

## ON THE MINIMIZING POINT OF THE INCORRECTLY CENTERED EMPIRICAL PROCESS AND ITS LIMIT DISTRIBUTION IN NONREGULAR EXPERIMENTS

DIETMAR FERGER<sup>1</sup>

**Abstract.** Let  $F_n$  be the empirical distribution function (df) pertaining to independent random variables with continuous df  $F$ . We investigate the minimizing point  $\hat{\tau}_n$  of the empirical process  $F_n - F_0$ , where  $F_0$  is another df which differs from  $F$ . If  $F$  and  $F_0$  are locally Hölder-continuous of order  $\alpha$  at a point  $\tau$  our main result states that  $n^{1/\alpha}(\hat{\tau}_n - \tau)$  converges in distribution. The limit variable is the almost sure unique minimizing point of a two-sided time-transformed homogeneous Poisson-process with a drift. The time-transformation and the drift-function are of the type  $|t|^\alpha$ .

**Mathematics Subject Classification.** 60E15, 60F05, 60F17, 62E20.

Received March 2, 2004. Revised March 24, 2005.

### 1. INTRODUCTION

Let  $X_1, \dots, X_n$ ,  $n \in \mathbb{N}$ , be independent random variables with common continuous distribution function  $F$ . Consider the empirical process

$$D_n(x) = F_n(x) - F_0(x), \quad x \in \mathbb{R}.$$

Here

$$F_n(x) = n^{-1} \sum_{i=1}^n 1_{\{X_i \leq x\}}, \quad x \in \mathbb{R},$$

is the empirical distribution function pertaining to the random variables  $X_1, \dots, X_n$  and  $F_0$  is an arbitrary continuous distribution function which in general needs not to coincide with the true underlying distribution function  $F$ . If however  $F_0 = F$  then  $D_n$  is the well-known classical empirical process, which has been intensively studied in the literature. Among many others there are the famous and fundamental results of Glivenko [15], Cantelli [4], Kolmogorov [22], Smirnov [28], Birnbaum and Tingey [3] or Donsker [5] to mention some of the pioneering work. A comprehensive and detailed treatment of the classical and related empirical processes is given by Shorack and Wellner [27]. A generalization of the theory to function-indexed empirical processes can be found in Dudley [8] or van der Vaart and Wellner [30].

---

*Keywords and phrases.* Rescaled empirical process, argmin-CMT, Poisson-process, weak convergence in  $D(\mathbb{R})$ .

<sup>1</sup> Department of Mathematics, Dresden University of Technology, Helmholtzstr. 10, 01062 Dresden, Germany;  
ferger@math.tu-dresden.de

© EDP Sciences, SMAI 2005

In this article we study the minimizing point  $\hat{\tau}_n$  of  $D_n$ , *i.e.*

$$\begin{aligned} \hat{\tau}_n &:= \operatorname{arg\,min}_{x \in \mathbb{R}} D_n(x) \\ &:= \min \left\{ t \in \mathbb{R} : D_n(t-) = \inf_{x \in \mathbb{R}} D_n(x) \right\}. \end{aligned}$$

This definition is in accordance with the usual one of the argmin-functional on the space  $D(\mathbb{R})$  of all right-continuous functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  with left-hand limits, see (1.3) below. Note that  $D_n \in D(\mathbb{R})$  and that between two successive observations it is continuous and monotone decreasing with positive jumps of height  $n^{-1}$  at every point  $X_i, 1 \leq i \leq n$ . This shows that

$$\hat{\tau}_n = X_{r:n} \quad \text{a.s.},$$

where  $X_{1:n} < \dots < X_{n:n}$  is our notation for the order statistics. The index  $r$  is random and given by

$$r = \operatorname{arg\,min}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - F_0(X_{i:n}) \right\},$$

where in case of ambiguity we have to take the smallest index. Indeed in general  $\hat{\tau}_n$  may not be unique but if  $F_0(X_1)$  has no atoms then with probability one  $\hat{\tau}_n$  is the only minimizer, confer Proposition A.1 in the appendix. If  $F_0 = F$  (correct centering) then  $\hat{\tau}_n$  has the same distribution as the original observations, that is  $P(\hat{\tau}_n \leq x) = F(x)$  for all  $x \in \mathbb{R}$ . This has been proved in the case of the uniform distribution  $F = U(0, 1)$  by several authors, confer Birnbaum and Pyke [2], Dwass [9], Kuiper [23], Takács [29] or Ferger [11]. For an arbitrary distribution function  $F$  the result follows by using the quantile transformation, confer Proposition A.2 in the appendix. Thus for  $F_0 = F$  the complete solution is known. Therefore we consider the situation when  $F_0$  and  $F$  differ. Assume  $F$  and  $F_0$  are (in some specified sense) local Hölder-continuous of order  $\alpha \in (0, \infty)$  at a point  $\tau$ , which is also the unique minimizer of the difference  $D = F - F_0 \neq 0$ . Then we will prove distributional convergence of the sequence  $n^{1/\alpha}(\hat{\tau}_n - \tau)$ ,  $n \in \mathbb{N}$ . The limit variable turns out to be the a.s. unique minimizing point of a time-transformed two-sided Poisson-process with positive drift.

Let us shortly outline our proof. First we show that  $n^{1/\alpha}(\hat{\tau}_n - \tau)$  is stochastically bounded, *i.e.*

$$n^{1/\alpha}(\hat{\tau}_n - \tau) = O_P(1) \text{ as } n \rightarrow \infty. \tag{1.1}$$

We will manage this by deriving a sufficiently sharp upper bound for the tail-probabilities  $P(|\hat{\tau}_n - \tau| > a)$ ,  $a > 0$ . Next introduce the rescaled and renormalized process

$$Y_n(t) = n \left\{ D_n(\tau + n^{-1/\alpha}t) - D_n(\tau) \right\}, \quad t \in \mathbb{R},$$

which yields the representation

$$n^{1/\alpha}(\hat{\tau}_n - \tau) = \operatorname{arg\,min}(Y_n). \tag{1.2}$$

Here the argmin-functional on  $D(\mathbb{R})$  is defined by

$$\operatorname{arg\,min}(f) := \min \left\{ t \in \mathbb{R} : \min(f(t-), f(t)) = \inf_{x \in \mathbb{R}} f(x) \right\} \tag{1.3}$$

for every  $f \in D(\mathbb{R})$  such that the set on the right hand side in (1.3) is not empty and bounded. Since this is a closed subset of the real line (confer Lem. 6.1 (i) of Ferger [12], the minimum exists in  $\mathbb{R}$ . Also notice that our definition of  $\operatorname{arg\,min}(f), f \in D(\mathbb{R})$ , follows the definition of Kallenberg [20], p. 226. The final step in our proof is to derive a functional limit theorem for  $Y_n$ :

$$Y_n \xrightarrow{\mathcal{L}} Y \text{ in } D(\mathbb{R}) \text{ as } n \rightarrow \infty, \tag{1.4}$$

where  $D(\mathbb{R})$  is endowed with Lindvall's [24] extension of the Skorokhod-metric. The limit  $Y$  is the above mentioned two-sided Poisson-process. We show that the minimizing point

$$\operatorname{argmin}(Y) \text{ is unique a.s.} \tag{1.5}$$

Thus we are in a position to apply the Argmin-Continuous Mapping Theorem (Argmin-CMT) of van der Vaart and Wellner [30] in the version of Ferger [13] for  $D(\mathbb{R})$ -valued processes. It states that under (1.1), (1.2), (1.4) and (1.5) we can infer that

$$n^{1/\alpha}(\hat{\tau}_n - \tau) = \operatorname{argmin}(Y_n) \xrightarrow{\mathcal{L}} \operatorname{argmin}(Y), \tag{1.6}$$

which is our main result. The limit variable in (1.6) admits a rather simple representation, especially it is continuously distributed.

