The Annals of Statistics
1993, Vol. 21, No. 3, 1431–1451

NONPARAMETRIC ESTIMATION IN RENEWAL THEORY I: THE EMPIRICAL RENEWAL FUNCTION

BY RUDOLF GRÜBEL AND SUSAN M. PITTS

Delft University of Technology and University College London

We introduce a nonparametric estimator for the renewal function and discuss its properties, including consistency, asymptotic normality and asymptotic validity of bootstrap confidence regions. The underlying theme is that stochastic models can be regarded as functionals or nonlinear operators. This view leads to nonparametric estimators in a natural way and statistical properties of the estimators can be related to the local behaviour of the functionals.

1. Introduction. The standard—and christening—example of renewal theory involves "an electric bulb, fuse or other piece of equipment with a finite life span. As soon as the piece fails, it is replaced by a new piece of the like kind, which in due time is replaced by a third piece, and so on" [Feller (1968), page 311]. Formally, we have a sequence $(X_i)_{i \in \mathbb{N}}$ of i.i.d. random variables representing the successive lifetimes, with partial sums $S_0 = 0$, $S_n = \sum_{i=1}^{n} X_i$, and

$$N(t) = 1 + \sup\{n \in \mathbb{N}_0 \colon S_n \le t\}$$

is the number of renewals up to time t, including the initial installation at $t = 0$.

A structure of this type can often be found in more complex stochastic models in the form of "renewal points" where, loosely speaking, the process "forgets about its previous history." Renewal theory provides some basic techniques for the analysis of such models, the renewal function $U(t) = EN(t)$, the expected number of renewals up to and including time t, playing a key role. Typically, a decomposition with respect to the value of X_1, the random time of the first renewal, yields an integral equation of a type known as renewal equations whose solution can be given in terms of U. For a detailed discussion, including a range of examples, see Feller (1971), Chapter XI. The renewal function U itself is of interest in warranty analysis, for example; other applications include sequential analysis where the X variates can take negative values [see Frees (1986a)].

In the present paper we deal with the problem of estimating U on the basis of a sample $X_1, \dots, X_n$ from a distribution function F, later referred to as the lifetime distribution or step distribution. The former implies $F(0) = 0$ or at least $F(0-) = 0$, we will call this the one-sided case; the latter refers to the

Received July 1991; revised June 1992.
AMS 1991 *subject classifications*. Primary 60K05; secondary 62M09, 62G05.
Key words and phrases. Nonparametric estimation, renewal functions, functional central limit theorems, bootstrap.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

interpretation of the partial sums S_n as a random walk on the real line. We do not assume a parametric model. This problem has been considered earlier by Frees (1986a, b, 1988) and, in the one-sided case, by Schneider, Lin and O'Cinneide (1990); a detailed discussion is given in subsection 4.2.

The basic idea of the present paper is to regard the stochastic model as a nonlinear operator or functional Ψ, the *renewal functional*, which maps a lifetime or step distribution F on the associated renewal function U; technically, this functional is a convolution series with constant coefficients [see Grübel (1989a) for a nonstatistical application]. Given a random sample of size n from F and with $\hat{F}_n$ denoting the associated empirical distribution function this leads us to estimate U by applying Ψ to $\hat{F}_n$. In short, we estimate the renewal function U by the *empirical renewal function* $\hat{U}_n = \Psi(\hat{F}_n)$. Estimators for the solutions of renewal type equations are obtained by inserting the empirical renewal function for the unknown U in the solution formula provided by renewal theory; these will be dealt with in Part II of the paper.

Statistical properties of these estimators can be related to local analytic properties of the functional Ψ. Consistency follows from continuity. Both asymptotic normality and asymptotic validity of bootstrap confidence regions are based on a local linearization of Ψ, that is, a version of Hadamard or compact differentiation [see Gill (1989), Grübel (1988) and the references given there].

The paper is organized as follows. Section 2 gives the main results; the proofs are given in Section 3. Though the underlying ideas are quite simple it turns out that a number of technical details have to be observed in order to arrive at a precise frame. Much of this has to do with the choice of suitable spaces and topologies for the functionals. Consequently, Section 3 is divided into several subsections; see the beginning of that section for an overview. A final section compares our results to those of Frees (1986a, b) and Schneider, Lin and O'Cinneide (1990), indicates possible extensions and collects some concluding remarks.

The methods of the present paper can be applied to a variety of other stochastic models. In forthcoming work by S. M. Pitts it will be shown that nonparametric estimation methods in queueing theory can be based on and analyzed with similar ideas and techniques.

2. Main results. Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables on some probability space $(\Omega, \mathscr{A}, \mathbb{P})$ and let F be the distribution function (d.f.) of X_1. We consider the general, two-sided case, that is, we do not assume that the X values are nonnegative. We do assume that the random variables have finite second and positive first moment. Under these assumptions the associated renewal function U exists and is given by $U = \sum_{k=0}^{\infty} F^{\star k}$ where $\star$ denotes convolution. The value $U(x)$ is the expected number of visits to the interval $(-\infty, x]$ of a random walk with steps $\{X_n\}$ and start at 0.

Let $\hat{F}_n$ be the empirical distribution function associated with the first n of the X values, that is, with 1_A denoting the indicator function of the set A,

$\hat{F}_n(x) = n^{-1}\sum_{k=1}^{n} 1_{(-\infty,x]}(X_k)$ for all $x \in \mathbb{R}$. We define the empirical renewal function $\hat{U}_n$ by $\hat{U}_n = \sum_{k=0}^{\infty} \hat{F}_n^{\star k}$ if the sample mean $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$ is positive and $\hat{U}_n \equiv 0$ otherwise. From the assumptions on F and the strong law of large numbers it follows that, for $\mathbb{P}$-almost all $\omega \in \Omega$, $\hat{U}_n(\omega)$ is given by the convolution series for n large enough.

We wish to consider renewal functions and their estimates as a single object in contrast to a view which selects a fixed $t \in \mathbb{R}$ and then considers the real value $U(t)$ and the real valued random variables $\hat{U}_n(t)$. This raises the question of the natural space for these functions. Distribution functions and their empirical counterparts are usually considered as (random) elements of D_∞, the set of functions $f\colon [-\infty, \infty] \to \mathbb{R}$ which are right continuous and have left-hand limits. This space is endowed with the supremum norm,

$$\|f\|_\infty = \sup_{-\infty \le x \le \infty} |f(x)|$$

and the σ-algebra generated by the closed $\|\cdot\|_\infty$-balls. Renewal functions are not bounded (unless the associated lifetime distribution is defective, a case not considered in this paper) and are therefore not elements of this space. Further, by the elementary renewal theorem, $U(x) \sim x/EX_1$ as $x \to \infty$, and $\hat{U}_n(x) \sim x/\bar{X}_n$. Hence, unless expectation and sample mean coincide, we cannot have uniform convergence of $\hat{U}_n - U$ on the whole half-line; this means that a suitable weight function is needed. By the elementary renewal theorem again we see that such a weight function must decrease at least as fast as $1/x$ as $x \to \infty$.

These considerations lead us to define $D_{0,-1}$ as the space of all functions $f\colon \mathbb{R} \to \mathbb{R}$ which can be written as $f(x) = \rho(x)g(x)$, where g is an element of D_∞ and $\rho(x) = 1$ for $x < 0$, $\rho(x) = 1 + x$ if $x \ge 0$. If f and g are related in this way, we put $\|f\|_{0,-1} = \|g\|_\infty$ and equip $D_{0,-1}$ with the σ-field generated by the closed $\|\cdot\|_{0,-1}$-balls (the logic behind the choice of indices will become clear in the next section where a more general class of spaces is needed for the proofs).

