




# Estimating duration distribution aided by auxiliary longitudinal measures in presence of missing time origin

Yi Xiong<sup>1</sup> · W. John Braun<sup>2</sup> · X. Joan Hu<sup>1</sup> 

Received: 18 November 2019 / Accepted: 15 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

Understanding the distribution of an event duration time is essential in many studies. The exact time to the event is often unavailable, and thus so is the full event duration. By linking relevant longitudinal measures to the event duration, we propose to estimate the duration distribution via the first-hitting-time model (e.g. Lee and Whitmore in *Stat Sci* 21(4):501–513, 2006). The longitudinal measures are assumed to follow a Wiener process with random drift. We apply a variant of the MCEM algorithm to compute likelihood-based estimators of the parameters in the longitudinal process model. This allows us to adapt the well-known empirical distribution function to estimate the duration distribution in the presence of missing time origin. Estimators with smooth realizations can then be obtained by conventional smoothing techniques. We establish the consistency and weak convergence of the proposed distribution estimator and present its variance estimation. We use a collection of wildland fire records from Alberta, Canada to motivate and illustrate the proposed approach. The finite-sample performance of the proposed estimator is examined by simulation. Viewing the available data as interval-censored times, we show that the proposed estimator can be more efficient than the well-established Turnbull estimator, an alternative that is often applied in such situations.

**Keywords** First-hitting-time model · Interval censoring · Joint modelling · Mixed effects model · Wiener process with random drift

---

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10985-021-09520-w>.

---

✉ X. Joan Hu  
joanh@stat.sfu.ca

<sup>1</sup> Simon Fraser University, Burnaby, Canada

<sup>2</sup> University of British Columbia - Kelowna, Kelowna, Canada

## 1 Introduction

The patterns of event duration times are of primary interest in many research studies. A close follow-up of each study individual who potentially experiences the event of interest is commonly unrealistic. Although the occurrence of the event may be reported, its start time is often unavailable, especially with so-called “silent” event occurrence (e.g. Balasubramanian and Lagakos 2003).

For example, it is important for the prediction of future fire growth and the allocation of suppression resources to understand the distribution of the duration from the fire start time to the time when the work of suppressing the fire begins, i.e. the initial attack time. A wildland fire is usually reported to the fire management agency by look-out towers or people in the area, and the fire manager then dispatches fire-fighting resources (e.g. Martell 2007; Morin 2014). The exact time when the fire starts is often unknown; instead, recorded is the time when the fire is reported. As another example, many HIV/AIDS studies are concerned with the duration of HIV infection from the time of infection to the onset of an AIDS event; see, for example, Degruittola et al. (1991) and Doksum and Normand (1995). Often the infection is detected at a time considerably later after it takes place and thus the exact HIV infection time is usually unavailable. Time to COVID-19 infection is a more recent example of this phenomenon.

A practical approach to handling data with missing time origins ignores the *reporting delay* and performs inference on the duration distribution using the observed portion of the event duration, that is, the duration between the report time and event termination. This *naive* approach yields apparently biased inference when the time gap is nonignorable between the event onset to when it is reported. Another commonly used approach to handling times with missing time origin is to view the observation on the time origin subject to interval-censoring. Thus the lower limit of the interval is the length of time that the event has been observed (which we refer to as  $L^*$ ), and the upper limit is the sum of the observed duration and the longest possible reporting delay (which we refer to as  $R_{max}$ ). Turnbull’s nonparametric maximum likelihood estimator (NPML; Turnbull 1976) can then be employed to estimate the distribution of the actual duration (which we refer to as  $L$ ) using such manufactured interval-censored data. The resulting inference can be unsatisfactory, especially when the longest expected reporting delay is large relative to the observed portion of the event duration. In addition, the interval-censoring is likely informative in many situations, which invalidates Turnbull’s estimator. For example, fires can occur at varying distances from fire management resources. It results in a reporting delay  $S$  and an observed duration  $L^*$  varying together, and thus  $L^*$  and  $S$  may not be independent. That is, the interval  $[L^*, L^* + R_{max}]$  provides information on  $L$  additional to the fact of  $L \in [L^*, L^* + R_{max}]$ . This paper considers estimation of the event duration distribution with the aforementioned type of event time data under a *first-hitting-time* model using the event associated longitudinal measures.

Many studies have readily available longitudinal measures associated with the event of interest. In reliability, for example, Lu and Meeker (1993) use degradation data to estimate the distribution of a failure time, taking the failure time to be the time when the degradation path hits a critical level. The concept of *first-hitting-time* has been widely applied. Various models have been used to formulate longitudinal measures,

such as a Gamma process (e.g. Lawless and Crowder 2004), a Wiener process (e.g. Doksum and Hoyland 1992; Horrocks and Thompson 2004; Lee and Whitmore 2006; Pennell et al. 2009; Choi et al. 2014; Mulatya et al. 2016), and an inverse Gaussian process (e.g. Peng 2015).

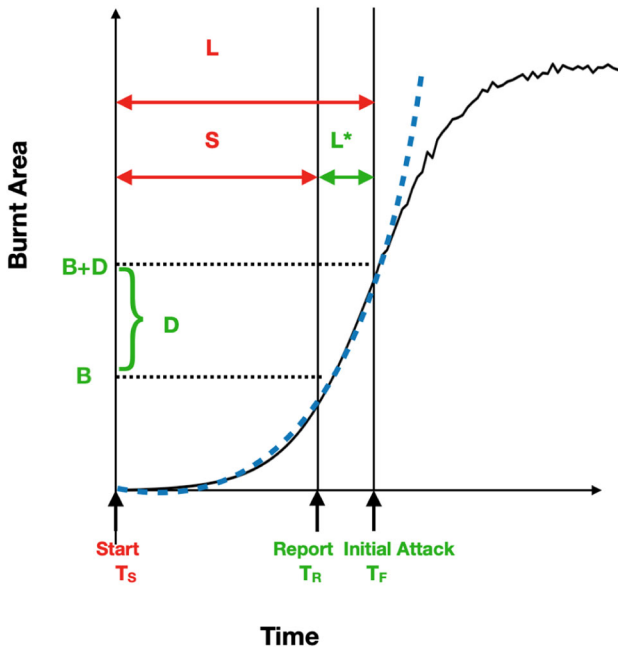
We aim to make inference on the population of all reported fires. Most of the *first-hitting-time* models formulate the event time via a hypothetical/underlying process. Our modelling is similar to the excellent exception presented in Mulatya et al. (2016). Using Brownian motion with random drift, we link the readily available longitudinal measures to the recorded times to locate the missing time origin. The issue of dependence between the reporting delay and the observed portion of duration is handled by conditioning on the random drift. We adapt the empirical distribution function, which requires independent and identically distributed (iid) observations, into an intuitive and easy-to-implement estimator for the distribution based on the event duration with missing time origin. Conventional smoothing techniques, such as kernel methods, can be straightforwardly applied to smooth the proposed estimator. A collection of wildland fire records from Alberta, Canada is used to motivate and present the proposed approach. However, potential applications of our approach are broad and not limited to wildland fire management studies.

The rest of the paper is organized as follows. Section 2 introduces a model for longitudinal measures of fire burnt area over time and proposes an estimator for the duration distribution aided by the longitudinal model using duration times with missing origins. It is straightforward to evaluate smoothed versions of the proposed estimator. We present procedures for estimating the parameters that are involved in the longitudinal model and required by the estimator for the duration distribution. We then derive the asymptotic properties of the distribution estimator and its variance estimation. Section 3 reports an analysis of wildland fire records with the proposed approach and Sect. 4 presents simulation studies conducted to examine finite-sample performance of the proposed estimator regarding its consistency, efficiency, and robustness. We also compare the performance of the proposed approach with that of the naive approach and of the Turnbull estimator. Some final remarks are given in Sect. 5.

## 2 Estimation of duration distribution in the presence of missing time origin

### 2.1 Notation and model

We formulate the aforementioned statistical problem in terms of wildland fire management. Following Parks (1964), Fig. 1 provides a description for the development of a hypothetical wildland fire via its progression of burnt area over time. The solid curve in the figure represents the burnt area overtime of a fire that is subject to suppression after detection. The time point when suppression begins is referred to as the time of *initial attack*. The dashed curve shows the expected trajectory of the fire's burnt area if it had continued to burn without any suppression or intervention. After ignition, the burnt area grows nonlinearly in time, and can be well approximated as exponential initially. Prior to initial attack, the dashed curve and the solid curve coincide. The



**Fig. 1** Hypothetical description of a wildfire progression through fire management phases following Parks (1964). The solid curve represents the burnt area of a fire that is reported and then dispatched with suppression resources. The dashed curve represents the burnt area of the fire that had continued to burn without any suppression

times  $T_S$ ,  $T_R$ , and  $T_F$  in Fig. 1 are the calendar times when a fire starts, is reported, and initial attack begins, respectively.

The *duration* time of interest, denoted by  $L$ , is the elapsed time from the fire’s start to the time of initial attack, i.e.,  $L = T_F - T_S$ . We are concerned with situations where the true fire start time  $T_S$  is unavailable, and thus the time origin of the event duration is missing. Denote the unobservable reporting delay by  $S = T_R - T_S$ . The observed portion of the duration is  $L^* = T_F - T_R = L - S$ , the period between report time and initial attack time. Let  $A(u)$  be the burnt area at time  $u$ , where we assume there is no burnt area at the start time, i.e.  $A(u) = 0$  when  $u = 0$ . Let  $B = A(S)$  and  $D = A(L) - A(S)$  denote the burnt area at the report time and the increase in area by the initial attack time, respectively.

Consider a collection of  $n$  independent wildland fires. We assume that the natural logarithm of fire  $i$ ’s burnt area is  $A_i(u) = g_i(u) + \sigma_i W_i(u)$  for  $i = 1, \dots, n$ , where  $g_i(u)$  is a nondecreasing function with  $g_i(0) = 0$  and  $W_i(\cdot)$  is the standard Wiener process. As a fire usually grows unhindered until initial attack, we suppose  $\sigma_i \equiv \sigma$  and use a linear approximation to  $g_i(u)$  with random drift  $v_i = \nu e^{\delta_i}$ , where the constant  $\nu$  is positive, and  $\delta_i$  is independent of  $W_i(\cdot)$  and following a distribution  $\phi(\cdot; \sigma_r)$  with  $E[\delta_i] = 0$  and  $Var[\delta_i] = \sigma_r^2$ . This results in the model considered in this paper:

$$A_i(u) = v_i u + \sigma W_i(u), \quad u \geq 0. \tag{1}$$

The random drift  $v_i$  characterizes the heterogeneity in fire growth among the individual fires. Note that  $v_i$  reduces to a constant drift when  $\sigma_r = 0$ . In the rest of this paper, we assume  $\delta_i \sim N(0, \sigma_r^2)$  with  $\sigma_r \geq 0$ .

