*Ergod. Th. & Dynam. Sys.* (2021), **41**, 461–470 doi:10.1017/etds.2019.64
© *Cambridge University Press, 2019*

# Continued fractions, the Chen–Stein method and extreme value theory

ANISH GHOSH†, MAXIM SØLUND KIRSEBOM‡ and PARTHANIL ROY§

† *School of Mathematics, Tata Institute of Fundamental Research, Mumbai 400005, India*
*(e-mail: ghosh.anish@gmail.com)*

‡ *Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany*
*(e-mail: maximkirsebom@gmail.com)*

§ *Theoretical Statistics and Mathematics Unit, Indian Statistical Institute, Bangalore 560059, India*
*(e-mail: parthanil.roy@gmail.com)*

(*Received* 17 *July* 2019 *and accepted in revised form* 4 *August* 2019)

*Abstract*. In this work we deal with extreme value theory in the context of continued fractions using techniques from probability theory, ergodic theory and real analysis. We give an upper bound for the rate of convergence in the Doeblin–Iosifescu asymptotics for the exceedances of digits obtained from the regular continued fraction expansion of a number chosen randomly from (0, 1) according to the Gauss measure. As a consequence, we significantly improve the best known upper bound on the rate of convergence of the maxima in this case. We observe that the asymptotics of order statistics and the extremal point process can also be investigated using our methods.

Key words: continued fractions, Chen–Stein method, Poisson approximation, rate of convergence, extreme value theory
2010 Mathematics Subject Classification: 60G70 (Primary); 11K50 (Secondary)

## 1. *Introduction*

This short paper establishes an upper bound for the Doeblin–Iosifescu asymptotics for exceedances (defined below) arising from the Gauss dynamical system. We briefly recall the basic facts about continued fraction expansions and the Gauss map. The reader is referred to the classic text [**12**] for more details. Let $X = (0, 1)$ and, for all $x \in X$, let $[A_1(x), A_2(x), \ldots]$ denote the regular continued fraction expansion. Define a transformation $T : X \to X$ by

$$T(x) = \{1/x\}, \tag{1.1}$$

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

where $\{\cdot\}$ denotes the fractional part. With the notation above, for all $x \in X \setminus \mathbb{Q}$,

$$A_1(x) = 1/x - T(x),$$
$$A_{j+1}(x) = A_j(T(x)) = A_1(T^j(x)) \quad \text{for all } j \in \mathbb{N}.$$

It is easy to check that $T$ defines a non-singular transformation on $(X, \lambda)$, where $\lambda$ denotes the Lebesgue measure. This means that, for all measurable $B \subseteq X$, we have $\lambda \circ T^{-1}(B) = 0$ if and only if $\lambda(B) = 0$.

Let $\widehat{T} : L^1(X, \lambda) \to L^1(X, \lambda)$ denote the dual operator (see, for example, [**1**, p. 33]) corresponding to $T$ that satisfies

$$\int_X \widehat{T}(f) g \, d\lambda = \int_X f(g \circ T) \, d\lambda$$

for all $f \in L^1(X, \lambda)$ and for all $g \in L^\infty(X, \lambda)$. It is easy to extend the domain of definition of $\widehat{T}$ to all non-negative measurable functions. Solving the functional equation $\widehat{T}(h) = h$, we get $h(x) = (1 + x)^{-1} \in L^\infty(X, \lambda)$. Hence, by [**1**, Proposition 1.4.1], the probability measure

$$P(dx) = ((1 + x) \log 2)^{-1} \, dx$$

on $X$ is $T$-invariant, making $T$ a positive transformation (see, for example, [**1**]). The measure $P$ is known as the Gauss measure.

