R_N(z) = P(\sqrt{N} \|S(\widehat{f}_{W_q}) - S(f_{W_q})\|_{\infty,b} \leq z) = P(\sqrt{N} \|\widehat{f}_W - f_W\|_{\infty,b} \leq z). \tag{4.24}$$

The choice of $\widehat{q}(\alpha)$ is justified if we can show that \widehat{R}_N and R_N are asymptotically equivalent in distribution in a suitable space of functions. This could be shown by following the approach found in the proof of [13], Proposition 3.15. The proof will not be given here. We should also note that the discussion above assumes that the conditions of Theorem 4.1 and Proposition 4.5 are met, that is, $f_W(w)$ or $f_{W_q}(w)$ has a sufficiently fast decay. In practice, these conditions cannot be verified and caution should be exercised with the resulting CI's. Another practical problem with Proposition 4.5 and a related bootstrap procedure is that b is not known in advance. In the simulation study found in Section 5.1 below, we shall explore CI's given by (4.21) and also use a bootstrap procedure, but where $\|\cdot\|_{\infty,b}$ is replaced by $\|\cdot\|_l$ with $\|x\|_l = \sup_{1 \leq n \leq l} |x_n|$ for some fixed l .

4.2. Distribution of interrenewal times

Here, we are interested in estimating the original distribution function F_D of interrenewal times. We focus on nonparametric estimators, in the spirit of Buchmann and Grübel [5] and Bøgsted and Pitts [3], and their basic properties (such as asymptotic normality).

Estimation of F_D is based on a theoretical inversion of the relation (3.3). (As mentioned in Section 3, forms of conditioning other than that on $W_q = s$ are possible and could be dealt with by following the approach developed below.) Note that, from (3.3), we would expect

$$\tilde{F}_D(v) = G_{a_s}(\tilde{F}_{D_{q,i}|s}(v)), \tag{4.25}$$

where $a_s = \{a_{s,n}\}_{n \in \mathbb{N}}$ is the reversion of the sequence $A_s = \{A_{s,n}\}_{n \in \mathbb{N}}$ (see Section 2). In terms of distribution functions, the relation (4.25) can be written as

$$F_D = \sum_{n=1}^{\infty} a_{s,n} F_{D_{q,i}|s}^{*n}. \tag{4.26}$$

The following result provides sufficient conditions for (4.26) to hold and follows directly from [3], Theorem 1. It uses the following notation. Let $r(G_{A_s})$ and $r(G_{a_s})$ be the radii of convergence of the corresponding power series G_{A_s} and G_{a_s} (in the sense found in, e.g., [16]). By Bøgsted and Pitts [3], Proposition 1, there exist

$$\sigma(G_{A_s}) \in (0, r(G_{A_s})], \quad \sigma(G_{a_s}) \in (0, r(G_{a_s})], \tag{4.27}$$

such that $|w| < \sigma(G_{a_s})$ implies that there exists a unique z where $|z| < \sigma(G_{A_s})$ and

$$G_{A_s}(z) = w, \quad z = G_{a_s}(w). \tag{4.28}$$

We also need a suitable space of functions. For $\tau \geq 0$, let $D_\tau[0, \infty) = \{f : [0, \infty) \mapsto \mathbb{R} : \|f\|_{\infty, \tau} := \sup_{t \geq 0} e^{-\tau t} |f(t)| < \infty\}$.

Theorem 4.2 ([3], Theorem 1). *Let $\tau > 0$ be such that*

$$\tilde{F}_{D_{q,i}|s}(\tau) < \sigma(G_{a_s}). \tag{4.29}$$

The series on the right-hand side of (4.26) then converges in $D_\tau[0, \infty)$. If, in addition,

$$\tilde{F}_D(\tau) < \sigma(G_{A_s}), \tag{4.30}$$

then the relation (4.26) holds.

As discussed in [3], the conditions (4.29) and (4.30) always hold for large enough τ . We now turn to estimation based on the inversion (4.26). It is natural to consider the estimator

$$\hat{F}_D = \sum_{n=1}^{\infty} \hat{a}_{s,n} \hat{F}_{D_{q,i}|s}^{*n}. \tag{4.31}$$

Here, $\hat{a}_s = \{\hat{a}_{s,n}\}_{n \in \mathbb{N}}$ is the reversion of the sequence $\hat{A}_s = \{\hat{A}_{s,n}\}_{n \in \mathbb{N}}$ defined as

$$\hat{A}_{s,n} = \frac{q^s}{\hat{f}_{W_q}(s)} \sum_{w=s+n-1}^{\infty} \hat{f}_W(w) \binom{w-n}{s-1} (1-q)^{w-s}, \tag{4.32}$$

where \widehat{f}_{W_q} and \widehat{f}_W are given in (4.8) and (4.10), respectively (i.e., \widehat{A}_s is defined as in (3.4), but where f_{W_q} and f_W are replaced by their sample counterparts). For the distribution function estimator, we take

$$\widehat{F}_{D_{q,i}|s}(t) = \frac{1}{N \widehat{f}_{W_q}(s)} \sum_{k=1}^N 1_{\{D_{q,i,k} \leq t, W_{q,k}=s\}}, \tag{4.33}$$

where $D_{q,i,k}$ are i.i.d. observations of $D_{q,i}$. An important difference between (4.31) and similar estimators considered in [3,5] is that (4.31) involves an estimator \widehat{a}_s of a_s (whereas a_s was assumed to be known in these related works). This obviously makes the analysis of the estimator (4.31) more involved. Also, note that the estimator \widehat{F}_D in (4.31) is defined for fixed s and i . In addition, the estimator involves averages which are conditional on $W_q = s$. In particular, the expressions (4.32) and (4.33) are only defined for $\widehat{f}_{W_q}(s) \neq 0$.

We will show below that the estimator (4.31) is asymptotically normal under suitable assumptions. Note that the estimator (4.31) can be viewed as a functional of \widehat{A}_s (via \widehat{a}_s) and $\widehat{F}_{D_{q,i}|s}$. We first need a suitable result on the asymptotic normality of the latter quantities. Let $(\zeta = \{\zeta_n\}_{n \in \mathbb{N}}, Z = \{Z(t)\}_{t \geq 0})$ be a zero-mean Gaussian process characterized by the following:

$$\zeta_n = -\frac{\eta A_{s,n}}{f_{W_q}(s)} + \frac{q^s}{f_{W_q}(s)} \sum_{w=n+s-1}^{\infty} S(\xi)_w \binom{w-n}{s-1} (1-q)^{w-s}, \quad n \geq 1, \tag{4.34}$$

$$Z(t) = -\frac{\eta F_{D_{q,i}|s}(t)}{f_{W_q}(s)} + \frac{1}{f_{W_q}(s)} B(f_{W_q}(s) F_{D_{q,i}|s}(t)), \quad t \geq 0, \tag{4.35}$$

where $S(\xi)$ appears in (4.14) of Theorem 4.1, η is a zero-mean Gaussian variable with $E\eta^2 = f_{W_q}(s) - f_{W_q}(s)^2$ and B is a standard Brownian bridge such that

$$E\eta S(\xi)_w = 1_{\{s \geq w\}} \binom{s}{w} (-1)^{s-w} q^{-s} (1-q)^{s-w} f_{W_q}(s) - f_{W_q}(s) f_W(w),$$

$$E\eta B(f_{W_q}(s) F_{D_{q,i}|s}(t)) = f_{W_q}(s) (1 - f_{W_q}(s)) F_{D_{q,i}|s}(t),$$

$$ES(\xi)_w B(f_{W_q}(s) F_{D_{q,i}|s}(t)) = 1_{\{s \geq w\}} \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1-q)^{s-w} f_{W_q}(s) (1 - f_{W_q}(s)) F_{D_{q,i}|s}(t).$$

The proof of the following result can be found in Appendix A.

Proposition 4.6. *Suppose that for some $z_0 \geq 1$,*

$$\sum_{w=1}^{\infty} \sqrt{R_{q,w}} (1-q)^w z_0^w w^s < \infty, \tag{4.36}$$

where $R_{q,w}$ is defined in (4.11). Then,

$$\sqrt{N} ((\widehat{A}_s, \widehat{F}_{D_{q,i}|s}) - (A_s, F_{D_{q,i}|s})) \xrightarrow{d} (\zeta, Z) \tag{4.37}$$

with the convergence in the space $(l_{\infty, z_0}, D_0[0, \infty))$, where the limit (ζ, Z) is characterized by (4.34) and (4.35) above.

We are now ready to prove the asymptotic normality result for the estimator \widehat{F}_D in (4.31). In addition to the notation introduced before Theorem 4.2, we shall also use the following. For a formal power series $G(z) = \sum_{n=0}^{\infty} a_n z^n$, let

$$\nu_G(\sigma) = \inf_{|z|=\sigma} |G(z)|. \tag{4.38}$$

If $r(G)$ denotes the radius of convergence of G as before, we also let

$$r_0(G) = \sup\{\sigma_0 : \sigma_0 \leq r(G), \nu_G(\sigma) > 0 \text{ for } 0 < \sigma < \sigma_0\}. \tag{4.39}$$

The notation (4.38)–(4.39) follows that found in [16]; also recall the notation $\mathcal{G}_a(z)$ from Section 2.

Theorem 4.3. *Suppose that the condition (4.36) holds with $z_0 \geq 1$. Also, suppose that for some $\tau > 0$, (4.29) and (4.30) hold, and*

$$\mathcal{G}_{a_s}(\widetilde{F}_{D_{q,i}|s}(\tau)) \leq z_0, \tag{4.40}$$

$$\mathcal{G}_{A_s}(\mathcal{G}_{a_s}(\widetilde{F}_{D_{q,i}|s}(\tau))) < \max_{0 < \sigma < r_0(G_{A_s}) \wedge z_0} \nu_{G_{A_s}}(\sigma). \tag{4.41}$$

Then,

$$\sqrt{N}(\widehat{F}_D - F_D) \xrightarrow{d} X \tag{4.42}$$

with the convergence in the space $D_{\tau}[0, \infty)$. The limit X is a zero-mean Gaussian process which can be expressed as

$$X = - \sum_{n=1}^{\infty} (a_s^{(1)} * (\zeta \circ a_s))_n F_{D_{q,i}|s}^{*n} + \left(\sum_{n=1}^{\infty} a_{s,n} n F_{D_{q,i}|s}^{*(n-1)} \right) * Z, \tag{4.43}$$

where (ζ, Z) is the limit process appearing in (4.37).

