Statistica Sinica **17**(2007), 1091-1114

THE MAXIMUM LIKELIHOOD METHOD WITH ESTIMATED NUISANCE PARAMETERS IN HAZARD RATE MODELS WITH DISCONTINUITIES

Dietmar Ferger

Dresden University of Technology

Abstract: Let $X_1, \ldots, X_n$ be i.i.d. with common hazard function, a step function with exactly one jump. The location of the jump is the parameter of interest and is to be estimated based on our sample. We prove consistency and convergence in law of our estimators with rate n and non-normal limit distribution. There is also L_p -convergence with exact rate n^{-1}. This statistical experiment is non-regular in the sense of Ibragimov and Has'minskii (1981). Our approach is extended to general hazard functions with one jump-point. The basic idea can also be used in a complete nonparametric framework.

Key words and phrases: Argmax of Poisson-process, change-point, maximum-likelihood, nonparametric hazard.

1. Introduction

In clinical studies one often is interested in the time to a relapse of a certain serious disease. For patients with, e.g., leukemia, there are good reasons to assume that the relapse rate changes abruptly (from a high constant level to a lower one) after an unknown period of length τ. Matthews and Farewell (1982) describe the time span as a random variable X with a piece-wise constant hazard rate function

$$h(t) = \begin{cases} \alpha, & 0 \le t \le \tau \\ \beta, & t > \tau \end{cases} \tag{1.1}$$

with positive parameters $\alpha \neq \beta$ and τ. Once the physician knows the value of τ he or she is able to decide whether a patient still belongs to a high-risk group or not. In another example, one observes that a technical component operates with a constant failure rate until it suffers from a shock. After this the component continues to function but with another constant failure rate. Again we can model the life time of the component by a random variable with the failure rate function h given in (1.1). Similarly in case of censored data, Wu, Zhao and Wu (2003) state that the model plays an important role in medical follow-up studies after an operation, e.g., bone marrow transplantation. Likewise it is used in industrial life testing experiments with changing conditions, e.g., temperature increases. Due to its relevance in the applied sciences there are many contributions to the subject, confer, e.g., Anderson and Senthilselvan (1982), Basu, Ghosh and Joshi (1988), Ghosh and Joshi (1992), Matthews and Farewell (1982), Matthews, Farewell and Pyke (1985), Nguyen, Rogers and Walker (1984), Pham and Nguyen (1990, 1993) and Yao (1986, 1987). The model and some variants thereof are also very popular in human life sciences and demography. Finkelstein (2003) states that populations can experience change points due to positive or negative "enviromental" influences. For example, the implementation of better healthcare in the former East Germany after the reunification can be considered as an example of such a positive influence, see Scholz and Maier (2003). Similarly the demographic situation during the transitional period in Russia after the collapse of the former USSR shows a negative impact on mortality rates.

If τ is known, (1.1) is known as "step-stress-model" in the literature. For instance Balakrishnan, Kundu, Ng and Kannan (2007) determine the exact distribution of the maximum likelihood estimator of (α, β) if the observations are type-II censored.

One easily checks that h uniquely determines a distribution with density

$$f(x|\tau,\alpha,\beta) = \begin{cases} \alpha e^{-\alpha x}, & 0 \le x \le \tau \\ \beta e^{(\beta-\alpha)\tau - \beta x}, & x > \tau \end{cases} \tag{1.2}$$

and distribution function

$$F(x|\tau,\alpha,\beta) = \begin{cases} 1 - e^{-\alpha x}, & 0 \le x \le \tau \\ 1 - e^{(\beta-\alpha)\tau - \beta x}, & x > \tau \end{cases} . \tag{1.3}$$

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with common density $f(\cdot|\tau,\alpha,\beta)$. Then the log-likelihood function is given by

$$l_n(\tau,\alpha,\beta) = \sum_{i=1}^n 1_{\{\tau < X_i\}} \Big[\log \frac{\beta}{\alpha} + (\beta-\alpha)(\tau - X_i) \Big] + \sum_{i=1}^n [\log \alpha - \alpha X_i] \tag{1.4}$$

for every $(\tau,\alpha,\beta) \in (0,\infty)^3$ with $\alpha \neq \beta$. Nguyen, Rogers and Walker (1984) observe that $l_n(\tau,\alpha,1/(X_{n:n}-\tau)) \longrightarrow \infty$ for every fixed α as $\tau \uparrow X_{n:n}$, where $X_{n:n}$ denotes the largest observation. Thus the maximum-likelihood estimator (mle) of the three-dimensional parameter (τ,α,β) does not exist. As a way out, Yao (1986) and Pham and Nguyen (1990) restrict the domain of l_n to regions of the type

$$\{(\tau,\alpha,\beta) \in (0,\infty)^3 : T_n' \le \tau \le T_n'', \quad \alpha \neq \beta\},$$

where the bounds $0 \le T_n' < T_n'' < X_{n:n}$ for τ may be random. For $T_n' = 0$ and T_n'' equal to the second largest observation $X_{n-1:n}$, one obtains Yao's (1986) estimator. More generally Pham and Nguyen (1990) only require that with probability one the compact interval $[T_n', T_n'']$ contains the true value of τ for eventually all $n \in \mathbb{N}$. They prove strong consistency of the constrained mle

$$(\tilde{\tau}_n \tilde{\alpha}_n, \tilde{\beta}_n) := \text{argmax}\{l_n(\tau, \alpha, \beta) : T_n' \leq \tau \leq T_n'', \quad \alpha \neq \beta\}. \tag{1.5}$$

In addition they show that $n(\tilde{\tau}_n - \tau)$ converges to a non-normal limit T which can be expressed in terms of a certain random walk on the integers. Yao (1986) actually proves that the random vector $(n(\tilde{\tau}_n - \tau), n^{1/2}(\tilde{\alpha}_n - \alpha), n^{1/2}(\tilde{\beta}_n - \beta))$ has a distributional limit (T, A, B), where T, A and B are independent and (A, B) is bivariate normally distributed. Finally, Pham and Nguyen (1993) show that a parametric bootstrap works, meaning that the bootstrap version $n(\tilde{\tau}_n^* - \tilde{\tau}_n)$ of $n(\tilde{\tau}_n - \tau)$ converge to the same limit T as do the originals.

In reality, quite often the data are not completely observable due to censoring. In this situation, Chang, Chen and Hsiung (1994) suggest the argmax of an Aalen-Nelson type process as estimate for τ. They prove weak consistency and convergence in distribution to the argmax of a Poisson process with linear drift. For a review of the literature we recommend Müller and Wang (1994), who also include a nonparametric extension of (1.1). A discussion about the impact of censoring is given in Loader (1991). For two recent contributions we refer to Antoniadis, Gijbels and Macgibbon (2000) and Wu, Zhao and Wu (2003).

In this article we focus on the parametric model (1.1), but our approach is such that it can be generalized to a hazard function which has a jump at a point τ but is otherwise smooth. (We discuss this generalization at the end.) The idea is to assume – for a moment – that α and β are known. Then the mle of τ is well-defined. Since the second summand in (1.4) does not depend on τ, the mle is given by

$$\tau_n(\alpha, \beta) = \text{argmax } S_n(t|\alpha, \beta), \tag{1.6}$$

where

$$S_n(t|\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{t < X_i\}} \Big[\log \frac{\beta}{\alpha} + (\beta - \alpha)(t - X_i) \Big], \quad t \geq 0.$$

Here we use the common convention

$$\text{argmax } f(t) := \min \Big\{ t \geq 0 : \max(f(t), f(t-)) = \sup_{s \geq 0} f(s) \Big\} \tag{1.7}$$

for all bounded functions $f : [0, \infty) \to \mathbb{R}$ right-continuous with left hand limits (rcll). In Abdel-Aty and Ferger (2003) we derive an explicit representation for the mle in terms of the order statistics pertaining to the sample $X_1, \ldots, X_n$. It allows a very easy computation of the mle. Moreover we prove strong consistency:

$$\tau_n(\alpha, \beta) \longrightarrow \tau \text{ a.s. as } n \to \infty \quad \forall\, \tau > 0, \quad \forall\, \alpha \neq \beta. \tag{1.8}$$

Next we drop the assumption that α and β need to be known. For that puropse consider any pair (α_n, β_n) of estimators for the unknown parameter vector (α, β). Assume that it is weakly consistent, i.e.,

$$(\alpha_n, \beta_n) \xrightarrow{P} (\alpha, \beta), \text{ as } n \to \infty \quad \forall\, \tau > 0, \quad \forall\, \alpha \neq \beta. \tag{1.9}$$

Then it is reasonable to replace the unknown (α, β) in the definition of the mle $\tau_n(\alpha, \beta)$ by (α_n, β_n). This leads to $\tau_n := \tau_n(\alpha_n, \beta_n) = \text{argmax } S_n(t|\alpha_n, \beta_n)$. The aim of this paper is to derive asymptotic properties of our estimator τ_n. The only requirement we make use of is (1.9). In Section 2 we prove consistency of τ_n. Section 3 deals with distributional and L_p-convergence of $n(\tau_n - \tau)$. Here we also present finite sample estimators for the bias and for the variance. Recall that we need a consistent estimator (α_n, β_n) for (α, β). One could use, e.g., $(\tilde{\alpha}_n, \tilde{\beta}_n)$ of the constrained mle (1.5). Alternative estimators for (α, β) are presented in Section 4. We prove strong consistency and also asymptotic normality. Section 5 contains a simulation study of the performance of $\tau_n = \tau_n(\alpha_n, \beta_n)$ for several estimators (α_n, β_n) of the parameter (α, β). It enables a comparison with Yao's (1986) estimator. Moreover we investigate the robustness of τ_n under certain deviations from model (1.1). We consider hazard functions h with a single jump at a point τ and which are continuous elsewhere in Section 6. Furthermore, in Section 7 we give an outlook on the nonparametric case in which the shape of h is completely unknown. Finally, Section 8 contains several proofs.