Our first result gives conditions for strong $\|\cdot\|_{0,-1}$-consistency of the empirical renewal function. For this we need measurability of $\sup_{x\in\mathbb{R}}|\hat{U}_n(x) - U(x)|/\rho(x)$. We argue as follows: For any fixed $x \in \mathbb{R}$, $k \in \mathbb{N}$,

$$\hat{F}_n^{\star k}(x) = n^{-k} \sum_{i_1,\dots,i_k=1}^{n} 1_{(-\infty,x]}(X_{i_1} + \cdots + X_{i_k})$$

is measurable, hence $\hat{U}_n(x)$ is a random variable for all $x \in \mathbb{R}$. This implies that $\hat{U}_n\colon \Omega \to D_{0,-1}$ is measurable, that is, a $D_{0,-1}$-valued random quantity. Further, by construction, the norm $\|\cdot\|_{0,-1}$ is measurable, which gives the required measurability of the composition. Alternatively, we can argue that all functions involved are right continuous so that it is enough to take the supremum over a suitable countable set of x values. (In the sequel, checking measurability will be left to the reader.)

THEOREM 2.1. *Let $F, U, \hat{U}_n$ be as above; assume that*

$$\int |x|^{\beta} F(dx) < \infty \quad \textit{for some } \beta > 2, \qquad \int x F(dx) > 0.$$

Then

$$\lim_{n \to \infty} \|\hat{U}_n - U\|_{0,-1} = 0 \quad \textit{almost surely}.$$

As to be expected, the asymptotic normality result will give conditions on F which ensure that $\sqrt{n}(\hat{U}_n - U)$ converges in distribution to a Gaussian process with mean function zero and covariance function depending on F. We say that a sequence $V_1, V_2, \dots$ of $D_{0,-1}$-valued random quantities converges in distribution to V, another $D_{0,-1}$-valued random quantity, or $V_n \to_{\text{distr}} V$ in $D_{0,-1}$ for short, if $Ef(V_n)$ tends to $Ef(V)$ for all bounded, continuous and measurable functions $f: D_{0,-1} \to \mathbb{R}$.

THEOREM 2.2. *Let $F, U, \hat{U}_n$ be as above; assume that*

$$\int |x|^{\beta} F(dx) < \infty \quad \textit{for some } \beta > 4, \qquad \int x F(dx) > 0.$$

Then, in $D_{0,-1}$,

$$n^{1/2}\big(\hat{U}_n - U\big) \to_{\text{distr}} Z^U,$$

where $Z^U = (Z_t^U)_{t \in \mathbb{R}}$ is a Gaussian process with mean $EZ_t^U = 0$ and, with $V = U \star U$,

$$\operatorname{cov}\big(Z_s^U, Z_t^U\big) = \int\int F((s-x) \wedge (t-y)) V(dx) V(dy) - F\star V(s) F\star V(t).$$

Suppose now that we wish to construct confidence regions for the unknown renewal function U. Let

$$R_n(z) = \mathbb{P}\big(\sqrt{n}\|\hat{U}_n - U\|_{0,-1} \le z\big), \qquad R(z) = \mathbb{P}(\|Z^U\|_{0,-1} \le z).$$

It follows from Theorem 2.2 that $\lim_{n\to\infty} R_n(z) = R(z)$ for all continuity points z of R so that an asymptotic level α confidence region would be given by the set of all renewal functions U with the property that $\|\hat{U}_n - U\|_{0,-1}$ does not exceed $q_\alpha / \sqrt{n}$, where R is continuous at q_α and $R(q_\alpha) \ge \alpha$. However, the distribution function R is not known. It depends on F in a complicated way; not even the shape of the distribution is known so that there is no hope for a "studentization procedure" to apply.

The bootstrap method provides us with an estimator for R_n. To motivate this estimator we first note that R_n is a quantity which depends on F only. In fact, we can write down a formula expressing this dependency if we use the following definition,

$$\mathbb{F}_n : \mathbb{R}^n \to D_\infty, \qquad \mathbb{F}_n(x) := n^{-1} \sum_{i=1}^{n} 1_{[x_i, \infty)}$$

for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. We also write $F^{\otimes n}$ for the nth (measure theoretic) power of F, a distribution function on $\mathbb{R}^n$. The dependency of R_n on F is then given by

$$R_n(z) = \int_{\mathbb{R}^n} 1_{[0,z]}\big(\sqrt{n}\, \|\Psi(\mathbb{F}_n(x)) - \Psi(F)\|_{0,-1}\big) F^{\otimes n}(dx),$$

where, again, Ψ is the functional which maps a distribution function on the associated renewal function. The bootstrap estimator for R_n is now obtained by replacing the unknown F by $\hat{F}_n$. As $\hat{F}_n$ assigns mass $1/n$ to each of the values $X_1, \dots, X_n$ we obtain

$$\hat{R}_n(z) = n^{-n} \sum_{i \in I_n} 1_{[0,z]}\Big(\sqrt{n}\, \big\|\Psi\big(\mathbb{F}_n(X_{i_1}, \dots, X_{i_n})\big) - \Psi\big(\hat{F}_n\big)\big\|_{0,-1}\Big),$$

where $i = (i_1, \dots, i_n)$, $I_n = \{1, \dots, n\}^n$; note that $\hat{F}_n = \mathbb{F}_n(X_1, \dots, X_n)$ in this notation. Replacing q_α in the above discussion by $\hat{q}_n(\alpha)$, the α quantile of $\hat{R}_n$, we obtain the level α bootstrap confidence region. The following theorem shows that, under essentially the same conditions as used for asymptotic normality, this results in an asymptotically correct procedure.

THEOREM 2.3. *Suppose* $0 < \alpha < 1$. *Let* F, U, $\hat{U}_n$ *and* $\hat{q}_n(\alpha)$ *be as above. Assume that* F *satisfies* $0 < F(x) < 1$ *for some* $x \in \mathbb{R}$ *and that*

$$\int |x|^\beta F(dx) < \infty \quad \textit{for some } \beta > 4, \qquad \int x F(dx) > 0.$$

Then

$$\lim_{n \to \infty} \mathbb{P}\big(\sqrt{n}\, \|\hat{U}_n - U\|_{0,-1} \le \hat{q}_n(\alpha)\big) = \alpha.$$

3. Proofs. We begin by introducing a class of function spaces in the first subsection. Subsection 3.2 deals with convolution and provides some useful inequalities, relating this operation to the function norms introduced earlier. The third subsection deals with renewal functions and collects some more inequalities; these are later used in conjunction with the convolution inequalities. Subsection 3.4 is on stochastic monotonicity a "sandwiching" technique is introduced. Empirical distribution functions and their weighted convergence are considered in subsection 3.5. In subsection 3.6 we investigate the local behaviour of the renewal functional, the nonlinear operator which maps a lifetime distribution function F on the associated renewal function U. Based on these preparations the proofs for Theorems 2.1, 2.2 and 2.3 are then given in subsection 3.7.

3.1. *Function spaces.* The basic objects of this paper are functions, defined on $\mathbb{R}$ or the compactified real line $[-\infty, \infty]$. Our function spaces are built on the familiar space D_∞ of bounded cadlag functions; to avoid trivialities we will assume that elements of D_∞ are left continuous at ∞. Endowed with the supremum norm $\|\cdot\|_\infty$, D_∞ is a Banach space.

We will consider renewal functions and their estimators as elements of a specific space of unbounded functions. In the various steps of our proofs we will also need subspaces of D_∞ which are characterized by certain rates of decrease at $\pm\infty$. We define all these spaces simultaneously. To do so, we first introduce operators $T_{\alpha\beta}$ which associate with a given function $f: \mathbb{R} \to \mathbb{R}$ a new function $T_{\alpha\beta} f: \mathbb{R} \to \mathbb{R}$ as follows:

$$T_{\alpha\beta} f(t) = \begin{cases} (1+t)^\beta f(t), & t \ge 0, \\ (1-t)^\alpha f(t), & t < 0 \end{cases}$$

and then define our function spaces and associated norms by

$$D_{\alpha\beta} = \{f: \mathbb{R} \to \mathbb{R}: T_{\alpha\beta} f \in D_\infty\}, \qquad \|f\|_{\alpha\beta} = \|T_{\alpha\beta} f\|_\infty.$$

Here the requirement $Tf \in D_\infty$ means that Tf can suitably be extended to $[-\infty, \infty]$: This means, for example, that for $f \in D_{0,-1}$, the limit of $f(x)/x$ as $x \to \infty$ exists (in $\mathbb{R}$). Evidently, these spaces are again Banach spaces and isomorphic to D_∞ via the associated T operator. Note the visual pun: $D_\infty = D_{00}$.