Remark 4.1. The conditions (4.40) and (4.41) hold for large enough τ . The presence of z_0 in (4.36) and (4.40) is of technical interest: if $f_W(w)$ has faster decay, then z_0 could possibly be taken larger in (4.36) and hence τ could be taken smaller in (4.40).

Proof of Theorem 4.3. For notational simplicity, we will not write the index s in A_s, a_s, \widehat{A}_s or \widehat{a}_s . By the Skorokhod representation theorem and using Proposition 4.6, we can suppose that

$$\sqrt{N}((\widehat{A}, \widehat{F}_{D_{q,i}|s}) - (A, F_{D_{q,i}|s})) \rightarrow (\zeta, Z) \quad \text{a.s.} \tag{4.44}$$

in the norm $(\|\cdot\|_{\infty, z_0}, \|\cdot\|_{\infty, 0})$. Write

$$\begin{aligned} \sqrt{N}(\widehat{F}_D - F_D) &= \sqrt{N} \sum_{n=1}^{\infty} \widehat{a}_n \widehat{F}_{D_{q,i}|s}^{*n} - \sqrt{N} \sum_{n=1}^{\infty} a_n F_{D_{q,i}|s}^{*n} \\ &= \sqrt{N} \sum_{n=1}^{\infty} (\widehat{a}_n - a_n) (\widehat{F}_{D_{q,i}|s}^{*n} - F_{D_{q,i}|s}^{*n}) + \sqrt{N} \sum_{n=1}^{\infty} (\widehat{a}_n - a_n) F_{D_{q,i}|s}^{*n} \\ &\quad + \sqrt{N} \sum_{n=1}^{\infty} a_n (\widehat{F}_{D_{q,i}|s}^{*n} - F_{D_{q,i}|s}^{*n}) =: T_1 + T_2 + T_3. \end{aligned}$$

By Bøgsted and Pitts [3], Theorem 2, we have

$$T_3 \rightarrow \left(\sum_{n=1}^{\infty} a_n n F_{D_{q,i}|s}^{*(n-1)} \right) * Z \quad \text{a.s.}$$

in the norm $\|\cdot\|_{\infty, \tau}$, which is the second term in the limit (4.43).

We next show that the term T_2 converges to the first term in the limit (4.43). Observe that by using [5], Lemma 6(b) (see also the inequality used in the proof of their Lemma 7), we have

$$\begin{aligned} \|E_2\|_{\infty, \tau} &:= \left\| \sqrt{N} \sum_{n=1}^{\infty} (\widehat{a}_n - a_n) F_{D_{q,i}|s}^{*n} - \sum_{n=1}^{\infty} (-a^{(1)} * (\zeta \circ a))_n F_{D_{q,i}|s}^{*n} \right\|_{\infty, \tau} \\ &\leq \sum_{n=1}^{\infty} |\sqrt{N}(\widehat{a}_n - a_n) + (a^{(1)} * (\zeta \circ a))_n| \|F_{D_{q,i}|s}^{*n}\|_{\infty, \tau} \\ &\leq C \sum_{n=1}^{\infty} |\sqrt{N}(\widehat{a}_n - a_n) + (a^{(1)} * (\zeta \circ a))_n| \widetilde{F}_{D_{q,i}|s}(\tau)^n. \end{aligned}$$

Writing $\sqrt{N}(\widehat{a} - a) + a^{(1)} * (\zeta \circ a) = (\sqrt{N}(\widehat{a} \circ A - I) + (a^{(1)} \circ A) * \zeta) \circ a$ and using the inequality $\mathcal{G}_{x \circ y}(z) \leq \mathcal{G}_x(\mathcal{G}_y(z))$, we further obtain that

$$\begin{aligned} \|E_2\|_{\infty, \tau} &\leq C \sum_{n=1}^{\infty} |(\sqrt{N}(\widehat{a} \circ A - I) + (a^{(1)} \circ A) * \zeta)_n| (\mathcal{G}_a(\widetilde{F}_{D_{q,i}|s}(\tau)))^n \\ &\leq C(R_1 + R_2 + R_3 + R_4), \end{aligned}$$

where, with $z_1 = \mathcal{G}_a(\widetilde{F}_{D_{q,i}|s}(\tau))$,

$$\begin{aligned} R_1 &= \sum_{n=1}^{\infty} |(\sqrt{N}(\widehat{a} \circ (\widehat{A} + A - \widehat{A}) - \widehat{a} \circ \widehat{A}) - (\widehat{a}^{(1)} \circ \widehat{A}) * (\sqrt{N}(A - \widehat{A})))_n| z_1^n, \\ R_2 &= \sum_{n=1}^{\infty} |((\widehat{a}^{(1)} \circ \widehat{A}) * (\sqrt{N}(A - \widehat{A})) - (a^{(1)} \circ \widehat{A}) * (\sqrt{N}(A - \widehat{A})))_n| z_1^n, \end{aligned}$$

$$R_3 = \sum_{n=1}^{\infty} |((a^{(1)} \circ \widehat{A}) * (\sqrt{N}(A - \widehat{A})) - (a^{(1)} \circ A) * (\sqrt{N}(A - \widehat{A})))_n| z_1^n,$$

$$R_4 = \sum_{n=1}^{\infty} |((a^{(1)} \circ A) * (\sqrt{N}(\widehat{A} - A)) - (a^{(1)} \circ A) * \zeta)_n| z_1^n.$$

We will next show that $R_k \rightarrow 0$ a.s., $k = 1, 2, 3, 4$.

For the term R_1 , by using Proposition B.1 in Appendix B, we first observe that

$$R_1 \leq \frac{\sqrt{N}}{2} \sum_{n=1}^{\infty} |((\widehat{a}^{(2)} \circ (\widehat{A}_+ + (A - \widehat{A})_+)) * (A - \widehat{A})_+ * (A - \widehat{A})_+)_n z_1^n|$$

$$= \frac{1}{2\sqrt{N}} \mathcal{G}_{\widehat{a}^{(2)}}(\mathcal{G}_{\widehat{A}}(z_1) + \mathcal{G}_{A-\widehat{A}}(z_1))(\mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1))^2.$$

By (4.44), and since $z_1 \leq z_0$ by the assumption (4.40), we have $\mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1) \rightarrow \mathcal{G}_{\zeta}(z_1)$ a.s. Next, we want to show that

$$\mathcal{G}_{\widehat{a}^{(2)}}(\mathcal{G}_{\widehat{A}}(z_1) + \mathcal{G}_{A-\widehat{A}}(z_1)) \leq C \quad \text{a.s.} \quad (4.45)$$

for some random constant C . For this, we further examine the radius of convergence of $\mathcal{G}_{\widehat{a}^{(2)}}$. By the Cauchy–Hadamard formula (see, e.g., [16], Theorem 2.2a, page 77),

$$r(\mathcal{G}_{\widehat{a}^{(2)}}) = r(\mathcal{G}_{\widehat{a}}) = r(G_{\widehat{a}}).$$

By applying the inequality after the proof of [16], Theorem 2.2b, page 99, we then have

$$r(\mathcal{G}_{\widehat{a}^{(2)}}) \geq \max_{0 < \sigma < r_0(G_{\widehat{a}})} \nu_{G_{\widehat{a}}}(\sigma) \geq \max_{0 < \sigma < r_0(G_{\widehat{a}}) \wedge z_0} \nu_{G_{\widehat{a}}}(\sigma),$$

where the notation $r_0(G)$, ν_G was introduced above. Since $G_{\widehat{A}}(z)$ converges to $G_A(z)$ a.s. and uniformly on $|z| \leq z_0$, we have

$$\max_{0 < \sigma < r_0(G_{\widehat{A}}) \wedge z_0} \nu_{G_{\widehat{A}}}(\sigma) \rightarrow \max_{0 < \sigma < r_0(G_A) \wedge z_0} \nu_{G_A}(\sigma) \quad \text{a.s.}$$

The relation (4.45) now follows from the relations above, the fact that $\mathcal{G}_{\widehat{A}}(z_1) + \mathcal{G}_{A-\widehat{A}}(z_1) \rightarrow \mathcal{G}_A(z_1)$ a.s. and the assumption (4.41). We can now conclude that $R_1 \rightarrow 0$ a.s.

For the term R_2 , note that for fixed $K \geq 1$,

$$R_2 \leq \sum_{n=1}^{\infty} |(\widehat{a}^{(1)} - a^{(1)})_n| (\mathcal{G}_{\widehat{A}}(z_1))^n \mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1)$$

$$\leq \left(\sum_{n=1}^K |(\widehat{a}^{(1)} - a^{(1)})_n| (\mathcal{G}_{\widehat{A}}(z_1))^n + \sum_{n=K+1}^{\infty} |\widehat{a}_n^{(1)}| (\mathcal{G}_{\widehat{A}}(z_1))^n \right.$$

$$\left. + \sum_{n=K+1}^{\infty} |a_n^{(1)}| (\mathcal{G}_{\widehat{A}}(z_1))^n \right) \mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1) =: R_{2,1} + R_{2,2} + R_{2,3}.$$

For a fixed K , $R_{2,1} \rightarrow 0$ a.s. by using the a.s. convergence of \widehat{A} in (4.44). On the other hand, arguing as for the term R_1 above, we can make sure that $R_{2,2}$ and $R_{2,3}$ are arbitrarily small for large enough K . For the term R_3 , by using Proposition B.1,

$$\begin{aligned} R_3 &\leq \sum_{n=1}^{\infty} |(a^{(1)} \circ \widehat{A} - a^{(1)} \circ A)_n| z_1^n \mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1) \\ &\leq \sum_{n=1}^{\infty} (a_+^{(2)} \circ (A_+ + (\widehat{A} - A)_+) * (\widehat{A} - A)_+)_n z_1^n \mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1) \\ &= \frac{1}{\sqrt{N}} \mathcal{G}_{a^{(2)}}(\mathcal{G}_A(z_1) + \mathcal{G}_{\widehat{A}-A}(z_1)) (\mathcal{G}_{\sqrt{N}(A-\widehat{A})}(z_1))^2 \end{aligned}$$

and we have $R_3 \rightarrow 0$ by arguing as for the term R_1 . By using the bound

$$R_4 \leq \mathcal{G}_{a^{(1)}}(\mathcal{G}_A(z_1)) \mathcal{G}_{\sqrt{N}(\widehat{A}-A)-\zeta}(z_1),$$

we get that $R_4 \rightarrow 0$ a.s.