We like to point out that the Argmin-CMT of van der Vaart and Wellner [30] deals with the bigger function space  $l^\infty(\mathbb{R}) \supset D(\mathbb{R})$  consisting of all locally bounded functions. This space is endowed with the topology of uniform convergence on compacta. Because of measurability reasons an extended theory of weak convergence is introduced. It goes back to Dudley [6,7] and Hoffman-Jørgenson [18]. Similarly as in the classical case weak convergence is established by showing convergence of the finite dimensional distributions and by proving tightness. However the tightness criteria there involve the modulus of continuity which (in our case) inevitably would result in a limit process that by Theorem 15.5 from Billingsley [1] almost surely had continuous trajectories. Since our limit process  $Y$  has jumps even with probability one the approach *via* weak convergence in  $l^\infty(\mathbb{R})$  necessarily must fail.

There is another version of an Argmin-CMT using the space of lower semicontinuous functions with the topology of epi-convergence, confer Theorem 1 of Knight [21]. His Argmin-CMT is very appropriate in stochastic optimization problems, when the underlying objective functions are minimized under constraints. To circumvent this difficulty one simply allows  $+\infty$  as possible value for the objective functions. In this case the Lindvall-Skorokhod-space  $D(\mathbb{R})$  is inappropriate and our Argmin-CMT does not work. But as long as no constraints are involved we prefer to work with  $D(\mathbb{R})$ . This is mainly for two reasons. Firstly, our objective functions  $Y_n$  automatically lie in  $D(\mathbb{R})$ . Secondly, weak convergence of stochastic processes in  $D(\mathbb{R})$  is equivalent with weak convergence of their restrictions in the traditional Skorokhod-space  $D[-a, a]$  for every  $a > 0$ . Consequently we can use the well-developed theory of weak convergence in  $D[-a, a]$  with compact interval  $[-a, a]$ . This theory yields a lot of convenient criteria for deriving functional limit theorems, confer, *e.g.*, Billingsley [1].

## 2. PROBABILITY INEQUALITIES

As was pointed out in the introduction the first important step is to derive tightness of the sequence  $n^{1/\alpha}(\hat{\tau}_n - \tau)$ . For that purpose we will derive probability inequalities for  $\hat{\tau}_n - \tau$ . The following proposition gives a first upper bound for the probability of the tail-event  $\{|\hat{\tau}_n - \tau| > r\}$ ,  $r > 0$ . It is valid under our general assumption that  $F$  and  $F_0$  are continuous.

**Proposition 2.1.** *Let  $F$  and  $F_0$  be continuous. Assume*

$$D := F - F_0$$

*has a unique minimizer  $\tau = \operatorname{argmin}(D)$  with  $D(\tau) < 0$ . Then there exists  $r_0 > 0$  so that for every  $0 < r \leq r_0$  there exists  $b = b(r) > 0$  such that*

$$P(|\hat{\tau}_n - \tau| > r) \leq 2 \exp\{-2nb^2\} \quad \forall n \in \mathbb{N}. \tag{2.1}$$

*A crucial auxiliary tool for deriving the probability inequality (2.1) is the following simple but very useful*

**Lemma 2.2.** *Let  $f \in D(\mathbb{R})$  have a unique real minimizer  $\tau$ , i.e.  $\tau = \operatorname{argmin}(f)$  is unique and  $f(\tau) = \min_{t \in \mathbb{R}} f(t)$ . If for  $r > 0$  the minimum*

$$a(r) := \min\{f(t) : |t - \tau| \geq r\}$$

*exists, then the following statements hold:*

- (1)  $b(r) := \frac{1}{3}(a(r) - f(\tau)) > 0$ ;
- (2) *let  $g \in D(\mathbb{R})$  be such that  $\operatorname{argmin}(g)$  is well-defined in the sense of (1.3). Then the following implication is true:*

$$\sup_{t \in \mathbb{R}} |f(t) - g(t)| \leq b(r) \Rightarrow |\operatorname{argmin}(f) - \operatorname{argmin}(g)| \leq r.$$

*Proof.* By assumption there exists  $t_0 \in \mathbb{R}$  with  $|t_0 - \tau| \geq r$  and  $f(t_0) = a(r) \in \mathbb{R}$ . Assume that  $a(r) \leq f(\tau)$ , whence  $f(\tau) \leq f(t_0) = a(r) \leq f(\tau)$ , so that  $t_0 \neq \tau$  is also a minimizer of  $f$  in contradiction to the uniqueness of  $\tau$ . This shows (1). To see (2) consider an arbitrary  $t \in \mathbb{R}$  with  $|t - \tau| \geq r$ . Then  $g(t) \geq f(t) - b(r) \geq a(r) - b(r)$  and  $g(\tau) \leq f(\tau) + b(r)$ . Consequently  $g(t) - g(\tau) \geq a(r) - f(\tau) - 2b(r) = b(r) > 0$  by (1). We obtain  $g(t) > g(\tau)$  and by taking the limit  $g(t-) \geq g(\tau) + b(r) > g(\tau)$  for all  $t \in \mathbb{R}$  with  $|t - \tau| > r$ . This shows for  $\sigma := \operatorname{argmin}(g)$  that  $|\sigma - \tau| \leq r$ , because otherwise  $\inf_{t \in \mathbb{R}} g(t) = \min(g(\sigma), g(\sigma-)) > g(\tau)$ , which is a contradiction.  $\square$

*Proof of Proposition 2.1.* By assumption  $D$  is continuous whence  $\tau$  is a unique and real minimizer. Moreover, since  $D(x) \rightarrow 0$  as  $|x| \rightarrow \infty$  one finds an  $r_0 > 0$  such that  $a(r) = \min\{D(x) : |x - \tau| \geq r\}$  exists in  $\mathbb{R}$  for all  $0 < r \leq r_0$ . Consequently we can apply (2) of Lemma 2.1 to  $f = D$  and  $g = D_n$ , which yields

$$\begin{aligned} P(|\hat{\tau}_n - \tau| > r) &\leq P\left(\sup_{x \in \mathbb{R}} |D_n(x) - D(x)| > b(r)\right) \\ &= P\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > b(r)\right) \\ &\leq 2 \exp\{-2nb^2\} \quad \forall n \in \mathbb{N}, \end{aligned}$$

where the last inequality follows from the Dvoretzky-Kiefer-Wolfowitz-inequality with the optimal constant of Massart [25].  $\square$

The next Corollary is an immediate consequence of Proposition 2.1 and the first Borel-Cantelli lemma.

**Corollary 2.3.** *Under the assumption of Proposition 2.1 we have that*

$$\hat{\tau}_n \rightarrow \tau \text{ a.s. as } n \rightarrow \infty.$$

Next we refine inequality (2.1) under the additional assumption of local Hölder-continuity in a specified sense. To be precise assume that  $F$  and  $F_0$  meet the following requirements:

There exist real numbers  $\alpha > 0, r > 0$  and  $\tau$  such that (2.2)

$$\begin{aligned} F(x) &= \begin{cases} F(\tau) + p(x)|x - \tau|^\alpha & , \quad \text{if } \tau \leq x \leq \tau + r \\ F(\tau) + q(x)|x - \tau|^\alpha & , \quad \text{if } \tau - r \leq x \leq \tau \end{cases} \\ F_0(x) &= \begin{cases} F_0(\tau) + p_0(x)|x - \tau|^\alpha & , \quad \text{if } \tau \leq x \leq \tau + r \\ F_0(\tau) + q_0(x)|x - \tau|^\alpha & , \quad \text{if } \tau - r \leq x \leq \tau \end{cases} \end{aligned}$$

where  $p$  and  $p_0$  are positive and continuous on  $[\tau, \tau + r]$ , whereas  $q$  and  $q_0$  are negative and continuous on  $[\tau - r, \tau]$ ,

$$F(\tau) < F_0(\tau), \tag{2.3}$$

and

$$D(\tau) - D(x) \leq -L|x - \tau|^\alpha \quad \forall x \in [\tau - r, \tau + r] \tag{2.4}$$

for some positive  $L = L(r)$ , so that especially  $\tau$  is a local minimizer of  $D$ . Observe that under (2.2) the condition (2.4) is fulfilled if

$$L_+ := \min \{p(x) - p_0(x) : \tau \leq x \leq \tau + r\} > 0 \tag{2.5}$$

and

$$L_- := \min \{q(x) - q_0(x) : \tau - r \leq x \leq \tau\} > 0$$

both are positive. Then  $L = L(r) = \min(L_+, L_-) > 0$  is appropriate. If  $\tau$  is not only a local but even the unique global minimizer of  $D$  then we obtain an inequality which is sufficiently good for proving (1.1).

