

Extended Geometric Processes: Semiparametric Estimation and Application to Reliability

Laurent Bordes, Sophie Mercier

Université de Pau et des Pays de l'Adour, Laboratoire de Mathématiques et de leurs Applications – Pau (UMR CNRS 5142), France.

Abstract. Lam (2007) introduces a generalization of renewal processes named Geometric processes, where inter-arrival times are independent and identically distributed up to a multiplicative scale parameter, in a geometric fashion. We here envision a more general scaling, not necessarily geometric. The corresponding counting process is named Extended Geometric Process (EGP). Semiparametric estimates are provided and studied for an EGP, which includes consistency results and convergence rates. In a reliability context, arrivals of an EGP may stand for successive failure times of a system submitted to imperfect repairs. In this context, we study: 1) the mean number of failures on some finite horizon time; 2) a replacement policy assessed through a cost function on an infinite horizon time.

Keywords. Imperfect repair, Markov renewal equation, replacement policy.

MSC: 60K15, 60K20.

Laurent Bordes(laurent.bordes@univ-pau.fr), Sophie Mercier(✉)(sophie.mercier@univ-pau.fr)

Received: June 20, 2011; Accepted: July 24, 2012

1 Introduction

For several years, many attention has been paid to the modeling of recurrent event data. Application fields are various and include medicine, reliability and insurance for instance. See [7] for an overview of models and their applications. In reliability, the events of interest typically are successive failures of a system submitted to instantaneous repair. In case of perfect repairs (As Good As New repairs), the underlying process describing the system evolution is a renewal process, which has been widely used in reliability, see [2]. In case of imperfect repairs, the successive times to failure may however become shorter and shorter, leading to some (stochastically) decreasing sequence of lifetimes. In the same way, in case of improving systems such as software releases e.g., successive times to failure may be increasing.

Such remarks have led to the development of different models taking into account such features, among which geometric processes introduced by [14]. In such a model, successive lifetimes $X_1, X_2, \dots, X_n, \dots$ are independent with identical distributions up to a multiplicative scale parameter: $X_n = a^{n-1}Y_n$ where $(Y_n)_{n \geq 1}$ is a sequence of independent and identically distributed random variables (the interarrival times of a renewal process). According to whether $a \geq 1$ or $0 < a < 1$, the sequence $(X_n)_{n \geq 1}$ may be (stochastically) non-decreasing or non-increasing, which is well adapted for modelling successive lifetimes. However, [5] point out that, *in the exponential case* [exponentially distributed Y_n 's], *the geometric process only allows for logarithmic growth or explosive growth, but nothing in between* (from the conclusion of [5]). In the same paper, *it is* [also] *shown that the expected number of counts at an arbitrary time does not exist for the decreasing geometric process* (from the abstract). Such drawbacks of geometric processes are linked to the fast increase or decrease in the successive periods, induced by the geometric progression. We here envision a more general scaling factor, where X_n is of the shape $X_n = a^{b_n}Y_n$ and $(b_n)_{n \geq 1}$ stands for a non decreasing sequence. This allows for more flexibility in the progression of the X_n 's. The corresponding counting process is named Extended Geometric Process (EGP) in the sequel. A similar extension is also considered in [10] where the author is only concerned with the case where the expected number of counts is not finite on any arbitrary time interval.

As a first step in the study of an EGP, we consider its semiparametric estimation based on the observation of the n first gap times. The sequence $(b_n)_{n \geq 1}$ is assumed to be known and we start with the estimation of the Euclidean parameter a . Following the regression method proposed by [14], several consistency results are obtained for the estimator \hat{a} , including convergence rates. We next proceed to the estimation of the unknown distribution of the underlying renewal process. The estimation method relies on a pseudo version $(\tilde{Y}_n)_{n \geq 1}$ of the points $(Y_n)_{n \geq 1}$ of the underlying renewal process, that is obtained by setting $\tilde{Y}_n = \hat{a}^{-b_n} X_n$. Again, several convergence results are obtained, such as strong uniform consistency.

We next turn to applications of EGPs to reliability, with the previous interpretation of arrivals of an EGP as successive failure times. A first quantity of interest then is the mean number of instantaneous repairs on some time interval $[0, t]$, which corresponds to the pseudo-renewal function associated to an EGP, seen as some pseudo-renewal process. The pseudo-renewal function is proved to fulfill a pseudo-renewal equation, and tools are provided for its numerical solving. In case $a < 1$, the system is aging and requires some action to prevent successive lifetimes to become shorter and shorter. In that case, a replacement policy is proposed: as soon as a lifetime is observed to be too short - below a predefined threshold -, the system is considered as too degraded and it is replaced by a new one. In case $a \geq 1$, the system is improving at each corrective action and no replacement policy is required. In case $a < 1$, the replacement policy is assessed through a cost function, which is provided in full form. The replacement policy proposed here is an alternative to the one considered by [17], where the failure times are modelled by a geometric process and the system is replaced by a new one once it has been repaired N times (with N fixed). Non negligible repair times are also considered by [17] (modelled by another geometric process), which we do not envision here.

This paper is organized as follows. Section 2 is devoted to the semiparametric estimation of an EGP. Applications and numerical examples are developed in Section 3 where the choice of $(b_n)_{n \geq 1}$ is also discussed. In Section 4 we consider applications to reliability together with numerical experiments. Concluding remarks end this paper in Section 5.

2 Estimation of extended geometric processes

2.1 The model

Let $(T_n)_{n \geq 0}$ be a sequence of failure times of a system. We have $0 = T_0 < T_1 < \dots < T_n < \dots$ and we set $X_n = T_n - T_{n-1}$ for $n \geq 1$. Assume that $(X_n)_{n \geq 1}$ satisfies $X_n = a^{b_n} Y_n$ where:

- $(Y_n)_{n \geq 1}$ are the interarrival times of a renewal process (RP), with $\mathbb{P}(Y_1 > 0) > 0$,
- $a \in (0, +\infty)$,
- $(b_n)_{n \geq 1}$ is a non decreasing sequence of non negative real numbers such that $b_1 = 0$ and b_n tends to infinity as n goes to infinity.

In [17], the sequence $(b_n)_{n \geq 1}$ is defined by $b_n = n - 1$ for $n \geq 1$. In the present paper, the sequence $(b_n)_{n \geq 1}$ is first assumed to be fully known. The case where b_n is only known up to an Euclidean parameter is further envisioned in Subsection 3.2. Unknown parameters hence are $a \in (0, +\infty)$ and the cumulative distribution function (c.d.f.) F of the underlying RP in a first step, plus the Euclidian parameter of the b_n 's in Subsection 3.2. Consequently, in each case, it is a semiparametric model.

2.2 Estimation

Assuming that $T_1 < \dots < T_n$ are observed, we consider the problem of estimating a and F (given the sequence b_n). The following estimation method was already considered by Lam in a series of papers, see [14, 16] and [17].

Lam's estimation method is based on a classical regression: writing $Z_n = \log X_n$ for $n \geq 1$, we have $Z_n = b_n \beta + \mu + e_n$ where $\beta = \log a$, $\mu = \mathbb{E}[\log Y_1]$ and $e_n = \log Y_n - \mu$ are independent and identically distributed (i.i.d.) centered errors. Parameters μ and β are next estimated by a least square method:

$$(\hat{\mu}_n, \hat{\beta}_n) = \arg \min_{\mu, \beta} \sum_{k=1}^n (Z_k - \beta b_k + \mu)^2.$$

Here, μ is a nuisance parameter and we concentrate on the estimation of β , or equivalently on the estimation of $a = \exp(\beta)$. We obtain

$$\hat{\beta}_n = \frac{n^{-1} \sum_{k=1}^n b_k Z_k - n^{-2} \sum_{k=1}^n Z_k \sum_{k=1}^n b_k}{n^{-1} \sum_{k=1}^n b_k^2 - (n^{-1} \sum_{k=1}^n b_k)^2},$$

and

$$\hat{\mu}_n = \bar{Z}_n - \hat{\beta}_n \bar{b}_n,$$

where $\bar{b}_n = (b_1 + \dots + b_n)/n$ and $\bar{Z}_n = (Z_1 + \dots + Z_n)/n$. Next, a is estimated by $\hat{a}_n = \exp(\hat{\beta}_n)$. Once a is estimated, we can obtain a pseudo version $(\tilde{Y}_n)_{n \geq 1}$ of the inter-arrival times $(Y_n)_{n \geq 1}$ by setting $\tilde{Y}_n = \hat{a}_n^{-b_n} X_n$. Then, we propose to estimate F by the empirical distribution function \hat{F}_n defined by

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n \mathbf{1}_{\{\tilde{Y}_k \leq x\}} \quad x \in \mathbb{R}^+,$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the set indicator function. The convergence of \hat{F}_n towards F is studied in Proposition 2.4, where a uniform strong consistency result is obtained.

Assuming that $\mathbb{E}(\log^2(Y_n))$ exists, let us define $\text{Var}(e_n) = \sigma^2$. We then have

$$\mathbb{E}(\hat{\beta}_n) = \beta,$$

and

$$\text{Var}(\hat{\beta}_n) = \frac{\sigma^2}{n\alpha_n^2}, \tag{1}$$

where

$$\alpha_n^2 = \frac{1}{n} \sum_{k=1}^n b_k^2 - \left(\frac{1}{n} \sum_{k=1}^n b_k \right)^2.$$

If a central limit theorem holds, its formulation can only be

$$\theta_n(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2),$$

where \xrightarrow{d} stands for the convergence in distribution and $\theta_n = \sqrt{n}\alpha_n$. Thus the convergence rate of $\hat{\beta}_n$ towards β necessarily is of order θ_n . Such a result is provided in Proposition 2.3.

2.3 Asymptotics

Asymptotic results are given with respect to $n \rightarrow +\infty$.

2.3.1 Euclidean parameters

We here make use of strong law of large numbers for weighted sum of i.i.d. random variables, as provided by [8, 3] and [4].