Finally, we shall prove that the term T_1 is asymptotically negligible. Note that, as in the proof of [5], Proposition 8,

$$\begin{aligned} \|T_1\|_{\infty, \tau} &\leq \sqrt{N} \sum_{n=1}^{\infty} |\widehat{a}_n - a_n| \|\widehat{F}_{D_{q,i}|s}^{*n} - F_{D_{q,i}|s}^{*n}\|_{\infty, \tau} \\ &\leq \|\sqrt{N}(\widehat{F}_{D_{q,i}|s} - F_{D_{q,i}|s})\|_{\infty, 0} \sum_{n=1}^{\infty} |\widehat{a}_n - a_n| \widetilde{H}_n(\tau), \end{aligned}$$

where

$$H_n = \sum_{j=0}^{n-1} \widehat{F}_{D_{q,i}|s}^{*j} * F_{D_{q,i}|s}^{*(n-1-j)}.$$

Using the fact that $\widetilde{F}_{D_{q,i}|s}(\tau) \rightarrow \widetilde{F}_{D_{q,i}|s}(\tau)$ a.s., we can further deduce that

$$\|T_1\|_{\infty, \tau} \leq C \|\sqrt{N}(\widehat{F}_{D_{q,i}|s} - F_{D_{q,i}|s})\|_{\infty, 0} \sum_{n=1}^{\infty} |\widehat{a}_n - a_n| ((1 + \delta) \widetilde{F}_{D_{q,i}|s}(\tau))^n$$

for any arbitrarily small $\delta > 0$ and all N large enough. The convergence of the latter series to 0 can be proven by arguing as for the term T_2 and using the fact that $\delta > 0$ can be chosen arbitrarily small. □

Despite its obvious theoretical interest, Theorem 4.3 is of limited relevance in practice where the conditions of the theorem cannot be verified and τ is unknown. Note also that the form of the limiting Gaussian process in (4.43) is not easily tractable. In a simulation study found in Section 5.2 below, we make qualitative observations on the performance of \widehat{F}_D and the use of

bootstrap for CI's. The CI's provided below will be based on the sup-norm on an interval $[0, T]$, with fixed T . Justification of the bootstrap procedure is beyond the scope of this paper.

5. Simulation study

Here, we provide a simulation study for the inference of the distributions of the number of renewals and interrenewal times.

5.1. Simulations for number of renewals

As mentioned in [Introduction](#), the estimator \widehat{f}_W in (4.9) is known to perform poorly as q gets smaller. Here, we shall re-examine this fact and the performance of the estimator \widehat{f}_W via the asymptotic variance $ES(\xi)_w^2$ in (4.16) of the normal limit in (4.14). In this regard, two regimes should be distinguished:

- stable $\left(\sup_{w \in \mathbb{N}} R_{q,w} < \infty \text{ or } \sup_{w \in \mathbb{N}} ES(\xi)_w^2 < \infty \right)$;
 - explosive $\left(\sup_{w \in \mathbb{N}} R_{q,w} = \infty \text{ or } \sup_{w \in \mathbb{N}} ES(\xi)_w^2 = \infty \right)$,
- (5.1)

where $R_{q,w}$ is defined in (4.11), that is,

$$R_{q,w} = \sum_{s=w}^{\infty} \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} f_{W_q}(s).$$

As will be seen below, the performance of the estimator \widehat{f}_W largely depends on which of the two regimes is considered. Before providing some simulations, we take a closer look at these regimes for the distributions f_W considered in Section 3.2.1.

Geometric distribution

Suppose that f_W follows a geometric distribution with parameter $c \in (0, 1)$, as in Section 3.2.1. We shall derive lower and upper bounds for $R_{q,w}$. For the lower bound, observe that

$$f_{W_q}(s) \geq \sum_{w=s}^{\infty} q^s (1-q)^{w-s} f_W(w) = C_1 (cq)^s$$

for some constant C_1 (which depends on c and q) and hence that

$$R_{q,w} \geq C_1 \sum_{s=w}^{\infty} \frac{(1-q)^{2(s-w)}}{q^{2s}} (cq)^s = C_2 \left(\frac{c}{q} \right)^w \tag{5.2}$$

for some constant C_2 (which depends on c and q). For the upper bound of $R_{q,w}$, we need a bound on binomial coefficients. We shall use the standard (but rough) bound given by $\binom{n}{k} \leq n^k/k!$. We shall also use the following auxiliary result, proved in [Appendix A](#).

Lemma 5.1. *Let $a \in (0, 1)$. Then, for some constant $C > 0$ and all $s \geq 1$,*

$$\int_s^\infty w^s a^w dw \leq C s^s a^s (1 + (\ln a^{-1})^{-s}), \tag{5.3}$$

$$\int_s^\infty w^{2s} a^w dw \leq C s^{2s} a^s \left(1 + \left(\frac{1}{2} \ln a^{-1}\right)^{-2s}\right). \tag{5.4}$$

Using the lemma, with generic constants C_k ,

$$\begin{aligned} f_{W_q}(s) &\leq \frac{C_1}{s!} q^s (1-q)^{-s} \sum_{w=s}^\infty w^s (c(1-q))^w \\ &\leq \frac{C_2}{s!} q^s (1-q)^{-s} \int_s^\infty w^s (c(1-q))^w dw \\ &\leq \frac{C_3}{s!} q^s (1-q)^{-s} s^s (c(1-q))^s (1 + (\ln(c(1-q))^{-1})^{-s}) \\ &\leq C_4 (ecq)^s (1 + (\ln(c(1-q))^{-1})^{-s}), \end{aligned}$$

where the last inequality follows from Stirling’s formula. Hence, by arguing similarly and using (5.4), we have

$$\begin{aligned} R_{q,w} &\leq C_1 \frac{(1-q)^{-2w}}{(w!)^2} \sum_{s=w}^\infty s^{2w} \frac{(1-q)^{2s}}{q^{2s}} \left((ecq)^s + \left(\frac{ecq}{\ln(c(1-q))^{-1}} \right)^s \right) \\ &\leq C_2 \frac{(1-q)^{-2w}}{(w!)^2} w^{2w} \left(\frac{(1-q)^{2w}}{q^{2w}} (ecq)^w \left(1 + \left(\frac{1}{2} \ln \frac{q}{(1-q)^2 ec} \right)^{-2w} \right) \right. \\ &\quad \left. + \frac{(1-q)^{2w}}{q^{2w}} \left(\frac{ecq}{\ln(c(1-q))^{-1}} \right)^w \right. \\ &\quad \left. \times \left(1 + \left(\frac{1}{2} \ln \frac{q \ln(c(1-q))^{-1}}{(1-q)^2 ec} \right)^{-2w} \right) \right) \\ &\leq C_3 \left(\frac{e^3 c}{q} \right)^w \left(1 + \left(\ln \frac{1}{c(1-q)} \right)^{-w} + \left(\frac{1}{2} \ln \frac{q}{(1-q)^2 ec} \right)^{-2w} \right. \\ &\quad \left. + \left(\frac{1}{2} \ln \frac{q \ln(c(1-q))^{-1}}{(1-q)^2 ec} \right)^{-2w} \right. \\ &\quad \left. + \left(\ln \frac{1}{c(1-q)} \right)^{-w} \left(\frac{1}{2} \ln \frac{q \ln(c(1-q))^{-1}}{(1-q)^2 ec} \right)^{-2w} \right) \\ &\leq C_4 \left(\frac{e^3 c}{q} \right)^w, \quad \text{if } \frac{e^3 c}{q} \leq 1, \end{aligned} \tag{5.5}$$

where, for the last inequality, we used the fact that all three log terms (before raising to the powers $-w$ and $-2w$) are bigger than 1 under $e^3 c/q \leq 1$.

The bounds (5.2) and (5.5) show that, for the geometric distribution f_W ,

$$\begin{aligned} \frac{c}{q} \leq e^{-3} &\Rightarrow \text{stable regime,} \\ \frac{c}{q} > 1 &\Rightarrow \text{explosive regime.} \end{aligned} \tag{5.6}$$

What happens in the range $c/q \in (e^{-3}, 1]$ remains an open question. Our experience in practice suggests that the critical point for c/q is closer to 1 and that e^{-3} is a very rough bound.

Heavy-tailed distribution

Suppose that f_W follows a heavy-tailed distribution with parameter α (Section 3.2.1). We will show that in this case, the distribution is always in the explosive regime (and this happens for all $\alpha > 0$, not just $\alpha \in (1, 2)$). Indeed, observe that

$$f_{W_q}(s) \geq C \sum_{w=s}^{\infty} q^s (1-q)^{w-s} w^{-\alpha-1} \geq C q^s s^{-\alpha-1}$$

and hence

$$R_{q,w} \geq C \sum_{s=w}^{\infty} \frac{(1-q)^{2(s-w)}}{q^{2s}} q^s s^{-\alpha-1} \geq C \frac{w^{-\alpha-1}}{q^w}. \tag{5.7}$$

It remains to observe that the lower bound in (5.7) diverges as $w \rightarrow \infty$ since $q \in (0, 1)$.

Remark 5.1. In the examples above, note that in the explosive regime, $R_{q,w}$ or $ES(\xi)_w^2$ diverges at an exponential rate. The term ‘‘explosive’’ was chosen to reflect this fact. Also, note that in the explosive regime, in order to have convergence of $\widehat{f}_W(w)$ for all w , we naturally need to consider weighted spaces, as in Proposition 4.5 above.

Remark 5.2 (The impact of small q). Note that $R_{q,w} \uparrow \infty$ as $q \downarrow 0$. Since $R_{q,w} \geq q^{-2w} f_{W_q}(w)$, when w is larger and q is small, $R_{q,w}$ (or $\text{Var}(\widehat{f}_W(w))$) can also be too large for practical purposes (e.g., with $w = 10$ and $q = 0.1$, $R_{q,w} \geq 10^{20} f_{W_q}(w)$). Moreover, as with the geometric distribution above, we may expect that most of distributions of interest (with unbounded support) belong to the explosive regime for small enough q . These observations reiterate the current understanding that inference of f_{W_q} becomes impractical for small q .