2. Weak and strong consistency

In this section we see that the consistency (1.8) of the mle $\tau_n(\alpha, \beta)$ carries over to the mle $\tau_n(\alpha_n, \beta_n)$ with estimated parameters, provided the estimators (α_n, β_n) are consistent.

Proposition 2.1. *Let* $(\alpha_n, \beta_n) \subseteq \mathbb{R}_+^2$ *be any sequence such that* (1.9) *holds. Then*

$$\tau_n(\alpha_n, \beta_n) \xrightarrow{P} \tau \text{ as } n \to \infty \quad \forall\, \tau > 0 \quad \forall\, \alpha \neq \beta, \tag{2.1}$$

and if (1.9) *and*

$$(\alpha_n, \beta_n) \to (\alpha, \beta) \text{ a.s. as } n \to \infty \quad \forall\, \tau > 0 \quad \forall\, \alpha \neq \beta, \tag{2.2}$$

$$\tau_n(\alpha_n, \beta_n) \to \tau \text{ a.s. as } n \to \infty \quad \forall\, \tau > 0 \quad \forall\, \alpha \neq \beta. \tag{2.3}$$

The proof of the proposition basically makes use of the following two lemmas due to Abdel-Aty and Ferger (2003) and Ferger (2005a). We restate them for the sake of convenience.

Lemma 2.2. *For every positive τ and $\alpha \neq \beta$ it follows that*

$$\sup_{t\geq 0} |S_n(t|\alpha,\beta) - S(t)| \to 0 \text{ a.s. as } n \to \infty. \tag{2.4}$$

Here the limit function S is given by

$$S(t) = \begin{cases} C_1 e^{-\alpha t} + D, & 0 \leq t \leq \tau, \\ C_2 e^{-\beta t}, & \tau < t, \end{cases}$$

with constants $C_1 = \log(\beta/\alpha) + 1 - \beta/\alpha < 0$, $C_2 = (\log(\beta/\alpha) - 1 + \alpha/\beta)e^{(\beta-\alpha)\tau} > 0$, and $D = (\alpha/\beta + \beta/\alpha - 2)e^{-\alpha\tau}$. Moreover, for every $\varepsilon > 0$,

$$S(\tau) - S(t) \geq L(\varepsilon)|\tau - t| \quad \forall \quad t \in [\tau-\varepsilon, \tau+\varepsilon] \cap \mathbb{R}_+ \tag{2.5}$$

with positive constant $L(\varepsilon) = \min\{-\alpha C_1 e^{-\alpha\tau}, \beta C_2 e^{-\beta(\tau+\varepsilon)}\} > 0$. Thus S is continuous with unique maximum at point τ, where the graph of S has a peak.

In the next lemma $D[0,\infty)$ denotes the set of all functions $f : [0;\infty) \longrightarrow \mathbb{R}$ which are right-continuous with left-hand limits (rcll).

Lemma 2.3. *Let $f \in D[0,\infty)$ have a unique* real maximizer *τ, i.e. $f(\tau) = \max_{t\geq 0} f(t)$. If for $r > 0$ the maximum $a(r) := \max\{f(t) : |t-\tau| \geq r\}$ exists, then*

(1) *$b(r) := \frac{1}{3}(f(\tau) - a(r)) > 0$, and*

(2) *if $g \in D[0,\infty)$ is such that the argmax of g is well-defined in the sense of (1.7), then*

$$\sup_{t\geq 0} |f(t) - g(t)| \leq b(r) \Rightarrow \Big| \text{argmax } f(t) - \text{argmax } g(t) \Big| \leq r.$$

Proof of Proposition 2.1. Put $c := \log(\beta/\alpha)$, $d := \beta - \alpha$, $c_n := \log(\beta_n/\alpha_n)$, $d_n := \beta_n - \alpha_n$ and let F_n denote the empirical distribution function pertaining to $X_1, \ldots, X_n$. Then for every $t \geq 0$,

$$S_n(t|\alpha_n,\beta_n) - S_n(t|\alpha,\beta) = (c_n - c)(1 - F_n(t)) + (d_n - d)n^{-1} \sum_{i=1}^{n} (t - X_i) 1_{\{X_i > t\}}.$$

Since for all $t \geq 0$, $|t - X_i| 1_{\{X_i > t\}} = (X_i - t) 1_{\{X_i > t\}} \leq X_i \ \forall\ 1 \leq i \leq n$, we can infer that

$$\sup_{t\geq 0} \Big| S_n(t|\alpha_n,\beta_n) - S_n(t|\alpha,\beta) \Big| \leq |c_n - c| + |d_n - d| n^{-1} \sum_{i=1}^{n} X_i.$$

The upper bound converges to zero P-stochastically or a.s. according as (1.9) or (2.2) holds. Conclude from (2.4) and the triangle-inequality that

$$\sup_{t\geq 0} \Big| S_n(t|\alpha_n,\beta_n) - S(t) \Big| \to 0 \text{ as } n \to \infty \tag{2.6}$$

P-stochastically or a.s. according as (1.9) or (2.2) holds. Now (2.1) and (2.3) follow from (2.6) and Lemma 2.3, upon noticing that $b(\epsilon) \to 0$ as $\epsilon \to 0$.

Of course Proposition 2.1 gives a desirable property of our estimator. But in addition it also serves as a technical tool in the next section about distributional convergence.

3. Distributional convergence

The aim of this section is to derive convergence in distribution of $n(\tau_n - \tau)$ where we again write $\tau_n := \tau_n(\alpha_n, \beta_n)$ for short. Analogously we use the abbreviation $S_n(t) := S_n(t|\alpha_n, \beta_n)$. Our starting point is a representation of $n(\tau_n - \tau)$ in terms of the localized process

$$Z_n(t) := n\Big\{S_n\Big(\tau + \frac{t}{n}\Big) - S_n(\tau)\Big\}, \quad t \in \mathbb{R},$$

which yields

$$n(\tau_n - \tau) = \operatorname*{arg\,max}_{t\in\mathbb{R}} Z_n(t). \tag{3.1}$$

The trajectories of Z_n are rcll and the argmax-functional in (3.1) is defined by analogy with the definition (1.7). We want to apply the Argmax-CMT of Ferger (2004). In short it states that convergence in distribution of Z_n in a functional sense to some limit process Z entails that of argmax of Z_n to the argmax of Z. More precisely we have to show

$$n(\tau_n - \tau) = O_P(1), \quad n \to \infty, \tag{3.2}$$

$$Z_n \xrightarrow{\mathcal{L}} Z \text{ in } D[-a,a] \text{ as } n \to \infty \text{ for all } a > 0, \tag{3.3}$$

$$\operatorname*{arg\,max}_{t\in\mathbb{R}} Z(t) \text{ is a.s. unique.} \tag{3.4}$$

Here $D[-a,a]$ denotes the Skorokhod-space of all rcll functions on $[-a,a]$, confer Billingsley (1968), Chapter 3. Once we have shown (3.2)−(3.4) we may apply the Argmax-CMT of Ferger (2004). Together with (3.1) it then follows

$$n(\tau_n - \tau) \xrightarrow{\mathcal{L}} \operatorname*{arg\,max}_{t\in\mathbb{R}} Z(t), \qquad n \to \infty. \tag{3.5}$$

Below we state (3.2)−(3.4) precisely in Propositions 3.1−3.3. The proofs are deferred to the appendix.

Proposition 3.1. *Let $(\alpha_n, \beta_n) \subseteq \mathbb{R}_+^2$ be an arbitrary sequence satisfying* (1.9)*. Then $n(\tau_n(\alpha_n, \beta_n) - \tau) = O_P(1)$ as $n \to \infty$.*

The exact formulation of (3.3) including the identification of the limit process Z is given in

Proposition 3.2. *If* (1.9) *holds then* $Z_n \xrightarrow{\mathcal{L}} Z$ *in* $D[-a,a]$ *as* $n \to \infty$ *for every* $a > 0$, *where*

$$Z(t) = N(t) + \Delta(t) \tag{3.6}$$

with

$$N(t) = \begin{cases} -\log(\frac{\beta}{\alpha})N_1(t), & t \geq 0 \\ \log(\frac{\beta}{\alpha})N_2(-t), & t < 0 \end{cases}$$

and $\Delta(t) = (\beta - \alpha)e^{-\alpha\tau}t$. *Here* N_1 *and* N_2 *are independent Poisson processes with parameters* $\lambda_1 = \beta e^{-\alpha\tau}$ *and* $\lambda_2 = \alpha e^{-\alpha\tau}$, *respectively. Moreover the second process* N_2 *is chosen such that its trajectories are left-continuous with right-hand limits. Thus* N *and* Z *are rcll.*

It remains to check the last condition (3.4) of our program for the proof of the distributional convergence (3.5).

