Author(s): Gasbarra, Dario & Arjas, Elja & Vehtari, Aki & Slama, Rémy & Keiding, Niels

Title: The current duration design for estimating the time to pregnancy distribution: a nonparametric Bayesian perspective

Year: 2015

Version: Post print

Please cite the original version:

Gasbarra, Dario & Arjas, Elja & Vehtari, Aki & Slama, Rémy & Keiding, Niels. 2015. The current duration design for estimating the time to pregnancy distribution: a nonparametric Bayesian perspective. Lifetime Data Analysis. Volume 21, Issue 4. 594-625. ISSN 1380-7870 (printed). DOI: 10.1007/s10985-015-9333-0.

Noname manuscript No.
(will be inserted by the editor)

The current duration design for estimating the time to pregnancy distribution: a nonparametric Bayesian perspective

Dario Gasbarra · Elja Arjas · Aki Vehtari ·
Rémy Slama · Niels Keiding

Received: date / Accepted: date

Abstract This paper was inspired by the studies of Niels Keiding and co-authors on estimating the waiting time-to-pregnancy (TTP) distribution, and in particular on using the current duration design in that context. In this design, a cross-sectional sample of women is collected from those who are currently attempting to become pregnant, and then by recording from each the time she has been attempting. Our aim here is to study the identifiability and the estimation of the waiting time distribution on the basis of current duration data. The main difficulty in this stems from the fact that very short waiting times are only rarely selected into the sample of current durations, and this renders their estimation unstable. We introduce here a Bayesian method for this estimation problem, prove its asymptotic consistency, and compare the method to some variants of the non-parametric maximum likelihood estimators (NPMLE), which have been used previously in this context. The properties of the Bayesian estimation method are studied also empirically, using both simulated data and TTP data on current durations collected by Slama et al. (2012).

1 Introduction

The time it takes for a couple from initiating attempts to become pregnant until conception leading to detected pregnancy, time-to-pregnancy, or TTP, is a key measure of

Dario Gasbarra E-mail: Dario.Gasbarra@helsinki.fi · Elja Arjas E-mail: Elja.Arjas@helsinki.fi
Department of Mathematics and Statistics, University of Helsinki,

Aki Vehtari E-mail: Aki.Vehtari@aalto.fi
Department of Computer Science, Aalto University

Rémy Slama E-mail: remy.slama@ujf-grenoble.fr
Inserm, French Institute of Health and Medical Research, Team of Environmental Epidemiology applied to Reproduction and Respiratory Health, Inserm-Univ. Grenoble Alpes

Niels Keiding E-mail: nike@sund.ku.dk
Department of Public Health, University of Copenhagen

natural fecundity (Baird et al., 1986). The two most obvious study designs for statistical inference on TTP are cohort (follow-up) study, where couples are followed-up in time from when they start attempting to become pregnant, and retrospective study of pregnant women where couples are retrospectively interviewed about the TTP of their past attempt(s) at pregnancy ("pregnancy-based design"). A third design, which has recently received attention (Weinberg and Gladen, 1986; Keiding et al., 2002, 2012; Slama et al., 2012), is to collect information on current durations: collect from a cross-sectional sample of women (couples) those that are currently attempting to become pregnant, and then obtain from each of these the time they have been attempting. The distribution of these times identifies the distribution of actually realized waiting times, either to pregnancy or to unsuccessful end of attempt, in the population. Other applications of the current duration approach include migration patterns, where Yamaguchi (2003) estimated length of stay in a house from current durations, and duration of psychiatric disorders McLaughlin et al. (2010). The purpose of the present paper is to explore the possibility of applying the Bayesian approach to the estimation problems connected to the current duration design, and to compare it to some variants of the non-parametric maximum likelihood estimator (NPMLE) which has previously been considered in this context by Keiding et al. (2012). These questions are studied both empirically, using data on current durations of unprotected intercourse from 867 French women collected by Slama et al. (2012), and theoretically, considering the asymptotic consistency properties of the estimators.

2 The current duration design: the main concepts

To set up the necessary notation, let

$Y =$ time elapsed from the start of current attempt to sampling = current duration;
$T =$ time (from the start of an attempt) to pregnancy;
$U =$ time (from the start of an attempt) to giving up;
$X = T \wedge U =$ duration $=$ time from the start to the end of an attempt.

Furthermore, let

$$H(x) = P(X \leq x) \quad \text{and} \quad S(x) = P(X > x) = 1 - H(x) \tag{1}$$

be the CDF and the survival function of the distribution of X, respectively, and $\mu = E(X)$. It is assumed throughout the paper that $E(X) < \infty$. By observing that the current duration takes the value $Y = y$, one has actually come to know that $X > y$. The data do not tell how the observed current duration in each case is going to end, that is, whether in pregnancy or in giving up. Therefore, from the perspective of durations X, current duration data are always right censored. On the other hand, the Y variables are not censored in the original data.

As in Keiding et al. (2002, 2012), statistical inference on the distribution of X, from current duration data $\{y_i : i = 1,\dots,n\}$, is here based on the key formula

$$f(y) = \frac{S(y)}{E(X)}, \quad y > 0\,, \tag{2}$$

between the density function f for the current duration variables Y and the survival function S for the duration variables X. Since clearly $S(0) = 1$, we have also the link $f(0) = 1/E(X)$.

This key formula is based on an adaptation of a modeling idea originally due to Brillinger (1986), in which initiations of pregnancies are viewed as being generated from the background population according to a Poisson process. If the Poisson process is homogeneous with rate μ, durations of length in $(x, x+dx)$ form a thinned Poisson process with rate $\mu H(dx)$. The result then follows by interpreting the current duration Y as a backward recurrence time from any fixed time point chosen independently of the generation process. The same approach has been earlier used in other similar contexts in Keiding (1991).

A slightly different way of arriving at this same result is to consider the problem from the perspective of length biased sampling. Suppose that individuals/couples are drawn from the background population to the sample independently and so that, for a couple whose attempt to become pregnant has duration X, has a chance of being included in the sample which is proportional to X, and that, moreover, the relative position of the time point at which the sampling is made on that interval is uniformly distributed:

$$Y|X \sim \text{Unif}([0,X])\,.$$

Then it is easy to see (Cox and Miller, 1965; Zelen, 2004; Van Es et al., 2000) that the link formula (2) holds: Let I = indicator of being included in the sample. Then

$$P(Y \in dy | I = 1) = \frac{P(Y \in dy, I = 1)}{P(I = 1)}\,,$$

where

$$P(I = 1) = E\big(P(I = 1|X)\big) = \int_0^\infty cx\, H(dx) = cE(X)\,,$$

for some suitable constant c , and

$$\begin{aligned} P(Y \in dy, I = 1) &= E\big(P(Y \in dy, I = 1|X)\big) = \int_0^\infty \mathbf{1}(y \le x)\frac{dy}{x}\, cx\, H(dx) \\ &= c\left(\int_y^\infty H(dx)\right) dy = c\, S(y)\, dy\,. \end{aligned}$$

As both these derivations show, we may think that result (2) is an approximation of reality, ignoring, for example, effects of seasonal variation on the initiation process, and of possible earlier pregnancies by the same woman.

On the other hand, in a situation in which there is no information that would enable distinguishing between individual women, and assuming random sampling from an infinite population, the observations $(y_i : i = 1,...,n)$ are therefore infinitely exchangeable from the point of view of a statistician examining these data. By the De Finetti representation theorem, this means that $(y_i : i = 1,...,n)$ are conditionally independent given the survival function S which is viewed as a common parameter, and in this sense the result (2) is exact. We close this section by some additional remarks.

One may argue that in the definition $X = T \wedge U$ the variables T and U do not exist separately: as soon as the value of one of these variables is realized, the other one loses its original meaning. Instead, the duration X could be paired with an indicator of either pregnancy or giving up, whichever occurs first. Either way, the hazard rate associated with (the termination of) the durations X can always be represented as the sum of two cause-specific hazard rates, one associated with pregnancy and the other with giving up. Because of selection mechanisms acting in the study population, the former is likely to be decreasing, and the latter increasing with respect to age.

The cause-specific hazard rates are not separately identifiable from current duration data, and it is not clear that their sum, corresponding to the duration variable X, would be monotone in either direction. When the follow-up time from the start of current attempt is short, the hazard rate of giving up is likely to be small and therefore, early on, the hazard rate corresponding to X should be a reasonable approximation of that corresponding to T.

An additional problem concerning the identifiability of the distribution for the time to pregnancy variable T is due to the fact that pregnancy is generally detected only after some delay after conception has occurred, then often in connection of absence of subsequent menstrual bleeding. Here we do not elaborate on this issue further, and simply define the variable T as the time of unprotected intercourse until pregnancy is detected.

As argued, e.g., in Keiding et al. (2012), it is common to focus on the beginning of the estimated TTP distribution and artificially censor it after about one or two years, motivated partly by lack of confidence in the precision of large retrospectively recalled current durations, partly by the less interesting nature and validity of the right tail of the TTP distribution.

3 Indentifiability

As before, let $H(t) = P(X \leq t)$, $E(X) = \int_0^\infty (1 - H(s))ds$, and the current duration density

$$f(t) = \frac{1 - H(t)}{E(X)}.$$

For $\alpha \in (0,1)$, consider the mixture distribution

$$H_{\alpha,0}(t) = P_\alpha(X \leq t) = \alpha \mathbf{1}(t = 0) + (1 - \alpha)H(t)$$

with a point mass of size α at 0. We have $E_{\alpha,0}(X) = (1-\alpha)E(X)$ and

$$f(t) = \frac{1-H(t)}{E(X)} = \frac{1-H_{\alpha,0}(t)}{E_{\alpha,0}(X)} \quad \forall\, t > 0$$

We see that moving probability mass to the origin in the TTP distribution $H(t)$ does not affect the current duration distribution $f(t)$, and $H(0)$ is not identifiable. The current duration setup is not limited to fecundity studies and may arise in different contexts, for example, in the survey we could ask questions like: Are you looking for a partner ? Or, for a job ? If so, how long have you been searching? Depending on the context, the event $\{X = 0\}$ is not always ruled out and possibly $H(0) = P(X = 0) > 0$. In such cases only the conditional distribution

$$P(X \le t | X > 0) = \frac{H(t) - H(0)}{1 - H(0)}$$

is identifiable. We show that, even when the mass α is large, moving it to a point sufficiently close to the origin will only have a small effect on the density $f(t)$. For $\alpha, \delta \in (0,1)$, let

$$H_{\alpha,\delta}(t) = \alpha \mathbf{1}\big(\delta E(X) \le t\big) + (1-\alpha)H(t)$$

with TTP expectation

$$E_{\alpha,\delta}(X) = \int_0^\infty (1 - H_{\alpha,\delta}(s))ds = E(X)(1 - \alpha + \alpha\delta) .$$

The corresponding current duration distribution has density

$$f_{\alpha,\delta}(t) = \begin{cases} f(t)\Big(1 + \alpha(1 - \delta(1-H(t))/\big\{(1-H(t))(1-\alpha+\alpha\delta)\big\}\Big) , & t < \delta E(X) \\ f(t)\big(1 - \alpha\delta/(1-\alpha+\alpha\delta)\big) , & t \ge \delta E(X) \end{cases}$$

which satisfies

$$\begin{aligned} & d_{TV}\left(f_{\alpha,\delta}, f\right) = \frac{1}{2}\int_0^\infty \big|f_{\alpha,\delta}(t) - f(t)\big| dt = \\ & \frac{\alpha}{2(1-\alpha+\alpha\delta)}\int_0^{\delta E(X)} \frac{1+\delta H(t) - \delta}{1-H(t)} dt + \frac{\alpha\delta}{2(1-\alpha+\alpha\delta)}\big(1 - H(E(X)\delta)\big) \\ & \le \frac{\delta}{2(1-\alpha)}\left\{\alpha + \frac{E(X)}{(1-H(E(X)\delta))}\right\} \end{aligned}$$

Therefore, $\forall\, \alpha \in (0,1)$, $\varepsilon > 0$, there is a $\delta \in (0,1)$ such that $d_{TV}\left(f_{\alpha,\delta}, f\right) < \varepsilon$, while

$$d_{TV}(H_{\alpha,\delta}, H) = \alpha(1 - \Delta H(E(X)\delta)) .$$

For fixed α, as $\delta \downarrow 0$ we need large and larger current duration samples in order to distinguish between the alternatives H and $H_{\alpha,\delta}$. Even if it is known that $H(0) = 0$, the TTP-distribution is only weakly identifiable from current duration data. In order to build a nonparametric estimator of the TTP-distribution it is necessary to regularize the estimators close to the origin, ruling out the bad alternatives $H_{\alpha,\delta}$ with small δ. A simple way to do this, which seems natural in the context of TTP data, is to assume a priori $H(t) \equiv 0$ in an interval $[0, t_0]$ (Section 4). An alternative to this, adopted here in Section 5, is to apply Bayesian nonparametrics.