In Figures 2 and 3, we illustrate the performance of the estimator $\widehat{f}_W(w)$ in the two regimes above. Figure 2 is for the geometric distribution with $c = 0.25$, $q = 0.6$, $N = 500$, and Figure 3 is for the Pareto distribution with $\alpha = 1.5$, $q = 0.7$, $N = 1000$. The left-hand plots in these figures depict the true p.m.f. $f_W(w)$, and the 5%, 50% and 95% percentiles of the distribution of the estimator $\widehat{f}_W(w)$, based on 1000 Monte Carlo (MC) replications. The right-hand plots

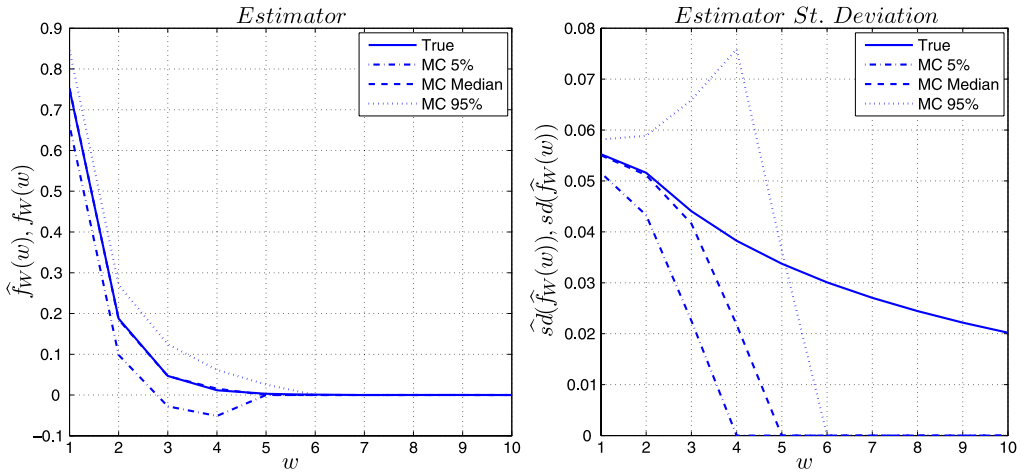


Figure 2. Performance of the estimator $\hat{f}_W(w)$ for geometric distribution with $c = 0.25, q = 0.6, N = 500$. Left plot: true $f_W(w)$ and MC-based 5%, 50%, 95% percentiles of $\hat{f}_W(w)$. Right plot: true asymptotic standard deviation $sd(\hat{f}_W(w))$ and MC-based 5%, 50%, 95% percentiles of $\hat{sd}(\hat{f}_W(w))$.

in these figures depict the true asymptotic standard deviation $(ES(\xi)_w^2/N)^{1/2}$ or $sd(\hat{f}_W(w))$, and the 5%, 50% and 95% percentiles of the distribution of the estimator $(\widehat{ES(\xi)_w^2}/N)^{1/2}$ or $\hat{sd}(\hat{f}_W(w))$, again based on 1000 MC replications.

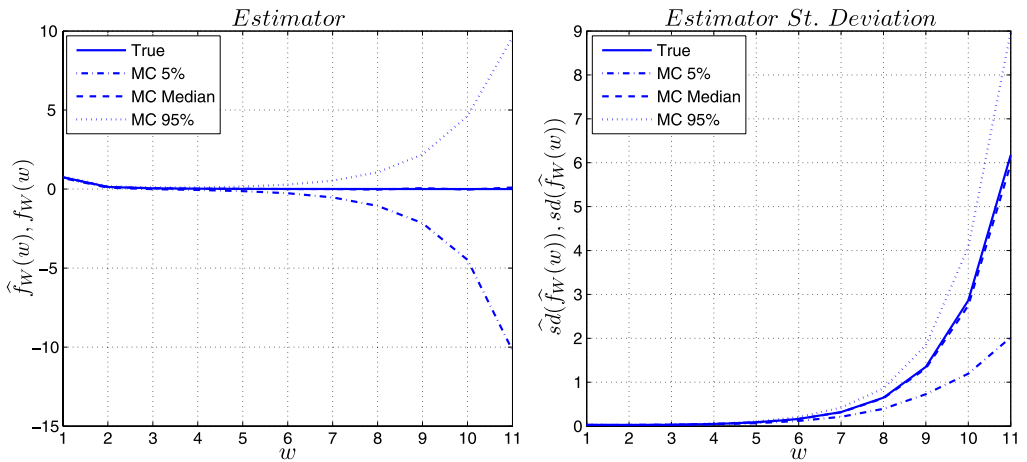


Figure 3. Performance of the estimator $\hat{f}_W(w)$ for Pareto distribution with $\alpha = 1.5, q = 0.7, N = 1000$. Left plot: true $f_W(w)$ and MC-based 5%, 50%, 95% percentiles of $\hat{f}_W(w)$. Right plot: true asymptotic standard deviation $sd(\hat{f}_W(w))$ and MC-based 5%, 50%, 95% percentiles of $\hat{sd}(\hat{f}_W(w))$.

Note that by the definition (4.10) of $\widehat{f}_W(w)$, we have $\widehat{f}_W(w) = 0$ for all w larger than the size of the largest sampled flow(s). The largest sampled flow(s) in the study corresponding to Figure 2 consists of 5 sampled packets. For this reason, the $\widehat{f}_W(w)$'s are all zero for $w \geq 6$ in Figure 2 (left plot) and obviously have zero estimated standard deviation for $w \geq 6$ in Figure 2 (right plot). Figure 3, on the other hand, does not depict zero $\widehat{f}_W(w)$'s because the size of the largest sampled flow(s) is much greater. Indeed, only a short range of $w = 1, \dots, 11$ is considered in Figure 3. As noted in Remark 5.1 above, the estimator standard deviation grows exponentially with increasing w . For larger w , the features of both plots in Figure 3 would be dominated by this exponential blow-up, even if there is a drop to zero in $\widehat{f}_W(w)$ and its estimated standard deviation after the largest sampled flow(s).

Figure 3 depicts typical estimation in the explosive regime: $\widehat{f}_W(w)$ and its estimated standard deviation blow up (before dropping to zero). In Figure 2, on the other hand, $\widehat{f}_W(w)$ and its estimated standard deviation are stable and reach zero. This is the situation associated with the stable regime. Thus, even though the parameters c and q in Figure 2 do not satisfy the sufficient condition for the stable regime in (5.6), we still expect that these parameters actually lead to the stable regime. (Recall that, as indicated earlier, the condition in (5.6) is likely to be very strong.) Also, observe that Figure 3 is for a sample size of only $N = 1000$, and roughly the first 5 frequencies can be estimated with confidence. With larger sample sizes, for example, with $N = 10\,000$ and $N = 100\,000$, roughly the first 10 and 12 frequencies, respectively, can be estimated with confidence.

In Figure 4, we give an idea of the appropriateness of confidence intervals (CI's). The left- and right-hand plots of Figure 4 correspond to the situations considered in Figures 2 and 3, respectively. The 95% MC-based percentile of $\widehat{f}_W(w)$ is depicted as in Figures 2 and 3. This percentile is, in fact, centered at the true $f_W(w)$. It can be thought of as an ideal upper 90% CI bound, corresponding to dealing with an exact distribution of $\widehat{f}_W(w)$. We also depict analogous bounds based on bootstrapping (BT) and the asymptotic normality (AN) result (4.14) for $\widehat{f}_W(w)$, with either true ("true") or estimated ("est.") asymptotic variance. In the cases of BT and AN (est.) bounds, what are depicted are, in fact, the median bounds obtained from MC realizations. As with MC, the BT, AN (true) and AN (est.) bounds are centered at $f_W(w)$. Note from Figure 4 that CI's based on both BT and AN (est.) approaches, which are the only practical alternatives, are quite satisfactory.

5.2. Simulations for interrenewal times

We present here a simulation study for the estimator \widehat{F}_D in (4.31) of the distribution of interrenewal times. Two cases are considered. In case 1, we take geometric f_W with $c = 0.25$ and $q = 0.6$, as in Section 5.1 (corresponding to the stable regime). In case 2, geometric f_W is taken with $c = 0.7$ and $q = 0.6$ (corresponding to the explosive regime; the simulation results are similar if the Pareto distribution is used as in Section 5.1). In both cases, D is supposed to be an exponential random variable with parameter 1. The respective sample sizes are $N = 500$ and $N = 1000$.

Figure 5 presents simulation results for case 1 in the left-hand plot and case 2 in the right-hand plot. We depict the true function F_D and the 5%, 50% and 95% percentiles of the distribution

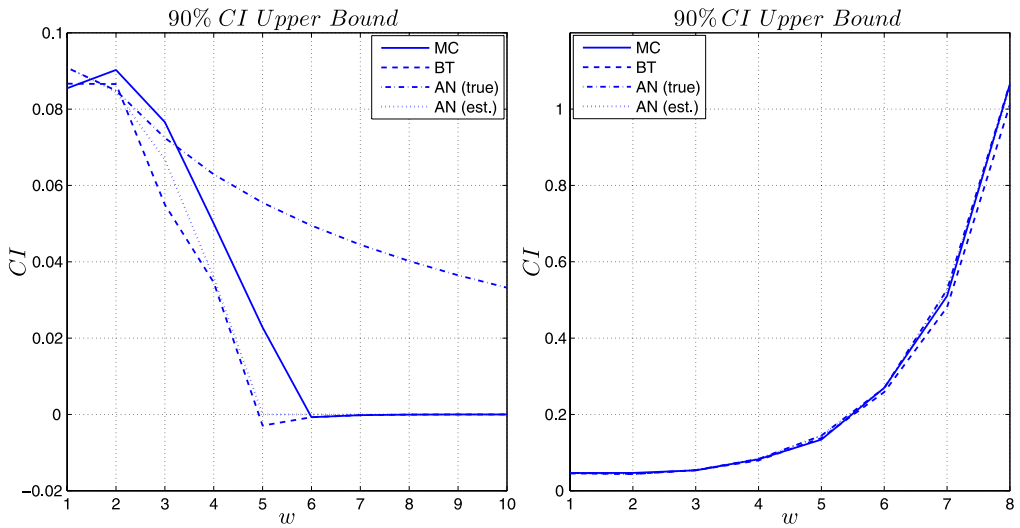


Figure 4. 95% CI upper bounds for $\widehat{f}_W(w) - f_W(w)$, that is, for $\widehat{f}_W(w)$ centered at true $f_W(w)$, for the two situations of Figures 2 and 3. The left plot corresponds to Figure 2 and the right plot corresponds to Figure 3. Both plots consist of the 95% MC-based percentile of $\widehat{f}_W(w) - f_W(w)$ (MC), median bound using bootstrap (BT) and bounds based on the asymptotic normality (AN) result (4.14) for $\widehat{f}_W(w)$ with true ("true") and estimated ("est.") asymptotic variance.