Proposition 2.1 (Strong consistency). *Suppose that $\mathbb{E}(Z_1^2) < +\infty$. Then $\alpha_n(\hat{\beta}_n - \beta) \xrightarrow{a.s.} 0$.*

Proof. Remember that $e_i = \log Y_i - \mu$, and let $S_n = \sum_{i=1}^n a_{i,n} e_i$, where weights $a_{i,n}$ are defined by

$$a_{i,n} = \frac{b_i - \bar{b}_n}{\alpha_n} \quad (\text{setting } \alpha_1 = 1).$$

Then, we have $n\alpha_n(\hat{\beta}_n - \beta) = S_n$. The e_i 's are i.i.d. centered random variables and have finite second order moment, because $\mathbb{E}(|Z_1|^2) < +\infty$. Moreover, following the notations in [4], we have

$$A_{n,2} = \left(\frac{1}{n} \sum_{i=1}^n a_{i,n}^2 \right)^{1/2} = 1$$

and hence $\limsup_n A_{n,2} = 1$. Applying Theorem 1.1 in [8], we obtain that $S_n/n = \alpha_n(\hat{\beta}_n - \beta) \rightarrow 0$ a.s.. ■

Remark 2.1. It is straightforward to verify that

$$\alpha_{n+1}^2 = \alpha_n^2 + \frac{n}{n+1} (b_{n+1} - \bar{b}_n)^2,$$

which implies that $(\alpha_n)_{n \geq 1}$ is a non decreasing sequence. This monotonicity plus the previous consistency result imply that $\hat{\beta}_n \xrightarrow{a.s.} \beta$.

Proposition 2.2. (Law of Iterated Logarithm) *If $\mathbb{E}[Z_1^2] < +\infty$ then*

$$\limsup_{n \rightarrow +\infty} \frac{\sqrt{n}\alpha_n^2}{b_n\sqrt{\log n}} |\hat{\beta}_n - \beta| \leq 2\sqrt{2}\sigma \quad a.s.$$

Proof. Let us consider again $S_n = \sum_{i=1}^n a_{i,n} e_i$, where the e_i 's are i.i.d. centered random variables with finite second order moment. Weights $a_{i,n}$ are now chosen equal to $(b_i - \bar{b}_n)/2b_n$ and satisfy

$$A_\infty = \sup_{n \geq i \geq 1} |a_{i,n}| \leq 1 \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n a_{i,n}^2 = A_{2,n} \leq 1.$$

[3] established in their Theorem 2.1 that

$$\limsup_{n \rightarrow +\infty} \frac{|S_n|}{\sqrt{n \log n}} \leq \sqrt{2} A_2 \sqrt{\mathbb{E}[e_1^2]} \quad \text{a.s.} \quad (2)$$

where $A_2 = \limsup_{n \rightarrow +\infty} A_{2,n}$. Because

$$\hat{\beta}_n - \beta = \frac{1}{n\alpha_n^2} \sum_{i=1}^n (b_i - \bar{b}_n) e_i = \frac{2b_n}{n\alpha_n^2} S_n$$

and since $A_2 \leq 1$, we have by (2)

$$\limsup_{n \rightarrow +\infty} \frac{\sqrt{n\alpha_n^2} |\hat{\beta}_n - \beta|}{b_n \sqrt{\log n}} \leq 2\sqrt{2}\sigma \quad \text{a.s.}$$

which proves the result. \blacksquare

Proposition 2.3. (Central Limit Theorem) *If $\mathbb{E}([Z_1^2]) < +\infty$ and $\sqrt{n}\alpha_n/b_n \rightarrow +\infty$, then*

$$\theta_n(\hat{a}_n - a) \xrightarrow{d} \mathcal{N}(0, a^2\sigma^2),$$

where we recall that $\theta_n = \sqrt{n}\alpha_n$.

Proof. We first prove that

$$\theta_n(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

applying the Lindeberg-Feller theorem (see [12]) to

$$\theta_n(\hat{\beta}_n - \beta) = \frac{1}{\theta_n} \sum_{k=1}^n (b_k - \bar{b}_n) e_k.$$

Using (1), we already know that

$$\text{Var}(\theta_n(\hat{\beta}_n - \beta)) = \sigma^2$$

for all $n \geq 1$ and the first condition in the theorem is fulfilled.

We now check the second condition: for all $\varepsilon > 0$, we have

$$\begin{aligned} & \sum_{k=1}^n \frac{(b_k - \bar{b}_n)^2}{\theta_n^2} \mathbb{E} \left(e_k^2 \mathbf{1}_{\left\{ |e_k| > \frac{\varepsilon \theta_n}{|b_k - \bar{b}_n|} \right\}} \right) \\ & \leq \mathbb{E} \left(e_1^2 \mathbf{1}_{\left\{ |e_1| > \frac{\varepsilon \theta_n}{\max_{1 \leq k \leq n} |b_k - \bar{b}_n|} \right\}} \right) \times \sum_{k=1}^n \frac{(b_k - \bar{b}_n)^2}{\theta_n^2} \\ & \leq \mathbb{E} \left(e_1^2 \mathbf{1}_{\{|e_1| > \varepsilon \theta_n / 2b_n\}} \right). \end{aligned} \quad (3)$$

Because $e_1 = Z_1 - \mu$, $\mathbb{E}(|Z_1|^2) < +\infty$ and $\theta_n/b_n \rightarrow +\infty$, we obtain by Lebesgue's dominated convergence theorem that $\mathbb{E}(e_1^2 \mathbf{1}_{\{|e_1| > \varepsilon \theta_n / 2b_n\}}) \rightarrow 0$. Expression (3) hence tends to zero and the second condition of the Lindeberg-Feller theorem holds.

We derive that $\theta_n(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, and next that $\theta_n(\hat{a}_n - a) \xrightarrow{d} \mathcal{N}(0, a^2 \sigma^2)$ by the δ -method theorem (see e.g. [22]). ■

Example 2.1. If $b_n = (n - 1)^\alpha$ with $\alpha > 0$, we have

$$\theta_n \stackrel{+\infty}{\sim} \frac{\alpha n^{\alpha+1/2}}{(\alpha + 1)\sqrt{2\alpha + 1}}$$

and $\theta_n/b_n \rightarrow +\infty$ as $n \rightarrow +\infty$. We hence get that $n^{\alpha+1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, (\alpha + 1)^2 (2\alpha + 1) \sigma^2 / \alpha^2)$. In the special case where $b_n = n - 1$, this is consistent with the central limit result from [16] which states that $n^{3/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, 12\sigma^2)$.

Example 2.2. If $b_n = \log n$, we have

$$\theta_n \stackrel{+\infty}{\sim} \sqrt{n}$$

and $\theta_n/b_n \rightarrow +\infty$ as $n \rightarrow +\infty$. We hence get that $n^{1/2}(\hat{\beta}_n - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$.

Remark 2.2. Note that from standard results on linear regression

$$\hat{\sigma}_n^2 = \frac{1}{n-2} \sum_{k=1}^n \left(Z_k - \hat{\beta}_n b_k - \hat{\mu}_n \right)^2$$

is an unbiased consistent estimator of σ^2 . Then, the asymptotic variance of $\theta_n(\hat{a}_n - a)$ is consistently estimated by $\hat{a}_n^2 \hat{\sigma}_n^2$.

2.3.2 Functional parameter

The cumulative distribution function F is now estimated by

$$\begin{aligned}\hat{F}_n(x) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\hat{Y}_i \leq x\}} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\log \hat{Y}_i \leq \log x\}} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\log Y_i + b_i(\beta - \hat{\beta}_n) \leq \log x\}} = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\log Y_i \leq \log x + b_i(\hat{\beta}_n - \beta)\}}\end{aligned}$$

for all $x \in (0, +\infty)$.

We also define \hat{F}_n^\pm by

$$\hat{F}_n^\pm(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\log Y_i \leq \log x \pm b_n |\hat{\beta}_n - \beta|\}} \quad \text{for all } x \in (0, +\infty),$$

and we have

$$\hat{F}_n^-(x) \leq \hat{F}_n(x) \leq \hat{F}_n^+(x) \quad (4)$$

for all $x \in (0, +\infty)$.

Define moreover \hat{G}_n and G by

$$\hat{G}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{\log Y_i \leq x\}}, \quad G(x) = \mathbb{P}(\log Y_1 \leq x) \quad \text{for all } x \in \mathbb{R}.$$

Proposition 2.4. (Uniform Strong Consistency) *Assume that Z_1 admits a bounded density g with respect to Lebesgue measure, that Z_1 has a second order moment and that*

$$\limsup_{n \rightarrow +\infty} \frac{b_n^2 \sqrt{\log n}}{\sqrt{n} \alpha_n^2} = 0. \quad (5)$$

Then $\|\hat{F}_n - F\|_\infty$ converges to 0 almost surely as n tends to infinity.

Proof. We have for all $x \in (0, +\infty)$:

$$\begin{aligned}|\hat{F}_n^+(x) - F(x)| &\leq |\hat{G}_n(\log x + b_n |\hat{\beta}_n - \beta|) - G(\log x + b_n |\hat{\beta}_n - \beta|)| \\ &\quad + |G(\log x + b_n |\hat{\beta}_n - \beta|) - G(\log x)| \\ &\leq \|\hat{G}_n - G\|_\infty + \|g\|_\infty b_n |\hat{\beta}_n - \beta|,\end{aligned} \quad (6)$$

where (6) is obtained by applying the mean value theorem to the second term in the right hand side of the first inequality. From the Glivenko-Cantelli theorem, we know that $\|\hat{G}_n - G\|_\infty \rightarrow 0$ a.s..