4 Semiparametric Maximum Likelihood Estimator

We modify in a simple way the Nonparametric Maximum Likelihood Estimator (NPMLE) discussed in Keiding et al. (2012), by imposing on the non-increasing current duration density the constraint $f(t) \equiv f(0)$ for $t \in [0,t_0]$. This is achieved by maximizing the loglikelihood

$$n(t_0)\log f(0) + \sum_{y_i > t_0} \log f(y_i), \qquad \text{with } n(t_0) = \sum_{i=1}^{n} \mathbf{1}(y_i \le t_0)\ ,$$

with respect to $f(0)$ and $(f(y_i) : y_i > t_0)$, under the constraints

$$f(0) \ge f(y_i) \ge f(y_{i+1}) > 0, \quad y_i > t_0$$

and

$$f(0)t_0 + \sum_i f(y_i)(y_i \vee t_0 - y_{i-1} \vee t_0) = 1$$

The corresponding *Semiparametric Maximum Likelihood Estimator* (SPMLE) of the current duration density is determined explicitly as

$$\widehat{f}_n(t;t_0) = \begin{cases} \widehat{f}_n(y_k;t_0)\ , & y_{k-1} < y \le y_k,\ k = 1,\dots,n \\ 0\ , & t > y_n \end{cases}$$

$$\widehat{f}_n(y_k;t_0) = \frac{1}{n} \min_{0 \le \ell \le k-1} \max_{k \le m \le n} \left\{ \frac{m-\ell}{y_m \vee t_0 - y_\ell \vee t_0} \right\}, \quad k = 1,\dots,n,$$

where $y_0 = 0$. The corresponding SPMLE of the TTP-distribution is then defined as

$$\widehat{H}_n(t;t_0) = 1 - \frac{\widehat{f}_n(t;t_0)}{\widehat{f}_n(0;t_0)} = 1 - \frac{\widehat{f}_n(t;t_0)}{\widehat{f}_n(t_0;t_0)}\ ,$$

The NPMLE discussed in Keiding et al. (2012) corresponds to

$$\check{H}_n(t;t_0) = 1 - 1 \wedge \frac{\widehat{f}_n(t;0)}{\widehat{f}_n(t_0;0)}\ ,$$

and it is consistent if and only if $t_0 > 0$.

5 Non-parametric Bayesian estimation of S

The non-parametric Bayesian approach to inference means that probability distributions are assigned to random (= uncertain) functions, where the probability represents a quantification of the uncertainty involved. Here f and S are the natural candidates for such functions. Since $f(y) = S(y)/E(X)$, where

$$E(X) = \int_0^\infty S(x)dx\ ,$$

knowledge of S clearly leads to knowledge of f. But also conversely: since $f(0) = 1/E(X)$, we get that $S(y) = f(y)/f(0)$. Note also that in the inferential problem of estimating S (or H), the mean $E(X)$, which is often viewed as a normalizing constant, cannot be treated as a constant (in S) because it is the integral of S.

Any prior distribution on the realizations of S can be readily required to satisfy the natural properties that, with prior probability 1, $S(0) = 1$, and S must be non-increasing. These properties are then automatically inherited from the prior to the posterior. Note first that, because of the key formula (2), the realizations of f must be non-increasing as well. This also implies that

$$\lim_{y \downarrow 0} f(y)/f(0) = \lim_{y \downarrow 0} S(y)/S(0) \leq 1 ,$$

with $E(X)$ cancelling from the latter ratio. Thus the inconsistency problem at $t = 0$ of the non-parametric maximum likelihood estimator (NPMLE) (Grenander, 1956; Denby and Vardi, 1986), which was pointed out by Woodroofe and Sun (1993) and has since then been considered by several authors, is solved here automatically for the realizations of the random function f. It is shown in Appendix 9 that our suggested Bayesian estimator of the density f is consistent, and thus it shares this asymptotic property with the penalized NPMLE introduced by Woodroofe and Sun (1993). However, it also turns out that, in spite of the consistency, the problematic behaviour of the original NPMLE near the origin continues to hamper the practical application of both the penalized NPMLE and the nonparametric Bayesian estimator in finite samples. As we will illustrate later by examples, the Bayesian estimators are quite sensitive to variations in the hyperparameter values specifying the prior, and a similar type of dependence holds on values of the parameter tuning the penalized NPMLE.

The likelihood contributions from different current duration data points y_i are here considered independent, conditionally on their assumed distribution function H, and therefore lead to the combined likelihood expression of the simple product form

$$L(S) = \prod_{i=1}^{n} f(y_i) = \left(\int_0^\infty S(t)dt \right)^{-n} \prod_{i=1}^{n} S(y_i) \tag{3}$$

From (3) the logarithmic likelihood $\ell(S) = \log L(S)$ can be written in the form

$$\begin{aligned} \ell(S) &= \sum_{i=1}^{n} \log(S(y_i)) - n \log\left(\int_0^\infty S(t)dt \right) \\ &= \sum_{i=1}^{n} R_i \log\left(\frac{S(\tau_i)}{S(\tau_{i-1})} \right) - n \log\left(\int_0^\infty S(t)dt \right) , \end{aligned}$$

where $(\tau_i : i = 1, \ldots, n)$ are the ordered individual observation times $(y_i : i = 1, \ldots, n)$, and

$$R_i = R(\tau_i), \quad \text{with } R(s) = \sum_{k=1}^{n} \mathbf{1}(\tau_k \geq s) ,$$

is the size of the risk set at τ_i.

In a nutshell, our Bayesian approach to inference can be stated as follows:

- Considering (realizations of) survival functions S, combine the consequent likelihood expression (3) with a chosen prior distribution on the set of these functions.
- This gives, via Bayes formula, the posterior, that is, the conditional distribution of S given the data.
- Finally apply some appropriate version of a Markov chain Monte Carlo (McMC) algorithm to do the numerical computations.

6 Logistic Process Priors

A Bayesian nonparametric model for non-increasing densities on $[0,\infty)$ has been studied by Hansen and Lauritzen (2002), by mapping a cumulative distribution function G to a non-increasing density

$$f(t) = \int_t^\infty s^{-1} G(ds), \quad t \geq 0 \, .$$

In this way a prior on the space of non-decreasing densities is obtained from a prior on the space of cumulative distribution functions, like the Dirichlet process prior.

Since we are interested in $S(t) = f(t)/f(0)$ rather than in the non-increasing density $f(t)$ by itself, we take an alternative approach, and modify the construction of the logistic Gaussian density process to produce a monotone density.

6.1 Logistic Gaussian Process Prior

Let $(Z(t) : t \in I)$ be a Gaussian random field indexed by $I \in \mathbb{R}^d$, with law P_0 under which $E_{P_0}\big(Z(t)\big) = \mu(t)$ and $E_{P_0}\big(Z(t)Z(s)\big) - \mu(t)\mu(s) = K(s,t)$. After changing the probability measure to a non-Gaussian probability P_ρ with density

$$\frac{dP_\rho}{dP_0} = C(\rho)^{-1} \left\{ \int_I \exp(-Z_s) ds \right\}^\rho \quad \text{with} \quad C(\rho) = E_{P_0}\left(\left\{ \int_I \exp(-Z_s) ds \right\}^\rho \right) ,$$

parametrized by $\rho \in \mathbb{R}$, Lenk (1988, 1991) introduced the *logistic Gaussian process*

$$f(t) := \exp(-Z_t) \left(\int_I \exp(-Z_s) ds \right)^{-1}, \quad t \in I \, ,$$

which is a random density function on I, with prior parametrized by (μ, K, ρ). It follows that, when $Y_1, \ldots, Y_n$, which are conditionally i.i.d. with density $f(t)$, given $(Z(t) : t \in I)$, the posterior distribution of $(Z(t), t \in I)$ given the data $(Y_1, \ldots, Y_n)$, is again a logistic Gaussian density process with posterior parameters $(\mu^*, K, \rho - n)$, where

$$\mu^*(t) = \mu(t) - \sum_{i=1}^n K(t, Y_i), \quad t \in I \, ,$$

see also Ghosh and Ramamoorthi (2003), Lemma 5.7.2.

6.2 Logistic Generalized Gamma Convolution Prior

In our case $d = 1$, $I = [0,\infty)$ and $f_Z(t)$ and the corresponding realizations of the survival function $S(t)$ should be monotone, which rules out Gaussian priors for $Z(t)$. Instead, we model $Z(t)$ as a right-continuous non-decreasing process with $Z(0) = 0$. In general, the cumulative hazard process corresponding to $S(t) = \exp(-Z(t))$ is given by

$$\Lambda(t) = Z^c(t) + \sum_{u \le t} \big(1 - \exp(-\Delta Z(u))\big), \quad t \ge 0, \tag{4}$$

where $\Delta Z(t) := Z(t) - Z(t-)$ denote the jumps and

$$Z^c(t) := Z(t) - \sum_{u \le t} \Delta Z(u) \; ,$$

denotes the continuous part of Z.

Note also that $f(t) < \infty \,\forall\, t \in [0,\infty)$, and we cannot obtain unbounded realizations of the random density. When $Z(t)$ is $\mathbb{R}$-valued, the corresponding density $f(t)$ and survival probability $S(t)$ will be strictly positive for all $t > 0$. In order to model distributions with support on a compact interval which is a priori unknown, we may add to the model the cemetery state $+\infty$ and an independent r-exponential killing time σ such that $Z(t) = +\infty$ and correspondingly $f(t) = 0$ for $t \ge \sigma$.

Here we actually use a purely discontinuous jump process with $Z^c(t) \equiv 0$, and then impose the following special structure:

$(Z(t) : t \ge 0)$ is a *generalized gamma convolution* (GGC) process (James et al., 2008), such that $Z(0) = 0$, the increments are independent and non-negative, driven by the Lévy measure

$$\nu(dz, ds) = \mathbf{1}(z > 0) \exp\big(-b(s)z\big) z^{-1}\, dz\, a(ds) \; , \tag{5}$$

which is also characterized by the Laplace transform

$$E_P\left(\exp\left(-\int_0^\infty \psi(s) Z(ds)\right)\right) = \exp\left(-\int_0^\infty \log\left(1 + \frac{\psi(s)}{b(s)}\right) a(ds)\right) .$$

We call the corresponding random density $f(s)$ a *logistic generalized gamma convolution process*, and denote its distribution by $\mathrm{LGGC}(b(s), a(ds), \rho)$.

When $(f(t) : t \ge 0)$ has the logistic subordinator prior $\mathrm{LGGC}(b(s), a(ds), \rho)$, the posterior distribution, given the observations $(Y_1, \ldots, Y_n)$, is again a LGGC process with parameters $(b(s) + R(s), a(ds), \rho - n)$.

In the implemention we use a logistic subordinator prior with $\rho = 0$ and driven by a time-homogeneous gamma process, such that the increments $(Z(t) - Z(s)), 0 \le s \le t$ are gamma distributed with shape parameter $(t - s)a$, and a constant rate parameter b.

Note that, in any finite interval $[0,T]$, a time-homogeneous Gamma process has the representation

$$Z(u) = Z(T)W(u),\ u \in [0,T]\ ,$$

where $(W(u) : u \in [0,T])$ is a Poisson-Dirichlet process on $[0,T]$ with driving measure $a\,du$, which is independent from the final value $Z(T)$ (Phadia, 2013). It follows that the distribution of W does not depend on the rate parameter b of the time-homogeneous Gamma process Z. We restrict the analysis to a finite time interval $[0,T]$, assuming $Z(u) = +\infty$ and $S(u) = 0$ for $u > T$. We use the decomposition

$$Z(u) - Z(\tau_{i-1}) = W_i(u)\Delta Z_i \quad u \in (\tau_{i-1}, \tau_i]\ ,$$

where $\Delta Z_i = Z(\tau_i) - Z(\tau_{i-1})$, and $W_i(u)$ is an homogeneous Poisson-Dirichelet process on the interval $(\tau_{i-1}, \tau_i]$, with driving measure $a\,du$. Realizations from the Poisson-Dirichlet prior of $W(u)$ are generated by using the Ferguson and Sethuraman stick-breaking algorithm (Sethuraman, 1994; Phadia, 2013).