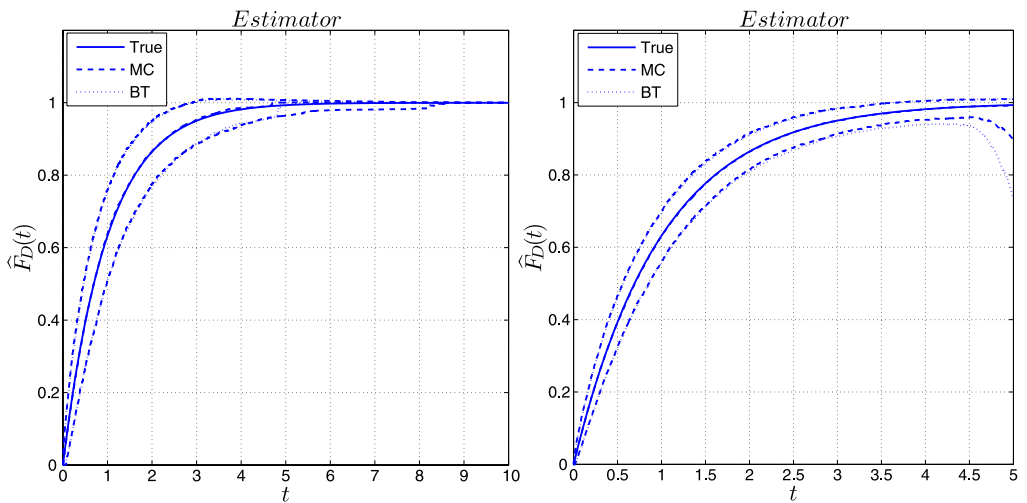


Figure 5. Performance of \widehat{F}_D and the bootstrap CI's for the exponential distribution with parameter 1. Left plot: $N = 500$, $q = 0.6$ and geometric f_W with $c = 0.25$. Right plot: $N = 500$, $q = 0.6$ and geometric f_W with $c = 0.7$. Plots include true F_D , MC-based 5%, 50% and 95% percentiles of \widehat{F}_D , and the medians of 90% bootstrap (BT) CI's.

of \widehat{F}_D , based on 1000 MC realizations. The plots also contain the medians of 90% bootstrap-based CI's, computed from 1000 MC replications. The latter suggests, in particular, that such CI's are quite appropriate. We should also note that the estimator \widehat{F}_D is here based only on the first sampled interrenewal times when $s = 2$, that is, $i = 1$ and $s = 2$ in (4.31). In addition, the estimator is computed by truncating the infinite sum in its definition (4.31) and by evaluating convolutions numerically through discretizing convolution integrals.

Note, from Figure 5, that estimation is satisfactory, even for case 2 which corresponds to the explosive regime. This seems quite surprising and needs further explanation. In fact, in the setting of case 2, only the first several values of $\widehat{f}_W(w)$ are really relevant for the estimator \widehat{F}_D . This can be seen from the definition (4.32) of the sequence \widehat{A}_s , which enters into \widehat{F}_D after reversion. Note, from (4.32), that in calculating $\widehat{A}_{s,n}$, the first series term $\widehat{f}_W(n + 1)$ is weighted by $(1 - q)^{n-1}$. This additional factor thus helps to keep the blow-up of \widehat{f}_W under control. Moreover, from the examples above, if the growth of $\widehat{f}_W(w)$ is thought to be of order q^{-w} , then the factor $(1 - q)^w$ annihilates this growth when $q > 0.5$. In practice, and in case 2, we also observe that, even though $\widehat{f}_W(w)$ is highly varying for larger w , the sequence $\widehat{A}_{s,n}$ decays to zero rapidly. Since the first few values of $\widehat{f}_W(w)$ can be estimated with confidence even in the explosive regime, the above explains the satisfactory performance of \widehat{F}_D in case 2 of Figure 5. On the other hand, note also that the t -range is smaller in the right-hand plot of Figure 5 and that the variance of the estimator starts to diverge at the boundary $T = 5$ of the range. In fact, this divergence would be more pronounced (and would dominate the plot) if we increased the range of t . Thus, the performance of the estimator is satisfactory, but only to some time point T .

In Figure 6, we also illustrate the performance of \widehat{F}_D when using several other forms of conditioning on W_q . The two plots in the figure correspond to the respective plots of Figure 5. They depict the 5%, 50% and 95% percentiles of \widehat{F}_D when conditioning on $W_q = s$ with $s = 2, s \geq 2$

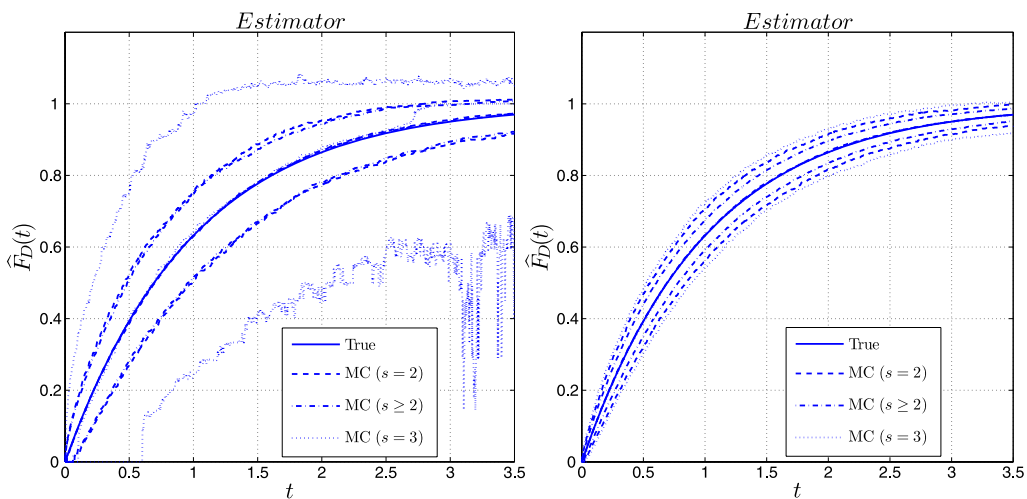


Figure 6. Performance of the estimator $\widehat{F}_D(t)$ for two cases from Figure 5. Plots include MC-based 5%, 50% and 95% percentiles of $\widehat{F}_D(t)$ when conditioning on $W_q = s$ with $s = 2, s \geq 2$ and $s = 3$.

and $s = 3$. (In the case of conditioning on $s \geq 2$, the formulas (3.10) and (3.11) are used.) The 5–95% interpercentile range is the smallest when conditioning on $s \geq 2$, although comparable to that when $s = 2$ is used. An obvious note to add here is that different types of conditioning (for the same N) lead to different effective sample sizes of the data used in the computation of \widehat{F}_D , with that for $s \geq 2$ being the largest and that for $s = 3$ being the smallest. The interpercentile range is much larger for $s = 3$ than for $s = 2$ in the left-hand plot because the difference between the corresponding effective sample sizes is much larger for case 1.

Appendix A: Proofs

Proof of Theorem 3.2. For $1 \leq i_1 < \dots < i_n < s$ and $t_1 \geq 0, \dots, t_n \geq 0$, we have

$$\begin{aligned}
 &P(D_{q,i_1} \leq t_1, \dots, D_{q,i_n} \leq t_n, W_q = s) \\
 &= \sum_{w=s}^{\infty} f_W(w) P(W_q = s | W = w) P(D_{q,i_1} \leq t_1, \dots, D_{q,i_n} \leq t_n | W = w, W_q = s).
 \end{aligned} \tag{A.1}$$

Let $M_j, j = 1, \dots, n$, be the number of renewals not sampled between the i_j th and $(i_j + 1)$ th sampled renewals plus 1. If $W = w$ and $W_q = s$ are fixed and $M_1 = m_1, \dots, M_{j-1} = m_{j-1}$, then M_j can take the values $1, \dots, w - m_1 - \dots - m_{j-1} - (s - j)$ and hence

$$\begin{aligned}
 &P(D_{q,i_1} \leq t_1, \dots, D_{q,i_n} \leq t_n | W = w, W_q = s) \\
 &= \sum_{m_1=1}^{w-(s-1)} \sum_{m_2=1}^{w-m_1-(s-2)} \dots \sum_{m_n=1}^{w-m_1-\dots-m_{n-1}-(s-n)} F_D^{*m_1}(t_1) \dots F_D^{*m_n}(t_n) \\
 &\quad \times P(M_1 = m_1, \dots, M_n = m_n | W = w, W_q = s).
 \end{aligned} \tag{A.2}$$

We shall next derive an expression for the probability $P(M_1 = m_1, \dots, M_n = m_n | W = w, W_q = s)$ in (A.2). The indices i_1, \dots, i_n are not necessarily consecutive, but still form separate blocks of consecutive indices (e.g., $i_1 = 1, i_2 = 2, i_3 = 4, i_4 = 8, i_5 = 9$ form three separate blocks of indices $\{i_1, i_2\}, \{i_3\}$ and $\{i_4, i_5\}$). We need additional notation to keep track of these separate blocks and the gaps between them. Thus, let $1 = j_1 < j_2 < \dots < j_k \leq n$ denote subindices j in i_j for which the corresponding i_j is the start of a separate block (e.g., with $i_1 = 1, i_2 = 2, i_3 = 4, i_4 = 8, i_5 = 9$ above, we have $j_1 = 1, j_2 = 3, j_3 = 4$). With this notation, note that $r_u = j_{u+1} - j_u, 1 \leq u < k - 1$ and $r_k = n - j_k + 1$ denote the sizes of the k separate blocks. Also, let $x_0 = i_{j_1} - 1, x_u = i_{j_{u+1}} - i_{j_u} - r_u - 1, 1 \leq u < k - 1$, and $x_k = s - i_{j_k} - r_k$ be the numbers of sampled renewals, respectively, before the i_{j_1} th sampled renewal, between the $(i_{j_u} + r_u)$ th and $i_{j_{u+1}}$ th sampled renewals, and after the $(i_{j_k} + r_k)$ th sampled renewal.