Under the Wiener process model for burnt-area (1), the reporting delay  $S_i$  (report time since the fire starts) can be viewed as a first-hitting-time: it is the time when fire  $i$ 's burnt area first reaches the threshold  $B_i$ , the burnt area at the report time:  $S_i = \sup\{u : u > 0, A_i(u) < B_i\}$ , which is the same as  $\inf\{u : u > 0, A_i(u) > B_i\}$  almost surely. According to Chhikara and Folks (1989), the first-hitting-time  $S_i$  follows the inverse Gaussian distribution given threshold  $B_i$ , with the cumulative distribution function (CDF)

$$G(x|B_i, v_i, \sigma) = \Phi \left( \sqrt{\frac{B_i^2}{\sigma^2 x}} \left[ \frac{x v_i}{B_i} - 1 \right] \right) + \exp \left( 2B_i v_i / \sigma^2 \right) \Phi \left( -\sqrt{\frac{B_i^2}{\sigma^2 x}} \left[ \frac{x v_i}{B_i} + 1 \right] \right), \tag{2}$$

where  $\Phi(\cdot)$  is the CDF of the standard normal distribution. Denote the observed data by

$$\text{Observed-data} = \bigcup_{i=1}^n \mathbf{O}_i = \{(T_{Ri}, T_{Fi}, B_i, D_i) : i = 1, 2, \dots, n\}. \tag{3}$$

This paper focuses on estimation of  $F(\cdot)$ , the CDF of the event duration using the observed data (3) with the following assumptions.

ASSUMPTION (A1).  $\{(T_{Ri}, T_{Fi}, B_i, D_i, L_i) : i = 1, 2, \dots, n\}$  is a collection of iid realizations of  $(T_R, T_F, B, D, L)$ , where  $L \sim F(\cdot)$ .

ASSUMPTION (A2).  $L_i^* = T_{Fi} - T_{Ri} = L_i - S_i$  and  $S_i = T_{Ri} - T_{Si}$  are independent conditional on  $v_i$  for  $i = 1, \dots, n$ .

The assumptions may hold to a reasonable level of approximation in many practical situations. In our wildland fire application, the assumption A2 assumes that the reporting delay ( $S_i$ ) of fire  $i$  and the time to the initial attack since being reported ( $L_i^*$ ) are independent conditional on the fire spread rate  $v_i$ . This is plausible since the fire agency often assesses a reported fire regarding its spread rate, and then arranges for the initial attack accordingly. That is,  $L_i^*$  depends likely on  $S_i$  solely through  $v_i$ .

## 2.2 Procedures for estimating $F(\cdot)$

### 2.2.1 Review of the existing approaches

If all duration times  $L_i$  for  $i = 1, \dots, n$  were observed, the empirical distribution function, the nonparametric maximum likelihood estimator (NPMLE) based on the iid observations, could be applied to estimate the duration distribution:  $F_n(t) = \sum_{i=1}^n I(L_i \leq t) / n$ , where  $I(\mathcal{E})$  is the indicator function of event  $\mathcal{E}$ . Since a fire is usually reported after a delay, only  $L_i^*$ , a portion of  $L_i = S_i + L_i^*$ , is recorded. The

aforementioned *naive* estimator is  $F_n^*(t) = \sum_{i=1}^n I(L_i^* \leq t)/n$ . It is clearly biased when  $P(S_i > 0) \neq 0$ .

Observe that  $L_i = L_i^* + S_i \in [L_i^*, L_i^* + R_{max}]$  with  $R_{max}$  the longest possible reporting delay as discussed in the existing literature. The current observations might then be cast as interval-censored event times. Turnbull’s self-consistent estimator (Turnbull 1976) can then be used to estimate the distribution with the interval-censored observations. Let  $0 = \tau_0 < \tau_1 < \dots < \tau_Q$  be the ordered unique values of  $\{L_i^*\}_{i=1}^n, \{L_i^* + R_{max}\}_{i=1}^n$ , and define  $\alpha_{iq} = I\{(\tau_{q-1}, \tau_q] \subseteq (L_i^*, L_i^* + R_{max}]\}$  and  $p_q = F(\tau_q) - F(\tau_{q-1})$ . Following Klein and Moeschberger (2003), the Turnbull estimator is the solution to the equations

$$\begin{aligned}
 p_q &= \frac{1}{n} \sum_{i=1}^n E\left\{I(L_i \in (\tau_{q-1}, \tau_q]) \mid L_i \in (L_i^*, L_i^* + R_{max}]\right\} \\
 &= \frac{1}{n} \sum_{i=1}^n \frac{\alpha_{iq} p_q}{\sum_{k=1}^Q \alpha_{ik} p_k} \tag{4}
 \end{aligned}$$

for  $q = 1, \dots, Q$ . However, the Turnbull estimator may not perform very well in the situations of particular interest in this paper. Note that the Turnbull estimator is not uniquely defined over the whole positive real line but up to an equivalence class of distributions that may only differ over gaps, i.e. the innermost intervals. Since  $R_{max}$  is large relative to  $L_i^*$  in our application, the data form relatively small number of innermost intervals and thus often give a quite noninformative estimate. Moreover, the mechanism of interval-censoring in wildland fire studies may be informative since the observed  $L_i^*$  is often dependent on the reporting delay  $S_i$  through the fire spread rate  $v_i$ . Those considerations motivate us to propose an alternative estimator for the duration distribution  $F(\cdot)$  using available observations on the burnt-area process, which contain information related to the reporting delay.

### 2.2.2 Proposed estimator of $F(\cdot)$

By Model (1) and Assumption (A2), note that  $E\{I(L_i \leq t) \mid \text{Observed-data}\} = P(S_i \leq t - L_i^* \mid \mathbf{O}_i)$  can be expressed as  $\int_{-\infty}^{\infty} G(t - L_i^* \mid B_i, v e^\delta, \sigma) \phi(\delta \mid \mathbf{O}_i; v, \sigma, \sigma_r) d\delta$ , where  $\phi(\cdot \mid \mathbf{O}_i; v, \sigma, \sigma_r)$  is the conditional distribution of  $\delta_i$  given the observed data associated with fire  $i$  as specified in (3). The consideration above suggests the following estimator, provided that the parameters  $v, \sigma, \sigma_r$  are known:

$$\tilde{F}_n(t; v, \sigma, \sigma_r) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} G(t - L_i^*; B_i, v e^\delta, \sigma) \phi(\delta \mid \mathbf{O}_i; v, \sigma, \sigma_r) d\delta. \tag{5}$$

We propose to replace parameters in (5) with their consistent estimators based on the available data. This results in a feasible distribution estimator,

$$\tilde{F}_n(t; \hat{v}, \hat{\sigma}, \hat{\sigma}_r) = \frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} G(t - L_i^* \mid B_i, \hat{v} e^\delta, \hat{\sigma}) \phi(\delta \mid \mathbf{O}_i; \hat{v}, \hat{\sigma}, \hat{\sigma}_r) d\delta, \tag{6}$$

abbreviated by  $\hat{F}_n(t)$  in the rest of this paper. In Sect. 2.3, we present procedures for consistently estimating parameters  $\nu, \sigma, \sigma_r$ . To compute (6) numerically, one may approximate  $\hat{F}_n(t)$  with

$$\frac{1}{nJ} \sum_{j=1}^J \sum_{i=1}^n G(t - L_i^* | B_i, \hat{\nu}e^{\delta_i^{(j)}}, \hat{\sigma}), \tag{7}$$

where  $\delta_i^{(1)}, \dots, \delta_i^{(J)}$  are sampled independently from the estimated conditional distribution  $\phi(\cdot | \mathbf{O}_i; \hat{\nu}, \hat{\sigma}, \hat{\sigma}_r)$  for  $i = 1, \dots, n$ .

The proposed estimator in (5) is adapted from the empirical distribution function. Analogously, we can obtain a smoothed distribution estimator of  $F(\cdot)$  by adapting the kernel distribution estimator with all the duration observed. Recall that a kernel estimator of  $F(\cdot)$  with iid observed duration is  $F_{n,h}(t) = \sum_{i=1}^n K\left(\frac{t-L_i}{h}\right)/n$ , where  $K(t) = \int_{-\infty}^t k(u)du$  with  $k(\cdot)$  a kernel function and  $h$  being the bandwidth (e.g., Rosenblatt 1956). Its projection onto the available data space,  $\sum_{i=1}^n E\left\{K\left(\frac{t-(L_i^*+S_i)}{h}\right) | \mathbf{O}_i\right\}/n$ , yields the following estimator with smooth realizations, denoted by  $\hat{F}_{n,h}(t)$ :

$$\frac{1}{n} \sum_{i=1}^n \int_{-\infty}^{\infty} \int_0^{\infty} K\left(\frac{t - (L_i^* + s)}{h}\right) \phi(\delta | \mathbf{O}_i; \hat{\nu}, \hat{\sigma}, \hat{\sigma}_r) dG(s | B_i, \hat{\nu}e^{\delta}, \hat{\sigma}) d\delta. \tag{8}$$

When one deals with the situation where no random effect is involved and  $S_i$  is assumed to be uniformly distributed over  $[0, R_{max}]$ , the estimator  $\hat{F}_{n,h}(t)$  reduces to the one discussed in Braun et al. (2005). Since the choice of bandwidth  $h$  is still under investigation, we focus on the estimator  $\hat{F}_n(t)$  given in (6) for the rest of the paper.

### 2.3 Procedures for estimating parameters in Model (1): $\theta = (\nu, \sigma, \sigma_r)$

The log-likelihood function based on the available data is

$$\log L_{obs}(\theta | \text{Observed-data}) = \sum_{i=1}^n \log L_{obs}(\theta; \mathbf{O}_i), \tag{9}$$

where the contribution from fire  $i$   $\log L_{obs}(\theta; \mathbf{O}_i)$  is  $\log \int_0^{\infty} \int_{-\infty}^{\infty} \{L_{obs,i|S_i,\delta}\} d[S, \delta]$  with  $L_{obs,i|S_i,\delta_i} = [D_i | L_i^*, \delta_i][B_i | S_i, \delta_i]$ . Here  $[D_i | L_i^*, \delta_i]$  and  $[B_i | S_i, \delta_i]$  are the conditional distribution of  $D_i$  given  $L_i^*, \delta_i$  and the conditional distribution of  $B_i$  given  $S_i$  and  $\delta_i$ , respectively. Under Model (1),  $[D_i | L_i^*, \delta_i]$  and  $[B_i | S_i, \delta_i]$  are both normal, denoted by  $N(\nu e^{\delta_i} L_i^*, \sigma^2 L_i^*)$  and  $N(\nu e^{\delta_i} S_i, \sigma^2 S_i)$ , respectively.