**Proposition 2.4.** *Assume that (2.2)–(2.4) hold and that  $\tau = \operatorname{argmin}(D)$  is unique. Then there exist positive constants  $C, L$  and  $b$  such that*

$$P(|\hat{\tau}_n - \tau| > a) \leq CL^{-2}n^{-1}a^{-\alpha} + 2 \exp\{-2nb^2\} \tag{2.6}$$

for every  $a > 0$  and all  $n \in \mathbb{N}$ .

**Corollary 2.5.** *Under the assumptions of Proposition 2.4*

$$n^{1/\alpha}(\hat{\tau}_n - \tau) = O_P(1) \text{ as } n \rightarrow \infty.$$

*Proof of Proposition 2.4.* By Proposition 2.1 and (2.3) there exists  $r_0 > 0$  such that (2.1) is valid. With  $r$  coming from (2.2) let  $\delta = \min(r, r_0)$ , which is positive, and take  $b = b(\delta)$  and  $L = L(\delta)$  from Proposition 2.1 and (2.4), respectively. We start with the simple estimate

$$P(|\hat{\tau}_n - \tau| > a) \leq P(a < |\hat{\tau}_n - \tau| \leq \delta) + P(|\hat{\tau}_n - \tau| > \delta) =: P_n + Q_n. \tag{2.7}$$

By Proposition 2.1 we have

$$Q_n \leq 2 \exp\{-2nb^2\} \quad \forall n \in \mathbb{N}, \tag{2.8}$$

so it remains to investigate the first probability  $P_n$ . Since  $P_n = 0$  if  $a \geq \delta$  we may assume that  $a < \delta$ . In this case we observe that

$$\{a < |\hat{\tau}_n - \tau| \leq \delta\} \subseteq \left\{ \inf_{a < |t - \tau| \leq \delta} D_n(t) \leq D_n(\tau) \right\}.$$

To see this assume that  $a < |\hat{\tau}_n - \tau| \leq \delta$  but  $\inf_{a < |t - \tau| \leq \delta} D_n(t) > D_n(\tau)$ , which means that  $D_n(t) > D_n(\tau)$  and by taking the limit that also  $D_n(t-) \geq D_n(\tau)$  for every  $t$  with  $a < |t - \tau| \leq \tau$ . For  $t = \hat{\tau}_n$  we therefore obtain that  $D_n(\hat{\tau}_n-) = D_n(\tau)$ . Since  $D_n$  does not attain its infimum this is a contradiction. Thus we can conclude that

$$\begin{aligned} P_n &\leq P\left(\inf_{a < |t - \tau| \leq \delta} D_n(t) \leq D_n(\tau)\right) \\ &\leq P\left(\bigcup_{\tau + a < t \leq \tau + \delta} \{D_n(t) \leq D_n(\tau)\}\right) \\ &\quad + P\left(\bigcup_{\tau - \delta \leq t < \tau - a} \{D_n(t) \leq D_n(\tau)\}\right) =: \tilde{P}_n + P_n^*. \end{aligned} \tag{2.9}$$

Note that for all  $t \in [\tau - \delta, \tau + \delta]$

$$\begin{aligned} D_n(\tau) - D_n(t) &= D_n(\tau) - D(\tau) - (D_n(t) - D(t)) + D(\tau) - D(t) \\ &= F_n(\tau) - F(\tau) - (F_n(t) - F(t)) + D(\tau) - D(t) \\ &\leq F_n(\tau) - F(\tau) - (F_n(t) - F(t)) - L|t - \tau|^\alpha \end{aligned} \tag{2.10}$$

by (2.4). Let  $\alpha_n$  denote the uniform empirical process, *i.e.*

$$\alpha_n(u) = n^{-1/2} \sum_{i=1}^n (1_{\{F(X_i) \leq u\}} - u), \quad 0 \leq u \leq 1.$$

Then (2.10) shows that

$$\begin{aligned} \tilde{P}_n &\leq P \left( \bigcup_{\tau+a < t \leq \tau+\delta} \left\{ \frac{\alpha_n(F(\tau)) - \alpha_n(F(t))}{(t - \tau)^\alpha} \geq Ln^{1/2} \right\} \right) \\ &\leq P \left( \sup_{\tau+a < t \leq \tau+\delta} \frac{|\alpha_n(F(t)) - \alpha_n(F(\tau))|}{(t - \tau)^\alpha} \geq Ln^{1/2} \right) \\ &\leq P \left( \sup_{\tau+a < t \leq \tau+\delta} \frac{|\alpha_n(F(t)) - \alpha_n(F(\tau))|}{F(t) - F(\tau)} \geq \frac{Ln^{1/2}}{\|p\|} \right), \end{aligned} \tag{2.11}$$

where the last inequality follows from (2.2), *i.e.*

$$F(t) - F(\tau) = p(t)(t - \tau)^\alpha \leq \|p\|(t - \tau)^\alpha$$

for all  $\tau \leq t \leq \tau + \delta$ , where  $\|p\| := \sup_{\tau \leq t \leq \tau+\delta} p(t)$ . For the last probability in (2.11) we obtain the upper bound

$$\begin{aligned} P \left( \sup_{F(\tau+a) \leq v \leq F(\tau+\delta)} \frac{|\alpha_n(v) - \alpha_n(F(\tau))|}{v - F(\tau)} \geq \frac{Ln^{1/2}}{\|p\|} \right) &\leq P \left( \sup_{\gamma \leq s \leq \lambda} \frac{|\alpha_n(s + F(\tau)) - \alpha_n(F(\tau))|}{s} \geq \frac{Ln^{1/2}}{\|p\|} \right) \\ &= P \left( \sup_{\gamma \leq s \leq \lambda} \frac{|\alpha_n(s)|}{s} \geq \frac{Ln^{1/2}}{\|p\|} \right), \end{aligned} \tag{2.12}$$

where  $\gamma = F(\tau + a) - F(\tau)$  and  $\lambda = F(\tau + \delta) - F(\tau)$  and the last equality follows from the differential property of the uniform empirical process:

$$\{\alpha_n(u + s) - \alpha_n(u) : 0 \leq s \leq 1 - u\} \stackrel{\mathcal{L}}{=} \{\alpha_n(s) : 0 \leq s \leq 1 - u\} \in D[0, 1 - u]$$

for every fixed  $u \in [0, 1]$ . This is a simple consequence of the stationarity of the increments of  $\alpha_n$  (confer, *e.g.*, Dudley [8], Lem. 1.14(b)) and of Theorem 14.5 in Billingsley [1]. An application of Lemma A.3 in the appendix yields

$$\begin{aligned} P \left( \sup_{\gamma \leq s \leq \lambda} \frac{|\alpha_n(s)|}{s} \geq \frac{Ln^{1/2}}{\|p\|} \right) &\leq \{F(\tau + a) - F(\tau)\}^{-1} \|p\|^2 L^{-2} n^{-1} \\ &\leq p(\tau + a)^{-1} \|p\|^2 L^{-2} n^{-1} a^{-\alpha} \text{ by (2.2)}. \end{aligned} \tag{2.13}$$