Given $W = w$ and $W_q = s$, the number of possible distinct locations of the sampled renewals satisfying $M_1 = m_1, \dots, M_n = m_n$ can be obtained by considering the possible initial locations of the sampled renewals i_{j_u} , along with the number of possible distinct locations of the sampled

renewals before the i_{j_1} th, between the $(i_{j_u} + r_u)$ th and $i_{j_{u+1}}$ th, and after the $(i_{j_k} + r_k)$ th sampled renewal, that is,

$$\begin{aligned} & \sum_{l_1=x_0+1}^{w-\widehat{m}_1-\dots-\widehat{m}_k-x_1-\dots-x_k-(k-1)} \binom{l_1-1}{x_0} \sum_{l_2=l_1+\widehat{m}_1+x_1+1}^{w-\widehat{m}_2-\dots-\widehat{m}_k-x_2-\dots-x_k-(k-1)} \binom{l_2-l_1-\widehat{m}_1-1}{x_1} \\ & \times \dots \times \sum_{l_{k-1}=l_{k-2}+\widehat{m}_{k-2}+x_{k-1}+1}^{w-\widehat{m}_{k-1}-\widehat{m}_k-x_{k-1}-x_k-1} \binom{l_{k-1}-l_{k-2}-\widehat{m}_{k-2}-1}{x_{k-2}} \\ & \times \sum_{l_k=l_{k-1}+\widehat{m}_{k-1}+x_{k-1}+1}^{w-\widehat{m}_k-x_k} \binom{l_k-l_{k-1}-\widehat{m}_{k-1}-1}{x_{k-1}} \binom{w-l_k-\widehat{m}_k}{x_k}, \end{aligned}$$

where l_u is the location of the i_{j_u} th sampled renewal and $\widehat{m}_u = \sum_{v=j_u}^{j_u+r_u-1} m_v$ is the number of renewals between the i_{j_u} th and $(i_{j_u} + r_u)$ th sampled renewals plus 1. By making the change of variables $l'_u = l_u - \sum_{v=1}^{u-1} \widehat{m}_v$, $u = 2, \dots, k$, the last expression becomes

$$\begin{aligned} & \sum_{l_1=x_0+1}^{w-m-x_1-\dots-x_k-(k-1)} \binom{l_1-1}{x_0} \sum_{l'_2=l'_1+x_1+1}^{w-m-x_2-\dots-x_k-(k-2)} \binom{l'_2-l'_1-1}{x_1} \\ & \times \dots \times \sum_{l'_{k-1}=l_{k-2}+x_{k-1}+1}^{w-m-x_{k-1}-x_k-1} \binom{l'_{k-1}-l'_{k-2}-1}{x_{k-2}} \\ & \times \sum_{l'_k=l'_{k-1}+x_{k-1}+1}^{w-m-x_k} \binom{l'_k-l'_{k-1}-1}{x_{k-1}} \binom{w-m-l'_k}{x_k} = \binom{w-m}{x_0+\dots+x_k+k}, \end{aligned}$$

where the last identity can be viewed as a generalization of the identity (3.8) used in the proof of Theorem 3.1. Since $x_0 + \dots + x_k + k = s - n$, we deduce that

$$P(M_1 = m_1, \dots, M_n = m_n | W = w, W_q = s) = \binom{w-m}{s-n} / \binom{w}{s}. \quad (\text{A.3})$$

By using (A.1)–(A.3), we have that

$$\begin{aligned} F_{D_{q,n}|s}(t) &= \frac{P(D_{q,i_1} \leq t_1, \dots, D_{q,i_n} \leq t_n, W_q = s)}{f_{W_q}(s)} \\ &= \frac{1}{f_{W_q}(s)} \sum_{w=s}^{\infty} f_W(w) q^s (1-q)^{w-s} \\ & \quad \times \sum_{m_1=1}^{w-(s-1)} \dots \sum_{m_n=1}^{w-m_1-\dots-m_{n-1}-(s-n)} \binom{w-m}{s-n} \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} & \times F_D^{*m_1}(t_1) \cdots F_D^{*m_n}(t_n) \\ & = \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} B_{s,(m_1,\dots,m_n)} F_D^{*m_1}(t_1) \cdots F_D^{*m_n}(t_n), \end{aligned}$$

where $B_{s,\mathbf{m}}$ is given by (3.13). The relation (3.12) is obtained by taking the LT in (A.4). □

Proof of Proposition 3.1. For $s \geq 2$ and $t \geq 0$, we have

$$P(V_q \leq t, W_q = s) = \sum_{w=2}^{\infty} f_W(w) P(W_q = s | W = w) P(V_q \leq t | W = w, W_q = s). \tag{A.5}$$

Now, letting M be the number of original renewals between the first and last sampled renewals plus 1, we have

$$\begin{aligned} P(V_q \leq t | W = w, W_q = s) & = \sum_{m=1}^{w-1} F_D^{*m}(t) P(M = m | W = w, W_s = s) \\ & = \sum_{m=1}^{w-1} F_D^{*m}(t) (w - m) \binom{m-1}{s-2} / \binom{w}{s}, \end{aligned} \tag{A.6}$$

where $(w - m)$ is the number of possible distinct locations of the first sampled renewal and $\binom{m-1}{s-2}$ is the number of possible locations of sampled renewals between the first and last sampled renewals when $M = m$. By using (A.5) and (A.6), we deduce that

$$F_{V_q|2^+}(t) = \frac{1}{P(W_q \geq 2)} \sum_{s=2}^{\infty} P(V_q \leq t, W_q = s) = \sum_{m=1}^{\infty} C_m F_D^{*m}(t), \tag{A.7}$$

which can also be written as in (3.14). □

Proof of Proposition 4.1. The equality in (4.2) follows from (3.1) and

$$\begin{aligned} & \sum_{s=n}^{\infty} \binom{s}{n} \frac{(1-q)^{s-n}}{q^s} \sum_{w=s}^{\infty} \binom{w}{s} q^s (1-q)^{w-s} f_W(w) \\ & = \sum_{w=n}^{\infty} f_W(w) (1-q)^{w-n} \sum_{s=n}^w \binom{s}{n} \binom{w}{s} \\ & = \sum_{w=n}^{\infty} f_W(w) (1-q)^{w-n} \binom{w}{n} \sum_{s=n}^w \binom{w-n}{s-n} = \sum_{w=n}^{\infty} f_W(w) (1-q)^{w-n} \binom{w}{n} 2^{w-n}. \end{aligned}$$

Then, by the assumption (4.2), we can substitute (3.1) into the right-hand side of (4.1) and use Fubini's theorem to obtain

$$\begin{aligned}
 & \sum_{s=w}^{\infty} \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1-q)^{s-w} \sum_{n=s}^{\infty} \binom{n}{s} q^s (1-q)^{n-s} f_W(n) \\
 &= \sum_{n=w}^{\infty} f_W(n) (1-q)^{n-w} \sum_{s=w}^n \binom{s}{w} \binom{n}{s} (-1)^{s-w} \\
 &= \sum_{n=w}^{\infty} f_W(n) (1-q)^{n-w} \binom{n}{w} \sum_{s=w}^n \binom{n-w}{s-w} (-1)^{s-w} \\
 &= \sum_{n=w}^{\infty} f_W(n) (1-q)^{n-w} \binom{n}{w} \sum_{k=0}^{n-w} \binom{n-w}{k} (-1)^k = f_W(w),
 \end{aligned}$$

where we have used the identity $\sum_{k=0}^K \binom{K}{k} (-1)^k = 0$ if $K \geq 1$, and $= 1$ if $K = 0$. □

Proof of Proposition 4.2. We consider only the case $l = 2$ (and $z_2 = 0$). Note that

$$\begin{aligned}
 T^{(2)}(x)_n &= \sum_{i=n}^{\infty} \binom{i}{n} T^{(1)}(x)_i (-z_1)^{i-n} \\
 &= \sum_{i=n}^{\infty} \binom{i}{n} (-z_1)^{i-n} \sum_{k=i}^{\infty} \binom{k}{i} \frac{x_k}{q^k} (z_1 - (1-q))^{k-i} \\
 &= \sum_{k=n}^{\infty} \frac{x_k}{q^k} \sum_{i=n}^k \binom{i}{n} \binom{k}{i} (-z_1)^{i-n} (z_1 - (1-q))^{k-i} \\
 &= \sum_{k=n}^{\infty} \frac{x_k}{q^k} \binom{k}{n} \sum_{i=n}^k \binom{k-n}{i-n} (-z_1)^{i-n} (z_1 - (1-q))^{k-i} \\
 &= \sum_{k=n}^{\infty} \frac{x_k}{q^k} \binom{k}{n} (-1 - q)^{k-n} = S(x)_n.
 \end{aligned}$$

The change of the order of summation above can be justified by using Fubini's theorem and the assumption (4.5). □

Proof of Proposition 4.3. The proof is similar to that of Propositions 4.1 and 4.2, and only the case $l = 2$ (and $z_2 = 0$) will be considered. By using (3.1), we observe that

$$\begin{aligned}
 T^{(1)}(f_{W_q})_n &= \sum_{i=n}^{\infty} \binom{i}{n} \frac{(z_1 - (1-q))^{i-n}}{q^i} \sum_{k=i}^{\infty} \binom{k}{i} q^i (1-q)^{k-i} f_W(k) \\
 &= \sum_{k=n}^{\infty} f_W(k) \sum_{i=n}^k \binom{i}{n} \binom{k}{i} (z_1 - (1-q))^{i-n} (1-q)^{k-i} \\
 &= \sum_{k=n}^{\infty} f_W(k) \binom{k}{n} \sum_{i=n}^k \binom{k-n}{i-n} (z_1 - (1-q))^{i-n} (1-q)^{k-i} \\
 &= \sum_{k=n}^{\infty} f_W(k) \binom{k}{n} z_1^{k-n},
 \end{aligned} \tag{A.8}$$

where the order of the summation above is changed using Fubini’s theorem. Indeed, the application of Fubini’s theorem is possible as long as $|z_1 - (1-q)| + (1-q) < 1$ or $z_1 \in C_0$, which is one of the conditions on z_1 . By using (A.8), we further get that

$$\begin{aligned}
 T^{(2)}(f_{W_q})_n &= \sum_{i=n}^{\infty} \binom{i}{n} (-z_1)^{i-n} \sum_{k=i}^{\infty} \binom{k}{i} z_1^{k-i} f_W(k) \\
 &= \sum_{k=n}^{\infty} f_W(k) \sum_{i=n}^k \binom{i}{n} \binom{k}{i} (-z_1)^{i-n} z_1^{k-i} = f_W(n),
 \end{aligned}$$

where the use of Fubini’s theorem is justified by the fact that $|-z_1| + |z_1| < 1$ or $z_2 = 0 \in C_1$, which is one of the conditions on z_2 . □