This clarifies the linear and the topological structure of the D spaces. For the measurability structure we take the σ-fields generated by the open (or closed) balls in the respective norm. Note that the T-operators and their inverses are measurable. A sequence $\{X_n\}$ of random elements is said to converge in distribution to X if $Ef(X_n) \to Ef(X)$ as $n \to \infty$ for all real-valued, bounded, measurable and continuous functions f on the respective D-space.

Finally, let $C_{\alpha\beta}(F)$ denote the subspace of $D_{\alpha\beta}$ which consists of all functions that are continuous at each continuity point of the distribution function F. These spaces are separable since the set of all discontinuity points of F is countable; in particular, their σ-fields coincide with the respective Borel system. For more information on weak convergence on nonseparable spaces we refer the reader to Pollard (1984) and the references given there.

3.2. *Convolution*. The convolution product $F \star G$ of two distribution functions F, G is given by $F \star G(x) = \int F(x-y) G(dy)$ for all $x \in \mathbb{R}$. We need an extension of this notion where the integrator is allowed to be unbounded and where the integrand is allowed to be of unbounded variation. In the extended notion the ordering of the factors will be important.

Let $\mathscr{H}$ be the set of all nondecreasing and right continuous functions H: $\mathbb{R} \to \mathbb{R}$ with the property that $\lim_{x \to -\infty} H(x) = 0$. Integrals of the form $\int \cdots H(dx)$ for some $H \in \mathscr{H}$ are to be understood as Lebesgue integrals with respect to the (possibly infinite) measure which assigns mass $H(b) - H(a)$ to the interval $(a, b]$, $-\infty < a < b < \infty$. Let $H \in \mathscr{H}$ and let g: $\mathbb{R} \to \mathbb{R}$ be a measurable function. If the functions $y \to g(x-y)$ are H-integrable for all $x \in \mathbb{R}$, then we say that the convolution product $g \star H$ of g and H exists and we define its value at x by

$$g \star H(x) = \int g(x-y) H(dy).$$

It is easily seen that for $g, H \in \mathscr{H}$ this yields an element of $\mathscr{H}$ again, provided that the integrals exist. In short, if defined, $g \star H$ qualifies as a possible right-hand factor so that convolution powers $H^{\star k}$ can be defined inductively; we also put $H^{\star 0} = 1_{[0,\infty)}$.

Properties of convolution will be used below without further comment. Some of these are obvious from the definition, others might require a little argument, based on monotone convergence or on Fubini's theorem as in the following lemma. In either case, we leave the details to the reader. It should be kept in mind, however, that only elements of $\mathscr{H}$ are allowed to appear on the right-hand side of the convolution symbol: $g \star H_1 - g \star H_2$ may in general not be written as $g \star (H_1 - H_2)$ as $\mathscr{H}$ is not a linear space.

LEMMA 3.1. *Let g, H_1, H_2 be such that the convolution products $g \star H_1$ and $g \star H_2$ exist. Then, if one of the products $(g \star H_1) \star H_2$, $(g \star H_2) \star H_1$ exists, then the other one also exists and they are equal to each other.*

In the one-sided case, that is, if $H(x) = 0$ and $g(x) = 0$ for all $x < 0$, a sufficient condition for the convolution product to exist is that g be bounded on compact intervals. In the general case the respective rates of increase or decrease of g and H at $\pm\infty$ have to cancel each other. The following lemma collects some elementary inequalities which also give information on the asymptotic behaviour of convolution products. Implicitly, these inequalities also provide criteria for the existence of the respective convolution products. The following notation will be useful: For a function $f\colon \mathbb{R} \to \mathbb{R}$ and $h > 0$ let $\Delta_h f$ be the function given by

$$\Delta_h f(x) = f(x+h) - f(x).$$

We abbreviate Δ_1 by Δ. Using an obvious rescaling argument it is often enough to consider only the case $h = 1$ in statements referring to Δ_h.

LEMMA 3.2. *For each $\varepsilon > 0$ there exist constants $c_1(\varepsilon)$ and $c_2(\varepsilon)$ such that*

(i) $$\|f \star H\|_{1+\varepsilon,0} \le c_1(\varepsilon) \|f\|_{2+\varepsilon,1+\varepsilon} \|\Delta H\|_{1+\varepsilon,0},$$

(ii) $$\|f \star H\|_{0,-1} \le c_2(\varepsilon) \|f\|_{1+\varepsilon,0} (\|\Delta H\|_\infty + \|H\|_{0,-1}).$$

PROOF. We assume in the proof that the values appearing on the right-hand side of the inequality under consideration are finite.

(i) Let $x \ge 0$. We have

$$|f \star H(x)| \le \sum_{k \in \mathbb{Z}} \int_{(x+k,\, x+k+1]} |f(x-y)| H(dy).$$

The summands can be estimated by

$$\int_{(x+k,\,x+k+1\,]} |f(x-y)|H(dy) \le \begin{cases} \sup\limits_{y \le -k} |f(y)|\,\Delta H(x+k), & k \ge 0, \\ \sup\limits_{y \ge -k-1} |f(y)|\,\Delta H(x+k), & k < 0, \end{cases}$$

so that, for all $x \ge 0$,

$$|f\star H(x)| \le \|f\|_{1+\varepsilon,0}\|\Delta H\|_\infty \sum_{k=0}^{\infty}(1+k)^{-1-\varepsilon} + \|f\|_{0,1+\varepsilon}\|\Delta H\|_\infty \sum_{k=1}^{\infty} k^{-1-\varepsilon}.$$

For $x < 0$ we write

$$|f\star H(x)| \le \sum_{k\in\mathbb{Z}} \int_{(x/2+k,\,x/2+k+1\,]} |f(x-y)|H(dy).$$

We have

$$\int_{(x/2+k,\,x/2+k+1\,]} |f(x-y)|H(dy) \le \sup_{x/2-k-1 \le y < x/2-k} |f(y)|\,\Delta H(x/2+k)$$

so that, considering $k \ge 0$ and $k < 0$ separately,

$$\begin{aligned} &\sup_{x\le 0}(1+|x|)^{1+\varepsilon}|f\star H(x)| \\ &\quad\le \sup_{x\le 0}\Big((1+|x|)^{1+\varepsilon}\sup_{y\le x/2}\Delta H(y)\Big)\sum_{k\in\mathbb{Z}}\sup_{k-1<y\le k+1}|f(y)| \\ &\qquad + \|f\|_{2+\varepsilon,0}\|\Delta H\|_\infty \sup_{x\le 0}\Big((1+|x|)^{1+\varepsilon}\sum_{k=0}^{\infty}(1+|x/2|+k)^{-2-\varepsilon}\Big). \end{aligned}$$

The term $\sum_{k\in\mathbb{Z}}\sup_{k-1<y\le k+1}|f(y)|$ is easily seen to be bounded by a multiple of $\|f\|_{1+\varepsilon,1+\varepsilon}$. Some obvious inequalities between the various norms now complete the proof of (i).