We estimate  $\theta$  by maximizing  $\log L_{obs}(\theta | \text{Observed-data})$ . Denote the resulting estimator by  $\hat{\theta}_n$ . One can use  $\hat{\theta}_n$  together with the collection of  $\underline{\delta}^{(1)}, \dots, \underline{\delta}^{(J)}$  in the last iteration to compute (7) and (8) and to obtain  $\hat{F}_n(\cdot)$  and  $\hat{F}_{n,h}(\cdot)$ , respectively. Here  $\underline{\delta}^{(j)}$  for  $j = 1, \dots, J$  are the  $n$ -dimensional vectors with the  $i$ -th components  $\delta_i^{(j)}$ .

We apply the MCEM algorithm (Wei and Tanner 1990) to compute the MLE, and present details in ALGORITHM A below. The log-likelihood function of  $\theta$  based on the observed data (3) together with  $\underline{S}, \underline{\delta}$  is  $l_F(\theta|\text{Observed-data}, \underline{S}, \underline{\delta}) = l_{F_1}(\nu, \sigma|\underline{S}, \underline{\delta}) + l_{F_2}(\theta; \underline{S}, \underline{\delta})$ , where

$$l_{F_1}(\nu, \sigma|\underline{S}, \underline{\delta}) = -n \log \sigma^2 - \sum_{i=1}^n \frac{(D_i - \nu e^{\delta_i} L_i^*)^2}{2\sigma^2 L_i^*} - \sum_{i=1}^n \frac{(B_i - \nu e^{\delta_i} S_i)^2}{2\sigma^2 S_i}$$

and  $l_{F_2}(\theta; \underline{S}, \underline{\delta}) = \sum_{i=1}^n \log[S_i|\delta_i] + \sum_{i=1}^n \log \phi(\delta_i; \sigma_r)$ .

Algorithm A For  $m = 0, 1, 2, \dots$ , denote the estimate from the  $m$ th iteration by  $\theta^{(m)} = (\nu^{(m)}, \sigma^{(m)}, \sigma_r^{(m)})$ .

E-step. Approximate  $Q(\theta, \theta^{(m)}) = E\{l_F(\theta|\text{Observed-data}, \underline{S}, \underline{\delta})|\mathbf{O}, \theta^{(m)}\}$  as

$$\begin{aligned} \frac{1}{J} \sum_{j=1}^J l_F(\theta|\text{Observed-data}, \underline{S}^{(j)}, \underline{\delta}^{(j)}) &= \frac{1}{J} \sum_{j=1}^J l_{F_1}(\nu, \sigma|\underline{S}^{(j)}, \underline{\delta}^{(j)}) \\ &\quad + \frac{1}{J} \sum_{j=1}^J l_{F_2}(\theta; \underline{S}^{(j)}, \underline{\delta}^{(j)}), \end{aligned} \tag{10}$$

where for  $j = 1, 2, \dots, J$ ,  $(S_i^{(j)}, \delta_i^{(j)})$  is generated from the conditional distribution given the observed data with the current parameter estimate  $\theta^{(m)}$ ,

$$[S, \delta|\mathbf{O}_i; \theta^{(m)}] = \frac{L_{obs,i|S,\delta}(\nu^{(m)}, \sigma^{(m)}; S, \delta)[S, \delta; \theta^{(m)}]}{\int_0^\infty \int_{-\infty}^\infty L_{obs,i|S,\delta}(\nu^{(m)}, \sigma^{(m)}; S, \delta)d[S, \delta; \theta^{(m)}]}. \tag{11}$$

M-step. Maximize (10) with respect to  $\theta$  to obtain  $\theta^{(m+1)}$ .

Repeat Steps E and M until  $\|\theta^{(m+1)} - \theta^{(m)}\| < \epsilon$  for a pre-specified tolerance  $\epsilon$ . The limit of the sequence  $\{\theta^{(m)} : m = 1, 2, \dots\}$  is the MLE  $\hat{\theta}_n$ .

We follow the Metropolis–Hastings algorithm (Metropolis et al. 1953; Hastings 1970) to generate  $(S_i^{(j)}, \delta_i^{(j)})$  from the conditional distribution (11). The details are provided in Sect. S1.1 of the *Supplementary Material*. One should also note that  $[S_i|\delta_i]$  in  $l_{F_2}(\theta; \underline{S}, \underline{\delta})$  equals  $[S_i|B_i, \delta_i][B_i|\delta_i]/[B_i|S_i, \delta_i]$ , which does not have much additional information on the parameters  $\nu, \sigma$ . To ease the computational burden, one may replace (10) with the following to update  $\theta^{(m)}$ :

$$\tilde{Q}(\theta, \theta^{(m)}) = \frac{1}{J} \sum_{j=1}^J l_{F_1}(\nu, \sigma|\underline{S}^{(j)}, \underline{\delta}^{(j)}) + \frac{1}{J} \sum_{j=1}^J \sum_{i=1}^n \log \phi(\delta_i^{(j)}; \sigma_r). \tag{12}$$

The maximizing procedure based on (12) leads to a variant of ALGORITHM A and results in  $\tilde{\theta}_n$ , a close approximation to the MLE  $\hat{\theta}_n$ . For the numerical studies presented in this paper, we choose  $J = 200$  and the algorithm converges with  $J = 200$ .



We may obtain another estimator by maximizing the conditional likelihood function using only the observations on  $D$  and  $L^*$ . The log conditional likelihood function is  $\log L_{obs}^c(\theta | \text{Observed-data}) = \sum_{i=1}^n \log \int_{-\infty}^{\infty} [D_i | L_i^*, \delta] \phi(\delta; \sigma_r) d\delta$ , and can be written as

$$\sum_{i=1}^n \log \left( \int_{-\infty}^{\infty} \left[ \frac{1}{\sqrt{2\pi\sigma^2 L_i^*}} \exp \left\{ -\frac{(D_i - ve^\delta L_i^*)^2}{2\sigma^2 L_i^*} \right\} \right] \phi(\delta; \sigma_r) d\delta \right). \tag{13}$$

The estimator obtained by maximizing (13) is likely less efficient but easier to implement. We describe the procedure to obtain the maximizer of (13) in Sect. S1.2 of the *Supplementary Material*. We refer to the second algorithm as ALGORITHM B and denote estimators obtained from the two algorithms by  $\hat{\theta}_{nA}$  and  $\hat{\theta}_{nB}$  for the remainder of the paper. The estimate obtained from ALGORITHM B may be used as the initial estimate  $\theta^{(0)}$  for ALGORITHM A.

### 2.4 Asymptotic properties of $\hat{F}_n(t)$ and variance estimation

The proposed estimator  $\hat{F}_n(t)$  using  $\hat{\theta}_{nA}$  from ALGORITHM A in Sect. 2.3 has the following asymptotic properties.

**Theorem 1** *Under Assumptions (A1) and (A2) and Conditions (C1)-(C5) for the log-likelihood function in (9), the estimator  $\hat{F}_n(t)$  has the following properties:*

- (i) Strong Consistency.  $\sup_{t \in [0, \tau]} |\hat{F}_n(t) - F(t)| \xrightarrow{P} 0$  as  $n \rightarrow \infty$ .
- (ii) Weak Convergence. For  $t \in [0, \tau]$ , as  $n \rightarrow \infty$ ,  $\sqrt{n}(\hat{F}_n(t) - F(t))$  converges weakly in  $\ell^\infty([0, \tau])$  to a tight, mean-zero Gaussian process  $\mathcal{G}$  with covariance  $Cov(\mathcal{G}(t), \mathcal{G}(s))$  given by

$$\begin{cases} \int_0^\infty \int_0^\infty M(t, l^*, b; \theta_0) M(s, l^*, b; \theta_0) h(l^*, b) dl^* db - F(t)F(s), & t \neq s \\ \int_0^\infty \int_0^\infty M(t, l^*, b; \theta_0)^2 h(l^*, b) dl^* db - F^2(t) \\ + E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta_0) / \partial \theta]' \Pi^{-1}(\theta_0) E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta_0) / \partial \theta], & t = s, \end{cases} \tag{14}$$

where  $\theta_0$  is the true parameter,  $\Pi(\theta_0) = E\{-\partial^2 \log L_{obs}(\theta; \mathbf{O}_i) / \partial \theta \theta'\}$  is the same as  $\Sigma(\theta_0) = Var\{\partial \log L_{obs}(\theta; \mathbf{O}_i) / \partial \theta\}$  with  $\log L_{obs}(\theta; \mathbf{O}_i)$  the contribution from individual  $i$  to the log-likelihood function in (9),  $M(t, L_i^*, B_i; \theta) = \int_{-\infty}^{\infty} G(t - L_i^* | B_i, ve^\delta, \sigma) \phi(\delta | \mathbf{O}_i; v, \sigma, \sigma_r) d\delta$ , and  $h(l^*, b)$  is the joint probability density function of  $L_i^*$  and  $B_i$ .

A proof of Theorem 1 is outlined in the Appendix. It results in a consistent estimator of the covariance function in (14) substituting its unknown elements with their following estimators.

Note that  $\int_0^\infty \int_0^\infty M(t, l^*, b; \theta_0)^2 h(l^*, b) dl^* db$  can be approximated by  $n^{-1} \sum_{i=1}^n [\sum_{k=1}^K G(t - L_i^* | B_i, \hat{v}_n e^{\delta_i^{(k)}}, \hat{\sigma}_n) / K]^2$  with  $\delta_i^{(1)}, \dots, \delta_i^{(K)}$  obtained from the last iteration of ALGORITHM A in Sect. 2.3. We may similarly approximate  $E_{\theta_0}[\partial M(t, L_i^*, B_i;$

$\theta)/\partial\theta]$ . Further, note that  $\widehat{\Pi}_n(\theta_0) = -n^{-1}\partial^2 \log L_{obs}(\theta; \text{Observed-data})/\partial\theta\theta'$  converges in probability to  $\Pi(\theta_0) = \Sigma(\theta_0)$ , and so does  $\widehat{\Sigma}_n(\theta_0) = n^{-1}\text{Var}_{\theta_0}(\partial \log L_{obs}(\theta; \text{Observed-data})/\partial\theta)$ . Thus, either  $\widehat{\Pi}_n(\hat{\theta}_{n_A})$  or  $\widehat{\Sigma}_n(\hat{\theta}_{n_A})$  can be used to estimate  $\Pi(\theta_0) = \Sigma(\theta_0)$ .

### 2.5 Construction of confidence bands for $F(\cdot)$

Based on Theorem 1, we see that the process  $\sqrt{n}(\widehat{F}_n(t) - F(t))/\sqrt{\widehat{\text{var}}(t)}$  converges weakly to the standard Gaussian process, where  $\text{var}(t)$  is  $\text{Cov}(\mathcal{G}(t), \mathcal{G}(s))$  given in (14) with  $t = s$ . Denote the consistent estimator of  $\text{var}(t)$  obtained as described in Sect. 2.4 by  $\widehat{\text{var}}(t)$ . We employ the resampling approach in Hu and Lagakos (1999); Zhao et al. (2008) to construct the following confidence band (CB) for the distribution  $F(\cdot)$ .