Combining (2.11)–(2.13) and noticing that  $p(\tau + a) \geq \min_{\tau \leq t \leq \tau+\delta} p(t)$  is strictly positive by assumption we find a positive constant  $C_1$  such that

$$\tilde{P}_n \leq C_1 L^{-2} n^{-1} a^{-\alpha}. \tag{2.14}$$

In the same way one finds a positive constant  $C_2$  such that

$$P_n^* \leq C_2 L^{-2} n^{-1} a^{-\alpha} \tag{2.15}$$

taking into account that

$$\{\alpha_n(u - s) - \alpha_n(u) : 0 \leq s \leq u\} \stackrel{\mathcal{L}}{=} \{\alpha_n(s) : 0 \leq s \leq u\} \in D[0, u]$$

for every fixed  $u \in [0, 1]$ . Now the assertion (2.6) follows from (2.7)–(2.9), (2.14) and (2.15). □

### 3. A FUNCTIONAL LIMIT THEOREM FOR $Y_n$ IN $D(\mathbb{R})$

Recall the process

$$Y_n(t) = n\{D_n(\tau + n^{-1/\alpha}t) - D_n(\tau)\}, \quad t \in \mathbb{R},$$

introduced in section one. There we explained that our second task is to prove convergence in distribution of  $Y_n$  in the function space  $D(\mathbb{R})$  of Lindvall [24]. The process  $Y_n$  splits into a random and a purely deterministic part:

$$Y_n(t) = S_n(t) - \Delta_n(t), \quad t \in \mathbb{R}, \tag{3.1}$$

where

$$S_n(t) = \begin{cases} \sum_{i=1}^n \mathbf{1}_{\{\tau < X_i \leq \tau + n^{-1/\alpha}t\}}, & t \geq 0 \\ -\sum_{i=1}^n \mathbf{1}_{\{\tau + n^{-1/\alpha}t < X_i \leq \tau\}}, & t < 0 \end{cases}$$

and

$$\Delta_n(t) = n\{F_0(\tau + n^{-1/\alpha}t) - F_0(\tau)\}, \quad t \in \mathbb{R}.$$

Assumption (2.2) allows us to apply the Poisson approximation to the Binomial, which yields

$$S_n(t) \xrightarrow{\mathcal{L}} \begin{cases} \text{Poisson}(p(\tau)t^\alpha), & t \geq 0 \\ -\text{Poisson}(-q(\tau)(-t)^\alpha), & t < 0, \end{cases} \tag{3.2}$$

as  $n \rightarrow \infty$  for every fixed  $t \in \mathbb{R}$ . Here  $\text{Poisson}(\lambda)$  denotes a random variable which has a Poisson-distribution with parameter  $\lambda$ . As to the deterministic part it follows easily from (2.2) that

$$\Delta_n(t) \longrightarrow \Delta(t) = \begin{cases} p_0(\tau)t^\alpha, & t \geq 0 \\ q_0(\tau)(-t)^\alpha, & t < 0 \end{cases} \tag{3.3}$$

as  $n \rightarrow \infty$  where the convergence holds uniformly on every compact interval  $[-a, a]$   $a > 0$ . In view of (3.2) we introduce two independent Poisson processes  $N_1$  and  $N_2$  with parameters  $\lambda_1 = p(\tau)$  and  $\lambda_2 = -q(\tau)$ , respectively. The second process  $N_2$  is chosen such that its trajectories are left-continuous and possess right-hand limits. According to (3.2) and (3.3) it follows that

$$Y_n(t) \xrightarrow{\mathcal{L}} Y(t) \text{ as } n \rightarrow \infty \tag{3.4}$$

for every fixed  $t \in \mathbb{R}$ , where

$$Y(t) = \begin{cases} N_1(t^\alpha) - p_0(\tau)t^\alpha, & t \geq 0 \\ -N_2((-t)^\alpha) - q_0(\tau)(-t)^\alpha, & t < 0. \end{cases} \tag{3.5}$$

Notice that our choice of  $N_2$  ensures that the process  $Y$  possesses trajectories in  $D(\mathbb{R})$ . Next we extend (3.4) to the weak convergence of all finite-dimensional distributions.

**Lemma 3.1.** *If  $F$  satisfies assumption (2.2) then the finite-dimensional distributions of  $Y_n$  converge weakly to those of  $Y$ .*

*Proof.* For the sake of convenience we only consider the two-dimensional case, but the general case can be treated in the same manner. If  $0 \leq s < t$  then for every  $l, k \in \mathbb{N}_0$

$$P(S_n(s) = l, S_n(t) - S_n(s) = k) = \frac{n!}{l!k!(n-k-1)!} p_1^l p_2^k (1-p_1-p_2)^{n-k-l}, \tag{3.6}$$

where

$$\begin{aligned} np_1 &= n\{F(\tau + n^{-1/\alpha}s) - F(\tau)\} \\ &= p(\tau + n^{-1/\alpha}s)s^\alpha \text{ by (2.2)} \\ &\rightarrow p(\tau)s^\alpha \text{ as } n \rightarrow \infty \end{aligned}$$

and analogously

$$\begin{aligned} np_2 &= n\{F(\tau + n^{-1/\alpha}t) - F(\tau + n^{-1/\alpha}s)\} \\ &\rightarrow p(\tau)(t^\alpha - s^\alpha) \text{ as } n \rightarrow \infty. \end{aligned}$$

Thus with (3.6) we can conclude that  $(S_n(s), S_n(t) - S(s)) \xrightarrow{\mathcal{L}} (N_1(s^\alpha), N_1(t^\alpha) - N_1(s^\alpha))$  as  $n \rightarrow \infty$  and the Continuous Mapping Theorem (CMT) gives

$$(S_n(s), S_n(t)) \xrightarrow{\mathcal{L}} (N_1(s^\alpha), N_1(t^\alpha)) \text{ as } n \rightarrow \infty.$$

If  $s < 0 \leq t$  then one shows in the same way that  $(-S_n(s), S_n(t)) \rightarrow (N_2((-s)^\alpha), N_1(t^\alpha))$  as  $n \rightarrow \infty$ , whence by the CMT

$$(S_n(s), S_n(t)) \xrightarrow{\mathcal{L}} (-N_2((-s)^\alpha), N_1(t^\alpha)) \text{ as } n \rightarrow \infty.$$

Finally, if  $s < t < 0$  then one can proceed as in the case  $0 \leq s < t$ . Now the assertion follows from (3.1), (3.3) and Slutsky's theorem.  $\square$

In the sequel we consider the restrictions of our processes on intervals  $[-a, a]$ ,  $a > 0$ . These are random elements in the Skorokhod-space  $D[-a, a]$ . Let

$$\omega''(f, \delta) = \sup \min\{|f(t) - f(t_1)|, |f(t_2) - f(t)|\}, \quad \delta > 0,$$

where the supremum extends over  $t_1 \leq t \leq t_2 \in [-a, a]$  satisfying  $t_2 - t_1 \leq \delta$ . This modulus characterizes the compact sets in the Skorokhod topology. The following lemma guarantees the tightness of the sequence  $(S_n)$ .