Besides, by Proposition 2.2 and (5) we have

$$\begin{aligned} \limsup_{n \rightarrow +\infty} b_n |\hat{\beta}_n - \beta| &\leq \limsup_{n \rightarrow +\infty} \frac{\sqrt{n} \alpha_n^2}{b_n \sqrt{\log n}} |\hat{\beta}_n - \beta| \times \limsup_{n \rightarrow +\infty} \frac{b_n^2 \sqrt{\log n}}{\sqrt{n} \alpha_n^2} \\ &\leq 2\sqrt{2}\sigma \times 0 = 0 \text{ a.s..} \end{aligned}$$

Since g is bounded, we derive from (6) that $\|\hat{F}_n^+ - F\|_\infty$ converges to 0 almost surely. By similar arguments, we also get that $\|\hat{F}_n^- - F\|_\infty$ converges to 0 almost surely. Using (4), we have

$$\|\hat{F}_n - F\|_\infty \leq \max\left(\|\hat{F}_n^+ - F\|_\infty, \|\hat{F}_n^- - F\|_\infty\right)$$

which entails that $\|\hat{F}_n - F\|_\infty \rightarrow 0$ almost surely. Hence the proposition is proved. ■

Remark 2.3. The boundedness condition on g is satisfied whenever f belongs to several parametric families (Weibull, Gamma, log-normal, etc.). Condition (5) on the sequence $\left(\frac{b_n^2 \sqrt{\log n}}{\sqrt{n} \alpha_n^2}\right)_{n \geq 1}$ is satisfied for many non decreasing sequences $(b_n)_{n \geq 1}$ tending to infinity. For example:

- if $b_n^2 \sqrt{\log n} / \sqrt{n} \rightarrow 0$, then Condition (5) is true, using the non decreasingness of $(\alpha_n^2)_{n \in \mathbb{N}}$ (see Remark 2.1). As a special case, one can take $b_n = (\log n)^\alpha$ with $\alpha > 0$.
- if $b_n = (n-1)^\alpha$ with $\alpha > 0$ then

$$\frac{b_n^2 \sqrt{\log n}}{\sqrt{n} \alpha_n^2} \underset{+\infty}{\sim} \frac{(\alpha+1)^2 (2\alpha+1) \sqrt{\log n}}{\alpha^2 \sqrt{n}} \rightarrow 0$$

(see Example 2.1). Thus, Condition (5) is satisfied.

3 Numerical experiments

3.1 Monte Carlo study of the estimators

Figure 1 shows three boxplots obtained from estimates of $a \in \{0.85, 0.9, 0.95\}$ for various sequences $(b_n)_{n \geq 1}$ based on 1000 simulated samples of

size $n = 50$. Here, the underlying renewal process is generated using independent inter-arrival times that follow a Weibull distribution with shape parameter 2 and scale parameter 10. These boxplots show that the convergence rate of \hat{a}_n heavily depends on b_n . This is consistent with the fact that in Section 2, we showed that for $b_n = n - 1$, \sqrt{n} or $\log n$, the convergence rate of \hat{a}_n is proportional to $n^{3/2}$, n or \sqrt{n} , respectively.

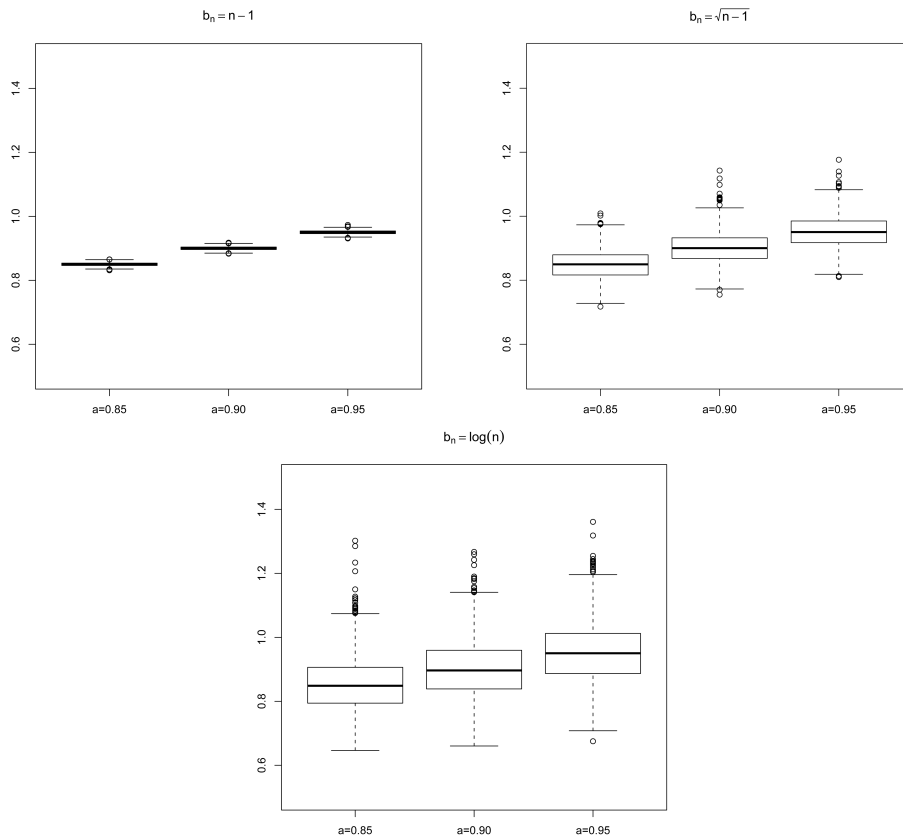


Figure 1: Comparison of boxplots of 1000 estimates of $a \in \{0.85, 0.9, 0.95\}$ obtained from samples of size 50 for $b_n = n - 1$, $\sqrt{n} - 1$ and $\log n$.

The estimator \hat{F}_n of F is based on the empirical distribution function obtained from the first n observations of the pseudo renewal process

$(\tilde{Y}_n)_{n \geq 1}$ defined by $\tilde{Y}_n = \hat{a}^{-b_n} X_n$. Figure 2 illustrates the uniform consistency result obtained in Proposition 2.4. The cumulative distribution function F (black solid line) is compared with 100 estimates \hat{F}_n (grey solid lines) for $n \in \{50, 100, 200, 400\}$.

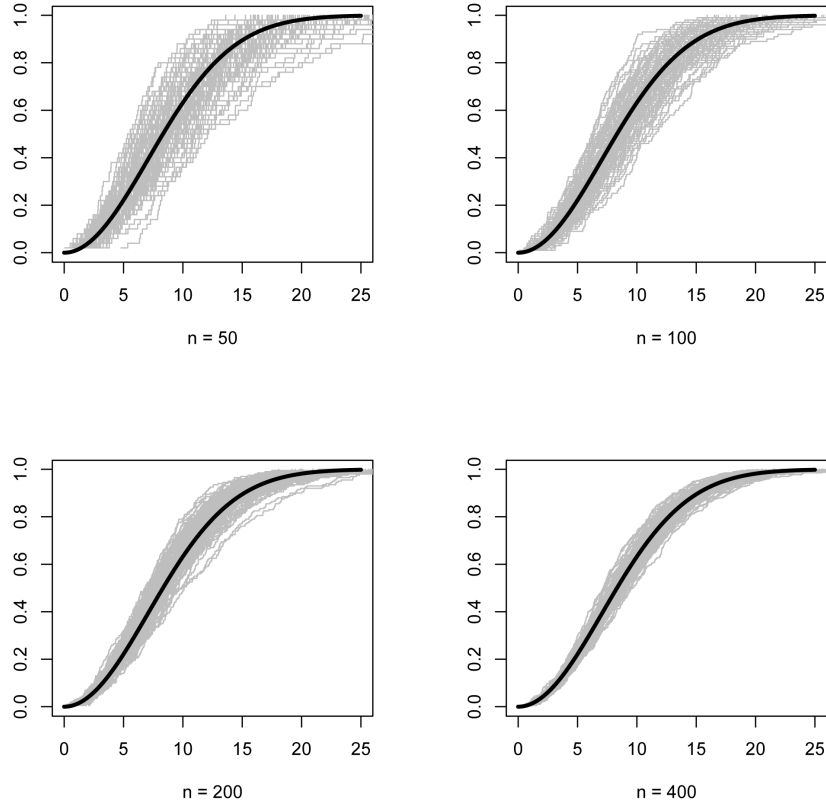


Figure 2: 100 estimates \hat{F}_n (grey solid lines) and the true F (black solid line) for various values of n .

To better illustrate the convergence of \hat{F}_n towards F , we now calculate the empirical mean of $N = 1000$ Mean Integrated Square Error (MISE) values. For one sample, the MISE equals

$$\frac{1}{n} \sum_{i=1}^n \left(\hat{F}_n(\tilde{Y}_{(i)}) - F(\tilde{Y}_{(i)}) \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(i/n - F(\tilde{Y}_{(i)}) \right)^2,$$

where $\tilde{Y}_i = \hat{a}_n^{-b_i} X_i$ for $1 \leq i \leq n$ and $\tilde{Y}_{(i)}$ is the i -th order statistic. F is the Weibull cdf with scale parameter 10 and shape parameter 2, and $a = 2$.

n	50	100	200	400
$b_n = \log n$	0.0252	0.0195	0.0119	0.0080
$b_n = \sqrt{n-1}$	0.0194	0.0106	0.0055	0.0028
$b_n = n-1$	0.0098	0.0054	0.0024	0.0013
$b_n = (n-1)^{3/2}$	0.0073	0.0039	0.0019	0.0010

Table 1: Mean of $N = 1000$ MISE values

3.2 On the choice of the b_n 's

We have assumed that the sequence $(b_n)_{n \geq 1}$ was known. A natural question hence is: how can we check the validity of the sequence $(b_n)_{n \geq 1}$? We here propose a residual analysis, based on the fact that, in case of a correct choice for b_n and of a "good" estimate \hat{a} of a , the residuals $k \mapsto \hat{a}^{-b_k} X_k$ should be nearly i.i.d.. Such residuals and the corresponding estimated cdf \hat{F}_n are plotted for different situations in Figures 3 and 4, with $b_n = n^{\theta_0}$. Such figures clearly illustrate the consequences of a bad choice for b_n .

