Partial Sum Process for Records

ARUP BOSE
Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India
E-mail: abose@isical.ac.in

SREELA GANGOPADHYAY
Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India
E-mail: res9616@isical.ac.in

ANISH SARKAR
Theoretical Statistics and Mathematics Unit, Indian Statistical Institute (Delhi Centre), 7 S.J.S.S. Marg, New Delhi 110016, India
E-mail: anish@isid.ac.in

Abstract. Suppose the upper records $\{X_{L_n}\}$ from a sequence of i.i.d. random variables is in the domain of attraction of a normal distribution. Consider the $D(0,1]$-valued process $\{Z_n(\cdot)\}$ constructed by usual interpolation of the partial sums of the records. We prove that under some mild conditions, $\{Z_n\}$ converges to a limiting Gaussian process in $D(0,1]$. As a consequence, the partial sums of records is asymptotically normal.

Key words. Records, domain of attraction, regularly varying function, slowly varying function, $D(0,1]$-valued process

AMS 2000 Subject Classification. Primary—60F17
Secondary—60G70

1. Introduction and main results

Let F be a continuous distribution. Let $\{X_1, X_2, \ldots\}$ be i.i.d. observations from F. Let

$$L_n = \inf\{n > L_{n-1} : X_n > X_{L_{n-1}}\}, \quad L_1 = 1.$$

Then $\{X_{L_n} : n \geq 1\}$ is called the sequence of (upper) records. Thus X_i is a record if $X_i > \max\{X_1, X_2, \ldots X_{i-1}\}$ and X_1 is taken to be a record by convention.

The study of records has been of much interest since the time of Gnedenko (1943). See Arnold et al. (1998) for an extensive bibliography and many interesting results. A crucial observation in the study of record values is the following: Let $\{Y_i\}$ be i.i.d. with exponential distribution having mean one and let ψ be defined as:

$$\psi(x) = F^{-1}(1 - \exp(-x)). \tag{1}$$

Then, the joint distribution of $\{X_{L_1}, X_{L_2}, \ldots, X_{L_n}\}$ is same as that of $\{\psi(Y_1), \psi(Y_1+Y_2,), \ldots \psi(\sum_{i=1}^{n} Y_i)\}$ [see, for example, Resnick (1964)]. So in particular, $\sum_{i=1}^{n} X_{L_i} = \sum_{i=1}^{n} \psi(\sum_{j=1}^{i} Y_j)$, in distribution.

Resnick (1964) has shown that the limiting distribution function, if it exists, of the properly centered and scaled sequence of records $(X_{L_n} - b_n)/a_n$ (where a_n and b_n are suitable sequences of constants) must be one of the following three distributions. Define $N(x)$ to be the standard normal distribution function. Then

- Case (i) The limiting distribution is $N(x)$. In this case, it turns out that

$$b_n = F^{-1}(1 - \exp(-n)) = \psi(n).$$

and

$$a_n = \psi(n + \sqrt{n}) - \psi(n).$$

- Case (ii) The limiting distribution is

$$N_{1\alpha}(x) = \begin{cases} 0 & x < 0, \\ N(\log x^{\alpha}) & x \geq 0. \end{cases}$$

- Case (iii) The limiting distribution is

$$N_{2\alpha}(x) = \begin{cases} N(\log(-x)^{-\alpha}) & x < 0, \\ 1 & x \geq 0. \end{cases}$$

On the other hand, Arnold and Villasenor (1999) obtained the following results on the asymptotic normality of partial sums of records:

- If $\psi(x) = x$,

$$\frac{\sum_{i=1}^{n} \psi\left(\sum_{j=1}^{i} Y_j\right) - n^2/2}{\sqrt{n^3/3}} \Rightarrow N(0,1).$$

- If $\psi(x) = \log x$,

$$\frac{\sum_{i=1}^{n} \psi\left(\sum_{j=1}^{i} Y_j\right) - (n+1)\log n + n}{\sqrt{2n}} \Rightarrow N(0,1).$$

They conjectured that these hold for a wider class of ψ functions. Bose et al. (2003) extended the above results to a class of functions ψ satisfying some technical growth conditions.

It may be noted that there is a lot of similarity between the limiting behavior of records and the sequence of maxima. Resnick (1987) made a thorough investigation of record values and maxima. Define $M_n = \max\{X_i : 1 \leq i \leq n\}$ for $n \geq 1$. Resnick proved that there exists a continuous time extremal process $\{Y(t) : t > 0\}$ so that $\{M_n : n \geq 1\} \overset{\mathcal{D}}{=} \{Y(n) : n \geq 1\}$. He studied $\{Y(t) : t > 0\}$ and his investigation throws light on the properties of $\{M_n : n \geq 1\}$. In particular, he showed that if the underlying distribution F is in the domain of attraction of an extreme-value distribution, then the sequence of maxima converges (in a stochastic process sense) to a limiting extremal process generated by the same extreme-value distribution. He uses point process based methods to obtain his results.

Our purpose in this article is to establish a functional limit theorem for the process obtained by the partial sums of records. Hence, considering case (i), for $t \in (0,1]$, define

$$Z_n(t) = \frac{\psi\left(\sum_{j=1}^{[nt]} Y_j\right) - \psi([nt])}{a_n} \tag{2}$$

where $a_n = \psi(n + \sqrt{n}) - \psi(n)$.

Point process based techniques do not seem to apply in case of records. Nevertheless, if case (i) holds, then in Theorem 1 we show that the process Z_n converges under mild restrictions on ψ.