From now on, we shall think of $\{A_n\}_{n \geq 1}$ as a sequence of random variables $A_n : X \to \mathbb{N}$ defined on the probability space $(X, P)$. The $T$-invariance of $P$ makes this a stationary sequence, that is, for all $k, l \in \mathbb{N}$, for all $m_1, m_2, \ldots, m_k \in \mathbb{N}$ and for all Borel subsets $B \subseteq X^k$,

$$P((A_{m_1}, A_{m_2}, \ldots, A_{m_k}) \in B) = P((A_{m_1+l}, A_{m_2+l}, \ldots, A_{m_k+l}) \in B).$$

We are interested in the extreme value theory for this stationary stochastic process. To the best of our knowledge, the first work in this direction was carried out by [**7**], who, among many other results, rightly observed that exceedances have Poissonian asymptotics: for all $u > 0$,

$$\mathcal{E}_n^u := \#\{1 \leq i \leq n : \ A_i \log 2 > nu\} \xrightarrow{d} \mathcal{E}_*^u \sim \text{Poi}(u^{-1}) \tag{1.2}$$

under $P$. Here $\xrightarrow{d}$ denotes convergence in distribution and the notation $\mathcal{E}_*^u \sim \text{Poi}(u^{-1})$ means that

$$P(\mathcal{E}_*^u = k) = \frac{u^{-k} e^{-u^{-1}}}{k!}, \quad k = 0, 1, 2, \ldots.$$

However, Doeblin's proof of (1.2) had a subtle error, which was corrected much later in [**10**, Theorem 2]. Therefore, we shall refer to (1.2) as the Doeblin–Iosifescu asymptotics; it forms the background of this paper.

Seemingly unaware of the work of [**7**], three decades later [**9**] showed that, for all $u > 0$,

$$P\left(\frac{\log 2 \ \max_{i=1}^n A_i}{n} \leq u\right) \to e^{-u^{-1}}, \tag{1.3}$$

which is a restatement of $P(\mathcal{E}_n^u = 0) \to P(\mathcal{E}_*^u = 0)$ and hence an easy consequence of (1.2). However, because of the subtle mistake of [**7**], the above result of [**9**] stands as

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

the first correctly proven result on extreme value theory of continued fractions. This has remained a topic of current interest; see, for example, the generalizations of (1.3) to fibred systems by [**13**] and to Oppenheim continued fractions by [**3**].

In view of the above, the following question arises naturally: what is the rate of convergence in the of the asymptotics in (1.2)? In this paper we give an upper bound on the rate of convergence using the Chen–Stein method of [**2**] (more specifically, Theorem 2.1 below). As far as we are aware, our work is the first to specifically employ the Chen–Stein method in the context of the Gauss map and continued fractions.

The Chen–Stein method is a very useful technique which yields an upper bound that is uniform in $u$ bounded away from zero; see Theorem 1.1 below. As a consequence, we also get a locally uniform (on $(0, \infty]$) upper bound for the convergence of distribution functions in (1.3) and this bound is much better than the best known bound given in [**14**] (we improve a slowly varying rate of convergence to a polynomial one; see Remark 1.4 below). In fact, we give a bound on the rate of convergence of the $k$th maxima, not just the maxima, and the Chen–Stein method is powerful enough to ensure that this locally uniform upper bound turns out to be uniform over $k \in \mathbb{N}$ as well (see Corollary 1.2).

Note that (1.3) implies that the $A_i$ are in the Fréchet(1) maximal domain of attraction. It is not difficult to observe that (1.3) holds because the $A_i$ enjoy a very strong exponential mixing property (see (1.7) below), and each $A_i$ (which are anyway identically distributed because of stationarity) is regularly varying with index $-1$, that is,

$$nP\left(\frac{A_1 \log 2}{n} \in \cdot\right) \xrightarrow{v} \nu \tag{1.4}$$

as measures on $(0, \infty]$. Here '$\xrightarrow{v}$' denotes vague convergence and $\nu$ is the unique measure on $(0, \infty]$ satisfying $\nu((u, \infty]) = u^{-1}$ for all $u \in (0, \infty)$. This was essentially the proof given in [**9**] except that Galambos did not use the language of vague convergence, and presented a direct proof instead.