Looking at the residuals can hence help to chose between several possible choices for b_n (between $b_n = n$, \sqrt{n} or $n^{3/2}$ in the previous examples). When the possible choices for b_n are unknown, another approach is required.

In case $b_n = g(n; \theta)$, where g is a known link function indexed by $\theta \in \Theta \in \mathbb{R}^p$, we can estimate θ in the following way. For $n \geq 1$, we have:

$$Z_n = \log X_n = g(n; \theta)\beta + \mu + e_n,$$

where $\beta = \log a$ and $\mu = \mathbb{E}[\log Y_1]$. Hence we can estimate μ , β and θ by minimizing the cost function c_n defined by

$$c_n(\mu, \beta, \theta) = \sum_{k=1}^n (Z_k - \beta g(k; \theta) - \mu)^2.$$

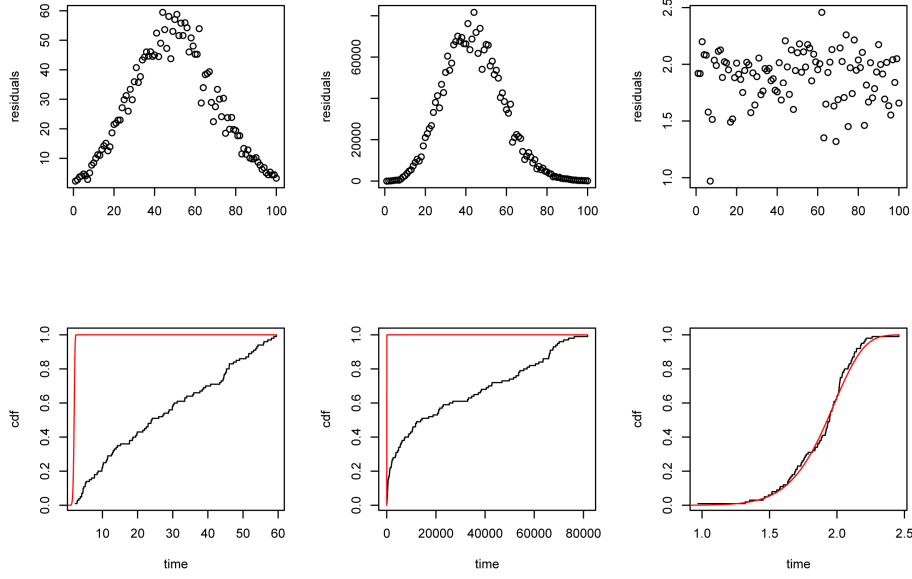


Figure 3: $n = 100$, $a = 0.98$, $\theta_0 = 1.5$, columns 1 to 3 correspond to $b_n = n$, $b_n = \sqrt{n}$ and $b_n = n^{3/2}$ (true) respectively. At the top are residuals $k \mapsto \hat{a}^{-b_k} X_k$ while at the bottom are both the estimated cdf of the renewal process (dotted) and the true cdf (solid).

It is easy to see that both optimal parameters $\mu_n(\theta)$ and $\beta_n(\theta)$ can be expressed as functions of θ , with:

$$\mu_n(\theta) = \frac{(\sum_{k=1}^n g(k; \theta)) (\sum_{k=1}^n y_k g(k; \theta)) - (\sum_{k=1}^n y_k) (\sum_{k=1}^n g^2(k; \theta))}{(\sum_{k=1}^n g(k; \theta))^2 - n (\sum_{k=1}^n g^2(k; \theta))}, \quad (7)$$

$$\beta_n(\theta) = \frac{(\sum_{k=1}^n g(k; \theta)) (\sum_{k=1}^n y_k) - n (\sum_{k=1}^n y_k g(k; \theta))}{(\sum_{k=1}^n g(k; \theta))^2 - n (\sum_{k=1}^n g^2(k; \theta))}. \quad (8)$$

Plugging these two functions into $c_n(\mu, \beta, \theta)$, we obtain a new cost function \mathcal{C}_n which only depends on θ :

$$\mathcal{C}_n(\theta) = \sum_{k=1}^n (Z_k - \beta_n(\theta)g(k; \theta) - \mu_n(\theta))^2.$$

We next minimize $\mathcal{C}_n(\theta)$ with respect to θ , which provides an estimate $\hat{\theta}_n$ for θ , and hence an estimate for b_n 's ($\hat{b}_n = g(n; \hat{\theta}_n)$).

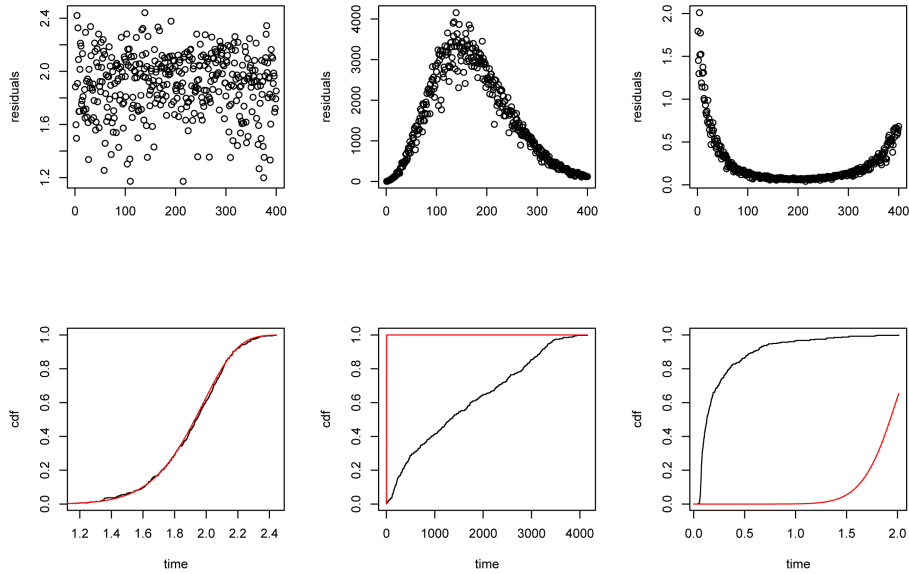


Figure 4: $n = 400$, $a = 0.95$, $\theta_0 = 1$, columns 1 to 3 correspond to $b_n = n$ (true), $b_n = \sqrt{n}$ and $b_n = n^{3/2}$ respectively. At the top are residuals $k \mapsto \hat{a}^{-b_k} X_k$ while at the bottom are both the estimated cdf of the renewal process (dotted) and the true cdf (plain).

This procedure is illustrated in Figures 5 and 6 for $g(k; \theta) = k^\theta$, which show its efficiency.

3.3 Aircraft data

We end this session with the study of a real data set of size $n = 29$. This data set contains successive times to failure (operating hours) of an air-conditioning equipment of a Boeing 720 aircraft and it is taken from data corresponding to 13 different aircraft. These data were studied in [21] and are available in [18].

Figure 7 shows the successive failure times (operating hours).

Optimizing the criterion $\theta \mapsto \mathcal{C}_n(\theta)$ for $b_n = (n - 1)^\theta$, we obtain $\hat{\theta} = 0.788$, see Fig. 8.

Table 2 summarizes the results obtained for the estimation of parameter a for various b_n 's. The estimate \hat{a} of a is given with a 95%

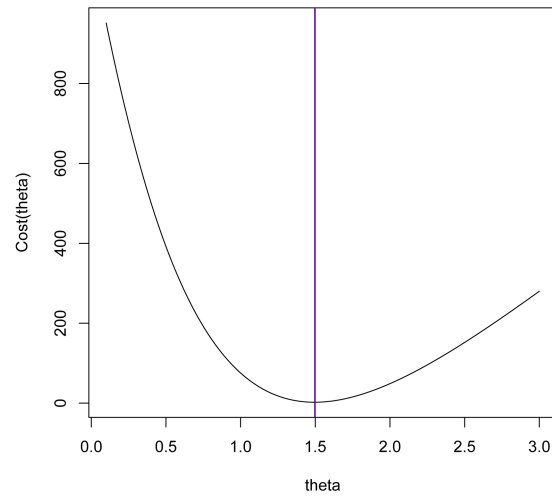


Figure 5: $n = 100$, $a = 0.98$, $\theta_0 = 1.5$, $\theta \mapsto \mathcal{C}_n(\theta)$ is the plain curve, θ_0 and $\hat{\theta}_n$ are superimposed vertical lines.

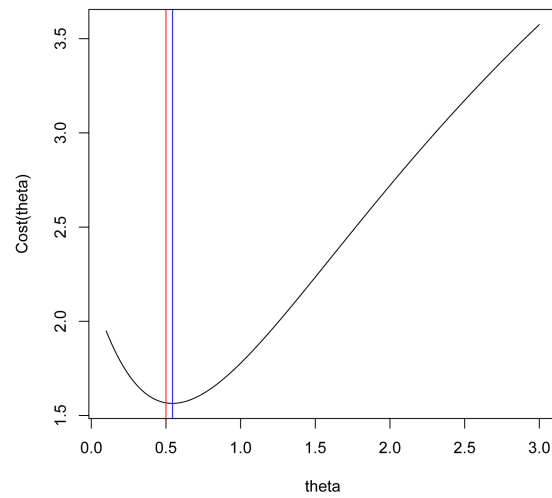


Figure 6: $n = 100$, $a = 0.90$, $\theta_0 = 0.5$, $\theta \mapsto \mathcal{C}_n(\theta)$ is the plain curve, θ_0 is the vertical plain line, $\hat{\theta}_n$ is the vertical dotted line.