As a consequence of the above result, we show that under minor restrictions on ψ, $R_n := \sum_{j=1}^{n}\left[\psi(\sum_{i=1}^{j} Y_i) - b_j\right]/(na_n)$ is asymptotically normal. The results of Arnold and Villasenor (1999), mentioned above, then follow as special cases.

Let us now look at cases (ii) and (iii). By Resnick (1964) case (ii) holds if and only if $\psi^{-1}(x) = (\frac{\alpha}{2}\log x + \log L(x))^2$ where $L(x)$ is a slowly varying function. Moreover, $b_n = 0$ and $a_n = \psi(n)$. So in this case, $\psi(x) \sim C\exp(2\sqrt{x}/\alpha)$. Now, $\frac{X_{L_n}}{\psi(n)} = \frac{\psi(\sum_{i=1}^{n} Y_i)}{\psi(n)} \Rightarrow W$ where $\alpha \log W \sim N(0,1)$. But $\frac{\psi\left(\sum_{i=1}^{[nt]} Y_i\right)}{\psi(n)} = \frac{\psi([nt])}{\psi(n)}\frac{\psi\left(\sum_{i=1}^{[nt]} Y_i\right)}{\psi([nt])} \Rightarrow \left(\lim_{n\to\infty}\frac{\psi([nt])}{\psi(n)}\right)W$ and $\frac{\psi([nt])}{\psi(n)} \to 0$ for all $t < 1$. So in this case $Z_n(\cdot)$; as defined cannot converge weakly to a valid process. Similarly, in case (iii), $\frac{a_{[nt]}}{a_n} \to 0$ for all $t < 1$ where a_n denotes the appropriate scaling (Resnick, 1964) and no nontrivial limit is possible.

So suppose that we are in case (i), that is, the records are in the domain of attraction of normal distribution and let Z_n be as defined in (2). Then $\{Z_n(t) : t \in (0, 1]\}$ is a $D((0,1])$-valued process. Also let for all $x \geq 0$,

$$a(x) = \psi(x + \sqrt{x}) - \psi(x).$$

We shall write a_n for $a(n)$. Resnick (1964) proved that case (i) holds if and only if

$$\lim_{n\to\infty} \frac{\psi(n + \sqrt{n}x) - \psi(n)}{a_n} = x, \; x \in \mathbb{R} \tag{3}$$

This condition is used crucially in our proof.

Theorem 1: *Assume (3) and $a(\cdot)$ is regularly varying function with index $\beta \in \mathbb{R}$. Then $Z_n \Rightarrow Z$ in $D((0,1])$ as $n \to \infty$ where $Z(t) = t^{\beta-1/2}B(t)$ for $t \in (0,1]$ and B is a standard Brownian motion.*

Remark 1: If $\beta > 0$, then the above convergence actually holds on [0, 1]. If the parent distribution F is standard normal, then $\beta = 0$ [see Arnold et al. (1998), p. 19]. So in this case, $Z_n(t) \Rightarrow B(t)/\sqrt{t}$ on (0, 1].

Remark 2: It should be noted that there are examples of functions, ψ which satisfy the condition (3), but a is not a regularly varying function. One such example is given by

$$\psi(y) = \int_e^{\exp(2\sqrt{y})} \frac{\exp\left[(\log x)^{\beta}\right]dx}{x}$$

where $0 < \beta < 1$. In these cases, though the records will have an asymptotically normal distribution, the process version does not converge to any process.

As a consequence of Theorem 1, we obtain

Theorem 2: *Assume (3) and that $a(x) = \psi(x + \sqrt{x}) - \psi(x)$ is a regularly varying function with exponent $\beta > -1$. Then*

$$R_n = \frac{1}{na_n}\sum_{i=1}^{n}\left[\psi\left(\sum_{j=1}^{i} Y_j\right) - \psi(i)\right] \Rightarrow N(0, g(\beta))$$

where $g(\beta) := \frac{2}{(3+2\beta)(1+\beta)}$.

Remark 3: Routine but tedious calculations show that if we set $\psi(x) = P_1(x)\, P_2(\log x)$ where P_1 and P_2 are polynomials, then all the required conditions in the above Theorems are satisfied. This in particular yields the asymptotic normality results of Arnold and Villasenor (1999).

2. Proofs

In this section we shall give all the proofs.

Proof of Theorem 1: For any process $U(\cdot)$, let $U(\cdot)|_{[a,b]}$ denote its restriction to $[a, b]$. Suppose $\{U_n : n \geq 1\}$ and U are $D((0,\infty))$ valued processes. Then $U_n \Rightarrow U$ in $D((0,\infty))$ if and only if $U_n|_{[a,b]} \Rightarrow U|_{[a,b]}$ in $D([a, b])$, for all $0 < a < b < \infty$ such that $\mathbb{P}(U(a) = U(a-)$ and $U(b) = U(b-)) = 1$ [see Proposition 4.18 of Resnick (1987), page 205].

In our case, it is enough to consider the restriction of the processes to $[\epsilon, 1]$ for every $0 < \epsilon < 1$ and prove that $Z_n|_{[\epsilon,1]} \Rightarrow Z|_{[\epsilon,1]}$ in $D([\epsilon, 1])$. Here is the brief outline of our approach.