The above vague convergence will play a very important role in this paper. Since $A_1$ is an integer-valued random variable, it follows that, for each $u > 0$,

$$P\left(\frac{A_1 \log 2}{n} > u\right) = P\left(A_1 \geq \left\lceil \frac{nu}{\log 2} \right\rceil\right),$$

from which, using [**9**, Lemma 1], we obtain

$$nP\left(\frac{A_1 \log 2}{n} > u\right) = \frac{n \log\left(1 + 1/\lceil nu/\log 2\rceil\right)}{\log 2} \to u^{-1} \tag{1.5}$$

as $n \to \infty$. From the above convergence, (1.4) follows by invoking [**16**, Theorem 3.6]. Further, using the inequality $\log(1 + x) \leq x$ whenever $x > 0$, we get the following upper bound, which will also be very useful in this paper: for all $u > 0$,

$$P\left(\frac{A_1 \log 2}{n} > u\right) = \frac{\log\left(1 + 1/\lceil nu/\log 2\rceil\right)}{\log 2} \leq \frac{\log\left(1 + (\log 2)/nu\right)}{\log 2} \leq \frac{1}{nu}. \tag{1.6}$$

In some sense, the $A_i$ behave very much like an independent and identically distributed (i.i.d.) sequence because of the following exponential mixing property: for all $m, n \in \mathbb{N}$, for all $F \in \sigma(A_1, A_2, \ldots, A_m)$, and for all $H \in \sigma(A_{m+n}, A_{m+n+1}, \ldots)$,

$$|P(F \cap H) - P(F)P(H)| \leq \psi(n)P(F)P(H) = C\theta^{-n}P(F)P(H), \tag{1.7}$$

where $\psi(n) = C\theta^{-n}$ for some $C > 0$ and $\theta > 1$ (see, for example, [**9**, Lemma 2]).

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

In order to state our main result and its corollary, we need to introduce some notation as described below. For each $n \in \mathbb{N}$ and for each $k \in \{1, 2, \ldots, n\}$, denote by $M_n^{(k)}$ the $k$th largest in the set $\{A_i \log 2 : 1 \le i \le n\}$. Then it follows from (1.2) that for all $u > 0$,

$$P\left(\frac{M_n^{(k)}}{n} \le u\right) = P(\mathcal{E}_n^u \le k-1) \to P(\mathcal{E}_*^u \le k-1) = e^{-u^{-1}} \sum_{i=0}^{k-1} \frac{u^{-i}}{i!}.$$

Obviously, the $k = 1$ case has already been taken care of in (1.3) above. Also, let $\{l_n\}$ be a sequence of positive real numbers such that

$$l_n \theta^{l_n} = n \tag{1.8}$$

for all $n \in \mathbb{N}$ (here $\theta$ is as in (1.7) above). Clearly, such a sequence exists by the intermediate value theorem and it increases to infinity at a rate strictly slower than $\log n$.

We are now ready to state our main result.

THEOREM 1.1. *With the notation as above, we have the following upper bound on the rate of convergence in (1.2): there exists $\kappa > 0$ such that, for all $\delta > 0$ and for all $n \in \mathbb{N}$,*

$$\sup_{u \in [\delta, \infty)} d_{\mathrm{TV}}(\mathcal{E}_n^u, \mathcal{E}_*^u) \le \frac{\kappa}{\min\{\delta, \delta^2\}} \frac{l_n}{n},$$

*where $d_{\mathrm{TV}}$ denotes the total variation distance.*

We would like to mention that we blend probability theory (namely, the Chen–Stein method; see Theorem 2.1), ergodic theory (specifically, the exponential mixing property (1.7)) and real analysis (more precisely, a second-order regular variation estimate; the second inequality in (2.11)) to prove the result above.

Theorem 1.1 has the following very strong consequence on the rate of weak convergence of scaled $k$th maxima. The upper bound here is uniform over $u$ bounded away from zero and uniform over $k \in \mathbb{N}$ at the same time.