(ii) For $x < 0$ we write

$$|f\star H(x)| \le \int_{(-\infty,0\,]} |f(x-y)|H(dy) + \sum_{k=1}^{\infty}\int_{(k-1,\,k\,]} |f(x-y)|H(dy).$$

The first term is bounded by $\|f\|_\infty H(0)$. On the second we use

$$\int_{(k-1,\,k\,]} |f(x-y)|H(dy) \le \Delta H(k-1)\sup_{y\le -k+1}|f(y)|,$$

which gives

$$\sum_{k=1}^{\infty}\int_{(k-1,\,k\,]} |f(x-y)|H(dy) \le \|\Delta H\|_\infty \|f\|_{1+\varepsilon,0}\sum_{k=1}^{\infty} k^{-1-\varepsilon}.$$

For $x > 0$ we use the same arguments,

$$|f\star H(x)| \le \int_{(-\infty, x]} |f(x-y)| H(dy) + \sum_{k=1}^{\infty} \int_{(x+k-1, x+k]} |f(x-y)| H(dy).$$

The first term is bounded by $\|f\|_\infty H(x)$. Further,

$$\begin{aligned}\int_{(x+k-1, x+k]} |f(x-y)| H(dy) &\le \Delta H(x+k-1) \sup_{y \le -k+1} |f(y)| \\ &\le k^{-1-\varepsilon} \|f\|_{1+\varepsilon, 0} \|\Delta H\|_\infty\end{aligned}$$

so that these terms sum to a finite value. □

3.3. *Renewal functions.* If F satisfies

$$\mu_2 := \int x^2 F(dx) < \infty, \qquad \mu_1 := \int x F(dx) > 0, \tag{3.1}$$

then the values $F^{\star k}(x)$, $k = 0, 1, \ldots$ sum to a finite limit [see, e.g., Gut (1988), page 89] so that we can define the renewal function U associated with F by

$$U(x) := \sum_{k=0}^{\infty} F^{\star k}(x) \quad \text{for all } x \in \mathbb{R}.$$

We need some bounds on U and its increments ΔU in order to be able to apply the inequalities from the preceding section. Suitable bounds for our purposes have been obtained by Daley (1980). We collect these in the following lemma, the first part being immediate from Theorem 1 in Daley (1980), the second following from Daley's equation (3.6) on noting that, for any random variable M,

$$-E((M+y) \wedge 0) = \int_{-\infty}^{-y} \mathbb{P}(M < z)\, dz \quad \text{for all } y > 0.$$

LEMMA 3.3 [Daley (1980)]. *Let U be the renewal function associated with a distribution function F for which* (3.1) *holds, and let M be the global minimum of a random walk with step distribution F. Then the following inequalities hold for all $x \in \mathbb{R}$:*

$$0 \le U(x) \le (x \vee 1)/\mu_1 + \mu_2/\mu_1^2,$$
$$\Delta U(x) \le \left(1/\mu_1 + \mu_2/\mu_1^2\right) \mathbb{P}(M < x+1).$$

These inequalities will be used in the form $\|U\|_{0,-1} < \infty$, $\|\Delta U\|_\infty < \infty$ below. We will also need the following variant: If $\int |x|^\alpha F(dx) < \infty$, then $E|M|^{\alpha-1} < \infty$ [see Gut (1988), Theorem 4.9] which implies $\|\Delta U\|_{\alpha-1,0} < \infty$ by the second inequality in the preceding lemma.

3.4. *Monotonicity.* An important tool in our proofs will be a "sandwiching technique" which makes use of stochastic ordering and the corresponding monotonicity of the renewal functional. Recall that a distribution function F

is stochastically smaller than or equal to a distribution function G, notation $F \preccurlyeq G$, if $1 - F(x) \le 1 - G(x)$ for all $x \in \mathbb{R}$. The first lemma relates $\|\cdot\|$-balls and $\preccurlyeq$ -intervals, the second gives monotonicity of the renewal functional. The third lemma bounds the variability of $g \star \Psi(F)$ as F ranges over a stochastic interval of distribution functions.

LEMMA 3.4. *Let α, β be such that $\alpha > \beta > 1$ and assume that the distribution function F satisfies*

$$\int |x|^{\alpha} F(dx) < \infty, \qquad \int x F(dx) > 0.$$

Then, for every $\varepsilon > 0$, there exist distribution functions G_1, G_2 such that

$$\|G_i - F\|_{\beta\beta} \le \varepsilon, \qquad \int |x|^{\beta} G_i(dx) < \infty, \qquad \int x G_i(dx) > 0 \quad \text{for } i = 1, 2,$$

and the property that, for some $\delta > 0$, the implication

$$\|H - F\|_{\alpha\alpha} < \delta \Rightarrow G_1 \preccurlyeq H \preccurlyeq G_2$$

holds for all distribution functions H.

PROOF. From the symmetry of the notions involved it is clear that it is enough to construct a suitable stochastic lower bound. For this, we first note that there exists a constant $c < \infty$ such that

$$\left|1_{(0,\infty)}(x) - F(x)\right| \le c(1 + |x|)^{-\alpha} \quad \text{for all } x \in \mathbb{R}.$$

For a given $\delta > 0$ with $0 < \delta < c \wedge (1 - F(0))$ let $x_0 := 1 - (2c/\delta)^{1/\alpha}$. Then G,

$$G(x) := \begin{cases} 1 \wedge (F(x) + \delta), & x \ge x_0, \\ 2c(1 + |x|)^{-\alpha}, & x < x_0, \end{cases}$$

is a distribution function. By construction, $\int |x|^{\beta} G(dx) < \infty$, and any distribution function H with $\|H - F\|_{\alpha\alpha} \le \delta$ satisfies $H(x) \le G(x)$ for all $x \in \mathbb{R}$. Further,

$$\begin{aligned} \|G - F\|_{\beta\beta} \le {} & \sup_{x < x_0} (1 + |x|)^{\beta} G(x) + \sup_{x_0 \le x \le -x_0} (1 + |x|)^{\beta} (G(x) - F(x)) \\ & + \sup_{x > -x_0} (1 + |x|)^{\beta} (1 - F(x)) \\ \le {} & 2c(1 + |x_0|)^{\beta - \alpha} + 2\delta(1 + |x_0|)^{\beta} + c(1 + |x_0|)^{\beta - \alpha}. \end{aligned}$$

From the definition of x_0 we see that $\|G - F\|_{\beta\beta} \to 0$ as $\delta \downarrow 0$; in particular, $\|G - F\|_{\beta\beta} \le \varepsilon$ if δ is chosen small enough. Since $\beta > 1$, $\|\cdot\|_{\beta\beta}$-convergence implies convergence of the first moments, that is, $\int x G(dx) > 0$ will also hold provided that δ is chosen small enough. Hence, for a suitable $\delta > 0$, $G_1 := G$ satisfies all the requirements of the lemma □

LEMMA 3.5. *Let F_i, $i = 1, 2$ be distribution functions satisfying* (3.1). *For $i = 1, 2$ let U_i be the renewal function associated with F_i. Then $F_1 \preccurlyeq F_2$ implies $U_1(x) \geq U_2(x)$ for all $x \in \mathbb{R}$.*

PROOF. Let $\{\xi_n\}$ be a sequence of independent and uniformly on $(0, 1)$ distributed random variables and put

$$X_n := F_1^{-1}(\xi_n), \qquad Y_n := F_2^{-1}(\xi_n) \quad \text{where } F_i^{-1}(x) := \inf\{y: F_i(y) \geq x\}.$$

Then $X_n \leq Y_n$ for all n, hence

$$\#\left\{n: \sum_{i=1}^{n} X_i \leq x\right\} \geq \#\left\{n: \sum_{i=1}^{n} Y_i \leq x\right\},$$

from which the statement follows on taking expectations. □

Let $\mathscr{I}$ be the space of linear combinations of indicator functions of intervals of the form $[a, b)$, $-\infty < a < b \leq \infty$. Remember that $\Psi(F)$ denotes the renewal function associated with the distribution function F.

LEMMA 3.6. *For every $g \in \mathscr{I}$ there exists a constant $c < \infty$ such that the following inequality holds for all distribution functions F_1, F_2, G_1, G_2 with $G_1 \preccurlyeq F_1$, $F_2 \preccurlyeq G_2$ satisfying* (3.1),

$$\|g \star \Psi(F_2) - g \star \Psi(F_1)\|_{0,-1} \leq c\|\Psi(G_2) - \Psi(G_1)\|_{0,-1}.$$

PROOF. It is enough to consider indicator functions, so let $g = 1_{[a,b)}$. Then, using Lemma 3.5,

$$\begin{aligned}
&|g \star \Psi(F_2)(x) - g \star \Psi(F_1)(x)| \\
&\quad = |\Psi(F_2)(x - a) - \Psi(F_2)(x - b) - \Psi(F_1)(x - a) + \Psi(F_1)(x - b)| \\
&\quad \leq (\Psi(G_1) - \Psi(G_2))(x - a) + (\Psi(G_1) - \Psi(G_2))(x - b),
\end{aligned}$$

where the last term vanishes if b is infinite. This suffices to establish the desired inequality as $x \to (1 + x + y)/(1 + x)$ is bounded on $x \geq 0$ by a constant depending only on y. □

3.5. *Empirical distribution functions.* Let $X, X_1, X_2, \ldots$ be a sequence of independent random variables on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ with distribution function F and let $\hat{F}_n$ be the empirical distribution function associated with $X_1, \ldots, X_n$. All limit statements in the remaining subsections refer to $n \to \infty$ unless stated otherwise.