We fix $0 < \epsilon < 1$ for the rest of this proof. First we define processes $V_n^{(\epsilon)}$ and $V^{(\epsilon)}$ such that $V_n^{(\epsilon)} \Rightarrow V^{(\epsilon)}$ on $[\epsilon, 1]$. Next, we define functions $g_n, g : D([\epsilon, 1]) \to D([\epsilon, 1])$

such that $Z_n|_{[\epsilon,1]} = g_n(V_n^{(\epsilon)})$ and $Z|_{[\epsilon,1]} = g(V^{(\epsilon)})$. Further, we will show that the functions g_n and g and the limiting process $V^{(\epsilon)}$ satisfy the condition of Theorem 5.5 of Billingsley (1968), thereby enabling us to conclude that $Z_n|_{[\epsilon,1]} = g_n(V_n^{(\epsilon)}) \Rightarrow g(V^{(\epsilon)}) = Z|_{[\epsilon,1]}$. We now proceed to do this:

Let $S_n = \sum_{i=1}^n Y_i - n$. For $t \in [\epsilon, 1]$, define

$$V_n^{(\epsilon)}(t) := S_{[nt]}/\sqrt{[nt]}$$

$$V^{(\epsilon)}(t) := B|_{[\epsilon,1]}(t)t^{-1/2}$$

where $B|_{[\epsilon,1]}$ is the restriction of the standard Brownian motion on $[\epsilon, 1]$.

Next, we define functions $g_n, g : D([\epsilon, 1]) \to D([\epsilon, 1])$ as follows: for any $f \in D([\epsilon, 1])$,

$$g_n(f)(t) = \frac{\psi([nt] + \sqrt{[nt]}f(t)) - \psi([nt])}{\psi(n + \sqrt{n}) - \psi(n)}$$

and

$$g(f)(t) = t^{\beta} f(t).$$

It is clear that we can write

$$Z_n|_{[\epsilon,1]} = g_n(V_n^{(\epsilon)}) \quad \text{and} \quad Z|_{[\epsilon,1]} = g(V^{(\epsilon)}).$$

Now, $V_n^{(\epsilon)}(t) = \left[\sqrt{n}/\sqrt{[nt]}\right] \times \left[S_{[nt]}/\sqrt{n}\right]$. Since $S_{[nt]}/\sqrt{n} \Rightarrow B|_{[\epsilon,1]}$ as $n \to \infty$ and $\sqrt{n}/\sqrt{[nt]} \to t^{-1/2}$ uniformly in $t \in [\epsilon, 1]$, we have, by Exercise 1, Billingsley (1968), page 28,

$$V_n^{(\epsilon)} \Rightarrow V^{(\epsilon)} \quad \text{as} \quad n \to \infty.$$

Finally to verify the condition of Theorem 5.5 of Billingsley (1968), we define

$$E = \{ f \in D([\epsilon, 1]) : g_n(f_n) \to g(f) \text{ fails to hold for some sequence } \{ f_n : n \geq 1\} \subseteq D([\epsilon, 1]) \text{ such that } f_n \to f \text{ in } D([\epsilon, 1])\}.$$

To conclude our result, we need to show that,

$$\mathbb{P}(V^{(\epsilon)} \in E) = 0.$$

Since the paths of $V^{(\epsilon)} (= B|_{[\epsilon,1]}(t)t^{-1/2})$ are almost surely continuous, $V^{(\epsilon)}$ has support only on $C([\epsilon, 1])$. Therefore, it is enough to prove that $E \subseteq D([\epsilon, 1]) \setminus C([\epsilon, 1])$.

In other words, it is enough to show that—for any $f \in C([\epsilon, 1])$ and any sequence $\{ f_n : n \geq 1\} \subseteq D([\epsilon, 1])$ such that $f_n \to f$ in $D([\epsilon, 1])$,

$$g_n(f_n)(t) \to g(f)(t) \quad \text{uniformly in } t \in [\epsilon, 1].$$

To prove this, fix any $0 < \delta < 1$. For any $f \in C([\epsilon, 1])$ and $f_n \in D([\epsilon, 1])$, such that $f_n \to f$ in $D([\epsilon, 1])$, we must have $f_n(t) \to f(t)$ uniformly in $t \in [\epsilon, 1]$ [see Billingsley

(1968), page 112]. Therefore, there exists N_1 such that $\sup\{|f_n(t)-f(t)| : t\in[\epsilon,1]\} < \delta$ for all $n \geq N_1$. Thus, we have $\sup\{|f_n(t)| : t\in[\epsilon,1]\} \leq \delta + \sup\{f(t) : t\in[\epsilon,1]\} = M_1$ (say) for all $n \geq N_1$.

Let $M_2 := \sup\{t^\beta : t\in[\epsilon,1]\}$. Since $a_{[nt]}/a_n \to t^\beta$ uniformly in $t\in[\epsilon,1]$, we may choose N_2 so that, for all $n \geq N_2$, we have

$$\sup\left\{\left|\frac{a_{[nt]}}{a_n} - t^\beta\right| : t\in[\epsilon,1]\right\} < \delta$$

and

$$\sup\left\{\left|\frac{a_{[nt]}}{a_n}\right| : t\in[\epsilon,1]\right\} < \delta + M_2 =: M_3(\text{say}).$$

Further for any positive integer n, $[\psi(n+\sqrt{n}x)-\psi(n)]/a_n$ is non-decreasing in x, we must have that $[\psi(n+\sqrt{n}x)-\psi(n)]/a_n$ converges to x uniformly on compact sets. Therefore, we can choose N_3 so that for all $n \geq N_3$, we have

$$\sup\left\{\left|\frac{\psi(n+\sqrt{n}x)-\psi(n)}{a_n} - x\right| : |x| \leq M_1\right\} < \delta.$$