The  $(1 - \alpha)$  confidence band for  $F(\cdot)$  is

$$\left\{q(\cdot) : \text{for all } t \in [0, \tau], q(t) \in \left[\widehat{F}_n(t) - c_\alpha \sqrt{\frac{\widehat{\text{var}}(t)}{n}}, \widehat{F}_n(t) + c_\alpha \sqrt{\frac{\widehat{\text{var}}(t)}{n}}\right]\right\},$$

where the critical value  $c_\alpha$  is determined by the resampling scheme as follows. For  $t \in [0, \tau]$ , define

$$C_n(t) = \sqrt{\frac{1}{n\widehat{\text{var}}(t)}} \sum_{i=1}^n \left[ \int_{-\infty}^{\infty} G(t - L_i^* | B_i, \hat{\nu}e^\delta, \hat{\sigma})\phi(\delta | \mathbf{O}_i; \hat{\nu}, \hat{\sigma}, \hat{\sigma}_r)d\delta - \widehat{F}_n(t) \right] Z_i,$$

where  $Z_1, \dots, Z_n \sim N(0, 1)$  iid and are independent of the data. We compute  $c_\alpha$  as follows:

- Step (i) . Generate  $M$  sets of independent realizations of  $(Z_1, \dots, Z_n)$  and, with each of sets, compute  $C_n^{(m)}(\cdot)$  for  $m = 1, \dots, M$ .
- Step (ii) . Choose  $c_\alpha$  as the  $(1 - \alpha)\%$  quantile of  $\sup_{t \in [0, \tau]} |C_n^{(1)}(t)|, \dots, \sup_{t \in [0, \tau]} |C_n^{(M)}(t)|$ .

### 3 Analysis of Alberta Forest fire data

We now apply the proposed approach to analyze the wildland fire data that motivated this research. Alberta Agriculture and Forestry collected records of 603 lightning-caused fires that occurred in 10 wildland fire management areas of Alberta, Canada during the fire season from May to August in 2004. Each fire record contains the fire progression information: the times and the fire burnt area at the time of report and at the time of initial attack. As expected, the records do not include the exact fire start times.

Figure 2 shows the burnt area at the report times and at initial attack times for the different regions. The 10 Alberta wildland fire management areas are classified into two

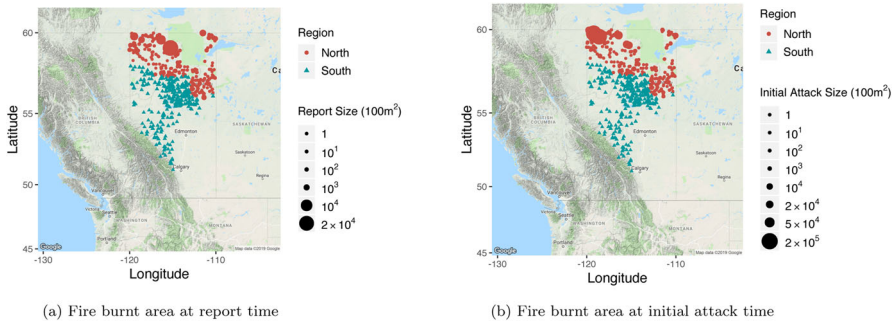


Fig. 2 Burnt area in each management region

groups: north and south. The north region includes Fort McMurray, High Level, Lac La Biche, Peace River, Slave Lake; the south region: Calgary, Edson, Grande Prairie, Rocky Mountain House and Whitecourt. Table S1 in Sect. S2 of the *Supplementary Material* summarizes burnt area for the two regions at the report times, the initial attack times, and at the time when fires were extinguished. Fires in the north region tend to have larger burnt area at the times of report and initial attack. The distributions of the burnt area are skewed so we use the transformed version  $\log_{10}(\text{burnt area} + 1)$  in the analysis.

The time of initial attack is when the first fire-fighting resource arrives at a wildland fire to prevent the fire from spreading, and to extinguish it if possible. It is believed that fires with a delayed initial attack may require a more substantial suppression effort.

Using the proposed approach, we estimate the distribution of the duration between the start of a fire to its initial attack. We consider two cases with Model (1): (i)  $\sigma_r = 0$  (i.e.,  $v_i = v$  for  $i = 1, 2, \dots, n$ ), and (ii)  $\sigma_r \geq 0$ . Table 1 presents the parameter estimates and the corresponding standard error estimates obtained by ALGORITHMS A and B in Sect. 2.3. The standard errors are estimated using both the inverse Fisher information matrix and the sandwich variance estimator. We also provide computing times for each algorithm in Table 1. ALGORITHMS B is computationally faster than ALGORITHMS A, but it yields less efficient estimator as the estimated standard errors are larger than those of ALGORITHMS A. The estimates of  $\sigma_r$  for the model with random drift are quite large, indicating considerable variation among the fires. This could be because the fire spread rates depend on location and local weather.

We estimate the distribution of duration by substituting the estimated model parameters into (7), and obtained the smoothed estimator based on (8). To make a comparison, we also evaluated the empirical distribution function based on the observed event duration, the naive estimator, and the Turnbull estimator viewing the fire data as interval-censored data with  $L_i \in [L_i^*, L_i^* + R_{max}]$ . We set  $R_{max} = 6, 12$  or 48 hours for illustration. In fact,  $R_{max}$  could be up to 2 weeks (Wotton and Martell 2005). Figure 3 presents the estimated distributions for the times to initial attack with ALGORITHMS A and B together with approximate 95% pointwise confidence intervals (CIs) calculated using the estimated asymptotic variance given in (14).

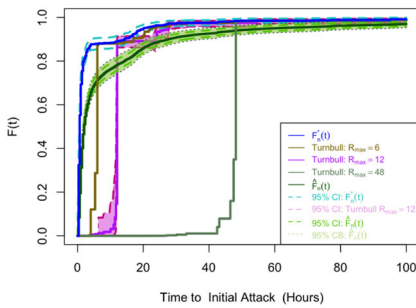
**Table 1** Estimates of parameters in model (1) with Alberta Wildfire Data

Parameter		Constant drift ( $\sigma_r = 0$ )		Random drift ( $\sigma_r \geq 0$ )		
		$\nu$	$\sigma$	$\nu$	$\sigma$	$\sigma_r$
ALGORITHM A	<sup>a</sup> $E_{St}$	0.090	1.283	0.093	1.216	0.122
	<sup>b</sup> $SE_F$	0.013	0.026	0.013	0.025	0.003
	<sup>c</sup> $SE_R$	0.013	0.030	0.017	0.027	0.003
	Computing Times	0.168 seconds		9.47 seconds		
ALGORITHM B	$E_{St}$	0.091	1.272	0.089	1.258	0.108
	$SE_F$	0.024	0.037	0.023	0.037	0.004
	$SE_R$	0.026	0.046	0.025	0.042	0.004
	Computing Times	0.080 seconds		8.67 seconds		

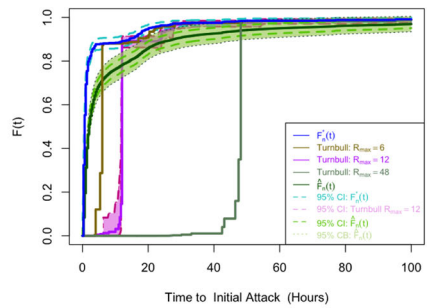
<sup>a</sup> $E_{St}$ : the estimates of parameters

<sup>b</sup> $SE_F$ : the standard error estimates using the Fisher information

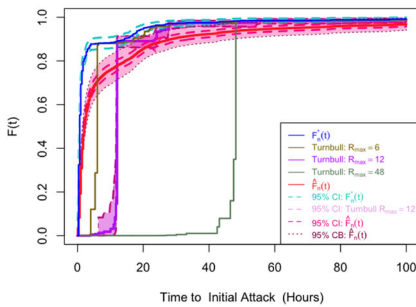
<sup>c</sup> $SE_R$ : the standard error estimates by the sandwich variance estimator



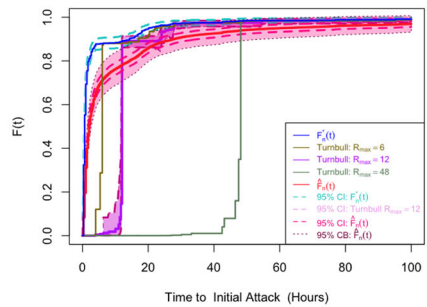
(a) Model with  $\sigma_r = 0$ , ALGORITHM A



(b) Model with  $\sigma_r = 0$ , ALGORITHM B



(c) Model with  $\sigma_r \geq 0$ , ALGORITHM A



(d) Model with  $\sigma_r \geq 0$ , ALGORITHM B

**Fig. 3** Estimated distributions for times to initial attack with Alberta Data

Figure 3 shows that the naive distribution estimate and the Turnbull's estimates are different from the proposed estimates. We see that Turnbull's estimates deviate more from the proposed estimates as  $R_{max}$  increases. This is because a larger  $R_{max}$  can lead to a wider interval  $(L_i^*, L_i^* + R_{max})$  for  $L_i$ . As a result, there are fewer disjoint innermost intervals within which the survivor function estimated by Turnbull's method can jump. Comparing the estimates by the two algorithms, we see that ALGORITHMS A can produce a more efficient estimator. We also evaluate the kernel-smoothed estimator (8) presented in Sect. 2.2. The distribution estimates and their corresponding 95% CIs/CBs are in close agreement with those un-smoothed estimates.

Figure 4 presents scatterplots of the final burnt area vs the estimated duration times. The estimated duration is calculated by  $\tilde{L}_i = L_i^* + \tilde{S}_i$ , where  $\tilde{S}_i$  is generated from the posterior distribution of reporting delay  $[S|\delta, \mathbf{O}_i; \theta^{(m)}]$  at the last iteration of the MCEM procedure in ALGORITHMS A, for  $i = 1, \dots, n$ . We present scatterplots using three realizations of  $\tilde{L}_i$  together with the scatterplot in Fig. 4a using the observed portion of the duration  $L_i^*$ . The pattern of association between the final burnt area and duration is apparently more obvious with the estimated duration. This suggests that the duration between fire start and initial attack may be more predictive of the final burnt area. Accounting for the reporting delay time is worthwhile when using the duration as a predictor for the final burnt area.

We applied the proposed procedure to analyze the data of fires from the north region and the ones from the south region separately. Table 2 gives the model parameter estimates. Figure 5 shows the estimated duration distributions. The estimate of  $\sigma_r$  associated with north region is large, significantly different from zero. It indicates a larger variation across fires in the region. The south region has a smaller estimate of  $\sigma_r$ .