**Lemma 3.2.** *Suppose  $F$  satisfies (2.2). Then for every  $a > 0$  it follows that*

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P(\omega''(S_n, \delta) \geq \varepsilon) = 0 \quad \forall \varepsilon > 0.$$

*Proof.* Let  $t_1 \leq t \leq t_2 \in [-a, a]$  be arbitrary points. Assume, e.g., that  $t_1 \leq 0 \leq t \leq t_2$ . Then

$$S_n(t) - S_n(t_1) = \sum_{i=1}^n 1_{\{\tau + n^{-1/\alpha}t_1 < X_i \leq \tau + n^{-1/\alpha}t\}}$$

and

$$S_n(t_2) - S_n(t) = \sum_{i=1}^n 1_{\{\tau + n^{1-\alpha}t < X_i \leq \tau + n^{-1/\alpha}t_2\}}.$$



Since the intervals  $(\tau + n^{-1/\alpha}t_1, \tau + n^{-1/\alpha}t]$  and  $(\tau + n^{-1/\alpha}t, \tau + n^{-1/\alpha}t_2]$  are disjoint we obtain

$$\begin{aligned} E(|S_n(t) - S_n(t_1)||S_n(t_2) - S_n(t)|) &= n(n-1)\{F(\tau + n^{-1/\alpha}t) \\ &\quad - F(\tau + n^{-1/\alpha}t_1)\}\{F(\tau + n^{-1/\alpha}t_2) - F(\tau + n^{-1/\alpha}t)\} \\ &\leq \{H_n(t) - H_n(t_1)\}\{H_n(t_2) - H_n(t)\} \\ &\leq \{H_n(t_2) - H_n(t_1)\}^2, \end{aligned}$$

where

$$H_n(s) = n\{F(\tau + n^{-1/\alpha}s) - F(\tau)\}$$

is a nondecreasing and continuous function on  $[-a, a]$ . These inequalities also hold for all other choices of  $t_1 \leq t \leq t_2 \in [-a, a]$ . Now, the derivation of (15.30) in Billingsley [1] shows that for all  $n \in \mathbb{N}$

$$P(\omega''(S_n, \delta) \geq \varepsilon) \leq 2K\varepsilon^{-2}[H_n(a) - H_n(-a)]\omega(H_n, 2\delta),$$

where  $K$  is a positive constant and  $\omega(f, \delta) = \sup\{|f(s) - f(t)| : s, t \in [-a, a], |s - t| \leq \delta\}$  is the modulus of continuity for an element  $f \in D[-a, a]$ . Note that  $H_n$  converges uniformly on  $[-a, a]$  to the continuous function  $H(s) = 1_{\{s \geq 0\}}p(\tau)|s|^\alpha + 1_{\{s < 0\}}q(\tau)|s|^\alpha$ . Therefore, because the mapping  $f \rightarrow \omega(f, \delta)$  is continuous,

$$\limsup_{n \rightarrow \infty} P(\omega''(S_n, \delta) \geq \varepsilon) \leq 2K\varepsilon^{-2}[H(a) - H(-a)]\omega(H, 2\delta).$$

Continuity of  $H$  ensures that  $\omega(H, 2\delta)$  tends to zero as  $\delta \rightarrow 0$ , whence the assertion of the lemma follows.  $\square$

We are now in a position to prove the main result of this section: an invariance principle for the sequence  $(Y_n)$ .

**Theorem 3.3.** *Suppose that  $F$  satisfies (2.2). Then*

$$Y_n \xrightarrow{\mathcal{L}} Y \text{ in } D[-a, a] \text{ as } n \rightarrow \infty \text{ for every } a > 0. \tag{3.7}$$

*Proof.* Fix  $a > 0$  and put

$$S(t) := \begin{cases} N_1(t^\alpha), & t \geq 0 \\ -N_2((-t)^\alpha), & t < 0. \end{cases}$$

Then  $P(S(a-) \neq S(a)) = 0$ , because

$$\{S(a-) \neq S(a)\} \subseteq \bigcup_{i \geq 1} \{\xi_i = a^\alpha\},$$

where  $\xi_i, i \geq 1$ , are the arrival-times of the Poisson process  $N_1$ . In view of Lemmas 3.1 and 3.2 we can apply Theorem 15.4 of Billingsley [1] which yields

$$S_n \xrightarrow{\mathcal{L}} S \text{ in } D[-a, a] \text{ as } n \rightarrow \infty.$$

Thus (3.7) is a consequence of (3.1), (3.3) and Slutsky's theorem.  $\square$

Note that according to Lindvall [24] the convergence in (3.7) is equivalent with  $Y_n \xrightarrow{\mathcal{L}} Y$  in  $D(\mathbb{R})$  as  $n$  tends to infinity.

4. CONVERGENCE IN DISTRIBUTION

Our proof relies on the following extension of van der Vaart’s and Wellner’s [30] Argmin-Continuous Mapping Theorem (Argmin-CMT) for  $D(\mathbb{R})$ -valued stochastic processes, confer Theorem 3 and Corollary 1 of Ferger [13].

**Theorem 4.1** (Argmin-CMT). *Assume that every single of the  $D(\mathbb{R})$ -valued random variables  $M_n, n \in \mathbb{N}$ , and  $M$  has a set of minimizers which is bounded. Then  $\hat{h}_n = \operatorname{argmin}(M_n)$  and  $h = \operatorname{argmin}(M)$  are well-defined and Borel-measurable. Suppose that*

$$M_n \xrightarrow{\mathcal{L}} M \text{ in } D[-a, a] \text{ as } n \rightarrow \infty \text{ for all } a > 0, \tag{4.1}$$

$$\hat{h}_n = O_P(1) \text{ as } n \rightarrow \infty, \tag{4.2}$$

$$h \text{ is almost surely unique} \tag{4.3}$$

and

$$\mathbb{X} := \{x \in \mathbb{R} : P(M \text{ is continuous at } x) = 1\} \tag{4.4}$$

is dense in  $\mathbb{R}$ . Then  $\hat{h}_n \xrightarrow{\mathcal{L}} h$  as  $n \rightarrow \infty$ .

In view of (1.2) we will apply Theorem 4.1 to

$$\hat{h}_n = n^{1/\alpha}(\hat{\tau}_n - \tau) = \operatorname{argmin}(Y_n)$$

and then obtain our main result

**Theorem 4.2.** *Suppose  $F$  and  $F_0$  satisfy the conditions (2.2)–(2.5) and  $\tau = \operatorname{argmin}(F - F_0)$  is unique. Then*

$$n^{1/\alpha}(\hat{\tau}_n - \tau) \xrightarrow{\mathcal{L}} h = \operatorname{argmin}(Y) \text{ as } n \rightarrow \infty, \tag{4.5}$$

where  $Y$  is the two-sided time-transformed Poisson-process defined in (3.5).

*Proof.* In order to apply Theorem 4.1 we must verify the conditions (4.1)–(4.4) there. Theorem 3.3 and Corollary 2.5 guarantee that (4.1) and (4.2) are fulfilled. For the validity of (4.3) we refer to Lemma A.4 in the appendix. It remains to check (4.4). But it is easy to see that  $\mathbb{X} = \mathbb{R}$  so that this condition is trivially fulfilled. To sum up we see that all conditions of the Argmin-CMT 4.1 are satisfied, whence (4.5) is proved.  $\square$

The random variable  $h$  in (4.5) features a simple a.s. representation. To see this define the random integers

$$l = \operatorname{arg\,min}_{i \geq 1} \{i - 1 - c_1 \xi_i\}$$

and

$$m = \operatorname{arg\,min}_{i \geq 1} \{i - c_2 \rho_i\}.$$

Here  $0 =: \xi_0 < \xi_1 < \xi_2 < \dots$  and  $0 =: \rho_0 < \rho_1 < \rho_2 < \dots$  denote the arrival-times pertaining to the Poisson-process  $N_1$  and  $N_2$ , respectively and  $c_1 := p_0(\tau)$  and  $c_2 := q_0(\tau)$ . From (5)–(9) below (see the appendix) one can deduce that a.s.  $l$  and  $m$  are finite and unique and that

$$h = \begin{cases} \xi_l^{1/\alpha}, & \text{if } l - 1 - c_1 \xi_l < -m - c_2 \rho_m \\ -\rho_m^{1/\alpha}, & \text{if } l - 1 - c_1 \xi_l \geq -m - c_2 \rho_m. \end{cases}$$

As a simple consequence we obtain that  $h$  has a continuous distribution. Thus Pólya’s theorem sharpens (4.5) to uniform convergence:

$$\sup_{x \in \mathbb{R}} \left| P(n^{1/\alpha}(\hat{\tau}_n - \tau) \leq x) - P(h \leq x) \right| \rightarrow 0$$

as  $n \rightarrow \infty$ .