Now, for all $n \geq \max\{N_1, N_2, N_3/\epsilon\}$ and $t\in[\epsilon,1]$, we have

$$|g_n(f_n)(t) - g(f)(t)|$$

$$\leq \frac{a_{[nt]}}{a_n}\left|\frac{\psi([nt]+\sqrt{[nt]}f_n(t))-\psi([nt])}{a_{[nt]}} - f_n(t)\right| + |f_n(t)|\left|\frac{a_{[nt]}}{a_n} - t^\beta\right| + t^\beta|f_n(t)-f(t)|$$

$$\leq M_3 \sup\left\{\left|\frac{\psi([nt]+\sqrt{[nt]}x)-\psi([nt])}{a_{[nt]}} - x\right| : |x| \leq M_1\right\} + M_1\delta + M_2\delta$$

$$\leq (M_1+M_2+M_3)\delta.$$

Thus, $g_n(f_n)$ converges to $g(f)$ uniformly in $[\epsilon,1]$. This proves the result. □

To prove Theorem 2 we need the following three lemmas. Fix any $0 < \delta < 1$ and consider the integral function on $D([\delta,1])$, i.e., $I : D([\delta,1]) \to \mathbb{R}$ defined by,

$$I(f) := \int_\delta^1 f(u)du.$$

This is well defined for all $f \in D([\delta,1])$, since f is bounded and right continuous. The first lemma is a straightforward consequence of weak convergence in Skorokhod topology.

Lemma 1: *The mapping I is continuous at every $f \in C([\delta,1])$ under Skorokhod topology.*

Proof: Suppose that $f_n \in D([\delta, 1])$ and $f \in C([\delta, 1])$ such that $f_n \to f$ in $D([\delta, 1])$. Now, we know that in such a case $f_n(u) \to f(u)$ uniformly in $u \in [\delta, 1]$ [see Billingsley (1968), page 112]. Therefore DCT applies to show that $I(f_n) \to I(f)$. □

Next we require two lemmas on regularly varying functions. The following lemma is essentially part (ii) of Lemma 0.8 of Resnick (1987), page 22. We state it here in the form we require.

Lemma 2: *If u is a function taking strictly positive values and is regularly varying with index β, then for any $\alpha > 0$, there exists $N_0 \geq 1$ such that*

$$(i) \quad \frac{u(x)}{u(y)} \leq C_1 \left(\frac{x}{y}\right)^{\beta+\alpha} \quad \text{for all } x \geq y \geq N_0$$
$$(ii) \quad \frac{u(x)}{u(y)} \leq C_2 \left(\frac{x}{y}\right)^{\beta-\alpha} \quad \text{for all } y \geq x \geq N_0$$

where C_1, $C_2 > 0$ are constants, depending on α.

Next, we derive a bound on the behavior of ψ using the regular variation of $a(x) = \psi(x + \sqrt{x}) - \psi(x)$. This result may be known in the literature on regularly varying functions. We include a proof since we could not find it in the literature.

Lemma 3: *Let $a(x) = \psi(x + \sqrt{x}) - \psi(x)$ be regularly varying with index β. Then for any $0 < \alpha < 1/2$, there exists $N_0 > 0$ and $C_3 > 0$, such that for all $n \geq N_0$ and for $|x| \leq n^\alpha$,*

$$\left| \frac{\psi(n + x\sqrt{n}) - \psi(n)}{\psi(n + \sqrt{n}) - \psi(n)} \right| \leq C_3(1 + |x|).$$

We postpone the proof of Lemma 3 for the time being and prove Theorem 2 assuming the above Lemmas. The main idea is to consider the partial sum as the integral of the process $\{Z_n\}$ and use Theorem 1 and continuity of the integral function. However, this cannot be done directly as the convergence in Theorem 1 is only on $(0, 1]$. Therefore, we need to consider the terms near 0 separately. The Lemmas 2 and 3 will be used to estimate those terms, while Lemma 1 along with Theorem 1, will be applied on the remaining part to give us the limiting random variable.

Proof of Theorem 2: Let N_0 be a fixed positive integer. The exact choice of N_0 will be specified shortly. Also, fix $\delta > 0$. Now, we can split R_n into three sums as follows:

$$R_n = \frac{1}{na(n)} \sum_{i=1}^{N_0-1} \left[\psi\left(\sum_{j=1}^{i} Y_j\right) - \psi(i)\right] + \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} \left[\psi\left(\sum_{j=1}^{i} Y_j\right) - \psi(i)\right]$$
$$+ \frac{1}{na(n)} \sum_{i=[n\delta]+1}^{n} \left[\psi\left(\sum_{j=1}^{i} Y_j\right) - \psi(i)\right] =: R^{(1)}(n) + R^{(2)}(n) + R_\delta(n).$$

The third term $R_\delta(n)$: This is the main term. It is easy to see that $R_\delta(n)$ can be written as $R_\delta(n) = \int_\delta^1 Z_n(t)dt$. Using, the notations used in the proof of Theorem 1, we can write $\int_\delta^1 Z_n(t)dt = I(Z_n|_{[\delta,1]})$ where I is the integral function on $D([\delta, 1])$, defined in Lemma 1 and $Z_n|_{[\delta,1]}$ is the restriction of Z_n to $[\delta, 1]$. By Theorem 1, we have $Z_n \Rightarrow Z$ on $(0, 1]$ where $Z(t) = t^{\beta-1/2}B(t)$ for $t > 0$. Using the converse part of Proposition 4.18 of Resnick (1987), page 205, we have, $Z_n|_{[\delta,1]} \Rightarrow Z|_{[\delta,1]}$, for every $\delta > 0$. By Lemma 1, the discontinuity points of I is a subset of $D([\delta,1])\backslash C([\delta,1])$. Since the paths of $Z|_{[\delta,1]}(= t^{\beta-1/2}B(t))$ are almost surely continuous, the measure of the set of discontinuities of I, under $Z|_{[\delta,1]}$ is zero. Therefore, using Theorem 5.1 of Billingsley (1968), we conclude that $R_\delta(n) = I(Z_n|_{[\delta,1]}) \Rightarrow I(Z|_{[\delta,1]}) =: R_\delta$ (say).