In order to be able to apply the results of the previous subsections we need a result which amplifies the Glivenko–Cantelli theorem to uniform convergence with a weight factor. Such a result can be obtained from the corresponding special result for the uniform distribution on the interval (0, 1) by applying the familiar time change using F [see e.g., Pollard (1984), page 97]. In the special case of uniformly distributed random variables a necessary and sufficient condition for the weight function is known [apparently due to Lai; see Shorack and Wellner (1986), Section 10.2]. For general F we obtain the following.

PROPOSITION 3.7. *Let $\alpha \geq 0$. Then $E|X|^\alpha < \infty$ implies*

$$\|\hat{F}_n - F\|_{\alpha\alpha} \to 0 \quad \textit{almost surely}.$$

We need an analogue of Proposition 3.7 for weak convergence to a Brownian bridge. Again, necessary and sufficient conditions on the weight function are known if F is uniform on $(0,1)$ [O'Reilly (1974); see also Shorack and Wellner (1986), Section 3.7]. To translate these to the case of general F we use an almost sure representation of the weak convergence result for uniform distributions and Chebyshev's inequality.

PROPOSITION 3.8. *If $E|X|^\alpha < \infty$ for some $\alpha > 0$, then with B a standard Brownian bridge and any $\beta < \alpha/2$,*

$$\sqrt{n}\left(\hat{F}_n - F\right) \to_{\text{distr}} B(F) \quad \textit{in } D_{\beta\beta}.$$

Since the Brownian bridge has continuous paths the distribution of $B(F)$ in the preceding proposition is concentrated on $C_{\beta\beta}(F)$.

3.6. *Local properties of the renewal functional.* Throughout this section $F_0, F_1, F_2 \dots$ are distribution functions satisfying (3.1) and $U_0, U_1, U_2 \dots$ are the associated renewal functions. The respective first and second moments are

$$\mu_{i,n} = \int x^i F_n(dx) \quad \text{for } n = 0, 1, \dots, \; i = 1, 2.$$

We study the local behaviour of the renewal functional Ψ which carries F to U, regarded as a mapping from (a subset of) a suitable D-space to $D_{0,-1}$. It might be interesting to note that this subsection follows its own "bootstrap principle": First, we prove local boundedness of the functional. This is used to obtain continuity which, in turn, is used to prove a differentiability property.

Local boundedness is obtained in Lemma 3.9. Lemma 3.10 contains a renewal type equation, Proposition 3.11 gives continuity.

LEMMA 3.9. *Let $\alpha > 2$ and assume that*

$$\int |x|^\alpha F_n(dx) < \infty \quad \textit{for } n = 0, 1, 2, \dots \quad \textit{and} \quad \lim_{n\to\infty} \|F_n - F_0\|_{\alpha\alpha} = 0.$$

Then

$$\sup_{n\geq 0} \|\Delta U_n\|_\infty < \infty, \qquad \sup_{n\geq 0} \|U_n\|_{0,-1} < \infty.$$

PROOF. Partial integration shows that $\|F_n - F_0\|_{\alpha\alpha} \to 0$ implies $\mu_{in} \to \mu_{i0}$, $i = 1, 2$. As F_0 and F_n satisfy (3.1) we obtain $\sup_n(1/\mu_{1,n} + \mu_{2,n}/\mu_{1,n}^2) < \infty$ and the statement of the lemma follows on using Lemma 3.3. □

LEMMA 3.10. *If* $\int |x|^\alpha F_0(dx) < \infty$ *and* $\int |x|^\alpha F_n(dx) < \infty$ *for some* $\alpha > 2$, *then*

$$U_0 - U_n = (F_0 - F_n)\star U_0 \star U_n.$$

PROOF. Our assumptions imply that, for $i = 0, n$, $\|F_i - 1_{[0,\infty)}\|_{\alpha\alpha} < \infty$ and, by the remarks following Lemma 3.3, $\|U_i\|_{0,-1} < \infty$ and $\|\Delta U_i\|_{\alpha-1,0} < \infty$, so that, by Lemma 3.2, the convolution products on the right-hand side of the asserted formula exist. Further, using monotone convergence, $1_{[0,\infty)} + F_i \star U_i = U_i$, and we obtain, using Lemma 3.1,

$$\begin{aligned}(F_0 - F_n)\star U_0 \star U_n &= \big((F_0 - F_0^{\star 0})\star U_0\big)\star U_n - \big((F_n - F_n^{\star 0})\star U_n\big)\star U_0 \\ &= U_0 - U_n. \qquad \square\end{aligned}$$

PROPOSITION 3.11. *Assume the conditions of Lemma* 3.9. *Then*

$$\lim_{n\to\infty} \|U_n - U_0\|_{0,-1} = 0.$$

PROOF. Again, $\mu_{i,n} \to \mu_{i,0}$, $i = 1, 2$. From Lemma 3.10, $U_n - U_0 = (F_n - F_0)\star U_0 \star U_n$. As explained at the end of subsection 3.3 we have $\|\Delta U_0\|_{\alpha-1,0} < \infty$, so that, by Lemma 3.2(i), $(F_n - F_0)\star U_0 \to 0$ in $D_{\alpha-1,0}$. Now apply Lemma 3.2(ii) and Lemma 3.9. □

For the proof of a differentiability property of the functional we need an auxiliary result which in turn requires the following lemma. Remember that $\mathscr{I}$ is the space of linear combinations of indicator functions of intervals of the form $[a, b)$, $-\infty < a < b \le \infty$. As the following lemma shows this set is dense in $D_{\alpha 0}$ with respect to a weaker norm. That such a weakening is necessary is shown by the function $g(x) = (1 + |x|)^{-\alpha}$ which cannot be approximated in $\|\cdot\|_{\alpha 0}$-norm by functions from $\mathscr{I}$.

LEMMA 3.12. *Let* α, β *be such that* $0 \le \beta < \alpha < \infty$ *and let* $g \in D_{\alpha 0}$. *Then, for each* $\varepsilon > 0$, *there exists a* $g_\varepsilon \in \mathscr{I}$ *such that* $\|g - g_\varepsilon\|_{\beta 0} < \varepsilon$.

PROOF. In $D_{\beta 0}$ $g(1 - 1_{[-c,\infty)})$ tends to 0 as $c \to \infty$. On a fixed interval $[-c, \infty)$, any D_∞-function can be uniformly approximated by functions from $\mathscr{I}$ since $g(x)$ tends to a finite limit as $x \to \infty$. □

PROPOSITION 3.13. *Assume the conditions of Lemma* 3.9. *Let further* $g \in D_{\beta 0}$ *for some* $\beta > \alpha - 1$. *Then* $\lim_{n\to\infty} \|g\star U_n - g\star U_0\|_{0,-1} = 0$.

PROOF. Let $\varepsilon > 0$ be given. From Lemma 3.12 we obtain a $g_0 \in \mathscr{I}$ such that, with c_2 as in Lemma 3.2 and $K := \sup_{n\ge 0}\|\Delta U_n\|_\infty + \sup_{n\ge 0}\|U_n\|_{0,-1}$,

$$\|g - g_0\|_{\alpha-1,0} \le \varepsilon/(3c_2(\alpha - 2)K);$$

K is finite by Lemma 3.9. Lemma 3.4, together with the continuity of Ψ established in Proposition 3.11, gives us distribution functions G_1, G_2 such that

$$\|\Psi(G_2) - \Psi(G_1)\|_{0,-1} \le \varepsilon/(3c),$$

where c is the constant from Lemma 3.6, depending on g_0 only; moreover, from some n_0 onwards, all F_n will satisfy $G_1 \preccurlyeq F_n \preccurlyeq G_2$. This means, using Lemma 3.6, that

$$\|g_0 \star U_n - g_0 \star U_0\|_{0,-1} < \varepsilon/3$$

for $n \geq n_0$, so that, on using Lemma 3.2,

$$\begin{aligned}\|g \star U_n - g \star U_0\|_{0,-1} &\leq \|(g - g_0)\star U_n\|_{0,-1} \\ &\qquad + \|g_0 \star U_n - g_0 \star U_0\|_{0,-1} + \|(g - g_0)\star U_0\|_{0,-1} \\ &\leq \varepsilon\end{aligned}$$

for all n large enough. □

The following result gives a more quantitative approximation and will later be referred to as a differentiability property of the renewal functional.