The idea is to parametrize the model by the Gamma-distributed increments ΔZ_i and a sequence of independent Poisson-Dirichlet processes $W_i(u)$ on each interval $(\tau_{i-1}, \tau_i]$. With this construction and notation, we can rewrite the log-likelihood corresponding to the the current duration data points $(y_i : i = 1, \dots, n)$ as

$$\begin{aligned} & -\sum_{i=1}^{n} R_i(Z(\tau_i) - Z(\tau_{i-1})) - n\log\left(\int_0^{\infty} \exp\big(-Z(u)\big)du\right) \\ &= -\sum_{i=1}^{n} R_i \Delta Z_i - n\log\left(\sum_{i=1}^{n} \exp\big(-\sum_{j<i} \Delta Z_j\big) \int_{\tau_{i-1}}^{\tau_i} \exp\big(-W_i(u)\Delta Z_i\big)du\right) \qquad . \end{aligned}$$

Note that $W_i(u)$ corresponds to a probability distribution which spreads the mass ΔZ_i on the interval $[\tau_{i-1}, \tau_i)$. The parameter a in the Gamma process prior tunes the deviation of the random probability distribution $Z(dt)/Z(T)$ from its mean, approaching the uniform distribution as $a \uparrow \infty$.

For comparison, the NPMLE is found by first setting $W_i(u) \equiv 1$, which assigns to each time point τ_{i-1} the point mass ΔZ_i, and then by maximizing the log-likelihood

$$-\sum_{i=1}^{n} R_i \Delta Z_i - n\log\left(\sum_{i=1}^{n} \exp(-\sum_{j\le i} \Delta Z_j)(\tau_i - \tau_{i-1})\right)\ ,$$

under the constraint $\Delta Z_i \ge 0$, $i = 1, \dots, n$.

Values of $(\Delta Z_i, i = 1, \dots, n)$ and of the Poisson-Dirichlet processes $W_i(u)$ can now be sampled from the posterior distribution by using a Metropolis-Hastings algorithm (Robert and Casella, 2004). This is done by alternating between updating $(W_i(u) : \tau_{i-1} < u \le \tau_i)$ and ΔZ_i, using the prior as the proposal distribution and keeping the other variables fixed, and by then accepting or rejecting each proposal according to the Metropolis-Hastings rule. This McMC updating scheme works also with more general LGGC priors with piecewise constant rate parameter $b(s)$ and without restrictions on the driving measure $a(ds)$.

6.3 Pseudo-posterior

Walker and Hjort (2001) proposed to combine a nonparametric prior $\Pi(df)$ on a space of density functions with an α-power of the likelihood, into the pseudo posterior

$$\widehat{\Pi}_n(df) \propto \prod_{i=1}^{n} f(Y_i)^{\alpha} \Pi(df) ,$$

and showed that when $\alpha \in (0,1)$ and the true density f^0 belongs to the Kullback-Leibler support of the prior, the pseudo-posterior is always strongly consistent (we discuss posterior consistency in the Appendix). When f^0 is non-increasing this follows also for $\alpha = 1$, and in this respect Walker and Hjort pseudo-posterior is not needed. We just remark that by starting with a LGGC$(b(s), a(ds), \rho)$ process prior for f, the resulting pseudo-posterior is the LGGC$(b(s)+\alpha R(s), a(ds), \rho - \alpha n)$ process, and the proposed McMC algorithm applies directly.

6.4 Choice of parameters

As noted above, the shape parameter a tunes the departure from uniformity of the random probability distributions $W_i(du)$ on the respective intervals $[\tau_{i-1}, \tau_i)$. Here we simply fix its value, but it would be also possible to assign a prior to it and compute its posterior distribution by adding a Metropolis update to the McMC. We assign a scale invariant improper prior to the rate parameter b as

$$\pi(b) \sim b^{-1} . \tag{6}$$

Since

$$p(b|Z,Y) \propto \pi(b) p(Z_T|b) \propto b^{aT-1} \exp(-bZ_T) ,$$

it turns out that, given Z_T, b is conditionally independent from the data $(y_i : i = 1,\dots,n)$, and Gamma distributed with shape parameter aT and rate parameter Z_T. In the McMC we update cyclically b by sampling from this full conditional Gamma distribution.

7 Bayesian Data Augmentation

An alternative approach consists in augmenting the model with the latent selected waiting times $X_1,\dots,X_n$. Note that conditionally on $\{I_i = 1\}$, the selected waiting time X_i has distribution

$$G(x) := P(X_n \le x | I_n = 1) = \int_0^x y\, H(dy) \Big/ \int_0^\infty y\, H(dy) . \tag{7}$$

We model the selected TTP distribution as

$$G(t) = 1 - \exp\big(-Z(t)\big) \tag{8}$$

where Z_t is a GGC-process with Lévy measure (5), which determines the waiting time distribution

$$H(x) = P(X_n \le x) = \int_0^x y^{-1}\, G(dy) \bigg/ \int_0^\infty y^{-1}\, G(dy)\ . \tag{9}$$

Under such a prior specification, given the data $(Y_1,\dots,Y_n)$, we sample the process $H(t)$ and the latent variables $X_1,\dots,X_n$ by using a two-step Gibbs-Metropolis algorithm:

1. Given the current durations $Y_1,\dots,Y_n$, the indicators $I_1,\dots,I_n$ of being included in the study, and the random waiting time distribution $H(t)$, the latent selected waiting times are sampled independently from the conditional distributions

$$P(X_i \in dx|Y_i, I_i = 1, H) \propto P(X_i \in dx|H)p(Y_i|X_i)p(I_i = 1|X_i) = H(dx)\frac{\mathbf{1}(x \ge Y_i)}{X_i}\ c\, X_i \tag{10}$$

$$\Longrightarrow P(X_i \in dx|Y_i, I_i = 1, H) = \frac{\mathbf{1}(x \ge Y_i)}{1 - H(Y_i-)} H(dx)\ . \tag{11}$$

2. We update the process $Z(t)$, which in turn determines the waiting time distributions $H(t)$ and $G(t)$ by (8) and (9), as follows:
Conditionally on the augmented data $(Y_i, X_i, I_i = 1, i = 1\dots,n)$ the GGC process $Z(t)$ is again a subordinator with decomposition

$$Z(t) = \hat{Z}(t) + \check{Z}(t),$$

into independent components, specified as follows:
   - $\hat{Z}(t)$ is a GGC-process with Lévy measure

$$\nu^*(dz,dt) = \exp\big(-z(b(t)+r(t))\big)\ z^{-1}\ a\ dzdt\ ,$$

   which has piecewise constant rate function when $b(s)$ is piecewise constant, where

$$r(t) = \#\big\{j : X_j > t\big\}$$

   denotes the size of the set of individuals at risk of becoming pregnant at time t.
   - $\check{Z}(t)$ has jumps $\Delta\check{Z}_i$ at the fixed discontinuities X_i, with respective densities

$$C_i^{-1} \exp\big(-(b(X_i)+r(X_i))z\big)\ z^{-1}\ (1-\exp(-z))^{\Delta n(X_i)} dz \tag{12}$$

   where

$$\Delta n(t) := \#\big\{j : X_j = t\big\}\ ,\ \text{and } C_i \text{ are normalizing constants.}$$

   In the Gibbs-Metropolis implementation, we update $\check{Z}_i$ by sampling its jumps from the proposal distribution

$$\Delta\check{Z}_i^* \sim \text{gamma}\big(\Delta n(X_i), b(X_i)+r(X_i)\big)$$

accepting the transition $\Delta\check{Z}_i \longrightarrow \Delta\check{Z}_i^*$ with probability

$$\min\left\{1 \wedge \left(\frac{\big(1-\exp(\Delta\check{Z}_i^*)\big)\Delta\check{Z}_i\big)}{\big(1-\exp(\Delta\check{Z}_i)\big)\Delta\check{Z}_i^*}\right)^{\Delta n(X_i)}\right\}.$$

We could also assign the prior distribution of $G(t)$ by modeling the corresponding cumulative hazard rate process (4) as a beta process (Hjort , 1990) or a neutral to the right process (Phadia, 2013).

Remark 1 Note that in the data augmentation scheme presented above, it was convenient to assign the nonparametric LGGC prior to the distribution G of length-selected waiting times, which determines by (9) also the prior of the genuine waiting time distribution H. In Section 6.2 we did not use data augmentation, and we assigned the nonparametric LGGC prior directly to the waiting time distribution H.

8 Some illustrations of the method

This section contains some numerical illustrations on applying the nonparametric Bayesian estimation method for estimating the waiting time distribution $H(t)$, or the corresponding survival distribution $S(t) = 1 - H(t)$, from current duration data. The purpose of the first illustration, based on simulated data, was to check how well the method calibrates with respect to the data generating distribution. For this purpose, we simulated independent samples of size $n = 10,000$ of waiting times X_i from the uniform distribution on $[0,1]$, with $S(t) = 1 - t$, $0 \le t \le 1$, together with the corresponding current duration times $Y_1, \ldots, Y_n$, with conditional density $f(y|x) = x^{-1}\mathbf{1}(y \le x)$ and marginal $f(t) = 2(1-t)$, $t \in [0,1]$, and tried then the method in estimating $S(t)$, based on the current duration data Y_i, $0 \le i \le n$.

Note that, although considering the uniform distribution as the basis of a simulation experiment may seem unduly restrictive, it has an obvious meaning, in terms of quantiles, for any continuous lifetime distribution $H(t)$.

For the Bayesian non-parametric prior, we assumed a gamma process prior for $Z(t) = -\log(1 - G(t))$, $0 \le t \le T = 2$, with driving measure $a(dt) = 1.5 \times dt$ and a non informative prior $\pi(b) \propto b^{-1}$ for the rate parameter. The nonparametric posteriors were computed by running the data augmented McMC algorithm of Section 7.

In Fig. 1 we compare the true current duration density $f(t)$ with the NPMLE and the nonparametric posterior in terms of the pointwise posterior mean, posterior median, and the 90 percent credible interval, while in Fig. 2 we compare the corresponding estimators for the waiting time distribution. The NPMLE estimator of the current duration density $f(t)$ is known to be inconsistent at $t = 0$, and $f(0)$ is overestimated. The corresponding survival probability $S(t)$ is underestimated by the NPMLE. The posterior credible intervals are quite wide closer to the origin, but become narrower as t increases, and contain the true generating $S(t)$ for all t.

The second illustration used current duration data from the French telephone survey on TTP, with sample size $n = 867$, discussed in detail in (Slama et al., 2012). In the specification of the Gamma process prior of $Z(t) = -\log(1 - G(t))$ we used the driving measure $a(dt) = a \times dt$, and assumed a non informative prior $\pi(b) \propto b^{-1}$ for the rate parameter. The nonparametric posteriors were computed either by running 10,000 cycles of the data-augmented McMC algorithm of section 7, or by running 50,000 cycles of the unaugmented McMC algorithm of section 6.2, using a time horizon of $T = 300$ months. In the illustrations shown in Figures 3-10, for clarity, we show only the first 36 months. Figures 3 and 4 display the current duration density estimators under the LGGC prior, based on the augmentation method and hyperparameter value $a = 0.1$, without cut-off (Fig. 3), and using a cut-off at two weeks (Fig. 4), which is understood as the minimal delay in the detection of pregnancy after conception. The effect of cut-off is clearly visible in the difference between the NPMLE and SPMLE close to the origin. The Bayesian posterior estimates seem quite insensitive to the use of the cut-off. Figures 5 and 6, again showing results from current duration density estimation, illustrate the effect of changing the value of the hyperparameter a from $a = 0.1$ to $a = 1$, while keeping the other settings as in Fig. 3. As expected, a larger value of a results in somewhat smoother curves, and also narrower pointwise credible intervals. Figures 7, 8, 9 and 10 correspond to each of Figures 3, 4, 5 and 6, but show the results in terms of the corresponding survival function estimates. Comparison of the Bayesian posterior estimates in the different settings follows closely the patterns that were already visible in the corresponding density estimates, but the NPMLE and SPMLE survival curves are still remarkably sensitive to the use of a cut-off at two weeks.