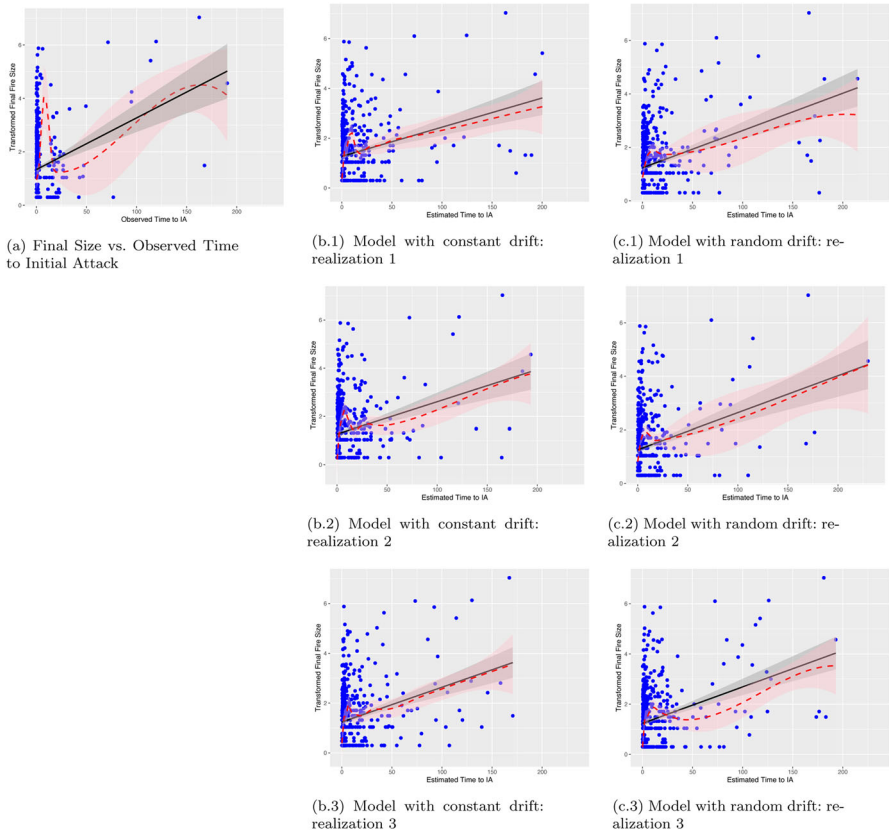
## 4 Simulation studies

We conducted two simulation studies to examine the finite-sample performance of the proposed approach and to verify the findings from the data analysis. Specifically, in the first simulation study, we generated data based on Model (1) to verify consistency and efficiency, and in the second simulation study, we assessed robustness of the approach against model misspecification.

### 4.1 Simulation A: Consistency and efficiency

To mimic the fire data, we simulated a study with  $n = 300$  independent fires with the data of fire  $i$  for  $i = 1, 2, \dots, n$  generated as follows.

- (i) Generate the burnt area process  $A_i(t), t \in [0, 30]$  based on Model (1) with the parameter values  $\nu = 2.0$  and  $\sigma = 0.5$ , and  $\delta_i \sim N(0, \sigma_r^2)$  with  $\sigma_r = 0, 0.5, \text{ or } 0.8$ .
- (ii) Generate the size at the report time  $B_i \sim \text{logNormal}(2.0, 0.1)$ , and determine the reporting delay as  $S_i = \max\{t | t \in [0, 30], A_i(t) \leq B_i\}$ .



**Fig. 4** Scatterplots of final burnt area and estimated times to initial attack with three realizations of  $\tilde{L}_i = L_i^* + \tilde{S}_i$ , where  $\tilde{S}_i$  is generated from the posterior distribution of reporting delay  $[S|\delta, \mathbf{O}_i; \theta^{(m)}]$  at the last iteration of the MCEM procedure in ALGORITHM A, for  $i = 1, \dots, n$ . The solid line represents the fitted linear regression line of the final burnt area and time to initial attack; the dashed curve is the local regression curve. The shaded areas define corresponding approximate 95% confidence bands

(iii) Generate  $L_i^* \sim \text{Exp}(3.0B_i^{-1})$ , calculate the duration  $L_i = S_i + L_i^*$ , and obtain the burnt area at the time of initial attack  $D_i = A_i(L_i)$ .

Using the simulated data, we evaluated the estimator  $\tilde{\theta}_{n_A}$ , the approximation to  $\hat{\theta}_{n_A}$ , by the variant of ALGORITHM A in Sect. 2.3. We then obtained the corresponding duration distribution estimates.

Table 3 summarizes the parameter estimates based on 200 simulation repetitions. The sample means of estimates obtained by Wiener process model with a constant drift are close to the true parameter values for the scenario of  $\sigma_r = 0$ ; the bias is evident when the true value of  $\sigma_r$  increases to 0.5 and 0.8. When we use a model with random drift, i.e.  $\sigma_r \geq 0$ , the sample means of estimates are close to the true parameter values for all three scenarios of  $\sigma_r$ . This provides an empirical verification of the consistency of the two estimators, and suggests that it may be acceptable not to assume  $\sigma_r = 0$  in practice. Further, we estimated  $\theta$  by maximizing the conditional likelihood given in

**Table 2** Estimates of parameters in model (1) with data of Wildfires in North region and South region

Parameter			Constant drift ( $\sigma_r = 0$ )		Random drift ( $\sigma_r \geq 0$ )		
			$\nu$	$\sigma$	$\nu$	$\sigma$	$\sigma_r$
North region	ALGORITHM A	<sup>a</sup> $Est$	0.085	1.637	0.074	0.859	0.670
		<sup>b</sup> $SE_F$	0.017	0.045	0.019	0.022	0.015
		<sup>c</sup> $SE_R$	0.023	0.052	0.020	0.029	0.015
	ALGORITHM B	$Est$	0.084	1.639	0.074	0.957	0.565
		$SE_F$	0.039	0.069	0.035	0.042	0.019
		$SE_R$	0.046	0.073	0.041	0.042	0.021
South region	ALGORITHM A	$Est$	0.102	1.272	0.105	0.995	0.058
		$SE_F$	0.015	0.023	0.015	0.016	0.002
		$SE_R$	0.011	0.029	0.015	0.017	0.003
	ALGORITHM B	$Est$	0.104	1.272	0.101	1.025	0.047
		$SE_F$	0.023	0.032	0.022	0.030	0.017
		$SE_R$	0.026	0.041	0.023	0.032	0.017

<sup>a</sup>  $Est$ : the estimates of parameters

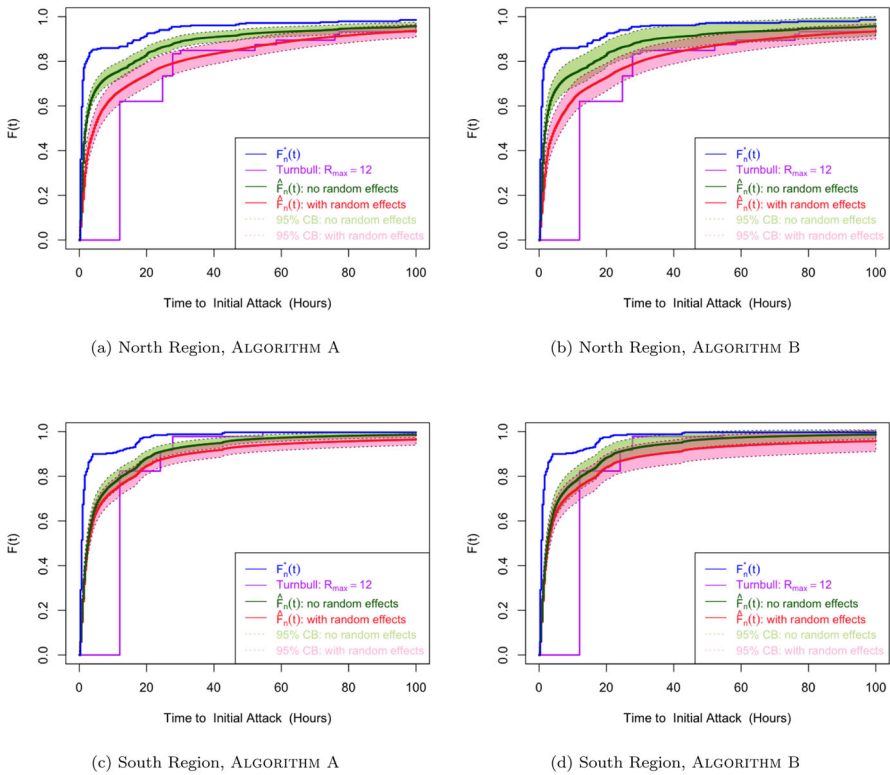
<sup>b</sup>  $SE_F$ : the standard error estimates using the Fisher information

<sup>c</sup>  $SE_R$ : the standard error estimates by the sandwich variance estimator

(13) which uses only observations on  $D$  and  $L^*$ . The results are presented in Table S2 of Sect. S3.1 of the *Supplementary Material*. While the parameter estimates are similar to those obtained by ALGORITHM A, the sample means of the estimated standard errors are larger. The sample means of the estimated standard errors associated with the robust sandwich variance estimator are similar to the corresponding sample standard deviations of the estimates for both algorithms, which suggests that the proposed variance estimator performs sufficiently well at the simulation settings, we conclude that maximizing (13) may yield less efficient estimators.

For each generated data set, we estimated the duration distribution by  $\hat{F}_n(t)$  using  $\theta$  obtained from ALGORITHM A, and used  $\tilde{F}_n(t; \nu, \sigma, \sigma_r)$  given in (5) with the true values of parameters as a reference. The consistent variance estimator of (14) given in Appendix C was evaluated to construct confidence intervals (CIs). Assuming the drift of Wiener process involves random effects, Fig. 6 shows the sample means of the 200 estimated distribution functions together with the approximate conventional 95% CIs and their 2.5%, 97.5% sample quantiles. To make a comparison, each plot in Fig. 6 also includes the sample means of the 200 evaluations of the empirical distribution function  $F_n(\cdot)$  using the true duration, the empirical distribution function  $F_n^*(\cdot)$  using the observed duration times (the naive approach), and the Turnbull estimator using  $R_{max}$  as the third quantile and maximum of the reporting delay in each generated data set.