PROPOSITION 3.14. *Assume, in addition to the conditions of Lemma* 3.9, *that*

$$\sqrt{n}\,(F_n - F_0) \to h \quad \text{in } D_{\alpha\alpha} \text{ for some } \alpha > 2,\ h \in D_{\alpha\alpha}.$$

Then

$$\sqrt{n}\,(U_n - U_0) \to h \star U_0 \star U_0 \quad \text{in } D_{0,-1}.$$

PROOF. Lemma 3.10 gives, with $g := h \star U_0$,

$$\begin{aligned}&\sqrt{n}\,(U_n - U_0) - h \star U_0 \star U_0 \\ &\qquad = \left(\sqrt{n}\,(F_n - F_0) - h\right)\star U_0 \star U_n + g \star U_n - g \star U_0.\end{aligned}$$

For the first term we obtain $\|\cdot\|_{0,-1}$-convergence in exactly the same manner as in the proof of Proposition 3.11, with $F_n - F_0$ replaced by $\sqrt{n}\,(F_n - F_0) - h$. For the second term we use Proposition 3.13. □

The proposition shows that the functional Ψ is differentiable along certain curves in its range of definition, and that its derivative at F_0 is the linear operator $T\colon D_{\alpha\alpha} \to D_{0,-1}$ which maps h onto $T(h) := h \star V_0$, where $V_0 := \Psi(F_0)\star\Psi(F_0) = U_0 \star U_0$. This operator is bounded: It follows from Lemma 3.2 and the remarks following Lemma 3.3 that

$$\|T\| \leq c_2(\alpha - 2)c_1(\alpha - 2)\|\Delta U_0\|_{\alpha-1,0}\left(\|\Delta U_0\|_\infty + \|U_0\|_{0,-1}\right) < \infty.$$

3.7. *Proofs of Theorems.*

PROOF OF THEOREM 2.1. This is now immediate from Proposition 3.7 and Proposition 3.11. □

PROOF OF THEOREM 2.2. Choose α such that $2 < \alpha < \beta/2$. From Proposition 3.8,

$$\sqrt{n}\left(\hat{F}_n - F\right) \to_{\text{distr}} B(F) \quad \text{in } D_{\alpha\alpha}.$$

We now apply a Skorohod–Dudley construction [see, e.g., Pollard (1984), page 71]: as the limit distribution is concentrated on the separable subspace $C_{\alpha\alpha}(F)$ we can find a probability space $(\Omega', \mathscr{A}', \mathbb{P}')$ and, on it, D_∞-valued random quantities F_n', $n \in \mathbb{N}$, and a Brownian bridge B' such that $F_n' =_{\text{distr}} \hat{F}_n$ for all $n \in \mathbb{N}$ and

$$\sqrt{n}\,(F_n' - F) \to B'(F) \quad \mathbb{P}'\text{-a.s. in } D_{\alpha\alpha}$$

(we write $X =_{\text{distr}} Y$ if the random quantities X and Y have the same distribution). Let U_n' be the renewal function associated with F_n'; clearly, $\hat{U}_n =_{\text{distr}} U_n'$. Proposition 3.14 implies

$$\sqrt{n}\,(U_n' - U) \to B'(F)\star U \star U \quad \mathbb{P}'\text{-a.s. in } D_{0,\,-1}.$$

The right-hand side is a linear transformation of a Gaussian process, hence Gaussian again. The second order structure of this process is easily seen to be of the form given in the theorem. □

PROOF OF THEOREM 2.3. We consider a slightly more general setup: Let ρ be a continuous and measurable nonnegative real function on $D_{0,\,-1}$ and put

$$R_n(z) = \mathbb{P}\Big(\rho\Big(\sqrt{n}\,\big(\hat{U}_n - U\big)\Big) \le z\Big), \qquad R(z) = \mathbb{P}\big(\rho(Z^U) \le z\big),$$

$$\hat{R}_n(z) = n^{-n} \sum_{i \in I_n} 1_{[0,z]}\Big(\rho\Big(\sqrt{n}\,\Big(\Psi\big(\mathbb{F}_n(X_{i_1}, \dots, X_{i_n})\big) - \Psi\big(\hat{F}_n\big)\Big)\Big)\Big).$$

In the theorem we have $\rho = \|\cdot\|_{0,\,-1}$. The sequence $\hat{R}_n$ of random distribution functions can be regarded as a sequence of D_∞-valued random elements. We first prove that this sequence converges in distribution to the constant R, the distribution function of $\rho(Z^U)$. This will be done on using an almost sure construction, that is, we construct D_∞-valued random quantities $R_1', R_2', \dots$ on some probability space $(\Omega', \mathscr{A}', \mathbb{P}')$ with the properties $R_n' =_{\text{distr}} \hat{R}_n$ for all $n \in \mathbb{N}$ and $R_n'(\omega) \to R$ for $\mathbb{P}'$-almost all $\omega \in \Omega'$; this representation implies the desired statement.

The construction is based on an idea of Shorack (1982). We essentially follow Section 4 in Gill (1989) but give a more detailed description of the construction; Gill also refers to Bickel and Freedman (1981). We need to base our functionals on more general domains, namely the $D_{\alpha\beta}$-spaces introduced in subsection 3.1. This gives rise to a small technical complication.

PROPOSITION 3.15. *Suppose that*

$$\int |x|^\beta F(dx) < \infty \quad \textit{for some } \beta > 4, \qquad \int x F(dx) > 0.$$

Then $\hat{R}_n \to_{\text{distr}} R$ *in* D_∞ *as* $n \to \infty$.

PROOF. Choose α and δ such that $2 < \alpha < \beta/2$, $\alpha/\beta < \delta < 1/2$ and let the function q: $(0,1) \to (0,1)$ be given by

$$q(t) = q(1-t) = t^{\delta}, \qquad 0 < t \le 1/2.$$

Remember that the $\hat{F}_n$'s are defined on $(\Omega, \mathscr{A}, \mathbb{P})$. From subsection 3.5 we know that

$$\hat{F}_n - F \to 0 \quad \mathbb{P}\text{-a.s. in } D_{\beta\beta}, \qquad \sqrt{n}\left(\hat{F}_n - F\right) \to_{\text{distr}} B(F) \quad \text{in } D_{\alpha\alpha},$$

where B is a Brownian bridge and $B(F)$ is short for $(t, \omega) \to B(F(t), \omega)$. Consider the pair $(\hat{F}_n, \sqrt{n}(\hat{F}_n - F))$. Its limiting distribution is concentrated on a separable space of pairs of functions, hence we can find a probability space $(\Omega', \mathscr{A}', \mathbb{P}')$ and, on it, D_∞-valued random quantities F_n', $n \in \mathbb{N}$, and a Brownian bridge B' such that $F_n' =_{\text{distr}} \hat{F}_n$ for all $n \in \mathbb{N}$ and

$$F_n' \to F \quad \mathbb{P}'\text{-a.s. in } D_{\beta\beta}, \qquad \sqrt{n}\left(F_n' - F\right) \to B'(F) \quad \mathbb{P}'\text{-a.s. in } D_{\alpha\alpha} \tag{3.2}$$