Before we consider the two remaining terms, we briefly explore the behaviour of R_δ. Since $R_\delta = I(Z|_{[\delta,1]}) = \int_\delta^1 t^{\beta-1/2}B(t)dt$, R_δ follows a normal distribution with mean 0 and variance given by $g_\delta(\beta)$ where

$$g_\delta(\beta) := \begin{cases} \dfrac{4}{3+2\beta}\left[\dfrac{1-\delta^{2+2\beta}}{2+2\beta} - \dfrac{2\delta^{3/2+\beta}(1-\delta^{1/2+\beta})}{1+2\beta}\right] & \text{if } \beta \neq -1/2 \\ 2(1-\delta) + 2\delta \log \delta & \text{if } \beta = -1/2. \end{cases}$$

Clearly, for $\beta > -1$, $g_\delta(\beta) \to g(\beta)$ as $\delta \to 0$. Therefore, we have that $R_\delta \Rightarrow R$ where R follows a normal distribution with mean 0 and variance $g(\beta)$.

To consider the terms $R^{(1)}(n)$ and $R^{(2)}(n)$, we have to make a formal choice of N_0. In order to do so, we first choose $\gamma > 0$ so that $\beta - \gamma > -1$. Applying Lemma 2, we choose N_1 such that,

$$\frac{a(x)}{a(y)} \leq C_4 \left(\frac{x}{y}\right)^{\beta-\gamma} \tag{4}$$

for all $y \geq x \geq N_1$.

Now, fix any $0 < \alpha < 1/2$. Applying Lemma 3, we choose N_2 so that

$$\left|\frac{\psi(n + x\sqrt{n}) - \psi(n)}{\psi(n + \sqrt{n}) - \psi(n)}\right| \leq C_3(1 + |x|) \quad \text{for } |x| \leq n^\alpha$$

for all $n \geq N_2$.

Further, set $P_n = \sum_{j=1}^n Y_j$ and fix any $\epsilon > 0$. Define the sequence of events $A_n = \{|P_n - n| \leq n^{1/2+\alpha}\}$. We want to show that all but finitely many of A_n's must occur. In order to show that, choose K so large that $K\alpha > 1$. Now, by Markov inequality, we have $\mathbb{P}(A_n^c) \leq \mathbb{E}(|(P_n - n)/\sqrt{n}|^K)/n^{K\alpha} \leq C_5/n^{K\alpha}$ where $C_5 = \sup\{\mathbb{E}(|(P_n - n)/\sqrt{n}|^K) : n \geq 1\} < \infty$. Hence, $\mathbb{P}(\limsup_{n\to\infty} A_n^c) = 0$, i.e., $\mathbb{P}(\liminf_{n\to\infty} A_n) = 1$. So, choose N_3 so large that $\mathbb{P}(B(N_3)) > 1 - \epsilon$ where $B(N_3) := \cap_{n \geq N_3} A_n$.

The first term $R^{(1)}(n)$: Let $N_0 = \max\{N_1, N_2, N_3\}$. Since the function $x \to a(x) = \psi(x + \sqrt{x}) - \psi(x)$ is regularly varying with index $\beta > -1$, we get that the function $x \to xa(x)$ is also regularly varying with index $1 + \beta > 0$. Hence, $na(n) \to \infty$ as $n \to \infty$. Since $R^{(1)}(n)$ is the ratio of a sum comprising of finitely many fixed terms and $na(n)$, we have that $R^{(1)}(n)$ converges to 0 in probability.

The second term $R^{(2)}(n)$: For this term, we have to work a little bit more. We will estimate the probability that it is bounded away from 0. For $n \geq N_0$, we have

$$\begin{aligned}
&\mathbb{P}\Big(| R^{(2)}(n) | \geq \epsilon \Big) \\
&= \mathbb{P}\Big(B(N_3) \cap \{ |R^{(2)}(n)| \geq \epsilon \} \Big) + \mathbb{P}((B(N_3))^c) \\
&\leq \mathbb{P}\left(1_{B(N_3)} \left| \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} [\psi(P_i) - \psi(i)] \right| \geq \epsilon \right) + \epsilon \\
&\leq \frac{\mathbb{E}\left(1_{B(N_3)} \left| \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} [\psi(P_i) - \psi(i)] \right| \right)}{\epsilon} + \epsilon.
\end{aligned} \tag{5}$$

Now, the above expectation is estimated in the following way:

$$\begin{aligned}
&\mathbb{E}\left(1_{B(N_3)} \left| \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} [\psi(P_i) - \psi(i)] \right| \right) \\
&\leq \mathbb{E}\left(\left| \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} 1_{B(N_3)} [\psi(P_i) - \psi(i)] \right| \right) \\
&\leq \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} \mathbb{E}(1_{B(N_3)} |\psi(P_i) - \psi(i)|) \\
&\leq \frac{1}{na(n)} \sum_{i=N_0}^{[n\delta]} \mathbb{E}(1_{A_i} |\psi(P_i) - \psi(i)|)
\end{aligned} \tag{6}$$

as $B(N_3) \subseteq A_n$ for all $n \geq N_0$.