5. APPLICATIONS IN STATISTICS

In this section we apply our results to a semi-parametric location-model. Our model turns out to be a non-regular statistical experiment. This notion goes back to Ibragimov and Has'minskii [19]. Such models are non-smooth in some specified sense. Here the asymptotic behavior of estimators differs from the classical one insofar that there is no  $\sqrt{n}$ -rate in the distributional convergence and moreover non-normal limit variables occur.

Let  $X_1, \dots, X_n$  be independent and identically distributed random variables (life times) with distribution function  $F(x) = F_0(x - \tau), x \in \mathbb{R}$ , where  $\tau > 0$  is the unknown shift-parameter we wish to estimate. Assume the error distribution  $F_0$  is known and has a density  $f_0$ , which vanishes on the negative real line. Furthermore it is supposed to be positive and strictly decreasing on the non-negative real line with limits  $f(0-) = 0$  and  $f(0+) > 0$ . Thus  $f_0$  is a non-smooth density with a jump at point zero. In this situation the parameter  $\tau$  is the unique minimizer of  $D(x) = F_0(x - \tau) - F_0(x)$ . To see this note that the derivative  $D'(x) = f_0(x - \tau) - f_0(x)$  is positive by monotonicity, if  $x > \tau$  and  $D'(x) = -f_0(x)$  is negative, if  $0 \leq x < \tau$ . Moreover  $D'(\tau-) = f(0-) - f(\tau) = -f(\tau) < 0$  and  $D'(\tau+) = f(0+) - f(\tau) > 0$ . Consequently  $D$  is strictly monotone decreasing or increasing, respectively, according as  $0 \leq x \leq \tau$  or  $x > \tau$ . Obviously  $D$  vanishes on  $(-\infty, 0)$  and

$$D(\tau) = -F_0(\tau) < 0, \tag{5.1}$$

whence indeed  $\tau$  is the unique minimizing point of  $D$ . Therefore Corollary 2.3 yields that  $\hat{\tau}_n$  is a strongly consistent estimator for  $\tau$ :

$$\hat{\tau}_n \rightarrow \tau \text{ a.s. as } n \rightarrow \infty \text{ for every } \tau > 0.$$

Use the Mean Value Theorem to show that

$$D(\tau) - D(x) \leq -L|\tau - x| \text{ for every } x \in [0, 2\tau], \tag{5.2}$$

where

$$L = \min \left\{ \inf_{0 \leq t \leq \tau} f_0(t), \inf_{0 \leq t \leq \tau} (f_0(t) - f_0(\tau + t)) \right\}.$$

This real number is positive if in addition, e.g.,  $f_0$  is continuous on  $[0, \infty)$ . Finally another application of the Mean Value Theorem gives that  $F$  and  $F_0$  satisfy (2.2) with  $\alpha = 1, p(\tau) = f_0(0+), q(\tau) = -f_0(0-) = 0, p_0(\tau) = f_0(\tau)$  and  $q_0(\tau) = -f_0(\tau)$ .

It follows that

$$Y(t) = \begin{cases} N_1(t) - f_0(\tau) t, & t \geq 0 \\ -f_0(\tau) t, & t < 0 \end{cases} \tag{5.3}$$

and  $N_1$  is a Poisson process with intensity  $f_0(0+) > 0$ . Note that the negative part of the limit process  $Y$  is a straight line through the origin with negative slope. This is because  $q(\tau) = 0$  and therefore  $N_2$  vanishes a.s. and because  $-f_0(\tau)$  is negative. Since we have verified all necessary conditions we may apply Theorem 4.2 and obtain convergence in distribution. We summarize our results in the following

**Corollary 5.1.** *Consider the location-model with life-time density  $f_0$  which is positive and strictly decreasing on its support  $[0, \infty)$ . Then*

$$\hat{\tau}_n \rightarrow \tau \text{ a.s. as } n \rightarrow \infty \text{ for every } \tau > 0.$$

*If in addition  $f_0$  is continuous on  $[0, \infty)$  then*

$$n(\hat{\tau}_n - \tau) \xrightarrow{\mathcal{L}} h = \operatorname{argmin}(Y) \text{ as } n \rightarrow \infty$$

*with  $Y$  given in (5.3). Observe that this  $Y$  is more simple than in the general case of Theorem 4.2 insofar that only a one-sided Poisson-process with linear drift occurs. For such processes Pflug [26] has determined the exact distribution of the minimizer  $h$ .*

We also have results for other kinds of densities  $f_0$  including those which are continuous, see Ferger [14]. For the sake of brevity we do not present them in the most general form. Nevertheless the following theorem makes it abundantly clear how the asymptotics change if we move from non-smooth models (like in Cor. 5.1 above) to smooth models. Indeed, cube-root asymptotics now come into play and instead of a Poisson process with linear drift the limit process turns out to be a Brownian motion with parabolic drift.

**Theorem 5.2.** *Assume that  $f_0$  is continuously differentiable on  $\mathbb{R} \setminus \{0\}$ , symmetric about zero and strictly decreasing on the positive real line with  $f_0(0+) \leq \infty$ . Then*

$$\hat{\tau}_n \rightarrow \tau/2 \quad \text{a.s. as } n \rightarrow \infty \quad \text{for every } \tau > 0.$$

Moreover,

$$n^{1/3}(\hat{\tau}_n - \tau/2) \xrightarrow{\mathcal{L}} h = \operatorname{argmin}(Y) \quad \text{as } n \rightarrow \infty,$$

with

$$Y(t) = B(t) - \frac{f'_0(\tau/2)}{\sqrt{f_0(\tau/2)}} t^2,$$

where  $B$  denotes a two-sided Brownian motion on  $\mathbb{R}$ .

Note that the parabolic drift is upwards, because  $f'_0(\tau/2)$  is negative by the monotonicity assumption on the density  $f_0$ . The distribution of the minimizer  $h$  of a Brownian motion with parabolic drift has been derived by Groeneboom [16] who gives an explicit expression of the pertaining density in terms of Airy functions. Dykstra and Carolan [10] and Groeneboom and Wellner [17] give numerical approximations.

### APPENDIX

We give here the outstanding proofs of statements made in the introduction and of an inequality used in the proof of Proposition 2.5. Recall that in the whole article we assume continuity of the involved distribution functions  $F$  and  $F_0$ .