[see Pollard (1984), page 71]. Let $(\Omega^\circ, \mathscr{A}^\circ, \mathbb{P}^\circ)$ be a third probability space, carrying an array $\{\xi_{ni}\colon n \in \mathbb{N},\ 1 \le i \le n\}$ of row-wise independent random variables, uniformly distributed on $(0,1)$, and a Brownian bridge B° such that, with G_n° denoting the empirical distribution function associated with $\xi_{n1}, \dots, \xi_{nn}$,

$$\sup_{0<t<1} \frac{1}{q(t)} \left|\sqrt{n}\left(G_n^\circ(t) - t\right) - B^\circ(t)\right| \to 0 \quad \mathbb{P}^\circ\text{-a.s.}; \tag{3.3}$$

this is possible since q satisfies the conditions of Theorem 2 in O'Reilly (1974) and an almost sure representation can again be found. Now fix $\omega' \in \Omega'$ and $\omega^\circ \in \Omega^\circ$. Let $x_1, \dots, x_n$ be the jumps of F_n', in increasing order. The function

$$t \to G_n^\circ\left(F_n'(t, \omega'), \omega^\circ\right)$$

is a discrete distribution function with jumps in $x_1, \dots, x_n$ only. The height of the jump in x_k is easily seen to be equal to n^{-1} times the number of values $\xi_{ni}(\omega^\circ)$, $1 \le i \le n$, in the interval $((k-1)/n, k/n)$. As the joint distribution of these frequencies is the same as the joint distribution of $i_1, \dots, i_n$ if $(i_1, \dots, i_n)$ is selected uniformly at random from I_n (both are multinomial with parameters n and $(1/n, \dots, 1/n)$) we see that, for fixed ω', the distribution of the above function, regarded as a mapping from Ω° to D, is given by

$$\mathbb{P}^\circ\left(G_n^\circ\left(F_n'(\cdot, \omega'), \cdot\right)\right) = n^{-n} \sum_{i \in I_n} \delta\left(\mathbb{F}_n(x_{i_1}, \dots, x_{i_n})\right), \tag{3.4}$$

where $\delta(a)$ denotes unit mass in a. We now define R_n': $\Omega' \to D$ by

$$\left(R_n'(\omega')\right)(z) = \mathbb{P}^\circ\Big(\Big\{\omega^\circ \in \Omega^\circ : \rho\Big(\sqrt{n}\left(\Psi\left(G_n^\circ\left(F_n'(\cdot, \omega'), \omega^\circ\right)\right)\right. \\ \left. - \Psi\left(F_n'(\cdot, \omega')\right)\right)\Big) \le z\Big\}\Big).$$

Then R_n' depends on ω' only through $F_n'(\omega')$, $\hat{R}_n$ depends on ω only through $\hat{F}_n(\omega)$. From (3.4) we see that these dependencies are the same. By construction, F_n' and $\hat{F}_n$ are equal in distribution, hence it follows that R_n' and R_n

have the same distribution. It remains to show that R'_n tends to R $\mathbb{P}'$-almost surely.

Fix $\omega' \in \Omega'$, $\omega^\circ \in \Omega^\circ$ satisfying (3.2) and (3.3), and drop these temporarily from the notation. We have

$$\begin{aligned}(3.5)\qquad &\sqrt{n}\,(\Psi(G_n^\circ(F'_n)) - \Psi(F'_n))\\ &\quad = \sqrt{n}\,(\Psi(G_n^\circ(F'_n)) - \Psi(F)) - \sqrt{n}\,(\Psi(F'_n) - \Psi(F)).\end{aligned}$$

From (3.2) and Proposition 3.14 we have

$$(3.6)\qquad \sqrt{n}\,(\Psi(F'_n) - \Psi(F)) \to T(B'(F)) \quad \text{in } D_{0,-1},$$

where $T(g) = g\star\Psi(F)\star\Psi(F)$ is the linear operator representing the derivative of Ψ at F (see the end of subsection 3.6). In order to be able to apply the same argument to the other term on the right-hand side of (3.5) we would need $D_{\alpha\alpha}$-convergence of $\sqrt{n}\,(G_n^\circ(F'_n) - F)$. We have

$$(3.7)\qquad \sqrt{n}\,(G_n^\circ(F'_n) - F) = \sqrt{n}\,(G_n^\circ(F'_n) - F'_n) + \sqrt{n}\,(F'_n - F).$$

Assume now that the following statement holds,

$$(3.8)\qquad \sqrt{n}\,(G_n^\circ(F'_n) - F'_n) \to B^\circ(F) \quad \text{in } D_{\alpha\alpha}.$$

Using (3.2) we then obtain the limit $B^\circ(F) + B'(F)$ in (3.7) so that we can apply differentiability of Ψ again to arrive at

$$\sqrt{n}\,(\Psi(G_n^\circ(F'_n)) - \Psi(F)) \to T(B^\circ(F) + B'(F)) \quad \text{in } D_{0,-1}.$$

Linearity of T, together with (3.6), now gives

$$(3.9)\qquad \sqrt{n}\,(\Psi(G_n^\circ(F'_n)) - \Psi(F'_n)) \to Z^\circ \quad \text{in } D_{0,-1},$$

where $Z^\circ := T(B^\circ(F))$. Note that Z° no longer depends on ω'. As a $D_{0,-1}$-valued random quantity Z° is equal in distribution to Z^U, the Gaussian process arising as the limit in distribution of $\sqrt{n}\,(\Psi(\hat{F}_n) - \Psi(F))$. Hence it follows from (3.9), the definition of R'_n and R and the assumptions on ρ that

$$R'_n(\omega')(z) \to \mathbb{P}^\circ(\rho(Z^\circ) \le z) = R(z)$$

as desired.

It remains to prove (3.8). Let $g_n(t) = \sqrt{n}\,(G_n^\circ(t) - t)$, $g(t) = B^\circ(t)$. Discarding another set of $\mathbb{P}^\circ$-measure zero if necessary we may assume that the underlying element of Ω° is such that, for some $\varepsilon > 0$,

$$(3.10)\quad 0 < s, t < 1, \qquad |s - t| \le \varepsilon \quad \Rightarrow \quad |g(s) - g(t)| \le q(|s - t|),$$

since q dominates the modulus of continuity of Brownian motion near 0. We have

$$\begin{aligned}(3.11)\qquad (1 + |t|)^\alpha |g_n(F'_n(t)) - g(F(t))| &\le (1 + |t|)^\alpha |g_n(F'_n(t)) - g(F'_n(t))|\\ &\quad + (1 + |t|)^\alpha |g(F'_n(t)) - g(F(t))|.\end{aligned}$$

For the first term in the decomposition we obtain

$$\sup_{t\in\mathbb{R}}(1+|t|)^{\alpha}|g_n(F_n'(t)) - g(F_n'(t))|$$
$$\leq \sup_{t\in\mathbb{R}}(1+|t|)^{\alpha}q(F_n'(t)) \sup_{0<s<1}\frac{1}{q(s)}|g_n(s) - g(s)|.$$

From the first part of (3.2) and the moment assumptions on F it follows that the sequence $\{1_{[0,\infty)} - F_n'\}$ is bounded in $D_{\beta\beta}$ so that, on using (3.3) and the definition of α and q, we see that this term tends to zero. For the second term in (3.11) we obtain on using (3.10)

$$(1+|t|)^{\alpha}|g(F_n'(t)) - g(F(t))| = (1+|t|)^{\alpha}O(q(|F_n'(t) - F(t)|)).$$

Taking the supremum over t and using $F_n' \to F$ in $D_{\beta\beta}$ and the definition of q and α settles this term too. □

Results of Tsirelson (1975) imply that R is continuous if $\rho = \|\cdot\|_{0,-1}$, except possibly in $r := \inf\{x\colon R(x) > 0\}$. By assumption the step distribution is not concentrated in one point which implies that the distribution of the supremum cannot have an atom at zero. On the other hand, we cannot have $r > 0$ since Z^U is the continuous image of a process that is known to stay inside $[-\varepsilon, \varepsilon]$ with positive probability, for all $\varepsilon > 0$. Hence R is continuous. Theorem 2.3 now follows from Proposition 3.15 and the following elementary lemma whose proof we leave to the reader.

LEMMA 3.16. *Let $X, X_1, X_2, \dots$ be real random variables with $X_n \to_{\text{distr}} X$ and assume that the distribution function R of X is continuous. Let further R_n be a sequence of random distribution functions which converge to R in probability as elements of D_∞, and let $\hat{q}_n(\alpha)$ be the α-quantile of $\hat{R}_n$; $0 < \alpha < 1$ fixed. Then*

$$\lim_{n\to\infty}\mathbb{P}(X_n \leq \hat{q}_n(\alpha)) = \alpha.$$

4. Comments. Here we collect some material dealing with various technical aspects and the literature. The final part attempts a summary.