For the expectation inside the summation in equation (6), we will use the bound given by Lemma 3. We have

$$\begin{aligned}
\mathbb{E}|1_{A_n}(\psi(P_n) - \psi(n))| &= \int_{-n^\alpha}^{n^\alpha} f_n(x) \left| \psi(n + x\sqrt{n}) - \psi(n) \right| dx \\
&\leq C_3 a(n) \int_{-n^\alpha}^{n^\alpha} f_n(x)(1 + |x|)dx \\
&\leq C_3 a(n) \int_{-\infty}^{\infty} f_n(x)(1 + |x|)dx \\
&\leq C_6 a(n)
\end{aligned} \tag{7}$$

where f_n is the density function of the random variable $(P_n - n)/\sqrt{n}$ and $C_6 = C_3 \sup\{1 + \mathbb{E}(|(P_n - n)/\sqrt{n}|) : n \geq 1\} < \infty$.

Now, putting together (5), (6) and (7), we have

$$\begin{aligned}\mathbb{P}\left(|R^{(2)}(n)| \geq \epsilon\right) &\leq \frac{C_6}{\epsilon n a(n)} \sum_{i=N_0}^{\lfloor n\delta \rfloor} a(i) + \epsilon \\ &\leq \frac{C_6 C_4}{\epsilon n} \sum_{i=N_0}^{\lfloor n\delta \rfloor} \left(\frac{i}{n}\right)^{\beta-\gamma} + \epsilon \qquad \text{using equation (4)} \\ &\to \frac{C_6 C_4}{\epsilon} \int_0^\delta y^{\beta-\gamma} dy + \epsilon \qquad \text{as } n \to \infty \\ &= \frac{C_6 C_4}{\epsilon(1+\beta-\gamma)} \delta^{1+\beta-\gamma} + \epsilon.\end{aligned}$$

We are now in a position to tackle R_n completely: Note that for $x \in \mathbb{R}$ and $\delta > 0$, we have

$$\mathbb{P}(R_n \leq x) \leq \mathbb{P}(|R^{(1)}(n)| \geq \epsilon) + \mathbb{P}(|R^{(2)}(n)| \geq \epsilon) + \mathbb{P}(R_\delta(n) \leq x + 2\epsilon).$$

Letting $n \to \infty$, we have

$$\limsup_{n\to\infty} \mathbb{P}(R_n \leq x) \leq \epsilon + \frac{C_6 C_4}{\epsilon(1+\beta-\gamma)} \delta^{1+\beta-\gamma} + \mathbb{P}(R_\delta \leq x + 2\epsilon).$$

Letting $\delta \to 0$, we have

$$\limsup_{n\to\infty} \mathbb{P}(R_n \leq x) \leq \epsilon + \mathbb{P}(R \leq x + 2\epsilon). \tag{8}$$

Conversely, for $x \in \mathbb{R}$ and $\delta > 0$, we have

$$\mathbb{P}(R_\delta(n) \leq x - 2\epsilon) \leq \mathbb{P}(|R^{(1)}(n)| \geq \epsilon) + \mathbb{P}(|R^{(2)}(n)| \geq \epsilon) + \mathbb{P}(R_n \leq x).$$

Letting $n \to \infty$, we have

$$\mathbb{P}(R_\delta \leq x - 2\epsilon) \leq \epsilon + \frac{C_6 C_4}{\epsilon(1+\beta-\gamma)} \delta^{1+\beta-\gamma} + \liminf_{n\to\infty} \mathbb{P}(R_n \leq x).$$

Letting $\delta \to 0$, we have,

$$\mathbb{P}(R \leq x - 2\epsilon) \leq \epsilon + \liminf_{n\to\infty} \mathbb{P}(R_n \leq x). \tag{9}$$

Now, letting $\epsilon \to 0$ in (8) and (9), we conclude,

$$\lim_{n\to\infty} \mathbb{P}(R_n \leq x) = \mathbb{P}(R \leq x)$$

for all $x \in \mathbb{R}$. This proves the result. □

Finally we prove Lemma 3.

Proof of Lemma 3: We divide the range of x into three separate regions:

Region A: $-n^{\alpha} \le x < 0$, Region B: $0 \le x \le 1$ and Region C: $1 < x \le n^{\alpha}$. It is enough to obtain separate bounds for each of these three regions.

Region B: This is the easiest to handle. On the set $0 \le x \le 1$, we have that $|\psi(n+\sqrt{n}x) - \psi(n)| = \psi(n+\sqrt{n}x) - \psi(n) \le \psi(n+\sqrt{n}) - \psi(n)$ using the fact that ψ is non-decreasing. Therefore, we have 1 as the upper bound for this region.

Region C: In this region, the idea is to write $\psi(n+\sqrt{n}x) - \psi(n)$ as a telescoping sum of terms of the form $\psi(u+\sqrt{u}) - \psi(u)$ and then estimate each of these terms using Lemma 2.