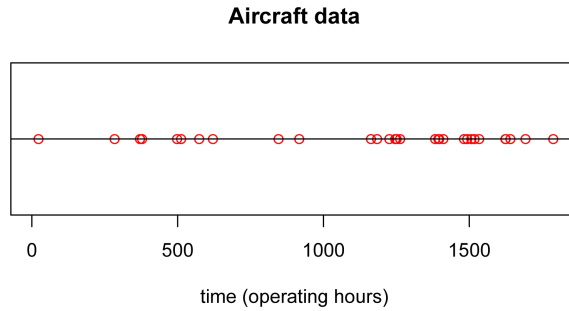


Figure 7: Successive failure times of the air-conditioning equipment.

asymptotic confidence interval $[\hat{a}_{\min}, \hat{a}_{\max}]$ which is computed via Proposition 2.3.

b_n	$(n - 1)^{0.788}$	$\log n$	$\sqrt{n - 1}$	$n - 1$	$(n - 1)^{3/2}$
\hat{a}	0.900	0.620	0.740	0.952	0.992
95% CI for a	[0.798,1.003]	[0.275,0.966]	[0.489,0.991]	[0.901,1.003]	[0.982,1.001]

Table 2: Estimates of a and various b_n 's for the aircraft data.

It is interesting to note that, whatever the choice for b_n , the estimate of a belongs to $(0, 1)$. This implies that the times between successive failures are stochastically non increasing. Note also that if we test $H_0 : a = 1$ by rejecting the hypothesis H_0 whenever the 95% Confidence Interval (CI) for a does not contain 1, then we do not reject H_0 when $b_n = (n - 1)^\theta$ with $\theta = 0.788, 1$ or 1.5 while this hypothesis is rejected when b_n is $\log n$ or $\sqrt{n - 1}$ (see Tab. 2). It however is highly likely that $a < 1$. Finally, Fig. 9 shows that the estimates of the cdf F also are sensitive to the choice of b_n : the further b_n is from the optimal sequence, the further the cdf estimates are from the empirical cdf obtained for $b_n = (n - 1)^{0.788}$.

4 Application to reliability

A repairable system is now considered, with instantaneous repairs at failure times and successive life-times modeled by an EGP. Once the process has been statistically estimated, it may be used for prediction purposes and/or optimization of replacement policies. As for prediction

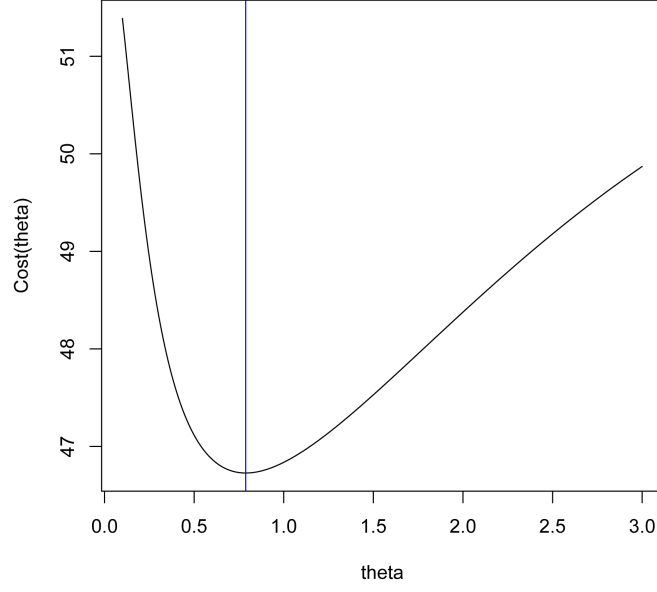


Figure 8: $\theta \mapsto C_n(\theta)$ for the aircraft data.

purpose, a typical quantity of interest is the mean number of failures on some time interval $[0, t]$. In case of non increasing lifetimes ($a \leq 1$), a replacement policy is next studied, where the system is renewed as soon as a lifetime is observed to be too short. We begin with some preliminary results.

4.1 Preliminary results

Lemma 4.1. *Setting $T_\infty = \lim_{n \rightarrow +\infty} T_n$, we have the following dichotomy:*

1. *If $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$, then $\mathbb{E}(T_\infty) < +\infty$ (and $T_\infty < +\infty$ a.s.).*
2. *If $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$, then $T_\infty = +\infty$ a.s. (and $\mathbb{E}(T_\infty) = +\infty$).*

Proof. In case $a \geq 1$ (which implies $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$), we clearly have: $T_n \geq S_n$, where $S_n = \sum_{j=1}^n Y_j$. As $S_\infty = +\infty$ a.s. (renewal case), we get $T_\infty = +\infty$ a.s..

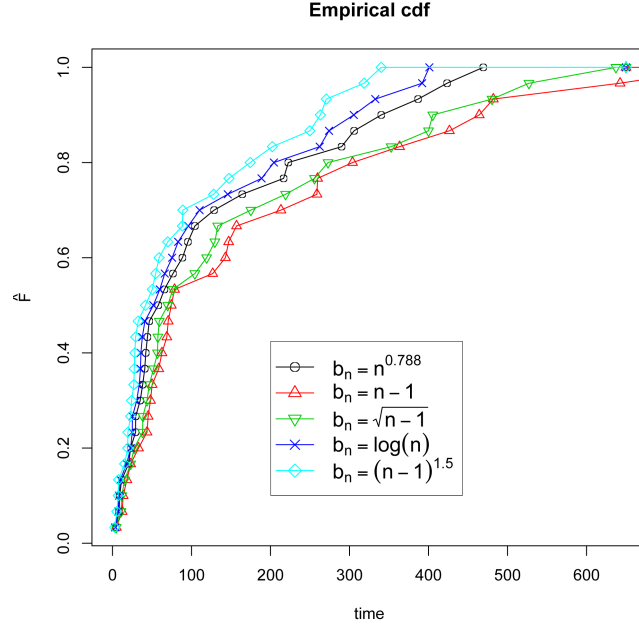


Figure 9: Empirical cumulative distribution function for the aircraft data.

Let us now assume $a \in (0, 1)$. If $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$, we easily derive the first point, due to

$$\mathbb{E}(T_n) = \sum_{i=1}^n \mathbb{E}(X_i) = \mathbb{E}(Y_1) \sum_{i=1}^n a^{b_i}. \quad (9)$$

As for the second point, let $c_n = \sum_{i=1}^n a^{b_i}$. As $c_n \geq na^{b_n}$, we have $a^{2b_n}/c_n^2 \leq 1/n^2$ and $\sum_{n=1}^{+\infty} \frac{a^{2b_n}}{c_n^2} < +\infty$. We derive that

$$\sum_{n=1}^{+\infty} \frac{\text{Var}(X_n)}{c_n^2} = \text{Var}(Y_1) \sum_{n=1}^{+\infty} \frac{a^{2b_n}}{c_n^2} < +\infty$$

and in case $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$, Theorem 6.7 from [20] implies that:

$$\frac{T_n - \mathbb{E}(T_n)}{c_n} = \frac{T_n}{c_n} - \mathbb{E}(Y_1) \rightarrow 0 \text{ a.s.}$$

so that $T_\infty = +\infty$ a.s.. ■

Remark 4.1. Such results extend similar results from [15] provided in the special case where $b_i = i - 1$.

We now look at an example.

Example 4.1. Let $b_n = n^\alpha (\log(n))^\beta$ with $\alpha \geq 0$ and $\beta \geq 0$, and let $a \in (0, 1)$. Then $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$ if and only if $\alpha = 0$ and one of the following conditions is fulfilled:

- $\beta < 1$,
- $\beta = 1$ and $a \geq 1/e$.

Proof. In case $\alpha > 0$, we have $0 \leq a^{b_n} = a^{n^\alpha (\log n)^\beta} \leq a^{n^\alpha}$ for all $n \geq 3$.

If $\alpha \geq 1$, then $0 \leq a^{b_n} \leq a^{n^\alpha} \leq a^n$, from which we derive that $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$.

If $0 < \alpha < 1$, we have:

$$\frac{a^{(n+1)^\alpha}}{a^{n^\alpha}} = 1 + \alpha \log(a) n^{\alpha-1} + o(n^{\alpha-1})$$

from where we derive that

$$\lim_{n \rightarrow +\infty} n \left(\frac{a^{(n+1)^\alpha}}{a^{n^\alpha}} - 1 \right) = \lim_{n \rightarrow +\infty} \alpha \log(a) n^\alpha = -\infty < -1.$$

This implies that $\sum_{n=1}^{+\infty} a^{n^\alpha} < +\infty$ using Raabe's rule, and hence $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$.

In case $(\alpha, \beta) = (0, 1)$, we have $a^{b_n} = n^{\log(a)}$, so that $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$ if and only if $a < 1/e$.

For $\alpha = 0$ and $\beta \neq 1$, the series $\sum_{i=1}^{+\infty} a^{b_i}$ has the same behavior as $\int_1^{+\infty} a^{(\log(u))^\beta} du$, with

$$\begin{aligned} \lim_{u \rightarrow +\infty} u^\theta a^{(\log(u))^\beta} &= \lim_{u \rightarrow +\infty} e^{((\log(u))^\beta - 1) \log(a) + \theta \log(u)} \\ &= \begin{cases} 0 & \text{if } \beta > 1, \\ +\infty & \text{if } \beta < 1, \end{cases} \end{aligned}$$

for all $\theta > 0$. We deduce that $\sum_{i=1}^{+\infty} a^{b_i} < +\infty$ if and only if $\beta > 1$, which completes this proof. ■

4.2 Mean number of failures

In order to get a "pseudo-renewal" equation for the "pseudo-renewal" function associated to the EGP, we here envision the case where the first interarrival time X_1 of the EGP is distributed as $X_k = a^{b_k} Y_k$, with $k \geq 1$. This means that at time $T_0 = 0$, the system has already been repaired $k - 1$ times. The successive interarrival times then are distributed as X_k, X_{k+1}, \dots . This situation is denoted by $\Phi_0 = k$.