The estimate associated with the proposed approach is very close to those based on  $\tilde{F}_n(t; \nu, \sigma, \sigma_r)$  using true  $\theta$ . At all simulation settings, both the approximate 95% CIs and the CIs using the 2.5% and 97.5% sample quantiles contain the empirical distribution functions  $F_n(\cdot)$  obtained with true duration. The naive estimates and the



**Fig. 5** Estimated distributions of time to initial attack in north region and south region

Turnbull’s estimates appear to be different from  $F_n(\cdot)$ . The Turnbull estimator is highly dependent on the assumed values of  $R_{max}$ , especially when  $R_{max}$  is much larger than  $L_i^*$ , the performance of the Turnbull estimator deteriorates substantially. Histograms for realizations of  $L_i^*$  presented in Fig. S1 of Sect. S3.1 of the *Supplementary Material* support this finding. For the scenario that the true value of  $\sigma_r$  is 0, the two values of  $R_{max}$  are relatively small and Turnbull’s estimates are close to the proposed estimate as shown in Fig. 6. When the true value of  $\sigma_r$  increases,  $R_{max}$  becomes much greater than the maximum of  $L_i^*$  when it is chosen as the maximum of the reporting delays in the generated dataset and the corresponding Turnbull’s estimates depart much further from the proposed estimates. This is consistent with the outcome seen in the data analysis. Moreover, we evaluated the distribution estimator using  $\hat{\theta}_{n_B}$  obtained from ALGORITHM B (See Fig. S2 of Sect. S3.1 in the *Supplementary Material*) and the kernel smoothed version of the proposed estimator. The behavior of these two estimates in comparison with the naive estimates and Turnbull’s estimates is similar to that observed in Fig. 6.

We computed the point-wise sample mean square errors of Turnbull’s estimates, the proposed estimates and the reference estimates based on  $\tilde{F}_n(t; \nu, \sigma, \sigma_r)$ . With any  $t \geq 0$ , the proposed estimator has the smallest sample mean squared error, which



**Table 3** Numerical Properties of Estimators for Model Parameters in Simulation A with ALGORITHM A

Parameter (true value)		Constant drift		Random drifts		
		$\nu$	$\sigma$	$\nu$	$\sigma$	$\sigma_r$
		2.0	0.5	2.0	0.5	
$\sigma_r = 0$						
ALGORITHM A	<sup>a</sup> <i>SMean</i>	1.999	0.499	1.998	0.497	0.001
	<sup>b</sup> <i>SSE</i>	0.014	0.016	0.015	0.015	0.002
	<sup>c</sup> <i>SMSE<sub>F</sub></i>	0.012	0.014	0.012	0.014	0.001
	<sup>d</sup> <i>SMSE<sub>R</sub></i>	0.012	0.014	0.012	0.014	0.001
$\sigma_r = 0.5$						
ALGORITHM A	<i>SMean</i>	2.243	1.887	1.992	0.490	0.504
	<i>SSE</i>	0.055	0.072	0.015	0.017	0.026
	<i>SMSE<sub>F</sub></i>	0.044	0.051	0.010	0.014	0.021
	<i>SMSE<sub>R</sub></i>	0.053	0.069	0.010	0.014	0.020
$\sigma_r = 0.8$						
ALGORITHM A	<i>SMean</i>	2.765	3.668	2.009	0.492	0.808
	<i>SSE</i>	0.112	0.122	0.062	0.027	0.035
	<i>SMSE<sub>F</sub></i>	0.100	0.105	0.048	0.024	0.033
	<i>SMSE<sub>R</sub></i>	0.105	0.118	0.048	0.025	0.032

<sup>a</sup>*SMean*: the sample mean of the estimates

<sup>b</sup>*SSE*: the sample standard error of the estimates.

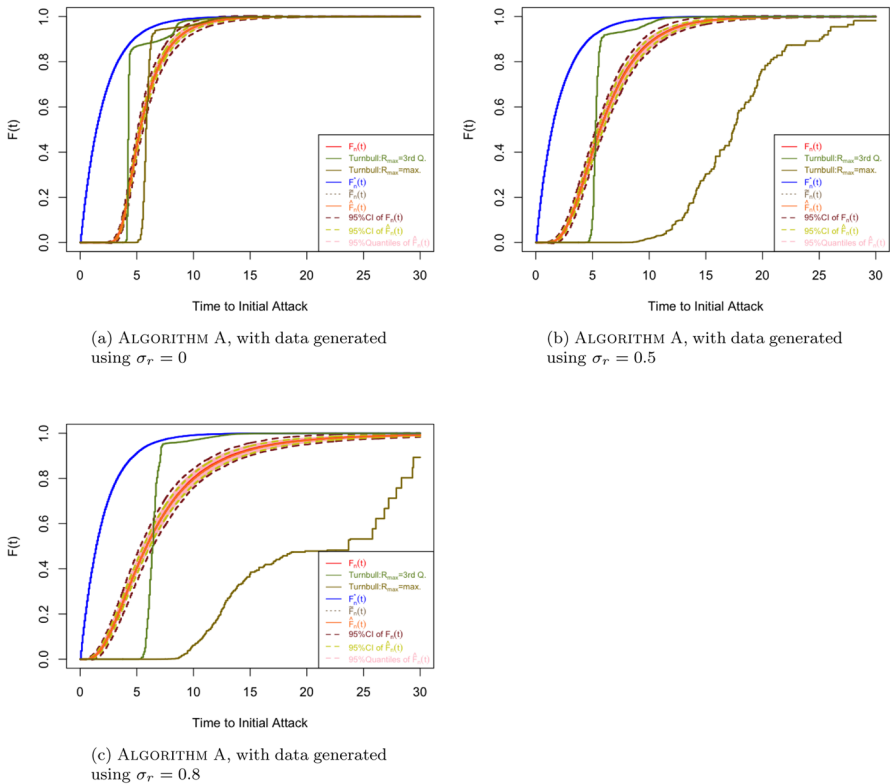
<sup>c</sup>*SMSE<sub>F</sub>*: the sample mean of the estimated standard errors using the Fisher information.

<sup>d</sup>*SMSE<sub>R</sub>*: the sample mean of the estimated standard errors using the sandwich variance estimator

demonstrates the relative efficiency of the proposed estimator over the naive estimator and the Turnbull estimator at all simulation settings. Figure S3 in Sect. S3.1 of the *Supplementary Material* presents the sample standard deviations and sample means of the estimated standard errors of the proposed distribution estimator with  $\hat{\theta}_{n_A}$  by ALGORITHM A, together with those associated with the empirical distribution function and  $\tilde{F}_n(\cdot; \nu, \sigma, \sigma_r)$ , which require more information than the data structure of interest. The plots in the figure show that the variation of the proposed estimator is comparable to that for  $\tilde{F}_n(\cdot; \nu, \sigma, \sigma_r)$ , and is, in some settings, smaller than that of the empirical function. This indicates that using the available information on fire growth can recover the efficiency loss due to the missing start times and even in some situations outperform the empirical distribution function, a nonparametric estimator for the duration distribution.

### 4.2 Simulation B: Robustness

We generated burnt area sample paths for a collection of simulated independent fires following the model  $A_i(t) = \nu_i t + \sigma_i W_i^*(t)$ ,  $i = 1, 2, \dots, n = 300$ , where  $\nu_i = \nu e^{\delta_i}$  with  $\delta_i \sim N(0, \sigma_r^2)$  and  $W_i^*(\cdot)$  is a process with correlated increments. Specifically, the increments  $W_i^*(t_k) - W_i^*(t_{k-1})$  for the partition  $t_k, k = 1, 2, \dots, K$  of the time



**Fig. 6** Simulation A: sample means of estimated distribution functions with corresponding 95% CI, based on model with random drift

period  $[0, 30]$  were generated from  $MN(\mathbf{0}, \Sigma)$  with the  $(k, k')$  entry as  $\Delta t \rho^{|k-k'|}$  for  $k \neq k'$  and  $\Delta t$  for  $k = k'$  with  $\rho = 0.2$ . The observations on variables  $B, S, L^*$ , and  $D$  were generated in the same way as in Simulation A.

We computed  $\theta$  estimates and then the duration distribution estimates as if the data were generated from Model (1). Table S3 in Sect. S3.2 of the *Supplementary Material* summarizes the simulation outcomes for 200 replicates of the estimates. The sample means of the  $\theta$  estimates are close to the true parameter values with the assumed Wiener process model using random drift. The sample means of the estimated standard errors from the robust estimator are fairly close to the sample standard deviations of the  $\theta$  estimates.

Figure S4 in Sect. S3.2 of the *Supplementary Material* presents the sample means of the estimated distribution functions from both ALGORITHMS A and B with the approximate 95% CIs and their 2.5% and 97.5% quantiles. In each plot, we also overlaid the sample means of the estimates by empirical distribution function,  $\tilde{F}_n(\cdot; \nu, \sigma, \sigma_r)$ , the naive estimator, and the Turnbull estimator. These plots indicate that the proposed estimator is close to the empirical function in the situation, even when Model (1) does not hold.

We also explored scenarios where the burnt-area process is generated following other models, such as  $A_i(t) = \nu_i t^2 + \sigma_i W_i(t)$ . The estimated duration distribution based on the proposed approach assuming Model (1) is also close to the empirical function using all the true duration times. This indicates that the proposed estimator  $\hat{F}_n(\cdot)$  can be quite robust to model misspecification. Further investigation could lead to a way of systematically checking the validity of Model (1).

## 5 Final remarks

We propose in this paper procedures for estimating the distribution of event duration with observations in the presence of missing time origins. By employing the distribution of the first-hitting-time with a Wiener process, we link the distribution of the event duration with associated longitudinal measures. Both simulation and real data analysis show that the proposed approach performs well in predicting the times to initial attack and also demonstrates the importance of taking into account the duration between the unobserved start time and the later report time.

The proposed approach is applicable to many situations where event duration is of interest but where the time origins in the duration observations are missing. Examples include predicting the length of period from the unknown HIV infection to detection of infection by making use of the longitudinal viral load measures (Doksum and Normand 1995), predicting the lifetime of trees by using longitudinal measures of the diameter at breast height for trees (Thompson 2011), and, as suggested by a referee, estimating the onset time of a disease by utilizing longitudinal medical expenditure data such as the usage of prescription drugs and the cost of skilled nursing facilities. The idea underlying our approach could readily be applied to analysis under a different model for longitudinal measures, e.g. Wang (2008); Wang and Xu (2010). It would be worth exploring the validity of the stochastic process for longitudinal measures.

Several other investigation would be worthwhile. The target population in the wildland fire application of this paper includes only the fires that are reported and have been dispatched with initial attack resources. When a study aims to explore the whole physical development process of wildland fire, the fires not reported should also be included in the population under consideration. The current available wildland fire records are then length-biased. We suggest to extend the idea of the proposed approach and adapt methods for estimating distributions with right-censored event times subject to length-biased sampling (e.g. Asgharian and Wolfson 2005; Huang and Qin 2011) to the situation where the origins of the duration times are missing.

Heterogeneity and correlation between fires should be accounted for. Applying the proposed approach to the data stratified by fire region has revealed that the event duration distributions are different for fires in different regions; see Table 2 and Fig. 5, for example. The duration is likely related to fuel type and moisture content as well as wind activity and local topography. To deal with this problem, as discussed in Wang (2010), we could follow (Lawless and Crowder 2004) and specify the drift parameter  $\nu_i$  of Model (1) as a function of covariates. In addition, due to potential correlation between wildland fires, it would be of interest to extend the approach to account for spatio-temporal correlation. A third possibility is to follow (Heitjan and Rubin 1990)

to accommodate semi-continuous data with the rounded burnt-area records shown in Table S1 of Sect. S2 of the *Supplementary Material*.