COROLLARY 1.2. *With $\kappa$ as in Theorem 1.1, we get that, for all $\delta > 0$ and for all $n \in \mathbb{N}$,*

$$\sup_{k \in \mathbb{N}} \sup_{u \in [\delta, \infty)} \left| P\left(\frac{M_n^{(k)}}{n} \le u\right) - e^{-u^{-1}} \sum_{i=0}^{k-1} \frac{u^{-i}}{i!} \right| \le \frac{\kappa}{\min\{\delta, \delta^2\}} \frac{l_n}{n}.$$

The above corollary follows from Theorem 1.1 by restricting the supremum in the definition of total variation distance to sets of the form $\{0, 1, \ldots, k-1\}$ with $k$ running over the set of all positive integers.

*Remark 1.3.* Note that if the $A_i$ were i.i.d. with same marginal distribution, then by [**17**] we would have obtained an upper bound of $O(1/n)$ on the rate of convergence of the maxima sequence. The Chen–Stein method gives the same rate in the i.i.d. case. In the Gauss dynamical system, we get an extra factor of $l_n$ because of the dependence of the $A_i$. However, since $l_n = o(\log n)$, it follows that our bound on the rate of convergence is $o((\log n)/n)$. Therefore, we almost attain the rate obtained in the i.i.d. case.

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

*Remark 1.4.* The best known rate of convergence for the maxima in our set-up was obtained by [**14**], who gave an upper bound of $O(e^{-(\log n)^{\delta}})$ with $\delta \in (0, 1)$ (the constant in $O$ depends on $\delta$). Note that $e^{-(\log n)^{\delta}}$ is a slowly varying function of $n$. Therefore, by the Potter bound (see, for example, [**16**, p. 32]), it follows that $n^{-\eta} = o(e^{-(\log n)^{\delta}})$ for all $\eta > 0$ and for all $\delta \in (0, 1)$. Hence, by Remark 1.3, it follows that

$$\frac{l_n}{n} = o(e^{-(\log n)^{\delta}})$$

for all $\delta \in (0, 1)$. Therefore, our bound on the rate of convergence is significantly better than the one obtained by [**14**]. More precisely, we improve a slowly varying rate of convergence to a polynomial one, bettering an error term that was used by Philipp in his proof of a conjecture of Paul Erdős.

Note that the $D$ and $D'$ conditions of [**5**] follow from (1.7). Therefore, by [**6**, Example 5.1], the following extremal point process weak convergence holds in the space $\mathcal{M}_p((0, \infty])$ of all Radon point measures (on $(0, \infty]$) equipped with the vague metric:

$$Q_n := \sum_{i=1}^{n} \delta_{(A_i \log 2)/n} \xrightarrow{d} Q_* \sim PRM((0, \infty], \nu). \tag{1.9}$$

Here the limit $Q_*$ is a Poisson random measure on $(0, \infty]$ with mean measure $\nu$; see [**19**, §4.1] for a direct proof of (1.9). In this paper, we observe that a tiny detour of our proof of Theorem 1.1 yields (1.9); see §2.3 below.

## 2. *Proofs*

As mentioned earlier, the proof of Theorem 1.1 relies on the Chen–Stein method of [**2**]. We first state their result and then present our proof. Finally, we observe how a tiny detour of the proof also establishes the weak convergence of the extremal point process of the digits arising in the continued fraction expansion.

### 2.1. *The Chen–Stein method of [2].*

Let $\mathcal{I}$ be an index set and $\{X_\alpha \sim \text{Ber}(p_\alpha)\}_{\alpha \in \mathcal{I}}$ be a collection of possibly dependent Bernoulli random variables. Suppose, for each $\alpha \in \mathcal{I}$, there exists a subset $B_\alpha \subseteq \mathcal{I}$ such that, roughly speaking, $X_\alpha$ is nearly independent of $\{X_\beta : \beta \in \mathcal{I} \setminus B_\alpha\}$. Arratia *et al* [**2**] called $B_\alpha$ the 'neighbourhood of dependence' of $X_\alpha$. Following their notation, we define