We also conducted a Monte Carlo study with $M = 4050$ replications of the French telephone survey data. As in Keiding et al. (2012), we assumed a generalized Gamma TTP distribution, with density

$$f_{\lambda,\sigma,\mu}(x) = \frac{|\lambda|\lambda^{-2\lambda^{-2}}}{|\sigma|\Gamma(\lambda^{-2})} x^{(\sigma^{-1}\lambda^{-1}-1)} \exp\left(-\frac{\mu}{\lambda\sigma} - \lambda^{-2} e^{-\mu\lambda/\sigma} x^{\lambda/\sigma}\right) \mathbf{1}(x \geq 0) \quad (13)$$

with parameters $\lambda, \sigma, \mu \in \mathbb{R}$. This is the density of the power $X(\omega) = \xi(\omega)^{\sigma/\lambda}$, where $\xi(\omega)$ is Gamma distributed with shape λ^{-2} and rate $\lambda^{-2}\exp(-\mu\lambda/\sigma)$. Note that

- $0 < f_{\lambda,\sigma,\mu}(0) < +\infty$ if and only if $\sigma\lambda = 1$.
- $f_{\lambda,\sigma,\mu}(0) = +\infty$ for $\sigma\lambda > 1$,
- $f_{\lambda,\sigma,\mu}(0) = 0$ for $\sigma\lambda < 1$.

The distribution of the selected TTP is the conditional density

$$g_{\lambda,\mu,\sigma}(x) = f_{\lambda,\mu,\sigma}(x|I=1) = \frac{x f_{\lambda,\mu,\sigma}(x)}{\int_0^\infty r f_{\lambda,\mu,\sigma}(r)dr} = f_{\widetilde{\lambda},\widetilde{\sigma},\widetilde{\mu}}(x) \quad (14)$$

which is the generalized gamma density with parameters $(\widetilde{\lambda}, \widetilde{\sigma}, \widetilde{\mu})$

$$\widetilde{\lambda} = \frac{\lambda}{\sqrt{1+\sigma\lambda}}, \quad \widetilde{\sigma} = \frac{\sigma}{\sqrt{1+\sigma\lambda}}, \quad \widetilde{\mu} = \mu + \frac{\sigma}{\lambda}\log(1+\sigma\lambda)\,. \quad (15)$$

The expected TTP in (14) is given by

$$E_{\lambda,\mu,\sigma}(X) = \frac{x f_{\lambda,\mu,\sigma}(x)}{f_{\widetilde{\lambda},\widetilde{\mu},\widetilde{\sigma}}(x)} = \exp\left(\mu + \left(\frac{1}{\lambda^2} + \frac{\sigma}{\lambda}\right)\log(1+\sigma\lambda)\right)\frac{\widetilde{\lambda}^{(2\widetilde{\lambda}^{-2})}\Gamma(\widetilde{\lambda}^{-2})}{\lambda^{(2\lambda^{-2})}\Gamma(\lambda^{-2})}\,, \quad \forall x > 0.$$

In the simulation we used parameter values $\lambda = 0.7, \mu = 1.4865, \sigma = 1.2857$, corresponding to 6.041 months mean TTP. When X is a selected TTP time with density (14) independent from $U \sim \text{Uniform}([0,1])$, the product $Y = UX$ follows the current duration distribution with density

$$\begin{aligned} h_{\lambda,\sigma,\mu}(y) &= \frac{P_{\lambda,\mu,\sigma}(X > y)}{E_{\lambda,\mu,\sigma}(X)} = \frac{\int_y^\infty f_{\lambda,\mu,\sigma}(r)dr}{\int_0^\infty x f_{\lambda,\mu,\sigma}(x)dx} \\ &= \frac{1 - \Gamma(\lambda^{-2}; \lambda^{-2} y^{\lambda/\sigma} e^{-\mu\lambda/\sigma})/\Gamma(\lambda^{-2})}{E_{\lambda,\sigma,\mu}(X)}, \quad y \geq 0\,, \end{aligned} \tag{16}$$

where

$$\Gamma(\alpha;t) = \int_0^t s^{\alpha-1} e^{-s} ds$$

is the incomplete Gamma function (see also Yamaguchi (2003)).

Each replicated dataset consists of $n = 867$ i.i.d. current durations sampled from the density (16). For each dataset we computed the nonparametetric Bayesian estimators as in section 7, assuming for the selected TTP distribution a nonparametric LGGC prior with hyperparameter $a = 0.05$. For each simulated dataset the posterior was computed in 10000 McMC cycles. Fig.11 shows the pointwise empirical median curve and the corresponding 90% confidence band of the pointwise Bayesian nonparametric estimate (median) of the current duration density, determined from the 4050 replicated samples by applying the nonparametric Bayesian estimation method. Also the empirical medians of the corresponding pointwise 90% posterior credible interval bounds are displayed. Note that the true current duration density is within the empirical confidence band. The corresponding statistics for the Bayesian nonparametric TTP-survival probabilities are displayed in Fig. 12. Typically the nonparametric TTP posterior is parsimonious, assigning a negligible mass to the neighbourhood of the origin where the TTP distribution identifies poorly.

9 Discussion

In the current duration design, the probability that an individual belonging to a large population will be included in the sample is proportional to the length of the considered waiting time or lifetime. As a consequence, typical data sets arising from such a design contain only few short durations, which renders the estimation of the distribution near the origin unstable. The problem is particularly evident in nonparametric maximum likelihood estimation, where even asymptotic consistency of the estimator at the origin does not hold without further constraints. As we have demonstrated in Section 8 above, nonparametric Bayesian modelling and inference gives an alternative approach to the analysis of current duration data. The empirical results provided there are complemented by Corollary 1, which shows that our Bayesian method is weakly consistent, with the proof given in the Appendix.

As always in the case of Bayesian methods, an issue of some concern is the dependence of the posterior inference on the specification of the hyperparameters controlling the prior distribution. While such influence in nonparametric estimation rarely exceeds the strong control imposed by the choice of a distribution family in parametric inference, it is true that often there are no clear rules as to how the values of such hyperparameters should be chosen. Here, larger values of the driving measure parameter lead to a greater degree of smoothness in the resulting estimates, somewhat reminiscent to choosing a wider bandwidth when applying nonparametric kernel smoothing. Apart from their behaviour very close to the origin, the estimates do not seem to depend much on whether a positive threshold parameter t_0 value has been used, and therefore our illustrations are mainly based on analyses where there was no such threshold.

A natural way to bypass the difficulty of specifying hyperparameters for the prior would be to combine, within the same statistical analysis, likelihood expressions from studies based on different designs and with consequently different strengths. For example, considering that the information content of TTP-data is weakest on short waiting times, it would make sense to complement such data with data from a retrospective survey study, conducted simultaneously on the same population. Couples who actually are pregnant or got children in a given time window are then asked how long they tried before getting pregnant, and couples who had tried to get pregnant without success and had given up, are asked how long they did try before giving up.

There are several alternatives to formulating a 'stochastic process' model for $S(t)$ to the one that was used here, employing a Markov jump process to model a piecewise constant (Arjas and Gasbarra, 1994) or piecewise linear functions for the hazard rate

$$\lambda(t) = -\frac{d\log(S(t))}{dt} .$$

Various extensions of the present basic version of the model would also be possible, for example, by postulating a Cox type proportionality property for the hazard rates corresponding to different covariates.

References

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

Baird DD, Wilcox AJ, Weinberg CR. (1986). Use of time to pregnancy to study environmental exposures. *Am J Epidemiol* 124 (3) 470-80.

Brillinger D.R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* 42 (4) 693-734.

Cox, D.R. and Miller, H.D. (1965). *The theory of stochastic processes*. Chapman.

Diaconis P. and Freedman D. (1986). On the consistency of Bayes estimates (with discus- sion). *Annals of Statistics* 14 1-67.

Denby L. and Vardi Y. (1986). The survival curve with decreasing density. *Technometrics* 28 (4) 359-367.

Ghosh J.K., Ramamoorthi R.V. (2003). *Bayesian nonparametrics*. Springer.

Grenander U. (1956). On the theory of mortality measurement II. *Skand. Aktuarietidskr.* 39 125-153.

Hansen M.B. and Lauritzen S.L. (2002). Nonparametric Bayes inference for concave distribution functions. *Statistica Neerlandica* 56 (1) 110-127.

Hjort N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* 18 (3) 1259-1294.

Hjort N.L. Holmes C. Müller P. Anderson M.D. and Walker S.G. (Editors) (2010). *Bayesian nonparametrics*. Cambridge University Press.

Ishikawa Y. (2013). *Stochastic calculus of variations for jump processes*. De Gruyter.

James L.F., Roynette B., Yor M. (2008). Generalized gamma convolutions, Dirichlet means, Thorin measures, with explicit examples. *Probability Surveys* 5 346-415.

Kallenberg O (2002). *Foundations of modern probability, 2nd edition*. Springer.

Keiding, N. (1991). Age-specific incidence and prevalence: a statistical perspective (with discussion). *J.Roy.Statist.Soc. A* 154 371-412.

Keiding, N., Kvist, K., Hartvig, H., Tvede, M. & Juul, S. (2002). Estimating time to pregnancy from current durations in a cross-sectional sample. *Biostatistics* 3 565-578.

Keiding, N., Hansen, O.H.H., Srensen, D.N. & Slama, R. (2012) The current duration approach to estimating time to pregnancy. *Scand. J. Statistics* 39 185-204.

Lenk P.J. (1988). The logistic normal distribution for Bayesian nonparametric predictive densities. *JASA* 83 (402) 509-516.

Lenk P.J. (1991). Towards practicable Bayesian nonparametric density estimator. *Biometrika* 78 (3) 531-543.

McLaughlin, K.A., Green, J.G., Gruber, M.J., Sampson, N.A., Zaslavsky, A.M. & Kessler, R.C. (2010). Childhood adversities and adult psychiatric disorders in the National Comorbidity Survey Replication II. *Arch. Gen. Psychiatry* 67 124-132.

Phadia, E.G. (2013). *Prior processes and their applications, nonparametric Bayesian estimation*. Springer.

Robert C., Casella G. (2004). *Monte Carlo statistical methods, 2nd edition*. Springer.

Schwartz L. (1965). On Bayes procedures. *Z. Wahrsch. Verw. Gebiete.* 4 10-26.

Sethuraman, J. (1994). A constructive definition of the Dirichlet process prior. *Statistica Sinica* 2 639-65.

Simon Th. (2000). Support theorem for jump processes. *Stochastic Process. Appl.* 89 (1) 1-30.

Slama, R., Hansen OK, Ducot B, Bohet A, Sorensen D, Giorgis Allemand L, Eijkemans MJ, Rosetta L, Thalabard JC, Keiding N, Bouyer J. (2012). Estimation of the frequency of involuntary infertility on a nation-wide basis. *Hum Reprod.* 27 (5) 1489-98.

Tokdar S.T, Ghosh J.K. (2007) Posterior consistency of logistic Gaussian process priors in density estimation. *Journal of statistical planning and inference* 137 34-42.

Van Es, B., Klaassen, C.A.J. and Oudshoorn, K. (2000). Survival analysis under cross sectional sampling: length bias and multiplicative censoring. *Journal of Statistical*

Planning and Inference 91 295-312.
Walker S.G. and Hjort N.L. (2001). On Bayesian consistency. *Journal of the Royal Statistical Society, Series B* 63 811-21.
Walker S.G. (2004). New approaches to Bayesian consistency. *Annals of Statistics* 32 (5) 2028-2043.
Walker S.G. Lijoi A. and Prünster I. (2005). Data tracking and the understanding of Bayesian consistency. *Biometrika* 92 (4) 765-778.
Weinberg, C.R. and Gladen, B.C. (1986). The beta-geometric distribution applied to comparative fecundability studies. *Biometrics* 42 547-560.
Williamson, R.E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Math. J.* 23 189-207.
Woodroofe M. and Sun J. (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is non-increasing. *Statistica Sinica* 3 501-515.
Yamaguchi K. (2003). Accelerated failure-time mover-stayer regression models for the analysis of last-episode data. *Sociol. Methodol.* 33 81-110.
Zelen, M. (2004). Forward and backward recurrence times and length biased sampling: age specific models. *Lifetime Data Analysis* 10 (4) 325-334.