For $k \geq 1$, we set \mathbb{P}_k to be the conditional probability measure given that $\Phi_0 = k$, with $k \geq 1$ and \mathbb{E}_k the associated conditional expectation. In case $k = 1$, we have: $\mathbb{P} = \mathbb{P}_1$ and $\mathbb{E} = \mathbb{E}_1$. For any interval $I \subset \mathbb{R}_+$, we also set $N(I)$ to be the number of failures (or arrivals of the EGP) on I , with

$$N(I) = \sum_{n=1}^{+\infty} \mathbf{1}_{\{T_n \in I\}}$$

In case $I = [0, t]$, we simply set: $N(t) = N([0, t])$.

Given that $\Phi_0 = k$, the "pseudo-renewal" function is

$$n_k(t) = \mathbb{E}_k(N(t)) = \sum_{n=1}^{+\infty} \mathbb{P}_k(T_n \leq t)$$

and $n_k(t)$ stands for the mean number of failures on $[0, t]$. In case $k = 1$, we set $n(t) = n_1(t)$.

A necessary condition for $n_k(t)$ to be finite for all $t \geq 0$ is $T_\infty = +\infty$ a.s. (see [6] in the more general set up of Markov renewal functions), which here writes $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$, see Lemma 4.1. We next provide a sufficient condition.

Proposition 4.1. *Assume $\mathbb{E}(Y_1) < +\infty$ and $\lim_{n \rightarrow +\infty} na^{b_n} > \frac{1}{\mathbb{E}(Y_1)}$ (and hence $\sum_{i=1}^{+\infty} a^{b_i} = +\infty$). Then $n_k(t) < +\infty$ for all $t \geq 0$ and all $k \geq 1$.*

Proof. In case $a \geq 1$, we have:

$$n_k(t) \leq n_1(t) = n(t) \leq U(t) < +\infty,$$

where $U(t)$ stands for the renewal function associated to the underlying renewal process.

In case $a \in (0, 1)$, let $t > 0$ and $k \geq 1$ be fixed. Due to the Markov inequality, we have:

$$n_k(t) = \sum_{n=1}^{+\infty} \mathbb{P}_k(e^{-T_n} \geq e^{-t}) \leq e^t \sum_{n=1}^{+\infty} u_{n,k}$$

with

$$u_{n,k} = \mathbb{E}_k(e^{-T_n}) = \prod_{i=k}^{k+n-1} \mathbb{E}(e^{-a^{b_i} Y_1})$$

and

$$\lim_{n \rightarrow +\infty} n \left(\frac{u_{n+1,k}}{u_{n,k}} - 1 \right) = - \lim_{n \rightarrow +\infty} n a^{b_{k+n}} \times \mathbb{E} \left(\frac{1 - e^{-a^{b_{k+n}} Y_1}}{a^{b_{k+n}}} \right)$$

As $\frac{1 - e^{-a^{b_{k+n}} Y_1}}{a^{b_{k+n}}}$ converges to Y_1 when $n \rightarrow +\infty$ and is bounded by Y_1 , we derive by Lebesgues's theorem that:

$$\lim_{n \rightarrow +\infty} n \left(\frac{u_{n+1,k}}{u_{n,k}} - 1 \right) = - \lim_{n \rightarrow +\infty} n a^{b_n} \times \mathbb{E}(Y_1) < -1$$

by assumption. We conclude with Raabe's rule. \blacksquare

Example 4.2. For $b_n = (\log(n))^\beta$ with $\beta \geq 0$ and $a \in (0, 1)$, we get that $n_k(t)$ is finite for all $t \geq 0$ and all $k \geq 1$ as soon as one of the following condition is fulfilled:

- $\beta < 1$,
- $\beta = 1$ and $a > \frac{1}{e}$,
- $\beta = 1$, $a = \frac{1}{e}$, and $\mathbb{E}(Y_1) > 1$.

Such results show that, contrary to classical geometric processes (see [5] and the introduction), it is possible to model decreasing successive lifetimes with extended geometric processes and get a finite expected number of counts at an arbitrary time.

Proposition 4.2. Assume that $\lim_{n \rightarrow +\infty} n a^{b_n} > \frac{1}{\mathbb{E}(Y_1)}$. The function n_k fulfills the following pseudo-renewal equation:

$$n_k = F_k + f_k * n_{k+1} \tag{10}$$

for all $k \geq 1$, where F_k (resp. f_k) stands for the cumulative (resp. probability) distribution function of X_k .

Proof. Using classical arguments ([6] e.g.), we have:

$$\begin{aligned} n_k(t) &= \mathbb{E}_k(N(t) \mathbf{1}_{\{X_1 \leq t\}}) \\ &= \mathbb{E}_k(\mathbb{E}_k(N(t) | X_1) \mathbf{1}_{\{X_1 \leq t\}}) \\ &= \mathbb{E}_k(\mathbb{E}_k(N(]0, X_1]) | X_1) \mathbf{1}_{\{X_1 \leq t\}}) + \mathbb{E}_k(\mathbb{E}_k(N(]X_1, t]) | X_1) \mathbf{1}_{\{X_1 \leq t\}}) \\ &= F_k(t) + \int_{[0,t]} n_{k+1}(t-u) f_k(u) du, \end{aligned}$$

which may be written as (10). ■

Remark 4.2. Setting $\Phi_{T_n} = k$ in case X_{n+1} is distributed as $a^{b_k} Y_k$ (with $k \geq n + 1$) and $\Phi_t = \Phi_{T_n}$ for $T_n \leq t < T_{n+1}$, the process $(\Phi_t)_{t \geq 0}$ then appears as a semi-Markov process with semi-Markov kernel provided by

$$q(i, j, dx) = \mathbf{1}_{\{j=i+1\}} dF_i(x).$$

Equation (10) then is the Markov renewal equation satisfied by the corresponding Markov renewal function.

We now provide practical tools for the numerical assessment of the pseudo (Markov) renewal function $n_k(t)$.

Corollary 4.1. *Assume $a \geq 1$. Setting*

$$u_n(t) = \mathbb{P}(T_n \leq t),$$

for all $n \geq 1$, we have:

$$0 \leq \frac{n(t) - \sum_{n=1}^N u_n(t)}{n(t)} \leq u_N(t), \tag{11}$$

for all $N \geq 1$. Also, $(u_n(t))_{n \geq 1}$ may be computed recursively using

$$\begin{aligned} u_1(t) &= F(t) \\ u_{n+1}(t) &= (f_{n+1} * u_n)(t) = \frac{1}{a^{b_{n+1}}} \int_0^t u_n(u) f\left(\frac{t-u}{a^{b_{n+1}}}\right) du \end{aligned} \tag{12}$$

for all $n \geq 1$, where F (resp. f) stands for the cumulative (resp. probability) distribution function of Y_1 .

Proof. We may write:

$$n(t) = \sum_{n=1}^N u_n(t) + \varepsilon_N(t)$$

where

$$\varepsilon_N(t) = \sum_{m=1}^{+\infty} \mathbb{P}(T_{m+N} \leq t).$$

Using similar arguments as [11], we have

$$\{T_{m+N} \leq t\} = \{T_N + (T_{m+N} - T_N) \leq t\} \subset \{T_N \leq t\} \cap \{T_{m+N} - T_N \leq t\}$$

where T_N and $T_{m+N} - T_N$ are independent. We derive:

$$\varepsilon_N(t) \leq \mathbb{P}(T_N \leq t) \sum_{m=1}^{+\infty} \mathbb{P}_N(T_m \leq t) = u_N(t) n_N(t) \leq u_N(t) n(t),$$

which implies (11). The remainder of the proof is straightforward. ■

Remark 4.3. This result allows to numerically assess the pseudo renewal function $n(t)$ up to a given precision $\varepsilon > 0$ by recursively computing $u_n(t)$ until $u_n(t)$ is smaller than ε . Note however that the $u_i(t)$'s are computed using discrete convolutions in (12), which induces numerical errors. Such errors might be quantified using similar methods as in [19].

In case $a < 1$, the previous result is not valid because $n_N(t) \geq n(t)$. In that case, Monte-Carlo simulations may be used to compute the pseudo-renewal function. A lower bound $n^c(t)$ may also be provided, which converges to $n(t)$ when c goes to zero. This bound is constructed via the following lemma.

Lemma 4.2. For $c > 0$ and $t \geq 0$, let

$$\tau^c = \inf(n \geq 1 : X_n < c) \tag{13}$$

and

$$n^c(t) = \mathbb{E} \left(\sum_{n=1}^{\tau^c-1} \mathbf{1}_{\{T_n \leq t\}} \right)$$

(0 in case of an empty sum).

Then $n^c(t) \leq n(t)$ and

$$\lim_{c \rightarrow 0^+} n^c(t) = n(t).$$

Proof. Using the fact that τ^c increases to infinity when c decreases to 0^+ , the result is a direct consequence of the monotone convergence theorem. ■

The following lemma provides tools for the numerical assessment of $n^c(t)$, which do not require $a \geq 1$.

Lemma 4.3. *Setting*

$$u_n^c(t) = \mathbb{P}(T_n \leq t, X_1 \geq c, \dots, X_n \geq c)$$

for all $n \geq 1$, we have:

$$n^c(t) = \sum_{n=1}^{\lfloor \frac{t}{c} \rfloor} u_n^c(t), \quad (14)$$

where $\lfloor \dots \rfloor$ stands for the floor function. Also, $(u_n^c(t))_{n \geq 1}$ may be computed recursively using

$$\begin{aligned} u_1^c(t) &= (F(t) - F(c))^+ \\ u_{n+1}^c(t) &= \frac{1}{a^{b_{n+1}}} \int_0^{(t-c)^+} u_n^c(u) f\left(\frac{t-u}{a^{b_{n+1}}}\right) du \end{aligned} \quad (15)$$

for all $n \geq 1$.

Proof. We have:

$$n^c(t) = \sum_{n=1}^{+\infty} \mathbb{P}(T_n \leq t, n < \tau^c) = \sum_{n=1}^{+\infty} \mathbb{P}(T_n \leq t, X_1 \geq c, \dots, X_n \geq c).$$

Noting that $X_1 \geq c, \dots, X_n \geq c$ implies $T_n \geq nc$, the summation may be restricted to $n \leq \lfloor \frac{t}{c} \rfloor$, which provides (14).