More investigation is required to systematically determine  $J$ , the number of Monte Carlo samples in ALGORITHM A. We can incorporate automated data-driven strategies (e.g. Levine and Casella 2001; Caffo et al. 2005) to the current algorithm to choose an appropriate  $J$  at each iteration. This paper evaluates integrals by monte carlo integration. As suggested by a referee, it can be interesting to compare the integral approximation with different numerical integration approaches, such as the Gaussian quadrature rule.

**Acknowledgements** We are grateful to N. McLoughlin of Alberta Wildfire Management Branch for providing the wildland fire data. We thank G.A. Whitmore for his many constructive suggestions, and the associate editor and two referees for their helpful comments and suggestions. The research was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and a CRT grant from the Canadian Statistical Sciences Institute (CANSSI).

**Funding** Funding was provided by Natural Sciences and Engineering Research Council of Canada (CA) (Grant No. NSERC DG 116860), Natural Sciences and Engineering Research Council of Canada (CA) (Grant No. NSERC DG and DAS 177430).

## Appendix: Technical details of Sect. 2.4

This appendix outlines our derivation for the asymptotic properties of the estimator  $\hat{F}_n(t)$  in Sect. 2.4. To simplify the notation, we define  $\boldsymbol{\theta} = (\nu, \sigma, \sigma_r)$  and use  $\boldsymbol{\theta}_0$  to represent the true values of the parameters. Denote  $\tilde{F}_n(t; \nu_0, \sigma_0, \sigma_{r0})$  by  $\tilde{F}_n(t)$ . We focus on the realization of the proposed estimator  $\hat{F}_n(\cdot)$  using  $\hat{\boldsymbol{\theta}}_{n_A}$  obtained by ALGORITHM A. The derivation can be adapted to handle  $\hat{F}_n(\cdot)$  using  $\hat{\boldsymbol{\theta}}_{n_B}$  by ALGORITHM B with little change.

The following are the regularity conditions assumed in Model (1). We adapt the regularity conditions for the MLE, which are summarized by Serfling (1980).

CONDITION (C1). *The parameter space  $\Theta$  is compact and the true parameter  $\boldsymbol{\theta}_0$  is an interior point of  $\Theta$ .*

CONDITION (C2). *The first, second, and third partial derivatives of the log-likelihood function given in (9) with respect to  $\boldsymbol{\theta}$  exist for each  $\boldsymbol{\theta} \in \Theta$ .*

CONDITION (C3). *Differentiation and integration are interchangeable for first, second and third partial derivatives of the log-likelihood function with respect to  $\boldsymbol{\theta}$ .*

CONDITION (C4). *The third partial derivative of the log-likelihood function with respect to  $\boldsymbol{\theta}$ , is dominated by a fixed integrable function  $M_2(\cdot)$  for every  $\boldsymbol{\theta} \in \Theta$ .*

CONDITION (C5).  $\Sigma(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}[(\partial \log L_{obs}(\boldsymbol{\theta}; \mathbf{O}_i) / \partial \boldsymbol{\theta})(\partial \log L_{obs}(\boldsymbol{\theta}; \mathbf{O}_i) / \partial \boldsymbol{\theta})']$  exists and is positive definite, where  $\log L_{obs}(\boldsymbol{\theta}; \mathbf{O}_i)$  is the  $i$ th term in (9). It is the same as

$$\Pi(\boldsymbol{\theta}_0) = -E_{\boldsymbol{\theta}_0}[\partial^2 \log L_{obs}(\boldsymbol{\theta}; \mathbf{O}_i) / \partial \boldsymbol{\theta} \boldsymbol{\theta}'].$$

**Appendix A: Proof of Theorem 1(i) Strong consistency**

To establish the consistency of the proposed estimator, we first note that the MLE  $\hat{\theta}_{n_A}$  is consistent and asymptotically normal. That is, under (C1)- (C5),  $\hat{\theta}_{n_A} \xrightarrow{a.s.} \theta_0$  as  $n \rightarrow \infty$ , and  $\sqrt{n}(\hat{\theta}_{n_A} - \theta_0) \xrightarrow{d} N(0, \Pi^{-1}(\theta_0))$ . Here  $\Pi(\theta_0) = \Sigma(\theta_0)$ .

Let  $M(t, L_i^*, B_i; \theta_0) = \int_{-\infty}^{\infty} G(t - L_i^* | B_i, \nu_0 e^{\delta_i}, \sigma_0) \phi(\delta_i | \mathbf{O}_i, \nu_0, \sigma_0, \sigma_r) d\delta$ . Define the class of functions  $\mathcal{F} = \{M(t, L_i^*, B_i; \theta_0) : t \in [0, \tau]\}$ . Let  $\mathbb{P}_n$  and  $P$  be the empirical measure and probability measure of  $n$  i.i.d. observations  $\{(L_i^*, B_i), i = 1, 2, \dots, n\}$ . Thus,  $\mathbb{P}_n M(t, L_i^*, B_i; \theta_0) = \sum_{i=1}^n M(t, L_i^*, B_i; \theta_0) / n$ ,  $PM(t, L_i^*, B_i; \theta_0) = E[M(t, L_i^*, B_i; \theta_0)]$ . Note that  $\hat{F}_n(t)$  can be written as  $\mathbb{P}_n M(t, L_i^*, B_i; \theta_0)$  and  $\tilde{F}_n(t)$  can be written as  $\mathbb{P}_n M(t, L_i^*, B_i; \hat{\theta}_{n_A})$ . Since  $G(t - L_i^* | B_i, \nu_0 e^{\delta_i}, \sigma_0)$  is a monotone process, the integral  $M(t, L_i^*, B_i; \theta_0)$  is also monotone. By Lemma 4.1 in Kosorok (2008), we have that  $\mathcal{F}$  is a  $P$ -Donsker. The Glivenko–Cantelli property of  $\mathcal{F}$  yields

$$\sup_{t \in [0, \tau]} \left| (\mathbb{P}_n - P) M(t, L_i^*, B_i; \theta_0) \right| \xrightarrow{P} 0. \tag{15}$$

Note that  $E[M(t, L_i^*, B_i; \theta_0)] = E[E\{G(t - L_i^* | B_i, \nu_0 e^{\delta_i}, \sigma_0) | \mathbf{O}_i\}] = F(t)$ . Thus  $\sup_{t \in [0, \tau]} |\tilde{F}_n(t) - F(t)| \xrightarrow{P} 0$ .

We then write  $\hat{F}_n(t) - \tilde{F}_n(t)$  as  $n^{-1} \sum_{i=1}^n [M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - M(t, L_i^*, B_i; \theta_0)]$ , which is

$$\frac{1}{n} \sum_{i=1}^n \partial E[G(t - L_i^* | B_i, \nu e^{\delta_i}, \sigma) | \mathbf{O}_i] / \partial \theta |_{\theta = \theta^*} (\hat{\theta}_{n_A} - \theta_0), \text{ and } \|\theta^* - \theta_0\| \leq \|\hat{\theta}_{n_A} - \theta_0\|.$$

For every  $t \in [0, \tau]$  and  $\nu, \sigma, \sigma_r$  belonging to a compact set  $\Theta$ ,  $|G(t - L_i^* | B_i, \nu e^{\delta_i}, \sigma) \phi(\delta_i | \mathbf{O}_i; \nu, \sigma, \sigma_r)|$  is bounded. Therefore, we can interchange integration and differentiation. Then,

$$\partial E[G(t - L_i^* | B_i, \nu e^{\delta_i}, \sigma) | \mathbf{O}_i] / \partial \theta \Big|_{\theta = \theta^*} E \left[ \partial G(t - L_i^* | B_i, \nu e^{\delta_i}, \sigma) / \partial (\nu, \sigma) | \mathbf{O}_i \right] \Big|_{\substack{\nu = \nu^* \\ \sigma = \sigma^*}}.$$

The derivatives  $\partial G(t - L_i^*|B_i, \nu e^{\delta_i}, \sigma)/\partial \nu$  and  $\partial G(t - L_i^*|B_i, \nu e^{\delta_i}, \sigma)/\partial \sigma$  are respectively

$$\begin{aligned} & \sqrt{\frac{t - L_i^*}{2\pi\sigma^2}} \left[ \exp\left(-\frac{A_1^2}{2} + \frac{2B_i\nu e^{\delta_i}}{\sigma^2} + \delta_i\right) - \exp\left(-\frac{A_2^2}{2} + \frac{2B_i\nu e^{\delta_i}}{\sigma^2} + \delta_i\right) \right] \\ & + \frac{2B_i e^{\delta_i}}{\sigma} \Phi(A_2) \text{ and } -\frac{1}{\sqrt{2\pi}\sigma} \left[ A_1 \exp\left(-\frac{A_1^2}{2}\right) + A_2 \exp\left(\frac{A_2^2}{2} + \frac{2B_i\nu e^{\delta_i}}{\sigma^2}\right) \right] \\ & + \exp\left(\frac{2B_i\nu}{\sigma^2}\right) \frac{-4B_i\nu e^{\delta_i}}{\sigma^3} \Phi(A_2), \end{aligned} \tag{16}$$

where  $A_1 = \sqrt{\frac{B_i}{\sigma^2(t-L_i^*)}} \left( \frac{(t-L_i^*)\nu e^{\delta_i}}{B_i} - 1 \right)$  and  $A_2 = -\sqrt{\frac{B_i}{\sigma^2(t-L_i^*)}} \left( \frac{(t-L_i^*)\nu e^{\delta_i}}{B_i} + 1 \right)$ .

The derivatives in (16) are uniformly bounded in probability for every  $t \in [0, \tau]$  and  $\nu, \sigma, \sigma_r$  belonging to a compact set  $\Theta$ . Therefore,  $n^{-1} \sum_{i=1}^n \partial E_{\theta^*} [G(t - L_i^*|B_i, \nu e^{\delta_i}, \sigma)|\mathbf{O}_i]/\partial \theta$  is uniformly bounded in probability. This together with the already observed fact that  $\hat{\theta}_{n_A} \xrightarrow{a.s.} \theta_0$  leads to  $\sup_{t \in [0, \tau]} |\hat{F}_n(t) - \tilde{F}_n(t)| \xrightarrow{P} 0$ , as  $n \rightarrow \infty$ . By (15), it follows that  $\sup_{t \in [0, \tau]} |\hat{F}_n(t) - F(t)| \xrightarrow{P} 0$ .