Proposition 3.3. *The process* Z *has almost surely a unique maximizing point* $T = \operatorname*{arg\,max}_{t\in\mathbb{R}} Z(t)$.

Proposition 3.1–3.3 allow us to apply the Argmax-CMT of Ferger (2004), confer Theorem 3 there, and notice the adjacent Remark 1(2).

Theorem 3.4. *Let* $(\alpha_n, \beta_n) \subseteq \mathbb{R}_+^2$ *be an arbitrary random sequence satisfying* (1.9). *Then*

$$n(\tau_n(\alpha_n, \beta_n) - \tau) \xrightarrow{\mathcal{L}} T = \operatorname*{arg\,max}_{t\in\mathbb{R}} Z(t), \tag{3.7}$$

where Z *is the two-sided Poisson process with linear drift given in* (3.6).

The random variable T admits a representation in terms of the arrival times $(\xi_i)_{i\geq 1}$ and $(\rho_i)_{i\geq 1}$ of the Poisson processes N_1 and N_2, respectively. Recall that $c = \log(\beta/\alpha)$ and put $m = (\beta - \alpha)e^{-\alpha\tau}$. Using the arguments of Section 4 of Ferger (2005a), it can be shown that $\gamma := \operatorname*{arg\,max}_{i\geq 1}\{m\xi_i - ci\}$ and $\sigma := \operatorname*{arg\,max}_{i\geq 1}\{c(i-1) - m\rho_i\}$ are a.s. finite. Moreover

$$T = \begin{cases} \xi_\gamma, & \text{if } \; m\xi_\gamma - c\gamma > c(\sigma - 1) - m\rho_\sigma \\ -\rho_\sigma, & \text{otherwise.} \end{cases} \tag{3.8}$$

Therefore T is a continuous random variable and, by Lemma 2.11 in Van der Vaart (1998), the distributional convergence (3.7) implies uniform convergence, i.e.,

$$\sup_{x\in\mathbb{R}} \|P(n(\tau_n - \tau) \leq x) - P(T \leq x)| \longrightarrow 0 \quad \text{ as } n \to \infty.$$

Remark 3.5. (1) Theorem 3.4 includes the mle $\tau_n(\alpha,\beta)$, because $(\alpha_n,\beta_n)=(\alpha,\beta)$ obviously fulfills (1.9). If α and β are known, the density $f(x|\tau,\alpha,\beta)$ depends on the one-dimensional parameter $\tau\in(0,\infty)$. Such parametrized densities with jumps (singularities) are studied in Chapter V of Ibragimov and Has'minskii (1981). A formal application of their Theorem 4.6 yields the same limit for $n(\tau_n(\alpha,\beta)-\tau)$ as our Theorem 3.4 in this special situation. (2) The limit variable T also coincides with that of Pham and Nguyen (1990). (Notice that there is a misprint concerning the reziprocals of the intensities λ_1 and λ_2.)

We end our main section with an application of Theorem 3.4 to *bias and variance estimation*. Here the following sharpening of Proposition 3.1 plays a key role. For the proof see Ferger (2005b).

Proposition 3.6. *For every even $m\in\mathbb{N}$ there exists a constant $C=C(m)$ such that for all $a\geq 1$, $\limsup_{n\to\infty}P(n|\tau_n-\tau|>a)\leq Ca^{-m/2}$.*

Corollary 3.7. *For every real number $p\geq 1$ the sequence $\{(n(\tau_n-\tau))^p : n\in\mathbb{N}\}$ is uniformly integrable.*

Proof. Put $\nu_n:=n(\tau_n-\tau)$. Then for every $\epsilon>0$,

$$\begin{aligned}\limsup_{n\to\infty}E(1_{\{|\nu_n|^p\geq a\}}|\nu_n|^p) &\leq a^{-\epsilon}\limsup_{n\to\infty}E(|\nu_n|^{p(1+\epsilon)})\\ &= a^{-\epsilon}\limsup_{n\to\infty}\int_0^\infty P(|\nu_n|^{p(1+\epsilon)}>x)dx\\ &\leq a^{-\epsilon}\limsup_{n\to\infty}\Big\{1+\int_1^\infty P\Big(|\nu_n|>x^{\frac{1}{p(1+\epsilon)}}\Big)dx\Big\}\\ &\leq a^{-\epsilon}\Big\{1+\int_1^\infty \limsup_{n\to\infty}P\Big(|\nu_n|>x^{\frac{1}{p(1+\epsilon)}}\Big)dx\Big\},\end{aligned}$$

where the last inequality holds by Fatou's Lemma. Now Proposition 3.6 with $m>2p(1+\epsilon)$ shows that the integral is bounded by a constant, whence the assertion follows.

Since T is a continuous random variable, Theorems 3.4 and 3.6 yield the following tailbound for $T : P(|T|>a)\leq Ca^{-m/2}$ for all $a\geq 1$ and for all even integers m. Consequently all moments of T exist and are finite. Moreover, another application of Theorem 3.4, and Corollary 3.7 in combination with Theorem 5.4 in Billingsley (1968), gives

Corollary 3.8. *For $p\geq 1$, let $\|\cdot\|_p$ denote the L_p -norm on our underlying probability space $(\Omega,\mathcal{A},P)$. Then $\|\tau_n-\tau\|_p\sim n^{-1}\|T\|_p$.*

For the same reason we also have $E((n(\tau_n-\tau))^p)\longrightarrow E(T^p)$ as $n\to\infty$ for every $p\geq 1$. Especially it follows for the bias that

$$E(\tau_n)-\tau\sim n^{-1}E(T) \tag{3.9}$$

and for the variance that

$$\mathrm{Var}\,(\tau_n) \sim n^{-2}\mathrm{Var}\,(T). \tag{3.10}$$

So once we have estimators for $E(T)$ and $\mathrm{Var}(T)$, the asymptotic expansions (3.9) and (3.10) yield finite sample estimators for the bias and for the variance of our estimator τ_n. Estimation of $E(T)$ and $\mathrm{Var}\,(T)$ can be done with the Monte-Carlo method using (3.8). Here the unknown quantities τ, α and β occurring in (3.8) have to be replaced by estimators τ_n, α_n and β_n, see the next section.

4. Estimation of the parameters α and β

Our estimator $\tau_n(\alpha_n, \beta_n)$ requires a weakly consistent sequence (α_n, β_n) for (α, β). One possibility is to use Yao' (1986) estimator $(\tilde{\alpha}_n, \tilde{\beta}_n)$ given in (1.5). In this section we present two alternative methods. For the first method we assume that there is a known interval $[\lambda, \rho]$ containing τ. The second method gets along without that assumption. Define for any fixed $t > 0$,

$$\alpha_n(t) := \frac{F_n(t)}{D_n(t) + t(1 - F_n(t))},$$

and for $0 < t < X_{n:n}$,

$$\beta_n(t) := \frac{1 - F_n(t)}{E_n(t) - t(1 - F_n(t))},$$

where $D_n(t) := 1/n \sum_{i=1}^n X_i 1_{\{X_i \le t\}}$ and $E_n(t) := 1/n \sum_{i=1}^n X_i 1_{\{X_i > t\}}$. Since $X_{n:n} \to \infty$ a.s. as $n \to \infty$, the second estimator $\beta_n(t)$ is well-defined for eventually all $n \in \mathbb{N}$. Notice that for fixed $\tau > 0$ the pair $(\alpha_n(\tau), \beta_n(\tau))$ coincides with the mle for (α, β).

The next result ensures that $(\alpha_n(\lambda), \beta_n(\rho))$ is strongly consistent for (α, β) whenever $\tau \in [\lambda, \rho]$. It is a simple consequence of the Strong Law of Large Numbers.

Proposition 4.1.

$$(\alpha_n(\lambda), \beta_n(\rho)) \longrightarrow (\alpha, \beta) \text{ a.s. as } n \to \infty \quad \forall\, \alpha \ne \beta \quad \forall\, \tau \in [\lambda, \rho]. \tag{4.1}$$

Besides strong consistency the estimator $(\alpha_n(\lambda), \beta_n(\rho))$ is asymptotically normal.

Proposition 4.2. *If* $0 < \lambda \le \tau \le \rho$, *then*

$$n^{\frac{1}{2}}\Big\{\begin{pmatrix}\alpha_n(\lambda)\\ \beta_n(\rho)\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\Big\} \xrightarrow{\mathcal{L}} \begin{pmatrix}U\\ V\end{pmatrix} \quad \text{as } n \to \infty,$$

where $U \sim N(0, \sigma^2)$ and $V \sim N(0, s^2)$ are independent normal variables. Furthermore $\sigma^2 = \sigma^2(\lambda) = \alpha^2\{1 - e^{-\alpha\lambda}\}^{-1}$ and $s^2 = s^2(\rho) = \beta^2 e^{\alpha\rho}$.