Equation (15) is a direct consequence of

$$\begin{aligned} u_{n+1}^c(t) &= \mathbb{E}\left(\mathbb{E}\left(\mathbf{1}_{\{T_n \leq t - X_{n+1}\}} \mathbf{1}_{\{X_1 \geq c, \dots, X_n \geq c\}} | X_{n+1}\right) \mathbf{1}_{\{X_{n+1} \geq c\}}\right) \\ &= \mathbb{E}\left(u_n^c(t - X_{n+1}) \mathbf{1}_{\{X_{n+1} \geq c\}}\right) \\ &= \frac{1}{a^{b_{n+1}}} \int_c^t u_n^c(t - u) f\left(\frac{u}{a^{b_{n+1}}}\right) du \end{aligned}$$

for all $t \geq c$. ■

4.3 A replacement policy

We here consider the case where $a < 1$ and the following renewal policy is considered: as soon as a lifetime X_i is observed to be shorter than the predefined threshold s ($s > 0$), the system is instantaneously replaced at some cost c_R . Between replacements, the cost of an instantaneous repair which follows a failure is denoted by c_F , with $c_R \geq c_F$. We set $c(s)$ to be the asymptotic unitary cost per time unit time.

The next proposition uses classical results from renewal theory to derive the existence of $c(s)$, and an expression for it.

Proposition 4.3. *Assume $a \in (0, 1)$. Setting $C([0, t])$ to be the cumulated cost on $[0, t]$, the asymptotic cost per unit time*

$$c(s) = \lim_{t \rightarrow +\infty} \frac{C([0, t])}{t} \text{ a.s.} \tag{16}$$

exists and is provided by

$$c(s) = \frac{c_R + c_F \mathbb{E}(\tau^s - 1)}{\mathbb{E}(T_{\tau^s})}, \tag{17}$$

where τ^s is defined as τ^c , see (13).

Furthermore,

$$\begin{aligned} \mathbb{E}(\tau^s - 1) &= \sum_{k=1}^{+\infty} v_k^s \\ \mathbb{E}(T_{\tau^s}) &= \mathbb{E}(Y_1) \left(1 + \sum_{k=1}^{+\infty} a^{b_{k+1}} v_k^s \right) \end{aligned}$$

with

$$v_k^s = \prod_{i=1}^k \bar{F} \left(\frac{s}{a^{b_i}} \right) \tag{18}$$

for all $k \geq 1$ and $\bar{F} = 1 - F$.

Proof. The evolution of the maintained system may be described by a regenerative process, with cycles delimited by the replacement of the

system and generic length T_{τ^s} . Moreover

$$\begin{aligned}\mathbb{E}(T_{\tau^s}) &= \sum_{k=2}^{+\infty} \mathbb{E}(T_k (\mathbf{1}_{\{\tau^s \geq k\}} - \mathbf{1}_{\{\tau^s \geq k+1\}})) + \mathbb{E}(T_1 \mathbf{1}_{\{\tau^s=1\}}) \\ &= \sum_{k=3}^{+\infty} \mathbb{E}((T_k - T_{k-1}) \mathbf{1}_{\{\tau^s \geq k\}}) + \mathbb{E}(T_2 \mathbf{1}_{\{\tau^s \geq 2\}}) + \mathbb{E}(X_1 \mathbf{1}_{\{\tau^s=1\}}) \\ &= \sum_{k=1}^{+\infty} w_k^s,\end{aligned}$$

with

$$\begin{aligned}w_k^s &= \mathbb{E}(X_k \mathbf{1}_{\{\tau^s \geq k\}}) \\ &= a^{b_k} \mathbb{E}(Y_k) \mathbb{P}(X_1 \geq s, \dots, X_{k-1} \geq s) \\ &= a^{b_k} \mathbb{E}(Y_1) v_{k-1}^s,\end{aligned}$$

for all $k \geq 2$ and $w_1^s = \mathbb{E}(Y_1)$.

Now, as

$$\lim_{k \rightarrow +\infty} \frac{w_{k+1}^s}{w_k^s} = \lim_{k \rightarrow +\infty} a^{b_{k+1}-b_k} \bar{F}\left(\frac{s}{a^{b_k}}\right) = 0,$$

the series with generic term w_k^s is convergent and $\mathbb{E}(T_{\tau^s}) < +\infty$.

We derive the existence of $c(s)$ and formula (17) (see [1] e.g.), noting that the mean cost on a generic cycle is $c_R + c_F \mathbb{E}(\tau^s - 1)$.

The quantity $\mathbb{E}(\tau^s)$ may finally be computed via:

$$\mathbb{E}(\tau^s) = \sum_{i=1}^{+\infty} \mathbb{P}(\tau^s \geq i) = 1 + \sum_{i=2}^{+\infty} \mathbb{P}(X_1 \geq s, \dots, X_{i-1} \geq s) = 1 + \sum_{i=2}^{+\infty} v_{i-1}^s.$$

■

We next provide tools for the numerical assessment of $c(s)$.

Proposition 4.4. *Assume $a \in (0, 1)$. We have the following bounds for $c(s)$:*

$$m_c^N(s) \leq c(s) \leq M_c^N(s),$$

where

$$\begin{aligned}m_c^N(s) &= \frac{c_R + c_F S_1^N(s)}{\mathbb{E}(Y_1) \left(1 + S_2^N(s) + a^{b_{N+2}} v_{N+1}^s / F\left(\frac{s}{a^{b_{N+2}}}\right)\right)}, \\ M_c^N(s) &= \frac{c_R + c_F \left(S_1^N(s) + v_{N+1}^s / F\left(\frac{s}{a^{b_{N+2}}}\right)\right)}{\mathbb{E}(Y_1) (1 + S_2^N(s))},\end{aligned}$$

and

$$S_1^N(s) = \sum_{k=1}^N v_k^s,$$

$$S_2^N(s) = \sum_{k=1}^N a^{b_{k+1}} v_k^s$$

(with v_k^s defined by (18)). Moreover we have

$$\left| c(s) - \frac{m_c^N(s) + M_c^N(s)}{2} \right| \leq \Delta c_{\max}^N(s) := \frac{M_c^N(s) - m_c^N(s)}{2}.$$

Proof. We have

$$1 + \sum_{k=1}^N a^{b_{k+1}} v_k^s \leq \frac{\mathbb{E}(T_{\tau^s})}{\mathbb{E}(Y_1)} \leq 1 + \sum_{k=1}^N a^{b_{k+1}} v_k^s + \sum_{k=N+1}^{+\infty} a^{b_{k+1}} v_k^s,$$

with

$$\sum_{k=N+1}^{+\infty} a^{b_{k+1}} v_k^s \leq a^{b_{N+2}} \sum_{k=N+1}^{+\infty} v_k^s,$$

and

$$\sum_{k=N+1}^{+\infty} v_k^s \leq v_{N+1}^s \times \sum_{k=N+1}^{+\infty} \left(\bar{F} \left(\frac{s}{a^{b_{N+2}}} \right) \right)^{k-N-1} = \frac{v_{N+1}^s}{F \left(\frac{s}{a^{b_{N+2}}} \right)}.$$

We derive

$$\mathbb{E}(Y_1) (1 + S_2^N(s)) \leq \mathbb{E}(T_{\tau^s}) \leq \mathbb{E}(Y_1) \left(1 + S_2^N(s) + \frac{a^{b_{N+2}} v_{N+1}^s}{F \left(\frac{s}{a^{b_{N+2}}} \right)} \right).$$

A similar method is used for bounding $\mathbb{E}(\tau^s)$, which provides the result. ■

This proposition allows to numerically assess the cost function $c(s)$ up to a given precision ε by recursively computing $S_1^N(s)$ and $S_2^N(s)$ until $\Delta c_{\max}^N(s)$ is smaller than ε .

4.4 Numerical experiments

4.4.1 Computation of the pseudo-renewal function

We first consider the case where $a \geq 1$. The random variable Y_1 is Gamma distributed with shape parameter 1.2 and scale parameter 2.5 (which provides $\mathbb{E}(Y_1) = 3, \text{Var}(Y_1) = 7.5$). This distribution is denoted by $\Gamma(1.2, 2.5)$. We also take $b_n = n^{0.3}$ and $a = 1.2$. The approximation of the pseudo-renewal function $n(t)$ provided by Corollary 4.1 is plotted against t in Figure 10 for $N = 20$. The maximal relative error provided by the approximation is about 4.2×10^{-6} . We also plot $n(t)$ computed by Monte-Carlo simulations and the 95% confidence band for 10^3 trajectories in the same figure. The results are quite similar.

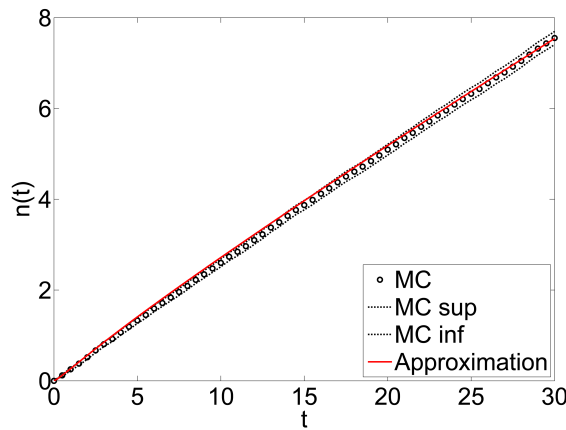


Figure 10: $n(t)$ with respect to t by Monte-Carlo simulations (MC) and by the approximation provided by Corollary 4.1.