### Appendix B: Proof of Theorem 1(ii) Weak convergence

Note that  $\sqrt{n}(\hat{F}_n(t) - F(t)) = \sqrt{n}(\mathbb{P}_n M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - PM(t, L_i^*, B_i; \theta_0))$  is

$$\begin{aligned} & \sqrt{n}(\mathbb{P}_n - P)M(t, L_i^*, B_i; \theta_0) + \sqrt{n}(\mathbb{P}_n - P)[M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - M(t, L_i^*, B_i; \theta_0)] \\ & + \sqrt{n}P[M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - M(t, L_i^*, B_i; \theta_0)]. \end{aligned} \tag{17}$$

Since the class of functions  $\mathcal{F} = \{M(t, L_i^*, B_i; \theta_0) : t \in [0, \tau]\}$  is *P-Donsker*, the first term  $\sqrt{n}(\mathbb{P}_n M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - PM(t, L_i^*, B_i; \theta_0))$  converges weakly to a tight, mean-zero Gaussian process in  $\ell^\infty([0, \tau])$ , whose variance is

$$\int_0^\infty \int_0^\infty M(t, l^*, b; \theta_0)M(s, l^*, b; \theta_0)h(l^*, b)dl^*db - F(t)F(s), \tag{18}$$

where  $h(l^*, b)$  is the PDF for the joint distribution function of  $L^*$  and  $B$ .

Also, for the second term on the right hand side (RHS) of (17), we have

$$\sup_{t \in [0, \tau]} \sqrt{n}(\mathbb{P}_n - P)[M(t, L_i^*, B_i; \hat{\theta}_{n_A}) - M(t, L_i^*, B_i; \theta_0)] \xrightarrow{P} 0.$$

The third term on RHS of (17) can be written as  $\sqrt{n}(\hat{\theta}_{n_A} - \theta_0)E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta] + o_p(1)$ , which is

$$\begin{aligned} & \sqrt{n}(\mathbb{P}_n - P)\Pi^{-1}(\theta_0)\partial \log L_{obs}(\theta; \text{Observed-data})/\partial \theta|_{\theta=\theta_0} \\ & E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta] + o_p(1). \end{aligned}$$

Therefore,  $\hat{F}_n(t)$  is asymptotically linear with influence function:

$$\begin{aligned} & M(t, L_i^*, B_i; \theta_0) + \Pi^{-1}(\theta_0)\partial \log L_{obs}(\theta; \text{Observed-data})/\partial \theta|_{\theta=\theta_0} \\ & E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta]. \end{aligned}$$

Since this influence function is *P-Donsker*,  $\sqrt{n}(\hat{F}_n(t) - F(t))$  converges weakly to a tight, mean-zero Gaussian process  $\mathcal{G}$ , as  $n \rightarrow \infty$ .

### Appendix C: Variance estimation in Theorem 1(ii)

To calculate the variance of  $\mathcal{G}$ , we write  $\text{Var}(\sqrt{n}(\hat{F}_n(t) - F(t)))$  as

$$\begin{aligned} & \text{Var}(\sqrt{n}(\hat{F}_n(t) - \tilde{F}_n(t))) + \text{Var}(\sqrt{n}(\tilde{F}_n(t) - F(t))) \\ & + \text{Cov}(\sqrt{n}(\hat{F}_n(t) - \tilde{F}_n(t)), \sqrt{n}(\tilde{F}_n(t) - F(t))). \end{aligned} \tag{19}$$

The second term on the RHS of (19) is given in (18) and the third term can be calculated as  $nE[(\hat{F}_n(t) - \tilde{F}_n(t))(\tilde{F}_n(t) - F(t))]$ . Noting that  $|\hat{F}_n(t)| \leq 1$ ,  $|\tilde{F}_n(t)| \leq 1$ ,  $|\hat{F}_n(t) - \tilde{F}_n(t)| \leq 1$ ,  $\sup_{t \in [0, \tau]} |\hat{F}_n(t) - \tilde{F}_n(t)| \xrightarrow{P} 0$ , it is clear to show that  $E[(\hat{F}_n(t) - \tilde{F}_n(t))(\tilde{F}_n(t) - F(t))] \xrightarrow{P} 0$ , by the bounded convergence theorem. Therefore,  $\text{Cov}(\sqrt{n}(\hat{F}_n(t) - \tilde{F}_n(t)), \sqrt{n}(\tilde{F}_n(t) - F(t))) \xrightarrow{P} 0$ .

The first term,  $\text{Var}(\sqrt{n}(\hat{F}_n(t) - \tilde{F}_n(t)))$ , can be calculated as

$$\left[ \frac{1}{n} \sum_{i=1}^n \partial M(t, L_i^*, B_i; \theta)/\partial \theta|_{\theta=\theta^*} \right]' \text{Var}(\sqrt{n}(\hat{\theta}_{n_A} - \theta_0)) \left[ \frac{1}{n} \sum_{i=1}^n \partial M(t, L_i^*, B_i; \theta)/\partial \theta|_{\theta=\theta^*} \right].$$

Note that  $\sum_{i=1}^n \frac{\partial M(t, L_i^*, B_i; \theta)}{\partial \theta} / n|_{\theta=\theta^*} \xrightarrow{P} E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta]$  and the asymptotic variance of  $\hat{\theta}_{n_A}$  is  $\Pi^{-1}(\theta_0)$ . Thus  $\text{Var}(\sqrt{n}(\hat{F}_n(t) - \tilde{F}_n(t)))$  converges in probability as  $n \rightarrow \infty$  to

$$E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta]' \Pi^{-1}(\theta_0) E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta)/\partial \theta].$$

The covariance of the limiting Gaussian process of  $\sqrt{n}(\hat{F}_n(t) - F(t))$  is as given in (14) in Theorem 1 (ii). Further, one can use  $\sum_{i=1}^n \frac{\partial M(t, L_i^*, B_i; \theta)}{\partial \theta} / n |_{\theta = \hat{\theta}_{n_A}}$  to estimate  $E_{\theta_0}[\partial M(t, L_i^*, B_i; \theta) / \partial \theta]$  and use  $\Pi(\hat{\theta}_{n_A})$  to estimate  $\Pi(\theta_0)$ .

## References

- Asgharian M, Wolfson DB (2005) Asymptotic behavior of the unconditional NPMLE of the length-biased survivor function from right censored prevalent cohort data. *Ann Stat* 33(5):2109–2131
- Balasubramanian R, Lagakos SW (2003) Estimation of a failure time distribution based on imperfect diagnostic tests. *Biometrika* 90(1):171–182
- Braun J, Duchesne T, Stafford JE (2005) Local likelihood density estimation for interval censored data. *Can J Stat* 33(1):39–60
- Caffo BS, Jank W, Jones GL (2005) Ascent-based Monte Carlo expectation-maximization. *J R Stat Soc Ser B (Stat Methodol)* 67(2):235–251
- Chhikara RS, Folks JL (1989) *The inverse Gaussian distribution: theory, methodology, and applications*. Marcel Dekker Inc, New York
- Choi S, Huang X, Cormier JN, Doksum KA (2014) A semiparametric inverse-gaussian model and inference for survival data with a cured proportion. *Can J Stat* 42(4):635–649
- Degruttola V, Lange N, Dafni U (1991) Modeling the progression of HIV infection. *J Am Stat Assoc* 86(415):569–577
- Doksum KA, Hoyland A (1992) Models for variable-stress accelerated life testing experiments based on wiener processes and the inverse gaussian distribution. *Technometrics* 34(1):74–82
- Doksum KA, Normand SLT (1995) Gaussian models for degradation processes-part I: methods for the analysis of biomarker data. *Lifetime Data Anal* 1(2):131–144
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109
- Heitjan DF, Rubin DB (1990) Inference from coarse data via multiple imputation with application to age heaping. *J Am Stat Assoc* 85(410):304–314
- Horrocks J, Thompson ME (2004) Modeling event times with multiple outcomes using the wiener process with drift. *Lifetime Data Anal* 10(1):29–49
- Hu XJ, Lagakos SW (1999) Interim analyses using repeated confidence bands. *Biometrika* 86(3):517–529
- Huang CY, Qin J (2011) Nonparametric estimation for length-biased and right-censored data. *Biometrika* 98(1):177–186
- Klein J, Moeschberger M (2003) *Survival analysis: statistical methods for censored and truncated data*. Springer-Verlag, New York
- Kosorok MR (2008) *Introduction to empirical processes and semiparametric inference*. Springer, Berlin
- Lawless J, Crowder M (2004) Covariates and random effects in a gamma process model with application to degradation and failure. *Lifetime Data Anal* 10(3):213–227
- Lee MLT, Whitmore GA (2006) Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Stat Sci* 21(4):501–513
- Levine RA, Casella G (2001) Implementations of the Monte Carlo EM algorithm. *J Comput Gr Stat* 10(3):422–439
- Lu CJ, Meeker WO (1993) Using degradation measures to estimate a time-to-failure distribution. *Technometrics* 35(2):161–174
- Martell DL (2007) Forest fire management. In: *Handbook of operations research in natural resources*. Springer, pp. 489–509
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6):1087–1092
- Morin AA (2014) A spatial analysis of forest fire survival and a marked cluster process for simulating fire load. Master's thesis, The University of Western Ontario, London, Ontario, Canada
- Mulatya CM, McLain AC, Cai B, Hardin JW, Albert PS (2016) Estimating time to event characteristics via longitudinal threshold regression models-an application to cervical dilation progression. *Stat Med* 35(24):4368–4379



- Parks GM (1964) Development and application of a model for suppression of forest fires. *Manag Sci* 10(4):760–766
- Peng CY (2015) Inverse gaussian processes with random effects and explanatory variables for degradation data. *Technometrics* 57(1):100–111
- Pennell ML, Whitmore G, Ting Lee ML (2009) Bayesian random-effects threshold regression with application to survival data with nonproportional hazards. *Biostatistics* 11(1):111–126
- Rosenblatt M (1956) Remarks on some nonparametric estimates of a density function. *Ann Math Stat* 27:832–837
- Serfling RJ (1980) *Approximation theorems of mathematical statistics*. Wiley, New Jersey
- Thompson DJ (2011) *Survival models for data arising from multiphase hazards, latent subgroups or subject to time-dependent treatment effects*. PhD thesis, Simon Fraser University, Burnaby, BC, Canada
- Turnbull BW (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. *J R Stat Soc Ser B (Methodol)* 38(3):290–295
- Wang X (2008) A pseudo-likelihood estimation method for nonhomogeneous gamma process model with random effects. *Stat Sin* 18:1153–1163
- Wang X (2010) Wiener processes with random effects for degradation data. *J Multivar Anal* 101(2):340–351
- Wang X, Xu D (2010) An inverse gaussian process model for degradation data. *Technometrics* 52(2):188–197
- Wei GC, Tanner MA (1990) A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J Am Stat Assoc* 85(411):699–704
- Wotton B, Martell DL (2005) A lightning fire occurrence model for Ontario. *Can J For Res* 35(6):1389–1401
- Zhao L, Hu XJ, Lagakos SW (2008) Statistical monitoring of clinical trials with multivariate response and/or multiple arms: a flexible approach. *Biostatistics* 10(2):310–323

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.