10 Appendix: Posterior consistency under logistic subordinator prior

Here we discuss the posterior consistency properties of the Bayesian estimator of the density $f(t)$ introduced in Section 6.2. This is done within the framework presented in the monographs by Ghosh and Ramamoorthi (2003) and Hjort et al. (2010), which contain a good account of the recent developments in Bayesian non-parametrics; see also Walker (2004); Walker et al. (2005). Before embarking on this topic in explicit terms, we make some general remarks concerning consistency in the Bayesian setting.

Statisticians applying Bayesian inferential methods in their work do not necessarily agree on the foundations of their approach: The subjectivist Bayesians, following De Finetti's ideas, do not want, and do not need, to assume that there exists in reality an underlying true probability generating the data. For them, probability has a purely operative meaning, quantifying the uncertainty about unknown quantities, with probability calculus providing a systematic framework for inductive learning and for making rational decisions. The frequentist minded users, on the other hand, reject the idea of subjective probabilities, and instead assume the existence of an unknown true probability generating the data, but nevertheless pragmatically choose to use Bayesian estimators because they work well in practice and have good properties also from the frequentist point of view. While the asymptotic properties of the posterior distribution under the subjective probability follow simply by Doob martingale convergence theorem, a frequentist minded statistician will accept a Bayesian procedure only if it is proved to be consistent under the hypothetical true probability model generating the data. For non-parametric models the frequentist consistency of Bayesian procedures is a subtle matter, since Diaconis and Freedman (1986) have constructed examples where non-parametric priors lead to a non-consistent posterior.

Definition 1 Let $\mathbb{F}$ be a space of probability densities on $\mathbb{R}^d$, equipped with a topology $\mathscr{T}$, and let Π be a prior probability on $\mathbb{F}$ equipped with the Borel σ-algebra $\sigma(\mathscr{T})$.

1. A prior Π on $\mathbb{F}$ is said to achieve $\mathscr{T}$-consistency at $f^0 \in \mathbb{F}$ if for every $\mathscr{T}$-neighbourhood U of f^0 the posterior distribution $\Pi(U|Y_1,\dots,Y_n) \to 1$ $P^{\infty}_{f^0}$-almost surely, where $(Y_i : i \in \mathbb{N})$ are i.i.d. observations with density f^0 under $P^{\infty}_{f^0}$.
2. When $\mathscr{T}$ is the topology of weak convergence of measures, we say that the prior Π achieves weak consistency atf^0, which means that $\forall\, \varepsilon > 0$ and bounded continuous test function g, $P^{\infty}_{f^0}$-almost surely
$$\Pi\left(\left\{f : \left|\int_{\mathbb{R}^d} (f^0(s) - f(s))g(s)ds\right| > \varepsilon\right\}\middle|Y_1,\dots,Y_n\right) \longrightarrow 0\,.$$
3. When $\mathscr{T}$ is the $L^1(\mathbb{R}^d)$-norm topology, we say that the prior Π achieves strong consistency atf^0, which means that $\forall\, \varepsilon > 0$, $P^{\infty}_{f^0}$-almost surely,
$$\Pi\left(\left\{f : \int_{\mathbb{R}^d} |f^0(s) - f(s)|ds > \varepsilon\right\}\middle|Y_1,\dots,Y_n\right) \longrightarrow 0\,.$$

Definition 2 The Kullback-Leibler entropy of the probability density f relative to f^0 is defined as

$$KL(f^0, f) := \int_{\mathbb{R}^d} \log\left(\frac{f^0(s)}{f(s)}\right) f^0(s)ds\,.$$

We say that f^0 is in the KL-support of the prior Π denoted by $KL(\Pi)$ if

$$\Pi\big(f : K(f^0, f) < \varepsilon\big) > 0, \quad \forall\, \varepsilon > 0\,. \tag{17}$$

Theorem 1 *Schwartz (1965), see also Ghosh and Ramamoorthi (2003). If $f^0 \in KL(\Pi)$, then Π achieves weak posterior consistency at f^0.*
In the follow-up we consider the nonparametric class

$$\mathbb{F} = \{ \text{ non-increasing densities on } [0,\infty) \}.$$

It is shown in Walker et al. (2005) that, when the true density is non-increasing, weak and strong consistency of the prior Π are equivalent. The next lemma is a reformulation of their argument.

Lemma 1 *Let $f(x)$, $f_n(x), n \in \mathbb{N}$ denote non-increasing probability densities on $[0,\infty)$, with respective cumulative distribution functions F and F_n . We assume that F is a proper probability distribution, $F(+\infty) = 1$. Then*

$$a) \quad \int_0^{\infty} |f_n(x) - f(x)|dx \to 0 \quad \iff \quad b) \quad F_n \stackrel{w}{\to} F$$

Proof *a)* $\Longrightarrow$ *b)* is obvious, since for any bounded and continuous test function ψ,

$$\left| \int_0^\infty \psi(x)(f_n(x) - f(x))dx \right| \leq \parallel \psi \parallel_\infty \parallel f_n - f \parallel_{L^1(\mathbb{R}^+)} \ .$$

We show *b)* $\Longrightarrow$ *a)*. A non-increasing probability density has representation

$$f(x) = \int_x^\infty y^{-1} G(dy)$$

where

$$G(x) = -\int_0^x y f(dy) = F(x) - x f(x)$$

is a cumulative distribution function with $G(\infty) = 1$ (Williamson, 1956). Note that the distribution G is in one-to-one correpondance with the distribution H in (1):

$$H(dx) = \mu x^{-1} G(dx),\ x > 0, \text{ with } \mu = \int_0^\infty x H(dx) = \left(\int_0^\infty x^{-1} G(dx) \right)^{-1} . \quad (18)$$

We show that *b)* $\Longleftrightarrow$ *c)* $\Longrightarrow$ *a)*, where

$$c) \quad G_n \stackrel{w}{\longrightarrow} G \ .$$

with $G_n(x) = F_n(x) - x f_n(x)$.

For $x > 0$ consider the function $\eta_x(y) = y^{-1}\mathbf{1}(y > x)$. When $c)$ holds, we approximate η_x from above by the bounded continuous functions

$$\eta_x^\varepsilon(y) = \eta_x(y) + \frac{(y - x + \varepsilon)}{x\varepsilon} \mathbf{1}(x - \varepsilon < y < x), \quad 0 < \varepsilon < x$$

Since for $0 < \varepsilon < x$

$$f(x-\varepsilon) \geq \int_0^\infty \eta_x^\varepsilon(y) G(dy) \geq f(x) \geq \int_0^\infty \eta_{x+\varepsilon}^\varepsilon(y) G(dy) \geq f(x+\varepsilon) \ , \quad \text{and}$$
$$\int_0^\infty \eta_x^\varepsilon(y) G_n(dy) \geq f_n(x) \geq \int_0^\infty \eta_{x+\varepsilon}^\varepsilon(y) G_n(dy) \ ,$$

since $G_n \stackrel{w}{\to} G$, we obtain

$$f(x-\varepsilon) \geq \limsup_n f_n(x) \geq \liminf_n f_n(x) \geq f(x+\varepsilon), \quad \forall\, \varepsilon \in (0,x) \ .$$

This implies $f_n(x) \to f(x)$ at all continuity points of f. Then *a)* follows by Scheffé lemma, since the set of discontinuities is at most countable and

$$\int_0^\infty f_n(x) dx = \int_0^\infty f(x) dx = 1 \ .$$

For any $\psi(x) \in C_b^\infty(\mathbb{R}^+,\mathbb{R})$ (the space of bounded smooth functions), by Fubini's Theorem

$$\int_0^\infty \psi(x)F(dx) = \int_0^\infty \psi(x)f(x)dx = \int_0^\infty \psi(x)\left(\int_x^\infty y^{-1}G(dy)\right)dx = \tag{19}$$
$$\int_0^\infty \left(y^{-1}\int_0^y \psi(x)dx\right)G(dy) = \int_0^\infty \varphi(y)G(dy)$$

where

$$\varphi(y) := y^{-1}\int_0^y \psi(x)dx \in C_b^\infty(\mathbb{R}^+,\mathbb{R})$$

satisfies $\varphi(0) = \psi(0), \varphi'(0) = \frac{1}{2}\psi'(0)$ and $\varphi(x) + x\varphi'(x)$. Since the class $C_b^\infty(\mathbb{R}^+,\mathbb{R})$ determines weak convergence, we see that $G_n \stackrel{w}{\to} G \Longrightarrow F_n \stackrel{w}{\to} F$.

In the other direction, for $\varphi(x) \in C_0^\infty(\mathbb{R}^+,\mathbb{R})$ (the space of smooth functions with compact support), define

$$\psi(x) := \left(\varphi(x) + x\varphi'(x)\right) \in C_0^\infty(\mathbb{R}^+,\mathbb{R}) \ .$$

The functions ϕ and ψ satisfy the integral relation (19). Since $C_0^\infty(\mathbb{R}^+,\mathbb{R})$ is also a convergence determining class when the limit is a proper probability distribution, Kallenberg (2002) Lemma 5.20, we obtain the implication $F_n \stackrel{w}{\to} F \Longrightarrow G_n \stackrel{w}{\to} G$ □

Posterior consistency of logistic Gaussian process priors has been studied by Tokdar and Ghosh (2007), by using an upper bound for the Kullback information between $f^0, f \in C([0,T])$ involving the supremum norm. Since we work with discontinuous densities, we shall use different inequalities.

Lemma 2 *Consider functions* $Z, Z^0 : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ *satisfying*

$$\int_{\mathbb{R}^d} \exp(-Z_s)ds < \infty \ \textit{and} \ \int_{\mathbb{R}^d} \exp(-Z_s^0)ds < \infty \ .$$

The probability densities

$$f(t) := \exp(-Z_t)\left(\int_{\mathbb{R}^d} \exp(-Z_s)ds\right)^{-1}, \quad f^0(t) := \exp(-Z_t^0)\left(\int_{\mathbb{R}^d} \exp(-Z_s^0)ds\right)^{-1}$$

satisfy the inequalities

$$2\left(\int_{\mathbb{R}^d} |f(s) - f^0(s)|ds\right)^2 \le KL(f^0, f) \le \int_{\mathbb{R}^d} \left\{\exp(Z_s^0 - Z_s) - 1 + Z_s - Z_s^0\right\} f^0(s)ds \tag{20}$$

$$\le \frac{1}{2}\int_{\mathbb{R}^d} (Z_s - Z_s^0)^2 f^0(s)ds \ , \tag{21}$$

where (21) *holds under the additional condition*

$$Z_t^0 \le Z_t \quad \forall\, t \textit{ such that } Z_t^0 < \infty \ . \tag{22}$$

Proof: The left side of (20) is Pinsker's inequality. By using Jensen's inequality and the bounds $\log x \leq (x-1)$ and $\exp(-x) \leq (1-x+x^2/2)$ for $x \geq 0$,

$$KL(f^0,f) = \int_{\mathbb{R}^d} (Z_s - Z_s^0) f^0(s) ds + \log\left(\int_{\mathbb{R}^d} \exp\big(Z_s^0 - Z_s\big) f^0(s) ds \right) \leq$$
$$\int_{\mathbb{R}^d} \big\{ \exp(Z_s^0 - Z_s^0) - 1 + Z_s - Z_s^0 \big\} f^0(s) ds \leq \frac{1}{2} \int_{\mathbb{R}^d} (Z_s - Z_s^0)^2 f^0(s) ds \,,$$

where the last inequality holds under assumption (22) □

The next lemma is about the pathwise support of a purely discontinuous subordinator under the uniform norm, and we will use it combined with Lemma 2 to study the KL-support of the corresponding logistic subordinator prior. For related results on the pathwise support of jump processes in different topologies, see Simon (2000); Ishikawa (2013).

Lemma 3 *Let $(Z_t : t \in [0,T])$ a purely discontinuous subordinator with prior Π, such that $Z(0) = 0$ and the compensating measure satisfies $\nu((0,\varepsilon]) > 0 \; \forall \, \varepsilon > 0$. The support of Π in the supremum norm topology contains the continuous non-decreasing functions $Z^0 : [0,T] \to \mathbb{R}^+$ with $Z^0(0) = 0$.*

Proof. For $B \in \mathscr{B}(\mathbb{R}^+)$ we introduce the subordinator component

$$Z_t^B = \sum_{0<s\leq t} \mathbf{1}(\Delta Z_s \in B) \Delta Z_s \,,$$

and denote by Π^B the law of the subordinator component Z_t^B on the space of cadlag paths $D([0,T])$.