Proof. Put

$$Z_n := \frac{1}{n}\sum_{i=1}^{n}(1_{\{X_i \le \lambda\}}, X_i 1_{\{X_i \le \lambda\}}, 1_{\{X_i > \rho\}}, X_i 1_{\{X_i > \rho\}}). \tag{4.2}$$

By the Central Limit Theorem,

$$n^{\frac{1}{2}}\{Z_n - E(Z_1)\} \xrightarrow{\mathcal{L}} N_4(0, \Gamma) \text{ as } n \to \infty, \tag{4.3}$$

where the covariance matrix Γ is block-diagonal. Now a laborious but straightforward application of the Delta-method gives the result.

Remark 4.3. In the special case $\lambda = \tau = \rho$, we obtain Yao's (1986) asymptotic normality of the mle for (α, β) when τ is fixed (and known).

Remark 4.4. If $\tau \in [\lambda, \rho]$ it is reasonable to restrict the maximization of $S_n(t|\alpha_n, \beta_n)$ to the region $[\lambda, \rho]$, which leads to the modified estimator

$$\tau_n(\alpha_n, \beta_n, \lambda, \rho) := \underset{t \in [\lambda, \rho]}{\arg\max}\, S_n(t|\alpha_n, \beta_n).$$

Our results about consistency and distributional convergence in Sections 2 and 3 carry over to $\tau_n(\alpha_n, \beta_n, \lambda, \rho)$ without any problems.

Remark 4.5. In the situation of Remark 4.4, we do not know the true value of τ but a region $[\lambda, \rho]$ where in it lies. It is expected that the more we can narrow down this region the better the estimator should be concerning its variance. Indeed this phenomenon is reflected in Proposition 4.2. Namely the limit variances $\sigma^2(\lambda)$ and $s^2(\rho)$ are decreasing as $\lambda \uparrow \tau$ and $\rho \downarrow \tau$, respectively. On the other side these variances may be unacceptably large if the region $[\lambda, \rho]$ is too big, because $\sigma^2(\lambda) \to \infty$ and $s^2(\rho) \to \infty$ exponentially fast as $\lambda \to 0$ and $\rho \to \infty$, respectively. In this situation we recommend our second method, which also works in the general case where $\tau \in (0, \infty)$ is completely unknown and no specifying region $[\lambda, \rho]$ containing τ is available. Here we use a pilot-estimator, say τ_n^*, for τ, for instance $\tau_n^* = \tilde{\tau}_n$ as given in (1.5). Consider the corresponding plug-in-estimator $\alpha_n^* := \alpha_n(\tau_n^*)$ and $\beta_n^* := \beta_n(\tau_n^*)$. Recall that Yao (1986) shows distributional convergence of $n(\tilde{\tau}_n - \tau)$, which entails that $\tilde{\tau}_n - \tau = O_P(n^{-1})$. It turns out that this property is neccessary for proving asymptotic normality of (α_n^*, β_n^*), which exhibits the same minimal variances $\sigma^2(\tau)$ and $s^2(\tau)$ as the mle $(\alpha_n(\tau), \beta_n(\tau))$. Once more notice that this mle requires the true value of τ which we do not know.

Proposition 4.6. (1) *If τ_n^* is weakly consistent, then $(\alpha_n^*, \beta_n^*) \xrightarrow{P} (\alpha, \beta)$ as $n \to \infty$ $\forall\ \alpha \neq \beta\ \forall\ \tau > 0$.* (2) *If τ_n^* actually is strongly consistent, then $(\alpha_n^*, \beta_n^*) \longrightarrow (\alpha, \beta)$ a.s. as $n \to \infty$ $\forall\ \alpha \neq \beta\ \forall\ \tau > 0$.*

Proof. Observe that $|F_n(\tau_n^*)-F(\tau)| \le |F(\tau_n^*)-F(\tau)|+\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|$ and $|D_n(\tau_n^*)-D(\tau)| \le |D(\tau_n^*)-D(\tau)|+\sup_{x\in\mathbb{R}}|D_n(x)-D(x)|$. In the proof of Lemma 3.3 in Abdel-Aty and Ferger (2003) we show that $\sup_{x\in\mathbb{R}}|D_n(x)-D(x)| \longrightarrow 0$ a.s. as $n \to \infty$, which immediately yields the a.s. uniform convergence of E_n. Together with the Glivenko-Cantelli Theorem, we arrive at (1) and (2) upon noticing that F and D are continuous.

Proposition 4.7. *If*

$$\tau_n^* - \tau = O_P(n^{-1}), \tag{4.4}$$

then

$$n^{\frac{1}{2}}\left\{\begin{pmatrix}\alpha_n(\tau_n^*)\\ \beta_n(\tau_n^*)\end{pmatrix} - \begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right\} \xrightarrow{\mathcal{L}} \begin{pmatrix}U\\ V\end{pmatrix},$$

where $U \sim N(0,\sigma^2(\tau))$ and $V \sim N(0,s^2(\tau))$ are independent.

Proof. Let Z_n and Z_n^* be the random vector in (4.2) with $\lambda := \tau =: \rho$ and $\lambda := \tau_n^* =: \rho$, respectively. Then once we have shown that $n^{1/2}\{Z_n^* - Z_n\} \xrightarrow{P} 0$ as $n \to \infty$, the result follows from (4.3), Slutsky's Theorem and the Delta-Method. For that purpose note that

$$Z_n^* - Z_n = (F_n(\tau_n^*) - F_n(\tau), D_n(\tau_n^*) - D_n(\tau), F_n(\tau) - F_n(\tau_n^*), D_n(\tau) - D_n(\tau_n^*)).$$

Thus it suffices to prove

$$n^{\frac{1}{2}}\Big\{F_n(\tau_n^*) - F_n(\tau)\Big\} \xrightarrow{P} 0, \quad n \to \infty, \tag{4.5}$$

$$n^{\frac{1}{2}}\Big\{D_n(\tau_n^*) - D_n(\tau)\Big\} \xrightarrow{P} 0, \quad n \to \infty. \tag{4.6}$$

For the proof of (4.5), let $\varepsilon > 0$ be an arbitrary positive real number. Then for every $M > 0$,

$$\begin{aligned} &P\Big(n^{\frac{1}{2}}|F_n(\tau_n^*) - F_n(\tau)| > \varepsilon\Big)\\ &\quad \le P\Big(n^{\frac{1}{2}}|F_n(\tau_n^*) - F_n(\tau)| > \varepsilon, |\tau_n^* - \tau| \le Mn^{-1}\Big) + P(|\tau_n^* - \tau| > Mn^{-1})\\ &\quad =: A_n(\varepsilon, M) + B_n(M). \end{aligned}$$

Since $A_n(\varepsilon, M) \le P(S_n > n^{1/2}\varepsilon)$, where by the Poisson Limit Theorem has

$$S_n := \sum_{i=1}^{n} 1_{\{\tau - Mn^{-1} < X_i \le \tau + Mn^{-1}\}} \xrightarrow{\mathcal{L}} \text{Poisson}(\lambda), \quad n \to \infty,$$

with $\lambda = (f'(\tau+) + f'(\tau-))M$. It follows that $\lim_{n\to\infty} A_n(\varepsilon, M) = 0\ \forall\ M > 0$ $\forall\ \varepsilon > 0$. By Assumption (4.4) we have that

$$\lim_{M\to\infty}\limsup_{n\to\infty} B_n(M) = 0, \tag{4.7}$$

which shows (4.5). The derivation of (4.6) is similar. Namely,

$$
\begin{aligned}
&P\Big(n^{\frac{1}{2}}|D_n(\tau_n^*)-D_n(\tau)|>\varepsilon\Big)\\
&\quad\le P\Big(n^{\frac{1}{2}}|D_n(\tau_n^*)-D_n(\tau)|>\varepsilon, |\tau_n^*-\tau|\le Mn^{-1}, X_{n:n}\le n^{\frac{1}{3}}\Big)\\
&\qquad+P\Big(|\tau_n^*-\tau|>Mn^{-1}\Big)+P\Big(X_{n:n}>n^{\frac{1}{3}}\Big)\\
&\quad=: a_n(\varepsilon,M)+B_n(M)+C_n.
\end{aligned}
$$

Since $a_n(\varepsilon,M)\le P(S_n>\varepsilon n^{1/6})$, we may conclude that

$$\lim_{n\to\infty} a_n(\varepsilon,M)=0 \qquad \forall\ M>0 \quad \forall\ \varepsilon>0. \tag{4.8}$$

A routine application of the First Borel-Cantelli Lemma yields that $\lim_{n\to\infty} C_n = 0$, so that (4.6) follows in view of (4.7) and (4.8).

Our theoretical results lead us to the following recommendations for the applications. If it is sure that τ lies in an interval $[\lambda,\rho]$, then clearly one should use $\tau_n(\alpha_n(\lambda),\beta_n(\rho),\lambda,\rho)$. In view of Remark 4.5 one must be cautious if λ is close to zero or ρ is large. As a way out, replace these estimators by $\alpha_n(\tilde\tau_n)$ and $\beta_n(\tilde\tau_n)$, where $\tilde\tau_n$ denotes Yao's (1986) estimator. This results in $\tau_n(\alpha_n(\tilde\tau_n),\beta_n(\tilde\tau_n),\lambda,\rho)$. Finally, if no prior information about the location of τ is available then use the estimator $\tau_n(\alpha_n(\tilde\tau_n),\beta_n(\tilde\tau_n))$. The performance of these estimators in comparison to Yao's (1986) estimator is investigated in a simulation study presented in the next section.