Proof of Proposition 4.4. Write $T^{(1)}(\xi^{(1)}) = T^{(1)}(\xi_1^{(1)}) + T^{(1)}(\xi_2^{(1)})$, where

$$\xi_{1,s}^{(1)} = \begin{cases} \xi_s^{(1)}, & \text{if } s \leq j, \\ 0, & \text{if } s > j, \end{cases} \quad \xi_{2,s}^{(1)} = \begin{cases} 0, & \text{if } s \leq j, \\ \xi_s^{(1)}, & \text{if } s > j \end{cases}$$

and observe that, for fixed j , $T^{(1)}(\xi_1^{(1)})$ and $T^{(1)}(\xi_2^{(1)})$ are independent. We can then also write

$$T^{(2)}(\xi^{(1)})_{j,w} = T^{(2)}(\xi_1^{(1)})_{j,w} + T^{(2)}(\xi_2^{(1)})_{j,w} = T^{(2)}(\xi_1^{(1)})_w + T^{(2)}(\xi_2^{(1)})_{j,w}, \tag{A.9}$$

where $T^{(2)}(\xi_1^{(1)})_w$ and $T^{(2)}(\xi_2^{(1)})_{j,w}$ are independent for fixed j . By Proposition 4.2, we have that

$$T^{(2)}(\xi_1^{(1)})_w = S(\xi_1^{(1)})_w = \sum_{s=w}^j \binom{s}{w} \frac{(-1)^{s-w}}{q^s} (1-q)^{s-w} f_{W_q}(s)^{1/2} \eta_s.$$

By using the fact that the two terms in (A.9) are Gaussian and independent, we can then write

$$E e^{i\theta T^{(2)}(\xi^{(1)})_{j,w}} = e^{-(1/2)\theta^2\sigma_{1,w}^2} e^{-(1/2)\theta^2\sigma_{2,w}^2}, \tag{A.10}$$

where

$$\begin{aligned} \sigma_{1,w}^2 &= E(T^{(2)}(\xi_1^{(1)})_w)^2 = \sum_{s=w}^j \binom{s}{w}^2 \frac{(1-q)^{2(s-w)}}{q^{2s}} f_{W_q}(s), \\ \sigma_{2,w}^2 &= E(T^{(2)}(\xi_2^{(1)})_{j,w})^2. \end{aligned}$$

In view of (A.10), for $T^{(2)}(\xi^{(1)})_{j,w}$ to converge in distribution, it is necessary that the series $\sigma_{1,w}^2$ converges, that is, that (4.13) holds. (In other words, if (4.13) does not hold, then the limit of (A.10) is zero.) \square

Proof of Proposition 4.5. The proof is in two steps: (a) $\sqrt{N}(\widehat{f}_{W_q} - f_{W_q}) \rightarrow_d \xi$ in the space $l_{\infty,q^{-1}(b'+1-q)}$, where ξ is the process appearing in the proof of Theorem 4.1; (b) the mapping S in (4.4) is continuous from $l_{\infty,q^{-1}(b'+1-q)}$ to $l_{\infty,b}$.

For (a), it is enough to show the tightness of $\sqrt{N}(\widehat{f}_{W_q} - f_{W_q})$ in $l_{\infty,q^{-1}(b'+1-q)}$. By the results of [23], page 229 (see also [4], Theorem 2.5), this follows from the fact that for any $\varepsilon > 0$,

$$\lim_{m \rightarrow \infty} \sup_{N \geq 1} P\left(\sup_{s > m} (q^{-1}(b'+1-q))^s |\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))| > \varepsilon\right) = 0. \tag{A.11}$$

For the latter, observe that

$$\begin{aligned} &\varepsilon P\left(\sup_{s > m} \left(\frac{b'+1-q}{q}\right)^s |\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))| > \varepsilon\right) \\ &\leq \sum_{s=m+1}^{\infty} \left(\frac{b'+1-q}{q}\right)^s E|\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))| \\ &\leq \sum_{s=m+1}^{\infty} \left(\frac{b'+1-q}{q}\right)^s \sqrt{E|\sqrt{N}(\widehat{f}_{W_q}(s) - f_{W_q}(s))|^2} \leq \sum_{s=m+1}^{\infty} \left(\frac{b'+1-q}{q}\right)^s f_{W_q}(s)^{1/2} \end{aligned}$$

and that (A.11) follows from the assumption (4.20).

For (b), observe that

$$\begin{aligned} \|S(x) - S(y)\|_{\infty,b} &= \sup_{n \geq 0} b^n |S(x)_n - S(y)_n| \leq \sum_{n=0}^{\infty} b^n |S(x)_n - S(y)_n| \\ &\leq \sum_{n=0}^{\infty} b^n \sum_{i=n}^{\infty} \binom{i}{n} \frac{(1-q)^{i-n}}{q^i} |x_i - y_i| = \sum_{i=0}^{\infty} |x_i - y_i| \left(\frac{b+1-q}{q}\right)^i \end{aligned}$$

$$\begin{aligned} &\leq \sup_{n \geq 0} \left(\frac{b' + 1 - q}{q} \right)^n |x_n - y_n| \sum_{i=0}^{\infty} \left(\frac{b + 1 - q}{b' + 1 - q} \right)^i \\ &= \|x - y\|_{\infty, q^{-1}(b'+1-q)} \sum_{i=0}^{\infty} \left(\frac{b + 1 - q}{b' + 1 - q} \right)^i. \end{aligned} \quad \square$$

Proof of Proposition 4.6. With $X^N = \sqrt{N}(\widehat{A}_s - A_s)$ and $Y^N = \sqrt{N}(\widehat{F}_{D_{q,i}|s} - F_{D_{q,i}|s})$, it is enough to show that (a) the finite-dimensional distributions of (X^N, Y^N) converge to those of (ζ, Z) ; (b) the sequence (X^N, Y^N) is tight. The latter condition follows from the following two conditions: (b1) the sequence $X^N = \{X_n^N\}_{n \in \mathbb{N}}$ is tight; (b2) the sequence Y^N is tight. The tightness of Y_N in (b2) is a standard result, part of the empirical central limit theorem (see, e.g., [20], Section V.2) and will not be proven here. By the results of [23], the condition (b1) can be replaced by (b1) for any $\varepsilon > 0$, $\lim_{m \rightarrow \infty} \sup_{N \geq 1} P(\sup_{n > m} z_0^n |X_n^N| > \varepsilon) = 0$.

For the latter (b1), it is enough to show the same relation, but where X^N is replaced by \widetilde{X}^N , which is defined without $q^s/\widehat{f}_{W_q}(s)$ and $q^s/f_{W_q}(s)$ in \widehat{A}_s and A_s , respectively. With the sequence \widetilde{X}^N , observe that

$$\begin{aligned} \varepsilon P\left(\sup_{n > m} z_0^n |\widetilde{X}_n^N| > \varepsilon\right) &\leq \sum_{n=m+1}^{\infty} z_0^n E|\widetilde{X}_n^N| \\ &\leq \sum_{n=m+1}^{\infty} \sum_{w=n+s-1}^{\infty} z_0^n E|\sqrt{N}(\widehat{f}_W(w) - f_W(w))| \binom{w-n}{s-1} (1-q)^{w-s} \\ &= \sum_{w=m+s}^{\infty} E|\sqrt{N}(\widehat{f}_W(w) - f_W(w))| (1-q)^{w-s} \sum_{n=m+1}^{w-s+1} \binom{w-n}{s-1} z_0^k \\ &\leq \sum_{w=m+s}^{\infty} E|\sqrt{N}(\widehat{f}_W(w) - f_W(w))| (1-q)^{w-s} z_0^{w-s+1} \binom{w-m}{s}, \end{aligned}$$

where we have used the fact that $z_0 \geq 1$ and $\sum_{n=m+1}^{w-s+1} \binom{w-n}{s-1} = \binom{w-m}{s}$. Since $E|\sqrt{N}(\widehat{f}_W(w) - f_W(w))|^2 \leq R_{q,w}$, as in the proof of Theorem 4.1, we further get that

$$\begin{aligned} \varepsilon P\left(\sup_{n > m} z_0^n |\widetilde{X}_n^N| > \varepsilon\right) &\leq \sum_{w=m+s}^{\infty} \sqrt{E|\sqrt{N}(\widehat{f}_W(w) - f_W(w))|^2} (1-q)^{w-s} z_0^{w-s+1} \binom{w-m}{s} \\ &\leq \sum_{w=m+s}^{\infty} \sqrt{R_{q,w}} (1-q)^{w-s} z_0^{w-s+1} \binom{w-m}{s}. \end{aligned}$$

The condition (b1) now follows from the assumption (4.36).

For the convergence of the finite-dimensional distributions in (a), we show only the convergence of X_n^N (the general case can be considered along similar lines). For $j \geq n + s - 1$, define

$$X_{j,n}^N = \sqrt{N}(\widehat{A}_{s,n}^j - A_{s,n}^j),$$

where

$$\widehat{A}_{s,n}^j = \frac{q^s}{\widehat{f}_{W_q}(s)} \sum_{w=n+s-1}^j \widehat{f}_W(w) \binom{w-n}{s-1} (1-q)^{w-s}$$

and similarly with $A_{s,n}^j$, and also

$$\zeta_n^j = -\frac{\eta A_{s,n}}{f_{W_q}(s)} + \frac{q^s}{f_{W_q}(s)} \sum_{w=n+s-1}^j S(\xi)_w \binom{w-n}{s-1} (1-q)^{w-s}.$$

It is enough to show that:

- (i) $X_{j,n}^N \xrightarrow{d} \zeta_n^j$ as $N \rightarrow \infty$;
- (ii) $\zeta_n^j \xrightarrow{d} \zeta_n$ as $j \rightarrow \infty$;
- (iii) for any $\delta > 0$, $\limsup_{j \rightarrow \infty} \limsup_{N \rightarrow \infty} P(|X_{j,n}^N - X_n^N| > \delta) = 0$.