We next consider the case where $a < 1$ (and $\lim_{n \rightarrow +\infty} na^{b_n} > \frac{1}{\mathbb{E}(Y_1)}$): the random variable Y_1 follows $\Gamma(2.5, 1)$ with $\mathbb{E}(Y_1) = \text{Var}(Y_1) = 2.5$, $b_n = (\log n)^{0.7}$ and $a = 0.8$. The lower bound $n^c(t)$ for $n(t)$ is computed via the results of Lemma 4.3 for different values of c ($c = 0.05, c = 0.1, c = 0.25, c = 0.5$). The results are displayed in Figure 11. As expected (see Lemma 4.2), $n^c(t)$ is stabilizing when c goes to zero and the values for $c = 0.05$ and $c = 0.1$ are nearly super-imposed. We also plot $n(t)$ computed by Monte-Carlo simulations and the 95% confidence band for

10^3 trajectories in Figure 12, as well as $n^c(t)$ for $c = 0.05$. We observe that $n^c(t)$ is a good approximation of $n(t)$ for small c .

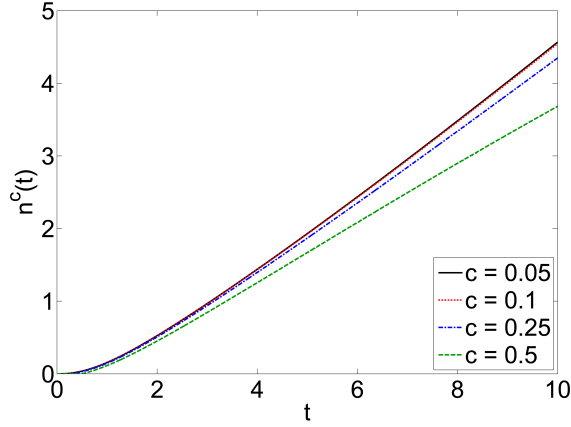


Figure 11: $n^c(t)$ with respect to t for different values of c .

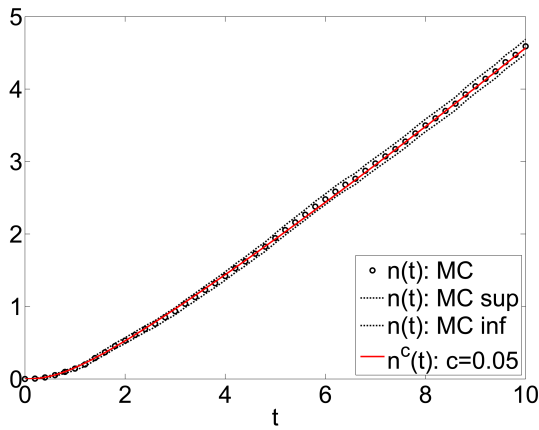


Figure 12: $n(t)$ by Monte-Carlo simulations and $n^c(t)$ for $c = 0.05$.

4.4.2 The replacement policy

The random variable Y_1 follows $\Gamma(2.5, 1)$ with $\mathbb{E}(Y_1) = \text{Var}(Y_1) = 2.5$, $b_n = (\log n)^{0.7}$, $a = 0.8$, $c_R = 1$ and $c_F = 0.5$.

For $N = 100$, the maximal absolute error $\Delta c_{\max}^N(s)$ decreases very quickly as s increases ($\Delta c_{\max}^N(0.4) \simeq 8 \times 10^{-5}$, $\Delta c_{\max}^N(0.7) \simeq 3 \times 10^{-12}$, beyond the machine precision for $s \geq 0.9$). The cost function $c(s)$ is plotted against s in Figure 13. The cost function reaches its minimum at $s^{opt} \simeq 1.70$, with $\min_{s>0} c(s) = c(s^{opt}) \simeq 0.17$.

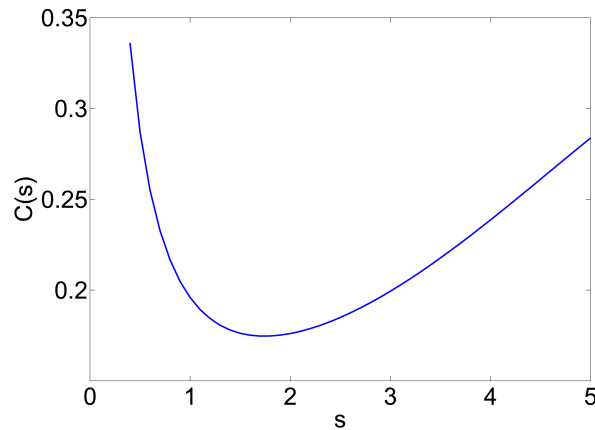


Figure 13: $c(s)$ with respect to s .

5 Concluding Remarks and Prospects

Contrary to renewal processes, geometric processes proposed by [17] and their present extension both allow successive inter-arrival times to be (stochastically) increasing or decreasing. From a modelling point of view, the extended version has however been seen to be more flexible. Also, in an applied context, the expected number of arrivals of the underlying counting process on some finite time interval is expected to be finite at any time. This had previously been seen by [5] to be incompatible with a decreasing geometric process. In contrast, GP's extended geometric processes do not suffer from this drawback. Extended geometric processes may hence be a simple alternative to the virtual age models proposed by [9] and [13] for the modeling of imperfect maintenance actions e.g..

From the estimation point of view, we saw that the convergence rate of the estimator of the Euclidean parameter a strongly depends on

the sequence $(b_n)_{n \geq 1}$. A miss-specification of the sequence $(b_n)_{n \geq 1}$ will naturally lead to biased estimates. To make the model more flexible, we hence considered a parametrized version of the sequence $(b_n)_{n \geq 1}$ by setting $b_n = g(n, \theta)$, where θ is an additional Euclidean parameter. Some procedure has been provided for its estimation.

Note the lack of a central limit theorem for the estimator \hat{F} of the underlying cumulative distribution function F . Indeed, standard methods cannot be used here, because of the deterministic nature of the b_n 's. This problem hence requires some more investigation along with the study of the properties of the estimator of θ for parametrized sequences $b_n = g(n; \theta)$. Such a result would however be useful for testing the hypothesis that the underlying cumulative distribution function F belongs to some parametric family. Another possible issue would be to include covariates in this model in order to describe (e.g.) the effect of the environment on the monotonicity of the EGP.

In case $a < 1$, a lower bound has been provided for the pseudo-renewal function, which is easy to compute using Lemma 4.3. However, we haven't been able to provide a computable upper bound, although it is necessary for the numerical assessment of the results precision. Indeed, the usual tools such as those used in case $a \geq 1$ are inappropriate here, and new tools should be developed. As for the replacement policy, because of the random character of the successive lifetimes, an alternate policy based on a predefined number m of consecutive lifetimes under a threshold s , might be better adapted than the present policy, based on a replacement at the first observation of a single lifetime below s .

References

- [1] Asmussen, S. (2003), Applied probability and queues. Applications of Mathematics, **51**, Second ed., New York: Springer-Verlag.
- [2] Barlow, R. E. and Proschan, F. (1996), Mathematical theory of reliability. Classics in Applied Mathematics, vol. 17, fifth ed., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1965, With contributions by Larry C. Hunter.
- [3] Bai, Z. D., Cheng, P. E., and Zhang, C. H. (1997), An extension of Hardy-Littlewood strong law. *Statistica Sinica*, 923–928.

- [4] Bai, Z. D., Cheng, P. E. (2000), Marcinkiewicz strong laws for linear statistics. *Statistics & Probability Letters*, **46**, 105–112.
- [5] Braun, W. J., Li, W., Zhao, Y. Q. (2005), Properties of the geometric and related processes. *Naval Research Logistics*, **52**, 607–616.
- [6] Çinlar, E. (1975), *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall.
- [7] Cook, R. J. and Lawless, J. F. (2007), *The Statistical Analysis of Recurrent Events*. Springer.
- [8] Cuzick, J. (1995), A strong law for large for weighted sums of I.I.D. random variables. *Journal of Theoretical Probability*, **8**(3), 625–641.
- [9] Doyen, L. and Gaudoin, O. (2004), Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety* **84**(1), 45–56.
- [10] Finkelstein, M. (2010), A note on converging geometric-type processes. *Journal of Applied Probability*, **47**, 601–607.
- [11] Feller, W. (1968), *An introduction to probability theory and its applications*. Vol. 1. New York: John Wiley & Sons Inc.
- [12] Feller, W. (1971), *An introduction to probability theory and its applications*. Vol. 2. New York: John Wiley & Sons Inc.
- [13] Kijima, M. (2002), Stochastic models in reliability and maintenance. Osaki, Shunji ed., ch. *Generalized Renewal Processes and General Repair Models*, 145–164, Berlin: Springer.
- [14] Lam, Y. (1992), Nonparametric inference for geometric processes. *Communications in Statistics-Theory and Methods*, **21**, 2083–2105.
- [15] Lam, Y., Zheng, Y. H., and Zhang, Y. L. (2003), Some limit theorems in geometric processes. *Acta Mathematicae Applicatae Sinica*, **19**, 1–12.
- [16] Lam, Y., Zhu, L. X., Chan, S. K. and Liu, Q. (2004), Analysis of data from a series of events by a geometric process model. *Acta Mathematicae Applicatae Sinica*, **20**, 263–282.

- [17] Lam, Y. (2007), *The Geometric Process and its Applications*. World Scientific.
- [18] Lindsey, J. K. (2004), *Statistical Analysis of Stochastic Processes in Time*, Cambridge University Press, Cambridge.
- [19] Mercier, S. (2008), Numerical bounds for semi-markovian quantities and application to reliability. *Methodology and Computing in Applied Probability*, **10**(2), 179–198.
- [20] Petrov, V. V. (1995), *Limit Theorems of Probability Theory: Sequences of Independent Random Variables*. Oxford University Press.
- [21] Proschan, F. (1963), Theoretical explanation of observed decreasing failure rate. *Technometrics*, **5**, 375–383.
- [22] Van der Vaart, A. W. (1998), *Asymptotic Statistics*. Cambridge: Cambridge University Press.