Let N_0 be a fixed positive integer. We will specify it shortly. Fix $n \ge N_0$ and $x > 1$ and define a sequence as follows: $r(1; n) = n$ and for $j \ge 1$,

$$r(j+1;n) = r(j;n) + \sqrt{r(j;n)}.$$

Define further,

$$R(x;n) = \min\{j : r(j;n) \ge n + x\sqrt{n}\}.$$

We have

$$\begin{aligned}
&|\psi(n+x\sqrt{n}) - \psi(n)| \\
&\quad = \psi(n+x\sqrt{n}) - \psi(n) \\
&\quad \le \psi(r(R(x;n);n)) - \psi(n) \\
&\quad = \psi(r(R(x;n);n)) - \psi(r(1;n)) \\
&\quad = \sum_{j=1}^{R(x;n)-1} \psi(r(j+1;n)) - \psi(r(j;n)) \\
&\quad = \sum_{j=1}^{R(x;n)-1} \psi(r(j;n) + \sqrt{r(j;n)}) - \psi(r(j;n)).
\end{aligned} \tag{10}$$

Next, we estimate the individual terms as well as the number of terms in the above summation in equation (10). To do this, we first choose a constant N_1, using Lemma 2, so large that

$$\frac{a(x)}{a(y)} \le C_7 \left(\frac{x}{y}\right)^{\beta+1} \quad \text{for all } x \ge y \ge N_1 \tag{11}$$

and N_2, so that $n^{\alpha-1/2} \le 1/2$ for all $n \ge N_2$. Fix $n \ge N_0 := \max\{N_1, N_2\}$.

Estimate of $R(x;\, n)$: Note that $r(j+1;n) \ge r(j;n)$ for all $j \ge 1$ and $r(j+1;n) - r(j;n) = \sqrt{r(j;n)} \ge \sqrt{r(1;n)} = \sqrt{n}$. Therefore, $r(j+1;n) \ge n + j\sqrt{n}$ for all $j \ge 1$. Hence, $R(x;\, n)$ must be finite and, in fact,

$$R(x;n) \le (2 + [x]) \le (2 + x) \tag{12}$$

where $[u]$ is the largest integer smaller or equal to u.

Estimation of the summands in (10): Since each of these terms is non-negative, we can estimate them separately. Using the estimate in (11) and the fact that $r(j;n) \geq n \geq N_1$, we have

$$\frac{\psi(r(j;n)+\sqrt{r(j;n)})-\psi(r(j;n))}{\psi(n+\sqrt{n})-\psi(n)} = \frac{a(r(j;n))}{a(n)} \leq C_7\left[\frac{r(j;n)}{n}\right]^{\beta+1}. \tag{13}$$

If $\beta+1<0$, each of these terms is bounded by 1. In case $\beta+1\geq 0$, we note that for all $j=1,2,\ldots,R(x;n)-1$, we must have $r(j;n)\leq n+x\sqrt{n}$. Thus, for all $n\geq N_2$, $r(j;n)/n \leq 1+x/\sqrt{n} \leq 1+n^{\alpha-1/2} \leq 3/2$. So, each of the terms are bounded by $(3/2)^{\beta+1}$. Therefore, using (12) and the above bound, we have that

$$\left|\frac{\psi(n+x\sqrt{n})-\psi(n)}{\psi(n+\sqrt{n})-\psi(n)}\right| \leq \sum_{j=1}^{R(x;n)-1}\frac{a(r(j;n))}{a(n)} \leq C_8(1+x)$$

where $C_8=\max\{C_7, C_7(3/2)^{1+\beta}\}$.

Region A: For this region, we employ a similar method, i.e., write down $|\psi(n+x\sqrt{n})-\psi(n)|$ as a telescoping sum and then estimate the number of terms in the summation and each of these terms in the summation.

Again assume that N_0' be a fixed positive integer, to be specified later. Let $n\geq N_0'$ and $-n^\alpha \leq x < 0$. Define, $s(1;n)=n$ and for $j\geq 1$

$$s(j+1;n)=\frac{2s(j;n)+1-\sqrt{4s(j;n)+1}}{2}.$$

Note that, by definition of $s(j+1;n)$, we have

$$s(j+1;n)+\sqrt{s(j+1;n)}=s(j;n)$$

for all $j\geq 0$. Define,

$$S(x;n)=\min\{j: s(j;n)\leq n+x\sqrt{n}\}.$$

As earlier, using the fact that ψ is non-decreasing, we can write,

$$\begin{aligned}
&|\psi(n+x\sqrt{n})-\psi(n)| \\
&\quad = \psi(n)-\psi(n+x\sqrt{n}) \\
&\quad \leq \psi(n)-\psi(s(S(x;n);n)) \\
&\quad = \psi(s(1;n))-\psi(s(S(x;n);n)) \\
&\quad = \sum_{j=1}^{S(x;n)-1}\psi(s(j;n))-\psi(s(j+1;n)) \\
&\quad = \sum_{j=2}^{S(x;n)}\psi(s(j;n)+\sqrt{s(j;n)})-\psi(s(j;n)).
\end{aligned} \tag{14}$$

As earlier, we have to estimate the number of terms in the above summation and each of these terms.

Fix, using Lemma 2, $N_3 \geq 1$ so that

$$\frac{a(x)}{a(y)} \leq C_9 \left(\frac{x}{y}\right)^{\beta-1} \quad \text{for all } y \geq x \geq N_3. \tag{15}$$

Next note that $s(j+1;n) - s(j;n) = (1 - \sqrt{4s(j;n)+1})/2$ for all $j \geq 1$. Choose N_4 so large that $-3/2 \leq (1 - \sqrt{4n+1})/(2\sqrt{n}) \leq -1/2$ for all $n \geq N_4$. Further, choose N_0' so large that $n - n^{1/2+\alpha} \geq n/2$ and $n - (n^\alpha + 3/2)n^{1/2} \geq \max\{N_3, N_4\}$ and $(n^\alpha + 3/2)/\sqrt{n} \leq 1/2$ for all $n \geq N_0'$.