**Lemma A.1.** *Suppose that  $F_0(X_1)$  has no atoms. Then*

$$\hat{\tau}_n = \operatorname{argmin}(F_n - F_0) \text{ is unique a.s.}$$

*Proof.* Let  $R_l$  denote the rank of  $X_l$  in the sample  $X_1, \dots, X_n$  and  $R_l^{-1}$  the pertaining anti-rank,  $1 \leq l \leq n$ . Then we obtain

$$\begin{aligned} \{\hat{\tau}_n \text{ is not unique}\} &\subseteq \bigcup_{1 \leq i < j \leq n} \left\{ \frac{i-1}{n} - F_0(X_{i:n}) = \frac{j-1}{n} - F_0(X_{j:n}) \right\} \\ &= \bigcup_{1 \leq i < j \leq n} \left\{ \frac{i-1}{n} - F_0(X_{R_i^{-1}}) = \frac{j-1}{n} - F_0(X_{R_j^{-1}}) \right\} \\ &\subseteq \bigcup_{1 \leq l \neq k \leq n} \left\{ \frac{R_l-1}{n} - F_0(X_l) = \frac{R_k-1}{n} - F_0(X_k) \right\} \\ &\subseteq \bigcup_{1 \leq m \leq n-1} \bigcup_{1 \leq l \neq k \leq n} \left\{ F_0(X_k) - F_0(X_l) = \frac{m}{n} \right\}, \end{aligned}$$

which yields the desired result. □

**Lemma A.2.** *If  $F_0 = F$ , then  $\hat{\tau}_n \stackrel{\mathcal{L}}{=} X_1$ , that is*

$$P(\hat{\tau}_n \leq x) = F(x) \text{ for all } x \in \mathbb{R}.$$

*Proof.* In view of the underlying independence and by the quantile-transformation we can assume that  $X_i = F^{-1}(U_i)$ , where  $U_1, \dots, U_n$  are i.i.d. with the uniform distribution function  $\bar{F}$  on  $(0,1)$  and  $G^{-1}$  is our notation for the quantile-function of any distribution function  $G$ . Let  $\bar{F}_n$  denote the empirical distribution function pertaining to  $U_1, \dots, U_n$ . It is well-known that  $F_n = \bar{F}_n \circ F$  and

$$F_n^{-1} = F^{-1} \circ \bar{F}_n^{-1}. \tag{1}$$

As was pointed out in the introduction

$$\hat{\tau}_n = \operatorname{argmin}(F_n - F) = X_{r:n} \tag{2}$$

where

$$\begin{aligned} r &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - F(X_{i:n}) \right\} \\ &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - F(F_n^{-1}\left(\frac{i}{n}\right)) \right\} \\ &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - F(F^{-1}(\bar{F}_n^{-1}\left(\frac{i}{n}\right))) \right\} \text{ by (1)} \\ &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - (\bar{F}_n^{-1}\left(\frac{i}{n}\right)) \right\} \text{ by continuity of } F \\ &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - U_{i:n} \right\} \\ &= \operatorname{argmin}_{1 \leq i \leq n} \left\{ \frac{i-1}{n} - \bar{F}(U_{i:n}) \right\}, \end{aligned}$$

whence by (2) (applied to  $F = \bar{F}$ )

$$U_{r:n} = \operatorname{argmin}(\bar{F}_n - \bar{F}). \tag{3}$$

On the other hand by (1)

$$\begin{aligned} \hat{\tau}_n = X_{r:n} &= F_n^{-1}\left(\frac{r}{n}\right) = F^{-1}\left(\bar{F}_n^{-1}\left(\frac{r}{n}\right)\right) \\ &= F^{-1}(U_{r:n}). \end{aligned} \tag{4}$$

As was mentioned in the introduction it is known that  $\operatorname{argmin}(\bar{F}_n - \bar{F})$  is uniformly distributed on  $(0, 1)$ , so that the assertion follows from (3) and (4).  $\square$

**Lemma A.3.** *Let  $\alpha_n = \sqrt{n}(\bar{F}_n - \bar{F})$  be the uniform empirical process. Then*

$$P\left(\sup_{\alpha \leq s \leq \beta} \frac{|\alpha_n(s)|}{s} \geq \lambda\right) \leq (\alpha^{-1} - \beta^{-1})\lambda^{-2}$$

for all  $\lambda > 0$  and for all  $0 < \alpha < \beta \leq 1$ .

*Proof.* We follow Shorack and Wellner [27], on p. 134 and p. 873. Start with the known fact that

$$\left\{ \beta_n(s) = \frac{\alpha_n(s)}{1-s} : 0 \leq s < 1 \right\}$$

is a centered martingale with respect to the filtration  $\mathcal{F}_n(s) = \sigma(\bar{F}_n(u) : 0 \leq u \leq s), 0 \leq s < 1$ . Let  $\{s_k = \alpha + (\beta - \alpha)k2^{-m} : 0 \leq k < 2^m\}, m \in \mathbb{N}$ , be the dyadic partition of  $[\alpha, \beta]$ . Then  $\{M_k = \beta_n(s_k) : 0 \leq k < 2^m\}$  is also a centered martingale with respect to the filtration  $\{\mathcal{G}_k = \mathcal{F}_n(s_k) : 0 \leq k < 2^m\}$ . Therefore

$$\begin{aligned} P(\alpha, \beta, \lambda) &:= P\left(\sup_{\alpha \leq s \leq \beta} \frac{|\alpha_n(s)|}{s} \geq \lambda\right) = P\left(\sup_{\alpha \leq s \leq \beta} \frac{1-s}{s} |\beta_n(s)| \geq \lambda\right) \\ &= \lim_{m \rightarrow \infty} P\left(\max_{1 \leq k < 2^m} \frac{1-s_k}{s_k} |M_k| \geq \lambda\right) \\ &\leq \lambda^{-2} \lim_{m \rightarrow \infty} \sum_{1 \leq k < 2^m} (1-s_k)^2 s_k^{-2} E(M_k - M_{k-1})^2 \\ &\quad \text{(by Hajek's martingale-inequality)} \\ &= \lambda^{-2} \lim_{m \rightarrow \infty} \sum_{1 \leq k < 2^m} (1-s_k)^2 s_k^{-2} (EM_k^2 - EM_{k-1}^2). \end{aligned}$$

Since  $E\beta_n(s) = \frac{s}{1-s} =: L(s)$  und  $L'(s) = (1-s)^{-2}$  it follows that

$$\begin{aligned} P(\alpha, \beta, \lambda) &\leq \lambda^{-2} \lim_{m \rightarrow \infty} \sum_{1 \leq k < 2^m} (1-s_k)^2 s_k^{-2} (L(s_k) - L(s_{k-1})) \\ &= \lambda^{-2} \int_{\alpha}^{\beta} (1-s)^2 s^{-2} L(ds) \\ &= \lambda^{-2} \int_{\alpha}^{\beta} s^{-2} ds, \end{aligned}$$

which gives the desired inequality. □

**Lemma A.4.** *Suppose (2.5) holds. Then  $h := \operatorname{argmin}(Y)$  exists and is a.s. unique.*

*Proof.* Introduce

$$Y_1(t) = N_1(t^\alpha) - c_1 t^\alpha, \quad t \geq 0$$

and

$$Y_2(t) = -N_2(t^\alpha) - c_1 t^\alpha, \quad t \geq 0,$$

where  $c_1 = p_0(\tau)$  and  $c_2 = q_0(\tau)$ . Thus  $Y$  admits the representation

$$Y(t) = 1_{\{t \geq 0\}} Y_1(t) + 1_{\{t < 0\}} Y_2(-t),$$

whence

$$\inf_{t \in \mathbb{R}} Y(t) = \min\{\inf_{t \geq 0} Y_1(t), \inf_{t \geq 0} Y_2(t)\}. \tag{5}$$

Let  $0 =: \xi_0 < \xi_1 < \xi_2 < \dots$  and  $0 =: \rho_0 < \rho_1 < \rho_2 < \dots$  denote the arrival-times pertaining to the Poisson-process  $N_1$  and  $N_2$ , respectively. The process  $N_1(t^\alpha)$  has a.s. jumps exactly at the points  $\xi_i^{1/\alpha}, i \geq 1$ , so that

$$\inf_{t \geq 0} Y_1(t) = \inf_{i \geq 1} Y_1(\xi_i^{1/\alpha} -) = \inf_{i \geq 1} \{i - 1 - c_1 \xi_i\} \text{ a.s.}$$