5. Simulation and robustness

In this section we present some results of a small simulation study. It yields the performance of the following five estimators: $\tau_n^{(1)} := \tau_n(\alpha,\beta)$, $\tau_n^{(2)} := \tau_n(\alpha_n(\lambda),\beta_n(\rho),\lambda,\rho)$, $\tau_n^{(3)}$:= Yao's estimator, $\tau_n^{(4)} := \tau_n(\alpha_n(\tau_n^{(3)}),\beta_n(\tau_n^{(3)}))$ and $\tau_n^{(5)} := \tau_n(\alpha_n(\tau_n^{(3)}),\beta_n(\tau_n^{(3)}),\lambda,\rho)$. The last estimator $\tau_n^{(5)}$ uses the obvious modification of Yao's estimator, where the maximization is restricted to the region $[\lambda,\rho]$.

From distribution theory established in Section 3, the five estimators do not differ asymptotically in the sense that $n(\tau_n^{(i)}-\tau)\longrightarrow T$ in law and in L_p, $1\le i\le 5$. But in finite sample situations it may happen that there are significant differences as our simulations results below demonstrate.

We simulated n independent random variables stemming from a hazard function (1.1) with $\alpha = 1.8$, $\beta = 0.9$ and $\tau = 1$, and computed the pertaining estimators $\tau_n^{(i)}, 1\le i\le 5$, based on that sample. This procedure was repeated 10^5-times. The following tables contain the mean and the square root of the mean squared error MSE of the five estimators. For the sake of brevity we display only a few tables, but our conclusions are based on much more.

Table 1. $n = 25$, $\lambda = 0.3$, $\rho = 2.4$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(1)}$	0.8156	0.4037
$\tau_n^{(2)}$	0.9363	0.5646
$\tau_n^{(3)}$	0.7560	0.8394
$\tau_n^{(4)}$	0.7614	0.8319
$\tau_n^{(5)}$	0.9291	0.4995

Table 2. $n = 500$, $\lambda = 0.3$, $\rho = 2.4$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(1)}$	0.9994	0.1059
$\tau_n^{(2)}$	1.0011	0.1769
$\tau_n^{(3)}$	0.9966	0.1513
$\tau_n^{(4)}$	0.9966	0.1513
$\tau_n^{(5)}$	0.9941	0.1119

Simulations show that, already for smaller sample sizes starting with $n \geq 50$, all estimators produce good results. The estimators $\tau_n^{(2)}$ and $\tau_n^{(5)}$ do this even for $n = 25$. This can be explained by the difference between α and β. If α and β are closer then one needs larger sample sizes. Note that $\tau_n^{(1)}$ requires the true values of α and β, which in our paper are assumed to be unknown. This means that we must consider this estimator as a non-official competitor. Among the remaining ones $\tau_n^{(5)}$ is superior in view of its MSE. The estimators $\tau_n^{(3)}$ and $\tau_n^{(4)}$ have similar performance. In contrast with $\tau_n^{(2)}$ and $\tau_n^{(5)}$, they get along without knowledge about the lower and upper bound λ and ρ. Nevertheless, as Table 2 shows, they are better than $\tau_n^{(2)}$ for larger sample sizes. The reason for this might lie in that the estimators $\alpha_n(\lambda)$ and $\beta_n(\rho)$ give poor approximations of the true parameters α and β if λ is close to zero and ρ is large. Indeed recall, e.g., the definition of $\alpha_n(\lambda)$. It involves only observation which lie to the left of λ, and there are only a few of those if the sample size is not large enough. The same is true for $\beta_n(\rho)$. Another theoretical explanation of this phenomenon has already been given in Remark 4.5 above. Tables 3-8 show the impact of λ and ρ. We let the length of the τ-covering interval $[\lambda, \rho]$ change from small to large. Of course for small intervals we obtain very nice results for all sample sizes and for both estimators, as was to be expected. However, for $n = 500$ we observe that the performance of the estimator $\tau_n^{(2)}$ becomes poorer the longer the intervals. This is reflected by the increasing MSE, whereas the MSE of $\tau_n^{(5)}$ remains rather stable. In Table 8 the difference of $\tau_n^{(2)}$ and $\tau_n^{(5)}$ is obvious.

Table 3. $n = 100$, $\lambda = 0.9$, $\rho = 1.1$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9732	0.0509
$\tau_n^{(5)}$	0.9733	0.0507

Table 4. $n = 100$, $\lambda = 0.1$, $\rho = 1.9$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9005	0.4474
$\tau_n^{(5)}$	0.8893	0.3664

Table 5. $n = 100$, $\lambda = 0.01$, $\rho = 19$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	1.0722	2.0051
$\tau_n^{(5)}$	1.0567	0.7902

Table 6. $n = 500$, $\lambda = 0.9$, $\rho = 1.1$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9910	0.0422
$\tau_n^{(5)}$	0.9913	0.0420

Table 7. $n = 500$, $\lambda = 0.1$, $\rho = 1.9$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9960	0.1446
$\tau_n^{(5)}$	0.9930	0.1081

Table 8. $n = 500$, $\lambda = 0.01$, $\rho = 19$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	1.2324	2.4798
$\tau_n^{(5)}$	0.9966	0.1448

The next tables show the performance of $\tau_n^{(2)}$, $\tau_n^{(3)}$ and $\tau_n^{(5)}$ in case the data stem from the hazard function

$$h(t) = \begin{cases} \alpha + A\sin(kt), & 0 \le t \le \tau \\ \beta + B\sin(kt), & t > \tau. \end{cases} \tag{5.1}$$

Here the hazard function of (1.1) is superimposed with a sinusoidal oscillation. Note that small values of A and B correspond to a small deviation. In our simulations we fixed $\alpha = 1.8$, $\beta = 0.9$, $\tau = 1$, $\lambda = 0.3$ and $\rho = 2.4$. Again each value of our tables below are based on 10^5 Monte-Carlo replicates.

Table 9. $n=50$, $A=B=0.1$, $k=5$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	1.0755	0.5955
$\tau_n^{(3)}$	0.8968	0.8455
$\tau_n^{(5)}$	0.9534	0.4647

Table 10. $n=50$, $A=B=0.5$, $k=5$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9876	0.5765
$\tau_n^{(3)}$	0.8065	0.7575
$\tau_n^{(5)}$	0.8111	0.4381

Table 11. $n=500$, $A=B=0.1$, $k=5$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.9634	0.1878
$\tau_n^{(3)}$	0.9611	0.1565
$\tau_n^{(5)}$	0.9590	0.1291

Table 12. $n=500$, $A=B=0.5$, $k=5$.

	mean	$\sqrt{\text{MSE}}$
$\tau_n^{(2)}$	0.7773	0.2874
$\tau_n^{(3)}$	0.7755	0.2662
$\tau_n^{(5)}$	0.7752	0.2650

For small deviations from (1.1) all estimators are robust, and $\tau_n^{(5)}$ again exhibits the smallest MSE. If the deviation increases, then even for large sample size the bias is rather poor.

We draw the following conclusions. If no information about the position of τ is available we recommend $\tau_n^{(4)}$, which shows almost the same performance as Yao's (1986) estimator $\tau_n^{(3)}$. Otherwise one should use $\tau_n^{(5)}$, which is better than $\tau_n^{(2)}$.

6. A generalized model

In reality h at (1.1) is an approximation of the true

$$h(t) = \begin{cases} h_1(t), & 0 \le t \le \tau \\ h_2(t), & \tau < t < \infty, \end{cases} \tag{6.1}$$

with $h_1, h_2 \in C([0,\infty))$ both positive and $h_1(\tau) \neq h_2(\tau)$.

For example Mair, Goymer, Pletcher and Partridge (2003) observe lifetimes of flies (drosophila). In their experiment the flies were fully fed with yeast and sugar until a (known) time point τ. After this the amount of food was abruptly and drastically reduced (dietatry restriction). The food medium after time τ contained roughly 35% less yeast and sugar. The estimated mortality rate indicates strongly that the true underlying mortality rate is $h_i(t) = \exp\{a_i + bt\}$, $i = 1, 2$, with $a_1 \neq a_2$ and $b > 0$.