4.1. *Numerical aspects.* The practical value of the procedures proposed and analyzed in this paper greatly depends on a good (i.e., fast, with negligible errors) algorithm that provides the renewal function for a given step distribution; note that this calculation has to be done several hundred times or more if a Monte Carlo approximation for $\hat{R}_n$ is used to obtain bootstrap confidence regions. Transform techniques lead to such a method which is based on the fast Fourier transform algorithm. A description is given in Grübel (1989b) for the one-sided case; the extension to the general case is straightforward.

4.2. *Other estimators.* In this subsection we relate our results to the work of Frees (1986a, b) and Schneider, Lin and O'Cinneide (1990). The situation in these papers is the same as the one considered here: $X_1, \dots, X_n$ is a sample from a distribution with (unknown) distribution function F and the renewal function $U = \sum_0^\infty F^{\star k}$ is to be estimated. (In some references, summation in the definition of the renewal function starts with $k = 1$, requiring trivial modifications only.) Schneider, Lin and O'Cinneide (1990) consider only the one-sided case.

Frees (1986a, b) obtains an estimator $\tilde{U}_{n,m}$ for U which differs from the empirical renewal function: First, the infinite convolution series is truncated at some $m \le n$, then unbiased estimators are inserted for the remaining convolution powers of F. Frees obtains uniform consistency on compact intervals of the form $[0, t]$, $0 \le t < \infty$, and asymptotic normality of $\tilde{U}_{n,m}(t)$ for a single fixed value t, under moment conditions that are weaker than our conditions for convergence of the whole process $\hat{U}_n$. His proofs are based on martingale arguments and other techniques from the theory of U-statistics and are, therefore, entirely different from the methods of the present paper.

We think that, in a variety of cases, the empirical renewal function is preferable to Frees' proposal, mainly because of the following reasons which have, at least partially, already been observed by Frees. First, $\tilde{U}_{n,m}$ involves the additional design parameter m which has to depend on n in order of the asymptotic results to become valid. No such parameter appears in our proposal. Second, computation of $\tilde{U}_{n,m}$ is awkward. Third, although $\tilde{U}_{n,m}$ is based on unbiased estimators for the convolution powers of F, it is itself not an unbiased estimator for U. Indeed, following Frees' proposal, we would estimate a renewal function by something which is not a renewal function. As $t \to \infty$, $\tilde{U}_{n,m}(t) \to m$, so, by design, the behaviour for large t-values will be poor.

At the end of his paper Frees (1986b) briefly compares his proposal to what we call the empirical renewal function. His main objection to the latter is its potential misbehaviour for two-sided distributions: If all sample values happen to be negative, then $\hat{U}_n \equiv \infty$ would arise. In our experience this is easily taken care of (see our definition of $\hat{U}_n$); also, the problem will arise with nonnegligible probability only if the sample size is small. Finally, if all data values are negative and no other information is available, the point could be made that one should indeed estimate the expected number of partial sums to the left of t to be infinite.

There is, however, one combination of circumstances where Frees' estimator might be the better choice: If the sample size is small, if interest is in small t-values only, and if the data are positive (such as lifetimes), then bias might be the overriding issue. This is amplified by the observation that, in the one-sided case, only sample values less than or equal to t contribute to the estimate of $U(t)$, so that the "effective sample size" might be even smaller than n. It should be mentioned here that in one of Frees' main applications, to warranty analysis, these three conditions are satisfied.

There is, of course, the obvious procedure for large t-values: Renewal theory describes the behaviour of $U(t)$ as $t \to \infty$ in terms of the first two moments of F. Replacing these by the sample moments leads to estimates of $U(t)$; see the introduction to Frees (1986b) for details. If t is indeed large, then this estimator should be used, maybe even just for its simplicity.

Schneider, Lin and O'Cinneide (1990) proffer a partial remedy for the computational problems associated with Frees' proposal and compare the behaviour of $\tilde{U}_{n,m}$ to that of $\hat{U}_n$ by means of a simulation study. Their findings seem to imply that even for small t-values the mean squared error of $\hat{U}_n$ will be only marginally bigger than that of $\tilde{U}_{n,m}$. Schneider, Lin and O'Cinneide (1990) also propose a method to obtain $\hat{U}_n$ numerically. This method is applicable, however, in the one-sided case only.

4.3. *One-sided distributions, finite intervals.* The reader patient enough to have followed us so far will have realized that the main part of the technical effort in Section 3 went into controlling the behaviour of the renewal functions and their estimates at plus and minus infinity. If we restrict ourselves to one-sided distributions, that is, if we assume that the underlying lifetime distribution function F satisfies $F(0-) = 0$ and if, moreover, we are interested in the behaviour of the empirical renewal function on some finite interval $[0, c]$ only, then a number of simplifications can be made. In particular, we do not need the $D_{\alpha\beta}$-spaces and we can regard Ψ, the renewal functional, as a mapping from $D[0, c]$ to $D[0, c]$, the spaces of cadlag functions on $[0, c]$, since $F(0-) = 0$ implies that the values of F on (c, ∞) are irrelevant to U on $[0, c]$. Again, a differentiability property such as Proposition 3.14 will give asymptotic normality via a Skorohod–Dudley construction. No moment conditions are required. Also, due to the monotonicity of the renewal functional (see subsection 3.4), it is possible in this simpler setting to obtain nontrivial confidence regions directly from the usual Kolmogorov–Smirnov confidence bands for the empirical distribution function by applying the renewal functional to the envelopes of these bands.

4.4. *Summary.* The overall strategy of this paper has been to consider the stochastic model as a nonlinear operator Ψ that has a differentiability property of the form

$$\sqrt{n}\,(F_n - F) \to f \quad \Rightarrow \quad \sqrt{n}\,(\Psi(F_n) - \Psi(F)) \to Tf,$$

where T is a linear operator. The topologies to which these limit statements refer depend on the model under consideration; conditions on F will imply that the left-hand side holds with probability 1 if F_n is the empirical distribution function associated with a sample of size n from F. This property also serves as the basis for a proof that "the bootstrap works," thereby solving an important practical problem which arises if functional central limit theorems are to be used to obtain confidence regions.

This strategy can be applied to other stochastic models, its success depending on the analytic tractability of the operator Ψ.

REFERENCES

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

DALEY, D. J. (1980). Tight bounds for the renewal function of a random walk. *Ann. Probab.* **8** 615–621.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

FREES, E. W. (1986a). Warranty analysis and renewal function estimation. *Naval Res. Logist. Quart.* **33** 361–372.

FREES, E. W. (1986b). Nonparametric renewal function estimation. *Ann. Statist.* **14** 1366–1378.

FREES, E. W. (1988). Estimating the cost of a warranty. *Journal of Business and Economic Statistics* **6** 79–86.

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part I). *Scand. J. Statist.* **16** 97–128.

GRÜBEL, R. (1988). The length of the shorth. *Ann. Statist.* **16** 619–628.

GRÜBEL, R. (1989a). Stochastic models as functionals: some remarks on the renewal case. *J. Appl. Probab.* **26** 296–303.

GRÜBEL, R. (1989b). The Fast Fourier Transform algorithm in applied probability theory. *Nieuw Arch. Wisk.* **7** 289–300.

GUT, A. (1988). *Stopped Random Walks, Limit Theorems and Applications*. Springer, New York.

O'REILLY, N. E. (1974). On the weak convergence of empirical processes in sup-norm metrics. *Ann. Probab.* **2** 642–651.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

SCHNEIDER, H., LIN, B.-S. and O'CINNEIDE, C. (1990). Comparison of nonparametric estimators for the renewal function. *J. Roy. Statist. Soc. Ser. C* **39** 55–61.

SHORACK, G. R. (1982). Bootstrapping robust regression. *Comm. Statist. Theory Methods* **11** 961–972.

SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.

TSIRELSON, V. S. (1975). The density of the distribution of the maximum of a Gaussian process. *Theory Probab. Appl.* **20** 847–856.

DEPARTMENT OF MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS

DEPARTMENT OF STATISTICAL SCIENCE
UNIVERSITY COLLEGE LONDON
GOWER STREET
LONDON WC1E 6BT
GREAT BRITAIN