Estimate of $S(x;n)$: Fix $n \geq N_0'$ For $j < S(x;n)$, we must have $s(j;n) \geq n + x\sqrt{n} \geq n - n^{\alpha+1/2} \geq N_4$. Therefore, by choice of N_4, we get

$$-\frac{3}{2} \leq \frac{s(j+1;n) - s(j;n)}{\sqrt{s(j;n)}} = \frac{1 - \sqrt{4s(j;n)+1}}{2\sqrt{s(j;n)}} \leq -\frac{1}{2}.$$

Hence, we have

$$\frac{3\sqrt{s(j;n)}}{2} \leq s(j+1;n) - s(j;n) \leq -\frac{\sqrt{s(j;n)}}{2}. \tag{16}$$

Further, for all $j < S(x;n)$, we have $s(j;n) \geq n - n^{\alpha+1/2} \geq n/2$. So, from (16), $s(j+1;n) - s(j;n) \leq -\sqrt{n}/(2\sqrt{2})$. Hence, $s(j+1;n) \leq s(1;n) - j\sqrt{n}/(2\sqrt{2}) = n - j\sqrt{n}/(2\sqrt{2})$. Using this, it is easy to see that $s(2 + [2\sqrt{2}|x|], n) \leq n + x\sqrt{n}$. Thus, we have the bound,

$$S(x;n) \leq 2 + [2\sqrt{2}|x|] \leq 2 + 2\sqrt{2}|x| \tag{17}$$

where $[u]$ is the largest integer smaller or equal to u.

Estimation of the summands in (14): Since each of the sumand is non-negative, so we estimate $[\psi(s(j;n) + \sqrt{s(j;n)}) - \psi(s(j;n))]/[\psi(n+\sqrt{n}) - \psi(n)]$ for $j=2,\ldots, S(x;n)$. Now, for terms $j = 2, 3, \ldots, S(x;n) - 1$, we have $s(j;n) \geq n + \sqrt{n}x \geq n - n^{\alpha+1/2} \geq N_3$, by our choice. So, we can apply equation (15) for each of these terms. For the term, $j = S(x;n)$, we note that

$$\begin{aligned} s(S(x;n);n) - s(S(x;n)-1;n) &= \frac{1 - \sqrt{4s(S(x;n)-1;n)+1}}{2} \\ &\geq -\frac{3\sqrt{s(S(x;n)-1;n)}}{2} \qquad \text{using (16)} \\ &\geq -\frac{3\sqrt{n}}{2}. \end{aligned}$$

So, we have $s(S(x;n);n) \geq s(S(x;n)-1;n) - 3\sqrt{n}/2 \geq n + (x - 3/2)\sqrt{n}$, since by definition of $S(x;n)$, $s(S(x;n)-1;n) \geq n + x\sqrt{n}$. Therefore, we obtain that for all

$j \leq S(x;n) s(j;n) \geq n + (x - 3/2)\sqrt{n} \geq n - (n^\alpha + 3/2)\sqrt{n} \geq N_3$ Thus, for $j = 2, 3, \ldots,$ $S(x;n)$, we get

$$\frac{\psi(s(j;n) + \sqrt{s(j;n)}) - \psi(s(j;n))}{[\psi(n + \sqrt{n}) - \psi(n)]} = \frac{a(s(j;n))}{a(n)} \leq C_9 \left(\frac{s(j;n)}{n} \right)^{\beta - 1}.$$

If $\beta - 1 \geq 0$, each of the terms are bounded by 1. If $\beta - 1 < 0$, $s(j;n)/n \geq 1 + (x - 3/2)/\sqrt{n} \geq 1 - (n^\alpha + 3/2)/\sqrt{n} \geq 1/2$. Thus, $(s(j;n)/n)^{\beta-1} \leq 2^{-\beta+1}$. Therefore, using (17) and the above bound, we have that

$$\left| \frac{\psi(n + x\sqrt{n}) - \psi(n)}{\psi(n + \sqrt{n}) - \psi(n)} \right| \leq \sum_{j=2}^{S(x;n)} \frac{a(r(j;n))}{a(n)} \leq C_{10}(1 + 2\sqrt{2}|x|) \leq C_{11}(1 + |x|)$$

where $C_{10} = \max\{C_9, C_9 2^{-\beta+1}\}$ and $C_{11} = C_{10} 2\sqrt{2}$. This completes the proof of the Lemma. □

Acknowledgment

We wish to thank the referee for his careful reading and for his detailed comments. It has significantly improved our presentation.

References

Arnold, B.C. and Villasenor, J.A., "The asymptotic distributions of sums of records," *Extremes* 1(3), 351–363, (1999).

Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N., *Records*, Wiley, New York, 1998.

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

Bose, A., Gangopadhyay, S., Sarkar, A., and Sengupta, A., "Asymptotic properties of sums of upper records," *Extremes* 6(2), 147–164, (2003).

Gnedenko, B., "Sur le distribution limite du terme maximum d'une serie aleatoire," *Ann. Math.* IInd Series 44, 423–453, (1943).

Resnick, S.I., Limit laws for record values, *Stoch. Process. their Appl.* 1, 67–82, (1964).

Resnick, S.I., *Extreme Values, Regular Variation, and Point Processes*, Springer-Verlag, New York, 1987.