The hazard function h uniquely determines a distribution function F with corresponding density

$$f(x) = \begin{cases} h_1(x)\exp\{-H_1(x)\}, & 0 \le t \le \tau \\ h_2(x)\exp\{H_2(\tau) - H_1(\tau) - H_2(x)\}, & \tau < x < \infty, \end{cases} \tag{6.2}$$

where

$$H_i(x) := \int_0^x h_i(t)dt, \quad x \ge 0, \quad i = 1, 2,$$

denotes the cumulative hazard functions. Let $X_1, \ldots, X_n$ be i.i.d. with density f. To begin with, assume that h_1 and h_2 are known. Then the mle of τ is $\tau_n = \operatorname{argmax}_{t \ge 0} S_n(t)$, where $S_n(t) = n^{-1}\sum_{i=1}^n 1_{\{X_i > t\}}[\log h_2/h_1(X_i) + (H_2 - H_1)(t) - (H_2 - H_1)(X_i)]$. If

$$E\Big|\log\frac{h_1}{h_2}(X_1)\Big| < \infty \text{ and } E|(H_2 - H_1)(X_1)| < \infty, \tag{6.3}$$

then by the Strong Law of Large Numbers, $S_n(t) \longrightarrow S(t)$ a.s. as $n \to \infty$ for all $t \ge 0$, where

$$S(t) = \int_t^\infty \Big[\log\frac{h_1}{h_2}(x) - (H_2 - H_1)(x)\Big] f(x)dx + (H_2 - H_1)(t)(1 - F(t)).$$

Check that for all $t \geq 0$, $S'(t) = -f(t) \log h_2/h_1(t) + (h_2(t) - h_1(t))(1 - F(t))$. Since by definition $h(t) = f(t)/(1 - F(t))$, division by $1 - F(t)$ shows that the derivative $S'(t)$ is non-negative or non-positive according as

$$-h(t) \log \frac{h_2}{h_1}(t) + (h_2(t) - h_1(t))$$

is non-negative or non-positive. Use the well-known inequalitites $1 - 1/x < \log x < x - 1$ for every positive $x \neq 1$ (with equality if and only if $x = 1$) to see that S is monotone increasing on $[0, \tau]$ and monotone decreasing on $[\tau, \infty)$. By continuity of h_1 and h_2, and since $h_1(\tau) \neq h_2(\tau)$ by assumption, the monotonicity is strict in a neighborhood of τ. Thus S has a unique maximum at point τ. This allows us to prove strong consistency of τ_n essentially in the same way as in Abdel-Aty and Ferger (2003). Here, however, some technical refinements in the arguments are necessary.

Theorem 6.1. *Suppose* (6.3) *holds. Moreover assume that there exists a positive sequence* (d_n) *with* $d_n \longrightarrow \infty$ *such that*

$$\sum_{n \geq 1} n \exp\{-H_2(d_n)\} < \infty, \tag{6.4}$$

$$H_i(d_n) n^{-\frac{1}{2}} \sqrt{\log \log n} \longrightarrow 0, \quad i = 1, 2, \tag{6.5}$$

$$|H_1(t) - H_2(t)| \exp\{-H_2(t)\} \longrightarrow 0 \text{ as } t \longrightarrow \infty. \tag{6.6}$$

Then $\hat{\tau}_n \to \tau$ *a.s. as* $n \to \infty$ *for every* $\tau > 0$.

The assumptions (6.4)−(6.6) concern the tail-behavior of the underlying distribution function F. They are easy to verify in the following two examples.

Example 6.2. (Generalized Exponential Distribution). Let $h_i(t) = \alpha_i + \lambda_i t$, $i = 1, 2$, with $\alpha_i, \lambda_i \geq 0$, $i = 1, 2$ and $\alpha_1 + \lambda_1 \tau \neq \alpha_2 + \lambda_2 \tau$. Finkelstein (2003) uses this model in life time data analysis for describing a change in the environment. He does not allow for a jump at point τ, but only a change in the slopes of the two linear function. Clearly $\lambda_1 = \lambda_2 = 0$ yields our model (1.1).

Example 6.3. (Gompertz Distribution). Recall the example with the fly drosophila. In general we have to do with $h_i(t) = \exp\{a_i + b_i t\}$, $i = 1, 2$, with $a_1, a_2 \in \mathbb{R}$, $b_1, b_2 > 0$ and $a_1 + b_1 \tau \neq a_2 + b_2 \tau$.

Note that the density f in (6.2) has a jump at point $\tau > 0$ (and also at point zero). If we ignore – as in Remark 3.5 – that condition (II) on p.242 in Ibragimov and Has'minskii (1981) is violated, a formal application of their

Theorem 4.6 yields convergence in distribution of $n(\hat{\tau}_n - \tau)$ with a similar limit as in (3.7).

7. An outlook

Clearly the requirement that h_1 and h_2 need to be known is very restrictive and in reality will rarely be fulfilled. So we close this paper with a discussion of the scenario in which h_1 and h_2 are both unknown. What we have in mind here is a general hazard function h as in (6.1), which however is not too far away from the simple h in (1.1). Figure 1 is an illustration.

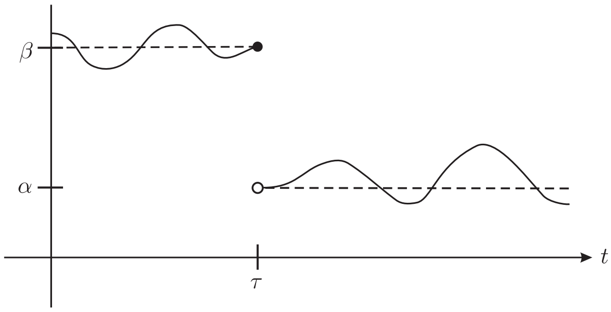

Figure 1. True hazard function (solid) and simple hazard function (dashed).

If we assume h_1 and h_2 are k-times continuously differentiable, then Taylor's Theorem brings us to parametrize h_1 and h_2 in the following way:

$$h_1(t) = \sum_{l=0}^{k} \alpha_l (t-\tau)^l, \quad t \geq 0, \tag{7.1}$$

$$h_2(t) = \sum_{l=0}^{k} \beta_l (t-\tau)^l, \quad t \geq 0, \tag{7.2}$$

with parameter vectors $\underline{\alpha} = (\alpha_0, \ldots, \alpha_k)$ and $\underline{\beta} = (\beta_0, \ldots, \beta_k)$ in $\mathbb{R}^{k+1}$. Let

$$\begin{aligned} l_n(\tau, \underline{\alpha}, \underline{\beta}) = \sum_{i=1}^{n} & 1_{\{X_i \leq \tau\}}[\log h_1(X_i) - H_1(X_i)] \\ & + 1_{\{X_i > \tau\}}[\log h_2(X_i) - H_2(X_i) + (H_2 - H_1)(\tau)] \end{aligned}$$

denote the log-likelihood function. For fixed τ the mle

$$(\underline{\hat{\alpha}}_n, \underline{\hat{\beta}}_n) := \text{argmax}\ \{l_n(\tau, \underline{\alpha}, \underline{\beta}) : \underline{\alpha}, \underline{\beta} \in \mathbb{R}^{k+1}\}$$

is well-defined. It is a solution of the pertinent system of normal equations. Now we recommend the following 3-step procedure.

Step 1. Start with $\tau_n^{(0)} := \tau_n(\alpha_n, \beta_n)$ as a pilot-estimator. (In a situation as depicted in Figure 1, we expect a good first shot due to robustness as was shown in Section 5.)

Step 2. Compute the mle $(\underline{\hat{\alpha}}_n, \underline{\hat{\beta}}_n)$ pertaining to $\tau_n^{(0)}$.

Step 3. Compute $\hat{\tau}_n$ pertaining to $\hat{h}_1(t) = \sum_{l=0}^{k} \hat{\alpha}_{ln}(t - \tau_n^{(0)})^l$ and $\hat{h}_2(t) = \sum_{l=0}^{k} \hat{\beta}_{ln}(t - \tau_n^{(0)})^l$, where $\hat{\alpha}_{ln}$ and $\hat{\beta}_{ln}$ denotes the l-th component of $\underline{\hat{\alpha}}_n$ and $\underline{\hat{\beta}}_n$, respectively.

Notice that in view of Theorem 3.4 the estimator $\tau_n^{(0)}$ should converge at rate n to τ, whereas the mle $(\underline{\hat{\alpha}}_n, \underline{\hat{\beta}}_n)$ should converge at rate $\sqrt{n}$ to $(\underline{\alpha}, \underline{\beta})$. Therefore we suppose that the mle in Step 2 produces a good estimate for $(\underline{\alpha}, \underline{\beta})$ and, in turn, $\hat{\tau}_n$ in Step 3 a good estimate for τ.

Of course, there are other parametrizations of h_1 and h_2 than those in (7.1) and (7.2). These may come from technical, non-mathematical considerations. However, in each case, the approach in principle remains the same.