For $0 < \delta < \varepsilon$ with $(0,\varepsilon) \subseteq \mathrm{supp}(\nu)$, the subordinator has decomposition

$$Z_t = Z_t^{(0,\delta]} + Z_t^{(\delta,\varepsilon)} + Z_t^{[\varepsilon,\infty)},$$

into independent components satisfying

$$\Pi\big(Z_t^{(0,\delta]} \leq \eta\big) > 0 \quad \text{and} \quad \Pi\big(Z_t^{[\varepsilon,\infty)} = 0\big) > 0 \quad \forall \, \eta, t > 0 \,.$$

Since the space of absolutely continuous functions is dense in $C([0,T])$ with respect to the supremum norm, without loss of generality we consider a function $Z^0 : [0,T] \to [0,\infty]$, which is non-decreasing and is absolutely continuous w.r.t. Lebesgue measure, with $Z^0(0) = 0$. We construct a tracking Markov jump process $\widetilde{Z}_t^{(\delta,\varepsilon)}$, with jumps $\Delta Z_t \in (\delta,\varepsilon)$, such that

$$\widetilde{\Pi}^{(\delta,\varepsilon)} \left(\sup_{t\in[0,T]} |\widetilde{Z}_t^{(\delta,\varepsilon)} - Z_t^0| \leq \varepsilon \right) = 1$$

and the law $\widetilde{\Pi}^{(\delta,\varepsilon)}$ of the tracking Markov process $\widetilde{Z}_t^{(\delta,\varepsilon)}$ is absolutely continuous with respect to the law $\Pi^{(\delta,\varepsilon)}$ of the subordinator component $Z_t^{(\delta,\varepsilon)}$.

Namely we specify the predictable measure compensating the jumps of $\widetilde{Z}_t^{(\delta,\varepsilon)}$ as

$$\widetilde{\nu}^{(\delta,\varepsilon)}(\omega; dz, dt) = \frac{\mathbf{1}\big(\delta < z < \varepsilon + Z^0 - \widetilde{Z}_{t-}^{(\delta,\varepsilon)}(\omega)\big)}{\big(\varepsilon + \widetilde{Z}_{t-}^{(\delta,\varepsilon)}(\omega) - Z_t^0\big) \; \nu\big((\delta, \varepsilon + Z_t^0 - \widetilde{Z}_{t-}^{(\delta,\varepsilon)}(\omega))\big)} \nu(dz) Z^0(dt) \,.$$

Let us have a closer look at the dynamics of the tracking process: it starts with

$$0 = Z_0^{(\delta,\varepsilon)}(\omega) \geq (Z_0^0 - \varepsilon) = -\varepsilon \ .$$

Between two consecutive jumps $T_0 = 0 < T_{k-1} < T_k$, the compensator of T_k, which is given by

$$\int_{T_{k-1}}^{T_k\wedge t} \big(\widetilde{Z}_{s-}^{(\delta,\varepsilon)} - Z_s^0 + \varepsilon\big)^{-1} Z^0(ds) = \log\big(\widetilde{Z}_{T_{k-1}}^{(\delta,\varepsilon)} - Z_{T_{k-1}}^0 + \varepsilon\big) - \log\big(\widetilde{Z}_{T_{k-1}}^{(\delta,\varepsilon)} - Z_{T_k\wedge t}^0 + \varepsilon\big) \ ,$$

explodes as the function $(Z_t^0 - \varepsilon)$ grows towards the piecewise constant process $\widetilde{Z}_{t-}^{(\delta,\varepsilon)}$. Consequently, with probability 1, $\widetilde{Z}_t^{(\delta,\varepsilon)}$ will jump upwards before being reached by the function $(Z_t^0 - \varepsilon)$. Since the jump size is constrained, after the jump necessarily

$$Z_t^0 - \varepsilon < \widetilde{Z}_t^{(\delta,\varepsilon)} < Z_t^0 + \varepsilon \ .$$

Note also that $\Delta\widetilde{Z}_t^{(\delta,\varepsilon)} > \delta > 0$, and the tracking Markov process will have at most $(Z_T^0 + \varepsilon)/\delta$ jumps in the time interval $[0,T]$. This implies that the law of $\widetilde{Z}_t^{(\delta,\varepsilon)}$ is absolutely continuous with respect to the law of the subordinator component $Z_t^{(\delta,\varepsilon)}$, with Radon-Nikodym derivatives satisfying $\widetilde{\Pi}^{(\delta,\varepsilon)}$-almost surely

$$\frac{d\widetilde{\Pi}_T^{(\delta,\varepsilon)}}{d\Pi_T^{(\delta,\varepsilon)}}(\omega) = \frac{\big(\widetilde{Z}_{T_{N_T}}^{(\delta,\varepsilon)}(\omega) - Z_T^0 + \varepsilon\big)}{\big(\widetilde{Z}_{T_{N_T}}^{(\delta,\varepsilon)}(\omega) - Z_{T_{N_T}}^0 + \varepsilon\big)} \times$$

$$\prod_{k=1}^{N_T(\omega)} \left\{ \frac{\mathbf{1}\big(\delta < \Delta\widetilde{Z}_{T_k}^{(\delta,\varepsilon)}(\omega) < \varepsilon + Z_{T_k}^0 - \widetilde{Z}_{T_{k-1}}^{(\delta,\varepsilon)}(\omega)\big)}{\big(\widetilde{Z}_{T_{k-1}}^{(\delta,\varepsilon)}(\omega) - Z_{T_{k-1}}^0 + \varepsilon\big)\nu\big((\delta, \varepsilon + Z_{T_k}^0 - \widetilde{Z}_{T_k-}^{(\delta,\varepsilon)}(\omega))\big)} \frac{dZ^0}{dt}(T_k) \right\} < \infty,$$

$$\text{and} \quad \frac{d\Pi_T^{(\delta,\varepsilon)}}{d\widetilde{\Pi}_T^{(\delta,\varepsilon)}}(\omega) = \left(\frac{d\widetilde{\Pi}_T^{(\delta,\varepsilon)}}{d\Pi_T^{(\delta,\varepsilon)}}(\omega)\right)^{-1} > 0 \ ,$$

where $N_t(\omega) = \#\big\{k : T_k(\omega) \leq t\big\}$ counts the jumps the tracking process. Consider now the event

$$A_T^{(\delta,\varepsilon)} = \left\{ \omega \in D([0,T]) : \sup_{t\in[0,T]} \big|Z_t^{(\delta,\varepsilon)}(\omega) - Z_t^0\big| \leq \varepsilon \right\}$$

with $\widetilde{\Pi}^{(\delta,\varepsilon)}\big(A_T^{(\delta,\varepsilon)}\big) = 1$. By taking the Lebesgue decomposition of the subordinator prior Π with respect to the law of the tracking process we deduce that

$$\Pi^{(\delta,\varepsilon)}\big(A_T^{(\delta,\varepsilon)}\big) =$$

$$\int_{D([0,T])} \left(\frac{d\Pi_T^{(\delta,\varepsilon)}}{d\widetilde{\Pi}_T^{(\delta,\varepsilon)}}(\omega)\right)\widetilde{\Pi}^{(\delta,\varepsilon)}(d\omega) + \Pi^{(\delta,\varepsilon)}\left(A_T^{(\delta,\varepsilon)} \cap \left\{\frac{d\widetilde{\Pi}_T^{(\delta,\varepsilon)}}{d\Pi_T^{(\delta,\varepsilon)}} = 0\right\}\right) > 0 \ .$$

To conclude, note that

$$\left\{\omega \in D([0,T]) : \sup_{t\in[0,T]} \big|Z_t(\omega) - Z_t^0\big| < 2\varepsilon \right\} \supseteq A_T^{(\delta,\varepsilon)} \cap \big\{Z_T^{(0,\delta]} \leq \varepsilon\big\} \cap \big\{Z_T^{[\varepsilon,\infty)} = 0\big\} \ ,$$

and the events on the right hand side are Π-independent with strictly positive probabilities $\square$

Finally we prove the main result of this appendix, which by Theorem 1 and by Lemma 1 implies strong consistency of the logistic subordinator prior when the true density is bounded, non-increasing, and with compact support.

Proposition 1 *Let $(Z_t : t \geq 0)$ a purely discontinuous subordinator with prior Π, with $Z_0 = 0$ and compensating measure satisfying $\nu((0,\varepsilon)) > 0 \ \forall\, \varepsilon > 0$. Then any bounded non-increasing probability density f^0 on $\mathbb{R}^+$ with compact support is in the KL-support of the logistic subordinator prior determined by Π.*

Proof: Let

$$T = \inf\{t \geq 0 : \ f^0(t) = 0\} < \infty$$

and for $\varepsilon > 0$, introduce the non-decreasing function

$$Z^0(t) = \big(\log f^0(0) - \log f^0(t) - \varepsilon\big) \geq -\varepsilon \ , \ t \in [0,T] \ .$$

Note that we could have either $f^0(T-) > 0$ and $Z^0(T-) < \infty$, or $f^0(T-) = 0$ and $Z^0(T-) = \infty$. Our argument covers both situations.

Since $\lim\limits_{x\to+\infty} e^x(e^{-x} - 1 + x) \to 1$, for some $T' = T'_\varepsilon \in (0,T)$ we have

$$\int_{T'}^{T} \big\{ \exp(Z_t^0 - Z_{T'}^0) + Z_{T'}^0 - Z_t^0 - 1\big\} f^0(t) dt \leq \varepsilon \ . \tag{23}$$

The convolution of Z^0 restricted to the interval $[0,T']$ with the distribution of a random variable uniform on the interval $[-\eta, 0]$ with $\eta > 0$, is an absolutely continuous non-decreasing function $\overline{Z}^0$ such that

$$Z^0(t) \leq \overline{Z}^0(t) \quad \forall\, t \in [0,T'], \quad Z^0(T') = \overline{Z}^0(T'),$$

and, when η is small enough,

$$\int_0^{T'} (\overline{Z}_t^0 - Z_t^0)^2 f^0(t) dt \leq \varepsilon \ .$$

By the construction of Lemma 3, and the independent increments property, the set

$$A_\varepsilon = \big\{\overline{Z}_t^0 \leq Z_t(\omega) \leq \overline{Z}_t^0 + \varepsilon, \, \forall\, t \in [0,T']\big\} \cap \big\{Z_t(\omega) - Z_{T'}(\omega) \leq \varepsilon, \, \forall\, t \in [T',T]\big\} \subset D([0,1])$$

has prior probability $\Pi(A_\varepsilon) > 0$. Since $Z_t^0 \leq \overline{Z}_t^0 \leq Z_t(\omega) \leq (\overline{Z}_t^0 + \varepsilon)$ for $\omega \in A_\varepsilon$ and $t \in [0,T']$, by using the inequalities from Lemma 2 together with (23), it follows that

$$\begin{aligned}
&KL\big(f^0, f(\omega)\big) \leq \\
&\int_0^{T'} \big(Z_t(\omega) - Z_t^0\big)^2 f^0(s) ds + \int_{T'}^{T} \big\{ \exp\big(Z_t^0 - Z_t(\omega)\big) + Z_t(\omega) - Z_t^0 - 1\big\} f^0(t) dt \\
&\leq 2\int_0^{T'} \big(Z_t(\omega) - \overline{Z}_t^0\big)^2 f^0(t) dt + 2\int_0^{T'} \big(\overline{Z}_t^0 - Z_t^0\big)^2 f^0(t) dt \\
&+ \int_{T'}^{T} \big\{ \exp\big(Z_t^0 - Z_{T'}^0\big) + Z_{T'}^0 - Z_t^0 - 1 + 2\varepsilon\big\} f^0(t) dt \leq 2\varepsilon^2 + 5\varepsilon \ , \quad \forall\, \omega \in A_\varepsilon \ ,
\end{aligned}$$

which implies $\forall\, \varepsilon > 0$

$$\Pi\big(K(f^0, f) \le 2\varepsilon^2 + 5\varepsilon\big) \ge \Pi(A_\varepsilon) > 0 \quad \square$$

Remark Proposition 1 does not apply to unbounded densities and densities with non-compact support. However the densities of form (2) are bounded, and in lifetime data studies one can safely restrict the analysis to densities with compact support.