Consider the interarrival-times  $\eta_j = \xi_j - \xi_{j-1}$ ,  $j \geq 1$ , which are known to be independent and to have a common exponential distribution with parameter  $\lambda_1 = p(\tau)$ . Therefore

$$Y_1(\xi_i^{1/\alpha} -) = i - 1 - c_1 \xi_i = \sum_{j=1}^i (1 - c_1 \eta_j) - 1, \quad i \geq 1,$$

which by the Strong Law of Large Numbers tends to infinity a.s. as  $i \rightarrow \infty$ , because  $E(1 - c_1 \eta_j) = 1 - c_1/\lambda_1$  is positive by (2.5). Consequently

$$\inf_{t \geq 0} Y_1(t) = \min_{i \geq 1} Y_1(\xi_i^{1/\alpha} -) = \min_{i \geq 1} \{i - 1 - c_1 \xi_i\} \text{ a.s.} \tag{6}$$

The first equality in (6) shows that the event  $F_1 := \{Y_1 \text{ has no minimizer}\}$  has probability zero. Next note that

$$E_1 := \{Y_1 \text{ has no unique minimizer}\} \subseteq F_1 \cup G_1,$$

where

$$\begin{aligned} G_1 &:= \{Y_1 \text{ has at least two minimizer}\} \\ &\subseteq \bigcup_{1 \leq i < j < \infty} \{Y_1(\xi_i^{1/\alpha} -) = Y_1(\xi_j^{1/\alpha} -)\} \\ &= \bigcup_{1 \leq i < j < \infty} \left\{ \xi_j - \xi_i = \frac{j-i}{c_1} \right\}. \end{aligned}$$

Since  $\xi_j - \xi_i$ ,  $j > i$ , have a continuous distribution (namely a Gamma-distribution) we can infer that  $P(G_1) = 0$ . It follows that

$$P(E_1) = P(Y_1 \text{ has no unique minimizer}) = 0. \tag{7}$$

Recall that  $N_2$  is left-continuous with right-hand limits. Then in the same way one proves that

$$\inf_{t \geq 0} Y_2(t) = \min_{i \geq 1} Y_2(\rho_i^{1/\alpha}) = \min_{i \geq 1} \{-i - c_2 \rho_i\} \text{ a.s.} \tag{8}$$

and that

$$P(E_2) = P(Y_2 \text{ has no unique minimizer}) = 0. \tag{9}$$

Conclude from (5), (6) and (8) that

$$\inf_{t \in \mathbb{R}} Y(t) = \min_{i \geq 1} \{Y(\xi_i^{1/\alpha} -), Y(-\rho_i^{1/\alpha})\} \text{ a.s.}$$

which in view of (1.3) immediately yields the a.s. existence of  $\operatorname{argmin}(Y)$ . As to the a.s. uniqueness consider

$$G := \{Y \text{ has at least two minimizer, } h_1 < h_2 \text{ say}\}$$

and observe that  $G \cap \{h_1 \geq 0\} \subseteq E_1$  and  $G \cap \{h_2 \leq 0\} \subseteq E_2$  both have probability zero by (7) and (9). Finally,

$$\begin{aligned} G \cap \{h_2 < 0 < h_1\} &\subseteq \bigcup_{1 \leq i, j < \infty} \{Y_1(\xi_i^{1/\alpha} -) = Y_2(\rho_j^{1/\alpha})\} \\ &= \bigcup_{1 \leq i, j < \infty} \{i - 1 - c_1 \xi_i = j - c_2 \rho_j\}, \end{aligned}$$

where the last event also has probability zero, because for every  $i, j \in \mathbb{N}$  the random variables  $\xi_i$  and  $\rho_j$  are independent and have a continuous distribution. This shows the assertion of the lemma.  $\square$

## REFERENCES

- [1] P. Billingsley, *Convergence of probability measures*. Wiley, New York (1968).
- [2] Z.W. Birnbaum and R. Pyke, On some distributions related to the statistic  $D_n^+$ . *Ann. Math. Statist.* **29** (1958) 179–187.
- [3] Z.W. Birnbaum and F.H. Tingey, One-sided confidence contours for probability distribution functions. *Ann. Math. Statist.* **22** (1951) 592–596.
- [4] F.P. Cantelli, Considerazioni sulla legge uniforme dei grandi numeri e sulla generalizzazione di un fondamentale teorema del sig. Paul Levy. *Giorn. Ist. Ital. Attuari* **4** (1933) 327–350.
- [5] J. Donsker, Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statist.* **23** (1952) 277–281.
- [6] R.M. Dudley, Weak convergence of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.* **10** (1966) 109–126.
- [7] R.M. Dudley, Measures on nonseparable metric spaces. *Illinois J. Math.* **11** (1967) 449–453.
- [8] R.M. Dudley, *Uniform central limit theorems*. Cambridge University Press, New York (1999).
- [9] M. Dwass, On several statistics related to empirical distribution functions. *Ann. Math. Statist.* **29** (1958) 188–191.
- [10] R. Dykstra and Ch. Carolan, The distribution of the argmax of two-sided Brownian motion with parabolic drift. *J. Statist. Comput. Simul.* **63** (1999) 47–58.
- [11] D. Ferger, The Birnbaum-Pyke-Dwass theorem as a consequence of a simple rectangle probability. *Theor. Probab. Math. Statist.* **51** (1995) 155–157.
- [12] D. Ferger, Analysis of change-point estimators under the null hypothesis. *Bernoulli* **7** (2001) 487–506.
- [13] D. Ferger, A continuous mapping theorem for the argmax-functional in the non-unique case. *Statistica Neerlandica* **58** (2004) 83–96.
- [14] D. Ferger, *Cube root asymptotics for argmin-estimators*. Unpublished manuscript, Technische Universität Dresden (2005).
- [15] V. Glivenko, Sulla determinazione empirica delle leggi die probabilita. *Giorn. Ist. Ital. Attuari* **4** (1933) 92–99.
- [16] P. Groneboom, Brownian motion with a parabolic drift and Airy Functions. *Probab. Th. Rel. Fields* **81** (1989) 79–109.
- [17] P. Groneboom and J.A. Wellner, Computing Chernov's distribution. *J. Comput. Graphical Statist.* **10** (2001) 388–400.
- [18] J. Hoffman-Jørgensen, *Stochastic processes on Polish spaces*. (Published (1991): Various Publication Series No. 39, Matematisk Institut, Aarhus Universitet) (1984).
- [19] I.A. Ibragimov and R.Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*. Springer-Verlag, New York (1981).
- [20] O. Kallenberg, *Foundations of Modern Probability*. Springer-Verlag, New York (1999).
- [21] K. Knight, *Epi-convergence in distribution and stochastic equi-semicontinuity*. Technical Report, University of Toronto (1999) 1–22.
- [22] A.N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione. *Giorn. Ist. Ital. Attuari* **4** (1933) 83–91.
- [23] N.H. Kuiper, Alternative proof of a theorem of Birnbaum and Pyke. *Ann. Math. Statist.* **30** (1959) 251–252.
- [24] T. Lindvall, Weak convergence of probability measures and random functions in the function space  $D[0, \infty)$ . *J. Appl. Prob.* **10** (1973) 109–121.
- [25] P. Massart, The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* **18** (1990) 1269–1283.
- [26] G.Ch. Pflug, On an argmax-distribution connected to the Poisson process, in *Proc. of the fifth Prague Conference on asymptotic statistics*, P. Mandl, H. Husková Eds. (1993) 123–130.
- [27] G.R. Shorack and J.A. Wellner, *Empirical processes with applications to statistics*. Wiley, New York (1986).
- [28] N.V. Smirnov, Näherungsgesetze der Verteilung von Zufallsveränderlichen von empirischen Daten. *Usp. Mat. Nauk.* **10** (1944) 179–206.
- [29] L. Takács, *Combinatorial Methods in the theory of stochastic processes*. Robert E. Krieger Publishing Company, Huntington, New York (1967).
- [30] A.W. van der Vaart and J.A. Wellner, *Weak convergence of empirical processes*. Springer-Verlag, New York (1996).