A. Appendix

Proof of Proposition 3.1. Fix $a > 0$ and put $x := x_n := an^{-1}$. Then for every $r \in (0, \tau)$ and for eventually all $n \in \mathbb{N}$ such that $x_n < r$, we have that

$$\begin{aligned} P(n|\tau_n - \tau| > a) &\le P(x < |\tau_n - \tau| \le r) + P(|\tau_n - \tau| > r) \\ &=: P_n(a, r) + Q_n(r). \end{aligned} \tag{A.1}$$

By Proposition 2.1

$$Q_n(r) \to 0 \ \text{ as } \ n \to \infty \quad \forall\, r > 0. \tag{A.2}$$

For the treatment of $P_n(a, r)$ we set $\bar{S}_n(t) := S_n(t|\alpha, \beta)$, $t \in \mathbb{R}$. Observe that

$$\begin{aligned} P_n(a, r) &\le P\Big(\sup_{x<|t-\tau|\le r} S_n(t) \ge S_n(\tau)\Big) \\ &\le P\Big(\sup_{x<t\le r} \{S_n(\tau + t) - S_n(\tau)\} \ge 0\Big) + P\Big(\sup_{x<t\le r} \{S_n(\tau - t) - S_n(\tau)\} \ge 0\Big) \\ &:= P_{n,1}(a, r) + P_{n,2}(a, r). \end{aligned} \tag{A.3}$$

Next notice that for all $t \in (x, r]$

$$\begin{aligned} S_n(\tau + t) - S_n(\tau) &= S_n(\tau + t) - S(\tau + t) - [S_n(\tau) - S(\tau)] - [S(\tau) - S(\tau + t)] \\ &\le S_n(\tau + t) - S(\tau + t) - [S_n(\tau) - S(\tau)] - L(r)t, \end{aligned}$$

where the last equality follows from (2.5). Recall that the constant $L(r)$ is positive. Consequently

$$\begin{aligned}
P_{n,1}(a,r) &\le P\Big(\sup_{x<t\le r} \frac{S_n(\tau+t)-S(\tau+t)-[S_n(\tau)-S(\tau)]}{t} \ge L(r)\Big)\\
&\le P\Big(\sup_{x<t\le r} \frac{\bar S_n(\tau+t)-S(\tau+t)-[\bar S_n(\tau)-S(\tau)]}{t} \ge \frac{L(r)}{2}\Big)\\
&\quad +P\Big(\sup_{x<t\le r} \frac{S_n(\tau+t)-\bar S(\tau+t)-[S_n(\tau)-\bar S(\tau)]}{t} \ge \frac{L(r)}{2}\Big)\\
&=: p_n(a,r)+q_n(a,r). \qquad\qquad \text{(A.4)}
\end{aligned}$$

Let us begin with with $p_n(a,r)$. Recall that by definition

$$\bar S_n(t) = c(1-F_n(t)) + dn^{-1}\sum_{i=1}^n (t-X_i)1_{\{X_i>t\}}, \quad t\in\mathbb{R},$$

with $c=\log(\beta/x)$ and $d=\beta-\alpha$, whence

$$\begin{aligned}
\bar S_n(\tau+t)-\bar S_n(\tau) &= -c\{F_n(\tau+t)-F_n(\tau)\} + dt\{1-F_n(\tau+t)\}\\
&\quad +dn^{-1}\sum_{i=1}^n (X_i-\tau)1_{\{\tau<X_i\le\tau+t\}}.
\end{aligned}$$

Since $S(t)=E\bar S_n(t)$, and taking expectation is linear, one has

$$\begin{aligned}
&\bar S_n(\tau+t)-S(\tau+t)-[\bar S_n(\tau)-S(\tau)]\\
&\quad = -c\{F_n(\tau+t)-F(\tau+t)-[F_n(\tau)-F(\tau)]\} - dt\{F_n(\tau+t)-F(\tau+t)\}\\
&\qquad +dn^{-1}\sum_{i=1}^n (X_i-\tau)1_{\{\tau<X_i\le\tau+t\}} - d\int_\tau^{\tau+t}(x-\tau)F(dx) \qquad \forall\, t\ge 0.
\end{aligned}$$

Check the simple inequalities

$$0 \le (X_i-\tau)1_{\{\tau<X_i\le\tau+t\}} \le t1_{\{\tau<X_i\le\tau+t\}} \le t1_{\{\tau<X_i\le\tau+r\}}, \qquad \text{(A.5)}$$

and $0\le\int_\tau^{\tau+t}(x-\tau)F(dx) \le t\{F(\tau+t)-F(\tau)\} \le t\{F(\tau+r)-F(\tau)\}$ for all $0<t\le r$. Moreover let $\alpha_n(t):=n^{1/2}\{F_n(t)-F(t)\}$, $t\in\mathbb{R}$, denote the empirical process pertaining to $X_1,\dots,X_n$. Then we obtain

$$\begin{aligned}
&\sup_{x<t\le r}\frac{\bar S_n(\tau+t)-S(\tau+t)-[\bar S_n(\tau)-S(\tau)]}{t}\\
&\quad\le |c|n^{-\frac12}\sup_{x<t\le r}\frac{|\alpha_n(\tau+t)-\alpha_n(\tau)|}{t} + |d|\sup_{s\in\mathbb{R}}|F_n(s)-F(s)|
\end{aligned}$$

$$+|d|\{F_n(\tau+r)-F_n(\tau)\}+|d|\{F(\tau+r)-F(\tau)\}.$$

Thus we can conclude

$$\begin{aligned}
p_n(a,r) \leq\ & P\Big(\sup_{x<t\leq r}\frac{|\alpha_n(\tau+t)-\alpha_n(\tau)|}{t} > \frac{1}{6}|c|^{-1}L(r)n^{\frac{1}{2}}\Big)\\
&+P\Big(\sup_{s\in\mathbb{R}}|F_n(s)-F(s)| > \frac{1}{6}|d|^{-1}L(r)\Big)\\
&+P\Big(\sup_{s\in\mathbb{R}}|F_n(s)-F(s)|+2\{F(\tau+r)-F(\tau)\} > \frac{1}{6}|d|^{-1}L(r)\Big)\\
=:\ & A_n(a,r)+B_n(r)+C_n(r). \qquad \text{(A.6)}
\end{aligned}$$

For the investigation of $A_n(a,r)$, it is convenient to put $y := 1/6|c|^{-1}L(r)n^{1/2}$ and to introduce the uniform empirical process

$$\bar{\alpha}_n(u) := n^{-\frac{1}{2}}\sum_{i=1}^{n}[1_{\{U_i\leq u\}}-u], \quad 0\leq u\leq 1,$$

where $U_1,\ldots,U_n$ are i.i.d. with uniform distribution on $[0,1]$. Then by the quantile-transformation $\alpha_n \overset{\mathcal{L}}{=} \bar{\alpha}_n\circ F$. Moreover it is easy to verify that

$$D(r)|t| \leq |F(\tau+t)-F(\tau)| \leq D|t| \quad \forall\, t\in[-r,r] \qquad \text{(A.7)}$$

with positive constants $D(r)=\min\{f(\tau-), f(\tau+r)\}$ and $D=\max\{f(\tau+), f(0-)\}$. Herewith we can infer that

$$\begin{aligned}
A_n(a,r) &= P\Big(\sup_{x<t\leq r}\frac{|\bar{\alpha}_n(F(\tau+t))-\bar{\alpha}_n(F(\tau))|}{t} > y\Big)\\
&\leq P\Big(\sup_{x<t\leq r}\frac{|\bar{\alpha}_n(F(\tau)+F(\tau+t))-\bar{\alpha}_n(F(\tau))|}{F(\tau+t)-F(\tau)} > \frac{y}{D}\Big) \text{ by (A.7)}\\
&= P\Big(\sup_{\alpha<s\leq\beta}\frac{|\bar{\alpha}_n(F(\tau)+s)-\bar{\alpha}_n(F(\tau))|}{s} > \frac{y}{D}\Big), \qquad \text{(A.8)}
\end{aligned}$$

where $\alpha = F(\tau+x)-F(\tau)$ and $\beta = F(\tau+r)-F(\tau)$. Note the differential property of the uniform empirical process, i.e.,

$$\{\bar{\alpha}_n(v+s)-\bar{\alpha}_n(v) : 0\leq s\leq 1-v\} \overset{\mathcal{L}}{=} \{\bar{\alpha}_n(s) : 0\leq s\leq 1-v\}$$

for every fixed $v\in[0,1]$. This follows from stationariy of the increments of $\bar{\alpha}_n$ (confer, e.g., Dudley (1999), Lemma 1.14) and from Theorem14.5 in Billingsley (1968). Therefore the last probability in (A.8) is less than or equal to

$$P\Big(\sup_{\alpha\leq s\leq\beta}\frac{|\bar{\alpha}_n(s)|}{s} > \frac{y}{D}\Big) \leq (\alpha^{-1}-\beta^{-1})D^2y^{-2},$$

where the inequality is ensured by Lemma A.3 of Ferger (2005a). Since $\alpha \geq D(r)x$ by (A.7), $\beta > 0$ and $x = a/n$ by definition, we obtain from (A.8) that

$$A_n(a,r) \leq 36c^2 L(r)^2 D(r)^{-1} a^{-1} \quad \forall\, n \in \mathbb{N} \quad \forall\, a > 0 \quad \forall\, r \in (0,\tau),$$

whence

$$\lim_{a \to 0} \limsup_{n \to \infty} A_n(a,r) = 0 \quad \forall\, 0 < r < \tau. \tag{A.9}$$