The convergence in (i) follows directly from Theorem 4.1. The convergence in (ii) follows using $E|S(\xi)_w| \leq \sqrt{ES(\xi)_w^2} \leq \sqrt{R_{q,w}}$ and the assumption (4.36). Condition (iii) can be proven as in part (b1) above using the assumption (4.36). □

Proof of Lemma 5.1. Using integration by parts, we obtain

$$\begin{aligned} \int_s^\infty w^s a^w dw &= \frac{s^s a^s}{\ln a^{-1}} + \frac{s}{\ln a^{-1}} \int_s^\infty w^{s-1} a^w dw \\ &\leq \frac{s^s a^s}{\ln a^{-1}} + \frac{s^s a^s}{(\ln a^{-1})^2} + \frac{s^2}{(\ln a^{-1})^2} \int_s^\infty w^{s-2} a^w dw \\ &\leq s^s a^s \sum_{k=1}^s \frac{1}{(\ln a^{-1})^k} + \frac{s^s}{(\ln a^{-1})^s} \int_s^\infty a^w dw \\ &= s^s a^s \sum_{k=1}^{s+1} \frac{1}{(\ln a^{-1})^k} = \frac{s^s a^s}{\ln a^{-1}} \cdot \frac{(\ln a^{-1})^{-s-1} - 1}{(\ln a^{-1})^{-1} - 1}, \end{aligned}$$

from which the bound (5.3) follows. The inequality (5.4) can be proven similarly. □

Appendix B: Bounds on remainder terms in Taylor expansions of compositions of power series

Here, we prove a result which was used several times in the proof of Theorem 4.3.

Proposition B.1. *For any formal power series $x(z) = \sum_{n=1}^{\infty} x_n z^n$, $y(z) = \sum_{n=1}^{\infty} y_n z^n$, $\varepsilon(z) = \sum_{n=1}^{\infty} \varepsilon_n z^n$ and $n \geq 1$, we have*

$$|(x \circ (y + \varepsilon) - x \circ y)_n| \leq ((x_+^{(1)} \circ (y_+ + \varepsilon_+)) * \varepsilon_+)_n, \quad (\text{B.1})$$

$$|(x \circ (y + \varepsilon) - x \circ y - (x^{(1)} \circ y) * \varepsilon)_n| \leq \frac{1}{2}((x_+^{(2)} \circ (y_+ + \varepsilon_+)) * \varepsilon_+ * \varepsilon_+)_n. \quad (\text{B.2})$$

Proof. For $N \geq 1$, consider new formal power series $x_N(z) = \sum_{n=1}^N x_n z^n$, $y_N(z) = \sum_{n=1}^N y_n z^n$ and $\varepsilon_N(z) = \sum_{n=1}^N \varepsilon_n z^n$. Since the power series $x_N \circ (y_N + \varepsilon_N) - x_N \circ y_N$ and $x \circ (y + \varepsilon) - x \circ y$, and $(x_{N,+}^{(1)} \circ (y_{N,+} + \varepsilon_{N,+})) * \varepsilon_{N,+}$ and $(x_+^{(1)} \circ (y_+ + \varepsilon_+)) * \varepsilon_+$ have the same first N elements, and since N is arbitrary, it is enough to prove (B.1) for any x_N , y_N and ε_N . Letting $f_m(u_1, \dots, u_N) = x_m(\sum_{k=1}^N u_k z^k)^m$ be a real-valued function on \mathbb{R}^N , and with Df_m denoting the gradient of f_m , observe that

$$\begin{aligned} (x_N \circ (y_N + \varepsilon_N) - x_N \circ y_N)(z) &= \sum_{m=1}^N x_m \left(\left(\sum_{k=1}^N (y_k + \varepsilon_k) z^k \right)^m - \left(\sum_{k=1}^N y_k z^k \right)^m \right) \\ &= \sum_{m=1}^N (f_m(y_1 + \varepsilon_1, \dots, y_N + \varepsilon_N) - f_m(y_1, \dots, y_N)) \\ &= \sum_{m=1}^N [\varepsilon_1 \cdots \varepsilon_N]^T Df_m(c_{m,1}, \dots, c_{m,N}) \\ &= \sum_{m=1}^N x_m m \left(\sum_{k=1}^N c_{m,k} z^k \right)^{m-1} \sum_{k=1}^N \varepsilon_k z^k, \end{aligned} \quad (\text{B.3})$$

where A^T is the transpose of the matrix A and $(c_{m,1}, \dots, c_{m,N}) = (1 - t_m)(y_1, \dots, y_N) + t_m(y_1 + \varepsilon_1, \dots, y_N + \varepsilon_N)$ for some $t_m \in [0, 1]$. On the other hand, observe that

$$((x_{N,+}^{(1)} \circ (y_{N,+} + \varepsilon_{N,+})) * \varepsilon_{N,+})(z) = \sum_{m=1}^N x_m^+ m \left(\sum_{k=1}^N (y_k^+ + \varepsilon_k^+) z^k \right)^{m-1} \sum_{k=1}^N \varepsilon_k^+ z^k. \quad (\text{B.4})$$

Since $|c_{m,k}| \leq y_k^+ + \varepsilon_k^+$, it is clear that the absolute value of the n th element of (B.3) is less than or equal to the n th element of (B.4). This completes the proof of (B.1).

The proof of (B.2) is similar, but involves the observation that

$$\begin{aligned}
 & (x_N \circ (y_N + \varepsilon_N) - x_N \circ y_N - (x_N^{(1)} \circ y_N) * \varepsilon_N)(z) \\
 &= \sum_{m=1}^N (f_m(y_1 + \varepsilon_1, \dots, y_N + \varepsilon_N) - f_m(y_1, \dots, y_N) - [\varepsilon_1 \cdots \varepsilon_N]^T Df_m(y_1, \dots, y_N)) \\
 &= \frac{1}{2} \sum_{m=1}^N [\varepsilon_1 \cdots \varepsilon_N]^T D^2 f_m(d_{m,1}, \dots, d_{m,N}) [\varepsilon_1 \cdots \varepsilon_N] \\
 &= \frac{1}{2} \sum_{m=1}^N x_m m(m-1) \left(\sum_{k=1}^N d_{m,k} z^k \right)^{m-2} \left(\sum_{k=1}^N \varepsilon_k z^k \right)^2,
 \end{aligned}$$

where $D^2 f$ denotes the Hessian of f_m and $(d_{m,1}, \dots, d_{m,N}) = (1 - t_m)(y_1, \dots, y_N) + t_m(y_1 + \varepsilon_1, \dots, y_N + \varepsilon_N)$ with $t_m \in [0, 1]$. □

Acknowledgements

The authors would like to thank the anonymous referees for useful comments and suggestions. The second author was supported in part by NSF Grant DMS-06-08669.

References

- [1] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. New York: Wiley. [MR1700749](#)
- [2] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular variation. Encyclopedia of Mathematics and Its Applications* **27**. Cambridge: Cambridge Univ. Press. [MR0898871](#)
- [3] Bøgsted, M. and Pitts, S.M. (2010). Decomposing random sums: A nonparametric approach. *Ann. Inst. Statist. Math.* **62** 855–872.
- [4] Buchmann, B. (2001). Decomposing: An estimation problem for the compound Poisson distribution. Ph.D. thesis, Univ. Hannover.
- [5] Buchmann, B. and Grübel, R. (2003). Decomposing: An estimation problem for Poisson random sums. *Ann. Statist.* **31** 1054–1074. [MR2001642](#)
- [6] Chabchoub, Y., Fricker, C., Guillemin, F. and Robert, P. (2010). On the statistical characterization of flows in Internet traffic with application to sampling. *Comput. Commun.* **31** 103–112.
- [7] Clegg, R.G., Landa, R., Haddadi, H., Rio, M. and Moore, A.W. (2008). Techniques for flow inversion on sampled data. In *2008 IEEE INFOCOM Workshops* 1–6.
- [8] Cox, D.R. and Isham, V. (1980). *Point Processes*. London: Chapman and Hall. [MR0598033](#)
- [9] Daley, D.J. and Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes. Vol. I*, 2nd ed. New York: Springer. [MR0950166](#)
- [10] Duffield, N., Lund, C. and Thorup, M. (2002). Properties and prediction of flow statistics from sampled packet streams. In *Proc. ACM SIGCOMM Internet Measurement Workshop* 159–171.
- [11] Duffield, N., Lund, C. and Thorup, M. (2005). Estimating flow distributions from sampled flow statistics. *IEEE/ACM Transactions on Networking* **13** 933–946.

- [12] Estan, C. and Varghese, G. (2002). New directions in traffic measurement and accounting. In *SIGCOMM'02. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* 323–336.
- [13] Grübel, R. and Pitts, S.M. (1993). Nonparametric estimation in renewal theory. I. The empirical renewal function. *Ann. Statist.* **21** 1431–1451. [MR1241273](#)
- [14] Hall, P. and Park, J. (2004). Nonparametric inference about service time distribution from indirect measurements. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **66** 861–875. [MR2102469](#)
- [15] Hansen, M.B. and Pitts, S.M. (2006). Nonparametric inference from the $M/G/1$ workload. *Bernoulli* **12** 737–759. [MR2248235](#)
- [16] Henrici, P. (1974). *Applied and Computational Complex Analysis*. New York: Wiley. [MR0372162](#)
- [17] Hohn, N. and Veitch, D. (2006). Inverting sampled traffic. *IEEE/ACM Transactions on Networking* **14** 68–80.
- [18] Hohn, N., Veitch, D. and Abry, P. (2003). Cluster processes: A natural language for network traffic. *IEEE Trans. Signal Process.* **51** 2229–2244. [MR2003784](#)
- [19] Karr, A.F. (1991). *Point Processes and Their Statistical Inference*, 2nd ed. *Probability: Pure and Applied* **7**. New York: Marcel Dekker. [MR1113698](#)
- [20] Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer. [MR0762984](#)
- [21] Robert, C.Y. and Segers, J. (2008). Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance Math. Econom.* **43** 85–92. [MR2442033](#)
- [22] Tune, P. and Veitch, D. (2008). Towards optimal sampling for flow size estimation. In *ACM SIGCOMM Internet Measurement Conference* 243–256.
- [23] Vakhania, N.N., Tarieladze, V.I. and Chobanyan, S.A. (1987). *Probability Distributions on Banach Spaces. Mathematics and Its Applications (Soviet Series)* **14**. Dordrecht: D. Reidel Publishing. [MR1435288](#)
- [24] Yang, L. and Michailidis, G. (2007). Sampled based estimation of network traffic flow characteristics. In *INFOCOM 2007. 26th IEEE International Conference on Computer Communications* 1775–1783.

Received January 2010 and revised August 2010