Corollary 1 *Consider the probability distribution function*

$$H^0(t) = 1 - \exp(-Z^0(t)), \quad t \ge 0, \ \textit{with} \quad Z^0(0) = H^0(0) = 0\,.$$

Under the assumptions of Proposition 1, the posterior distribution of $H(t) = (1 - \exp(Z(t)))$ *is weakly consistent at* f^0*, meaning that* $\forall\, \varepsilon > 0$ *and* $\psi \in C_b(\mathbb{R}^+, \mathbb{R})$*,*

$$\Pi\left(\left| \int_0^\infty \psi(x) H(dx) - \int_0^\infty \psi(x) H^0(dx) \right| > \varepsilon \middle| Y_1, \ldots, Y_n \right) \to 0$$

$P^\infty_{f^0}$*-almost surely, where the observations* $(Y_i : i \in \mathbb{N})$ *are i.i.d. with density* f^0*.*

Proof: It is enough to show that $G_n \stackrel{w}{\to} G^0 \implies H_n \stackrel{w}{\to} H^0$, where, in the notations of Lemma 1,

$$H^0(dx) = \mu^0 x^{-1} G^0(dx),\ H_n(dx) = \mu_n x^{-1} G_n(dx),\ x > 0, \tag{24}$$

$$\text{with } \mu^0 = \left(\int_0^\infty x^{-1} G^0(dx) \right)^{-1} \text{ and } \mu_n = \left(\int_0^\infty x^{-1} G_n(dx) \right)^{-1}. \tag{25}$$

We use Skorokhod representation, and construct on the same probability space $(\Omega, \mathscr{F}, \mathbb{P})$ random variables $U(\omega), \xi^0(\omega), \xi_n(\omega), n \in \mathbb{N}$ such that

$$\mathbb{P}(\xi^0 \le t) = G^0(t),\ \mathbb{P}(\xi_n \le t) = G_n(t), \text{ and } \xi_n(\omega) \to \xi(\omega)\ \mathbb{P}\text{-almost surely},$$

while U is uniformly distributed on $[0,1]$ and independent from $(\xi^0, \xi_n : n \in \mathbb{N})$.

Consider now the r.v. $\eta^0(\omega) = U(\omega)\xi^0(\omega)$ and $\eta_n(\omega) = U(\omega)\xi_n(\omega)$, which have non-increasing probability densities

$$p_{\eta^0}(y) = 1 - H^0(y) \quad \text{and} \quad p_{\eta_n}(y) = 1 - H_n(y), \quad \text{respectively.}$$

Now $\eta_n(\omega) \to \eta^0(\omega)$ $\mathbb{P}$-almost surely, which implies convergence in distribution, and by Lemma 1 we obtain

$$\int_0^\infty |H_n(y) - H^0(y)| dy \to 0 \qquad \text{as } n \to \infty\,.$$

Finally, $\forall\ \psi \in C_0^\infty(\mathbb{R}^+, \mathbb{R})$, (the space of smooth functions with compact support), integrating by parts we obtain

$$\left| \int_0^\infty \psi(x)(H_n(dx) - H^0(dx)) \right| = \left| \int_0^\infty \psi'(x)(H^0(x) - H_n(x)) dx \right|$$
$$\le \| \psi' \|_\infty \| H^0 - H_n \|_{L^1(\mathbb{R}^+)} \to 0\,,$$

which means $H_n \stackrel{w}{\to} H^0$, since $C_0^\infty(\mathbb{R}^+, \mathbb{R})$ is a convergence determining class when the limiting distribution is proper $\square$

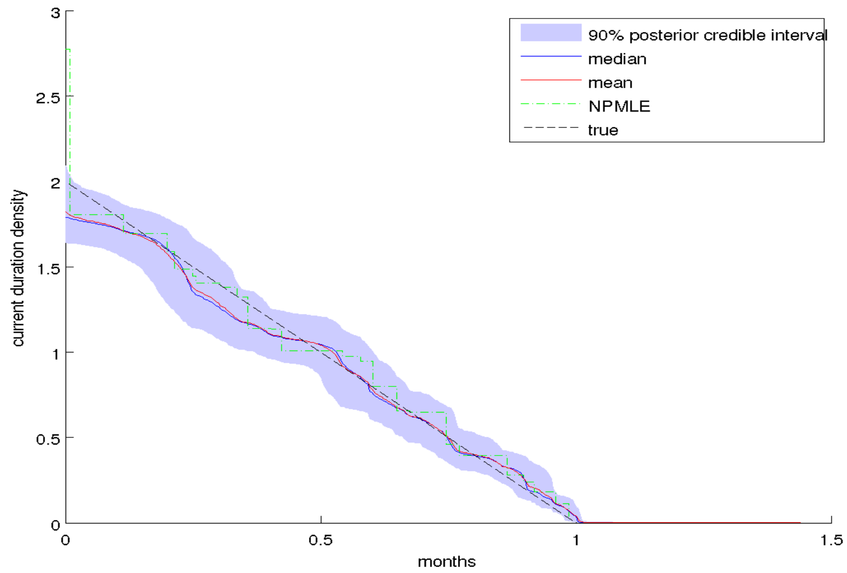

Fig. 1 current duration density estimators, under LGGC prior, $a = 1.5$, with augmentation, without cut-off, simulated data from $U([0,1])$, $n = 10000$.

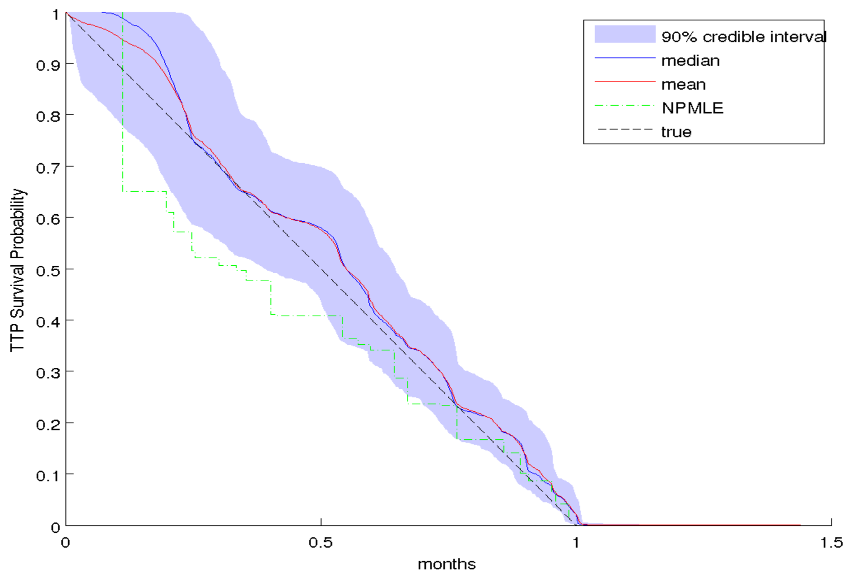

Fig. 2 TTP-survival estimators, under LGGC prior, $a = 1.5$, with augmentation, without cut-off, simulated data from $U([0,1])$, $n = 10000$.

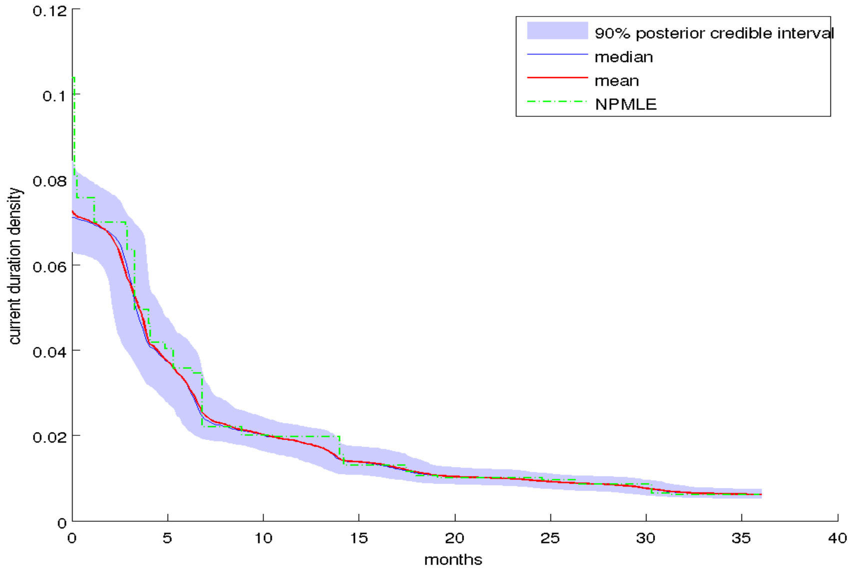

Fig. 3 current duration density estimators, under LGGC prior, $a = 0.1$, with augmentation, without cut-off.

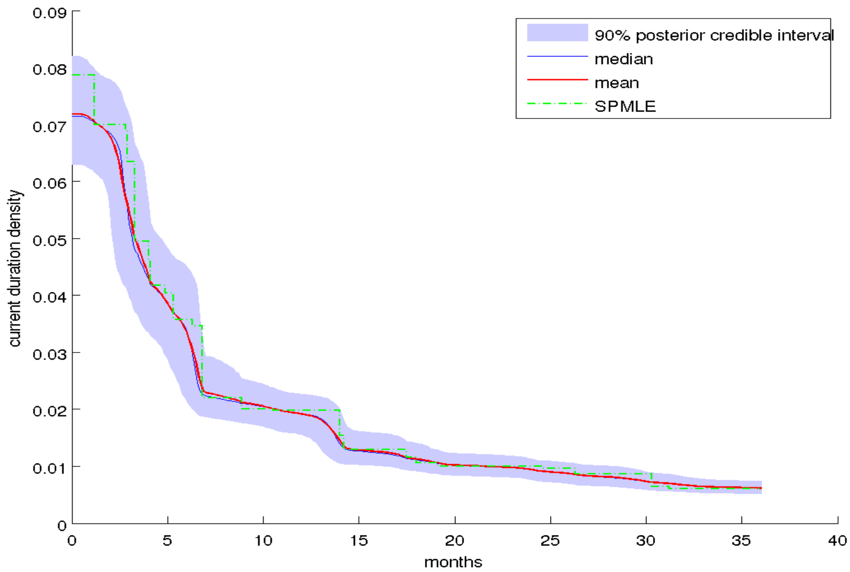

Fig. 4 current duration density estimators, under LGGC prior, $a = 0.1$, with augmentation, cut-off $t_0 = 2$ weeks.

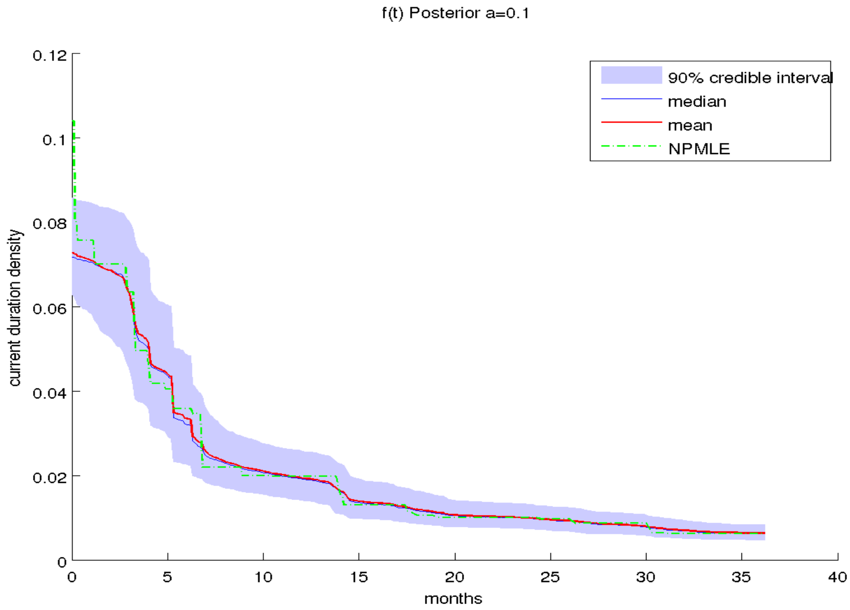

Fig. 5 current duration density, under LGGC prior, $a = 0.1$, without augmentation, without cut-off.