$$b_1 = \sum_{\alpha \in \mathcal{I}} \sum_{\beta \in B_\alpha} p_\alpha p_\beta, \tag{2.1}$$

$$b_2 = \sum_{\alpha \in \mathcal{I}} \sum_{\beta \in B_\alpha \setminus \{\alpha\}} E(X_\alpha X_\beta), \tag{2.2}$$

$$b_3 = \sum_{\alpha \in \mathcal{I}} E[|E(X_\alpha - p_\alpha | \mathcal{H}_\alpha)|], \tag{2.3}$$

where $\mathcal{H}_\alpha$ is the $\sigma$-field generated by $\{X_\beta : \beta \in \mathcal{I} \setminus B_\alpha\}$.

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

THEOREM 2.1. [**2**, Theorem 2] *Partition $\mathcal{I}$ into disjoint non-empty subsets $\mathcal{I}_1, \mathcal{I}_2, \ldots, \mathcal{I}_k$. Let $\{Y_\alpha \sim \mathrm{Poi}(p_\alpha)\}_{\alpha\in\mathcal{I}}$ be a collection of independent Poisson random variables. Set*

$$W_j = \sum_{\alpha\in\mathcal{I}_j} X_\alpha, \quad Z_j = \sum_{\alpha\in\mathcal{I}_j} Y_\alpha \quad \text{and} \quad \lambda_j = \sum_{\alpha\in\mathcal{I}_j} p_\alpha.$$

*Then*

$$\begin{aligned} &d_{\mathrm{TV}}(\mathcal{L}(W_1, W_2, \ldots, W_k), \mathcal{L}(Z_1, Z_2, \ldots, Z_k)) \\ &\quad \le \min\left\{2, 2.8 \max_{1\le j\le k} \lambda_j^{-1/2}\right\}(2b_1 + 2b_2 + b_3), \end{aligned} \tag{2.4}$$

*where $\mathcal{L}$ denotes the joint law.*

We would like to elaborate a bit on the phrase 'nearly independent' used above in the context of neighbourhood of dependence $B_\alpha$. In many examples (e.g., $m$-dependent time-series models, certain random graph asymptotics) where Theorem 2.1 is used, $X_\alpha$ is totally independent of $\{X_\beta : \beta \notin B_\alpha\}$, so that $b_3 = 0$. In our case, however, we need to bound $b_3$ tightly using the 'near-independence' property (1.7).

2.2. *Proof of Theorem 1.1.* Define a new Poisson random variable $\tilde{\mathcal{E}}_n^u$ with mean $nP(n^{-1}A_1 \log 2 > u)$. The basic strategy of the proof is to use the inequality

$$d_{\mathrm{TV}}(\mathcal{E}_n^u, \mathcal{E}_*^u) \le d_{\mathrm{TV}}(\mathcal{E}_n^u, \tilde{\mathcal{E}}_n^u) + d_{\mathrm{TV}}(\tilde{\mathcal{E}}_n^u, \mathcal{E}_*^u) \tag{2.5}$$

and to estimate each term separately. The bound on $d_{\mathrm{TV}}(\mathcal{E}_n^u, \tilde{\mathcal{E}}_n^u)$ will need the Chen–Stein method and the exponential mixing property (1.7), while the second term $d_{\mathrm{TV}}(\tilde{\mathcal{E}}_n^u, \mathcal{E}_*^u)$ will be estimated using a hard analytic bound on the second-order term of the convergence in (1.5). Thus, our proof combines tools from probability theory, ergodic theory and real analysis in a systematic manner.