An application of the Dvoretzky-Kiefer-Wolfowitz inequality yields

$$\lim_{n \to \infty} B_n(r) = 0 \quad \forall\, 0 < r < \tau. \tag{A.10}$$

Check that $L(r) \to L_0 = \min\{-\alpha C_1 e^{-\alpha\tau}, \beta C_2 e^{-\beta\tau}\}$ as $r \to 0$. Since $C_1 < 0 < C_2$, the limit L_0 is positive. Hence there exists $0 < r_0 < \tau$ such that $2\{F(\tau + r) - F(\tau)\} < 1/6|d|^{-1} L(r)\ \forall\, 0 < r < r_0$. Thus another application of the Dvoretzky-Kiefer-Wolfowitz inequality yields

$$\lim_{n \to \infty} C_n(r) = 0 \quad \forall\, 0 < r < r_0. \tag{A.11}$$

Combine (A.6) and (A.9) $-$ (A.11) to see that

$$\lim_{a \to \infty} \limsup_{n \to \infty} p_n(a,r) = 0 \quad \forall\, 0 < r < r_0. \tag{A.12}$$

As to the investigation of $q_n(a,r)$, first note that

$$\begin{aligned} &S_n(\tau + t) - \bar{S}_n(\tau + t) - [S_n(\tau) - \bar{S}_n(\tau)] \\ &\quad = (c - c_n)\{F_n(\tau + t) - F_n(\tau)\} + (d_n - d)t\{1 - F_n(\tau + t)\} \\ &\qquad + (d_n - d)n^{-1} \sum_{i=1}^{n} (X_i - \tau) 1_{\{\tau < X_i \leq \tau + t\}}. \end{aligned}$$

Using (A.5) and (A.7) gives

$$\begin{aligned} &\sup_{x < t \leq r} \frac{S_n(\tau + t) - \bar{S}_n(\tau + t) - [S_n(\tau) - \bar{S}_n(\tau)]}{t} \\ &\quad \leq |c_n - c| n^{-\frac{1}{2}} \sup_{x < t \leq r} \frac{|\alpha_n(\tau + t) - \alpha_n(\tau)|}{t} + D|c_n - c| \\ &\qquad + |d_n - d| + |d_n - d|\{F_n(\tau + r) - F_n(\tau)\}. \end{aligned}$$

Treating the sup-term in the first summand in the same way as in (A.8) the above inequality together with (1.9) yields $\lim_{a \to 0} \limsup_{n \to \infty} q_n(a,r) = 0$. Thus by (A.12) we arrive at

$$\lim_{a \to 0} \limsup_{n \to \infty} P_{n,1}(a,r) = 0 \quad \forall\, 0 < r < r_0. \tag{A.13}$$

A corresponding statement holds for $P_{n,2}(a,r)$. This can be shown with the same arguments as for (A.13) upon noticing that

$$\{\bar{\alpha}_n(v-s)-\alpha_n(v):0\leq s\leq v\}\stackrel{\mathcal{L}}{=}\{\bar{\alpha}_n(s):0\leq s\leq v\}$$

for every fixed $v\in[0,1]$. In view of (A.2) and (A.3), this finishes the proof.

Proof of Proposition 3.2. Fix $a>0$. Check that

$$Z_n(t)=-c_nN_n(t)+\Delta_n(t)+R_n(t), \tag{A.14}$$

where

$$N_n(t)=\begin{cases}\sum\limits_{i=1}^n 1_{\{\tau<X_i\leq\tau+\frac{t}{n}\}}, & t\geq 0\\ -\sum\limits_{i=1}^n 1_{\{\tau+\frac{t}{n}<X_i\leq\tau\}}, & t<0\end{cases},$$

$$\Delta_n(t)=d_nt\Big(1-F_n(\tau+\frac{t}{n})\Big),$$

$$R_n(t)=d_n\begin{cases}\sum\limits_{i=1}^n 1_{\{\tau<X_i\leq\tau+\frac{t}{n}\}}(X_i-\tau), & t\geq 0\\ \sum\limits_{i=1}^n 1_{\{\tau+\frac{t}{n}<X_i\leq\tau\}}(\tau-X_i), & t<0\end{cases}.$$

In Section 3 of Ferger (2005a), we prove that

$$N_n\xrightarrow{\mathcal{L}}N^*\text{ in }D[-a,a]\text{ as }n\to\infty\quad\forall\ a>0, \tag{A.15}$$

where $N^*(t)=1_{\{t\geq0\}}N_1(t)+1_{\{t<0\}}N_2(t)$.

From (1.9) and the Glivenko-Cantelli Theorem one can easily deduce that

$$\sup_{-a\leq t\leq a}|\Delta_n(t)-\Delta(t)|\xrightarrow{P}0\text{ as }n\to\infty. \tag{A.16}$$

Using (A.5) we find that

$$\sup_{-a\leq t\leq a}|R_n(t)|\leq d_nan^{-1}\max\{N_n(a),N_n(-a)\}.$$

Thus by (1.9), (A.15) and Slutsky's theorem, we obtain

$$\sup_{-a\leq t\leq a}|R_n(t)|\xrightarrow{P}0\text{ as }n\to\infty. \tag{A.17}$$

Thus (A.15) $-$ (A.17), and another application of Slutsky's theorem, yield the desired result.

Proof of Proposition 3.3. Use the same arguments as in the derivation of Lemma 4.2 of Ferger (2005a).

References

Abdel-Aty, Y. and Ferger, D. (2003). Estimation of the jump point in a hazard function. *Econ. Qual. Control* **18**, 251-261.

Anderson, J. A. and Senthilselvan, A. (1982). A two-step regression model for the hazard rate function. *Appl. Statist.* **31**, 44-51.

Antoniadis, A., Gijbels, I. and Macgibbon, B. (2000). Non-parametric estimation for the location of a change-point in an otherwise smooth hazard function under random censoring. *Scand. J. Statist.* **27**, 501-519.

Balakrishnan, N., Kundu, D., Ng, H. K. T. and Kannan, N. (2007). Point and interval estimation for a simple-stress model with type-II censoring. *Journal of Quality Technology* **39**, 35-47.

Basu, A. P., Ghosh, J. K. and Joshi, S. N. (1988). On estimating change points in a failure rate. *Statistical Decision Theory and Related Topics IV* **2** (Edited by S. S. Gupta and J. O. Berger), 239-252.

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

Chang, I.-S., Chen, C.-H. and Hsiung, C. A. (1994). Estimation in Change-Point Hazard Rate Models with Random Censorship. *IMS Lecture-Notes-Monograph Series* **23**, 78-92.

Dudley, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press, New York.

Ferger, D. (2004). A continuous mapping theorem for the argmax-functional in the non-unique case. *Statist. Neerlandica* **58**, 83-96.

Ferger, D. (2005a). On the minimizing point of the incorrect centered empirical process and its limit distribution in nonregular experiments. *ESAIM, Probab. Stat.* **9**, 307-322.

Ferger, D. (2005b). A Pyke-Shorack-type inequality for the uniform empirical process. Manuscript, University of Dresden.

Finkelstein, M. S. (2003). Modeling failure (mortality) rate with a change point. *Working paper WP* 2003-041 *of the Max Planck Institute for Demographic Research: http://www.demogr.mpg.de*, 1-9.

Ghosh, J. K. and Joshi, S. N. (1992). On the asymptotic distribution of an estimate of the change point in a failure rate. *Comm. Statist. Theory Methods* **21**, 3571-3588.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York.

Loader, C. R. (1991). Inference for a hazard rate change-point. *Biometrica* **78**, 749-757.

Mair, W., Goymer, P., Pletcher, S. D. and Partridge, L. (2003). Demography of dietary restriction and death in drosophila. *Science* **301**, 1731-1733.

Matthews, D. E. and Farewell, V. T. (1982). On testing for a constant hazard against a change-point alternative. *Biometrics* **38**, 463-468.

Matthews, D. E., Farewell, V. T. and Pyke, R. (1985). Asymptotic score statistic processes and tests for constant hazard against a change-point alternative. *Ann. Statist.* **13**, 583-591.

Müller, H.-G. and Wang, J. L. (1994). Change-Point Models for Hazard Functions. *IMS Lecture-Notes-Monograph Series* **23**, 224-241.

Nguyen, H. T., Rogers, G. S. and Walker, E. A. (1984). Estimation in change-point hazard rates models. *Biometrika* **71**, 299-304.

Pham, D. T. and Nguyen, H. T. (1990). Strong consistency of the maximum likelihood estimator in the change point hazard rate model. *Statistics* **21**, 203-216.

Pham, D. T. and Nguyen, H. T. (1993). Bootstrapping the change-point of a hazard-rate. *Ann. Inst. Statist. Math.* **45**, 331-340.

Scholz, R. and Maier, H. (2003). German unification and the plasticity of mortality at older ages. *Working paper WP* 2003-031 *of the Max Planck Institute for Demographic Research: http://www.demogr.mpg.de*, 1-21.

Van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, New York.

Wu, C. Q., Zhao, L. C. and Wu, Y. H. (2003). Estimation in change-point hazard function models. *Statist. Probab. Lett.* **63**, 41-48.

Yao, Y. C. (1986). Maximum likelihood estimation in hazard rate model with a change point. *Comm. Statist. Theory Methods* **15**, 2455-2466.

Yao, Y. C. (1987). A note on testing for constant hazard against a change point alternative. *Ann. Inst. Statist. Math.* **39**, 377-383.

Dresden University of Technology, Departments of Mathematics, Mommsenstr. 13, D-01062 Dresden, Germany.

E-mail: dietmar.ferger@tu-dresden.de

(Received August 2004; accepted November 2005)