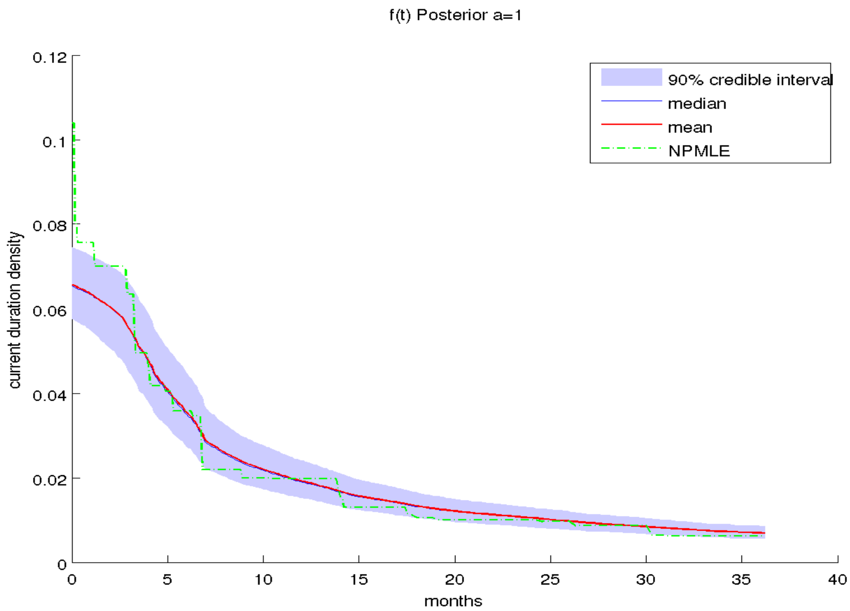

Fig. 6 current duration density, under LGGC prior, $a = 1$, without augmentation, without cut-off.

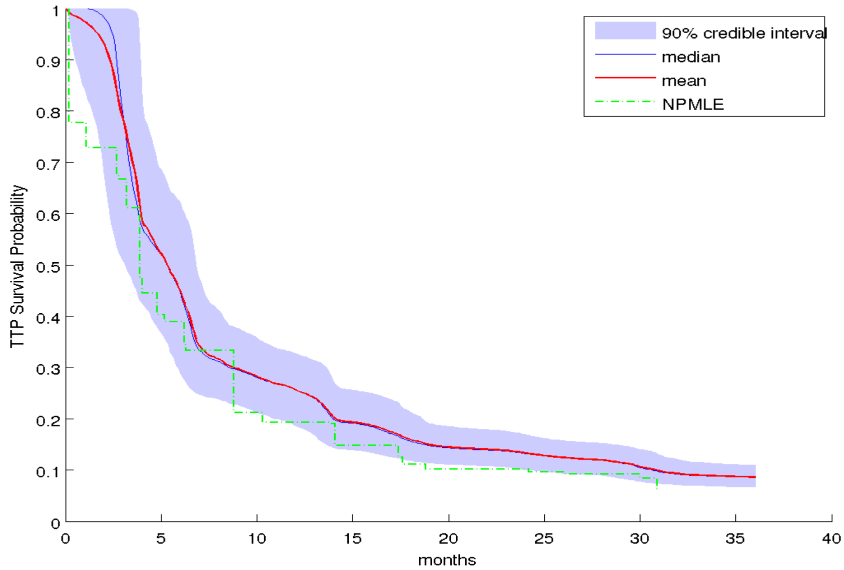

Fig. 7 TTP-survival estimators, under LGGC prior, with augmentation, $a = 0.1$, without cut-off.

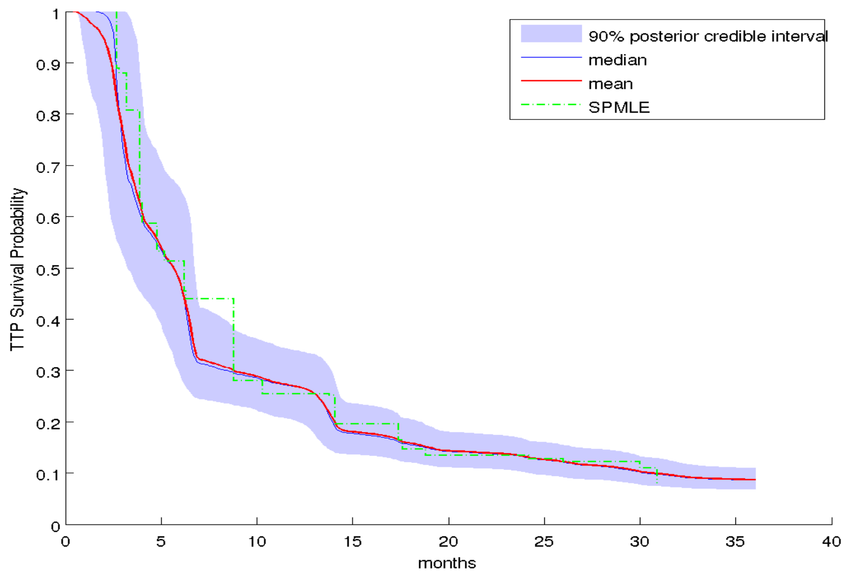

Fig. 8 TTP-survival estimators, under LGGC prior, with augmentation, $a = 0.1$, cut-off $t_0 = 2$ weeks.

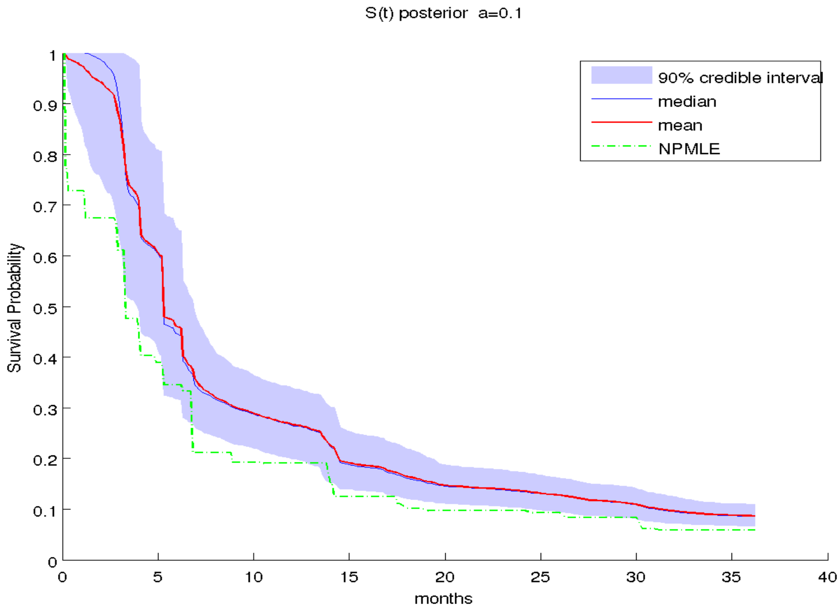

Fig. 9 TTP survival,under LGGC prior, without augmentation, $a = 0.1$, without cut-off.

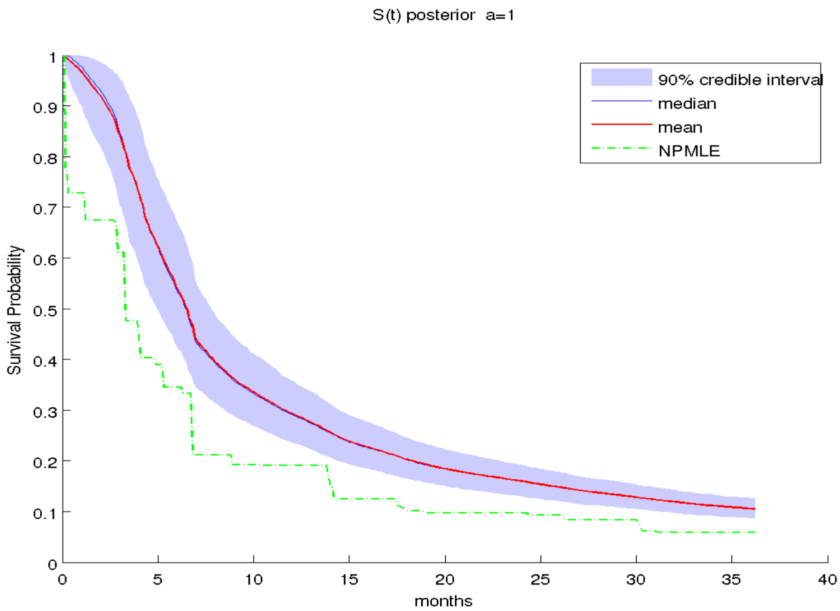

Fig. 10 TTP survival, under LGGC prior, without augmentation, $a = 1$, without cut-off.

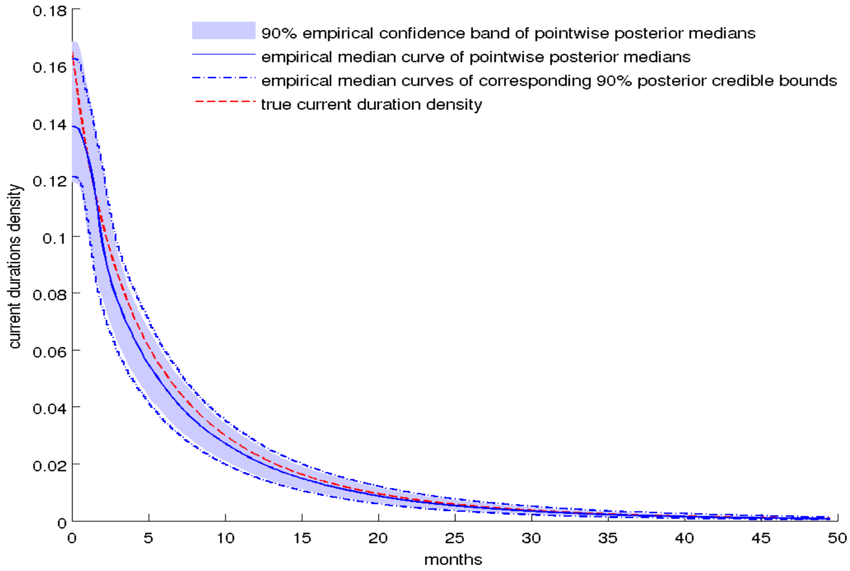

Fig. 11 Results from a simulation experiment on estimating the current duration density, based on $M = 4050$ replicated datasets of size $n = 867$, under LGGC prior with $a = 0.05$.

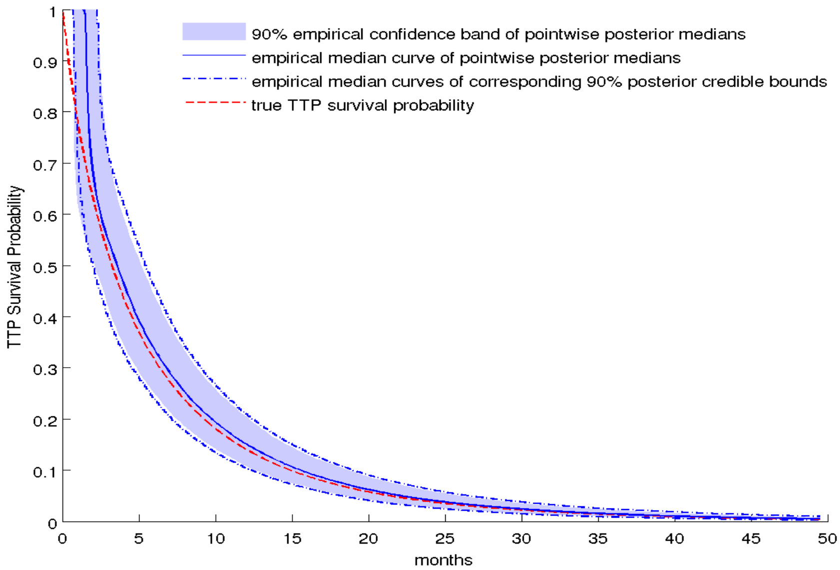

Fig. 12 Results from a simulation experiment on estimating the TTP distribution, based on $M = 4050$ replicated datasets of size $n = 867$, under LGGC prior with $a = 0.05$.