We will first show that there exists $\kappa_1 > 0$ such that, for all $u > 0$ and for all $n \ge 1$,

$$d_{\mathrm{TV}}(\mathcal{E}_n^u, \tilde{\mathcal{E}}_n^u) \le \kappa_1 \max\left\{\frac{1}{u}, \frac{1}{u^2}\right\}\frac{l_n}{n}, \tag{2.6}$$

where $l_n$ is as in (1.8). To this end, set

$$D = (u, \infty]. \tag{2.7}$$

We shall use Theorem 2.1 with $\mathcal{I} = \{1, 2, \ldots, n\}$, $k = 1$, $X_\alpha = I_{(n^{-1}A_\alpha \log 2 \in D)}$ (and hence $p_\alpha = E(X_\alpha) = E(X_1) = P(n^{-1}A_1 \log 2 \in D)$) and $B_\alpha = (\alpha - l_n, \alpha + l_n) \cap \mathcal{I}$ for each $\alpha \in \mathcal{I}$. Note that with these choices we have $W_1 = \mathcal{E}_n^u$ and $Z_1$ may be thought of, intuitively, as '$W_1$ if the $X_\alpha$ were independent'.

Because of stationarity, we get

$$\sum_{\alpha\in\mathcal{I}} p_\alpha = nP(n^{-1}A_1 \log 2 \in D) = nP(n^{-1}A_1 \log 2 > u).$$

In order to establish (2.6), we have to estimate the quantities defined by (2.1)–(2.3). For the first one, observe that

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

$$b_1 = \sum_{\alpha\in\mathcal{I}} \sum_{\beta\in B_\alpha} p_\alpha p_\beta$$
$$= \sum_{\alpha=1}^{n} \sum_{\beta\in B_\alpha} (P(n^{-1}A_1 \log 2 \in D))^2,$$

from which, using (1.6), we get

$$b_1 \le 2nl_n\left(\frac{1}{nu}\right)^2 = \frac{2}{u^2}\frac{l_n}{n}. \tag{2.8}$$

In order to bound the second term in (2.4), note that, for any $\alpha, \beta \in \mathbb{N}$ such that $\alpha \ne \beta$,

$$E(X_\alpha X_\beta) = P(n^{-1}A_\alpha \log 2 \in D,\, n^{-1}A_\beta \log 2 \in D)$$
$$\le (1 + \psi(|\alpha - \beta|))P(n^{-1}A_\alpha \log 2 \in D)P(n^{-1}A_\beta \log 2 \in D),$$

where the last step follows from (1.7). Applying stationarity, (1.6) and the inequality $\psi(n) \le C$, we obtain from the above bound that

$$E(X_\alpha X_\beta) \le (1 + C)(P(n^{-1}A_1 \log 2 \in D))^2 \le (1 + C)\left(\frac{1}{nu}\right)^2$$

for all $\alpha \ne \beta$. Hence

$$b_2 = \sum_{\alpha\in\mathcal{I}} \sum_{\beta\in B_\alpha\setminus\{\alpha\}} E(X_\alpha X_\beta) \le 2nl_n(1 + C)\frac{1}{u^2n^2} = \frac{2(1 + C)}{u^2}\frac{l_n}{n}. \tag{2.9}$$

Finally, we need to estimate (2.3). Fixing $\alpha \in \mathcal{I}$ and taking $F = (n^{-1}A_\alpha \log 2 \in D)$ with $D$ as in (2.7), we see that (1.7) yields

$$p_\alpha P(H)(1 - \psi(l_n)) \le P((n^{-1}A_\alpha \log 2 \in D) \cap H) \le p_\alpha P(H)(1 + \psi(l_n))$$

for all $H \in \mathcal{H}_\alpha = \sigma\{X_\beta :\ \beta \in \mathcal{I} \setminus B_\alpha\}$. The above pair of inequalities can be rewritten as

$$p_\alpha P(H)(1 - \psi(l_n)) \le \int_H X_\alpha\, dP = \int_H E(X_\alpha|\mathcal{H}_\alpha)\, dP \le p_\alpha P(H)(1 + \psi(l_n)),$$

yielding

$$-\int_H p_\alpha\psi(l_n)\, dP \le \int_H E(X_\alpha - p_\alpha|\mathcal{H}_\alpha)\, dP \le \int_H p_\alpha\psi(l_n)\, dP,$$

which holds for all $H \in \mathcal{H}_\alpha$, and hence

$$|E(X_\alpha - p_\alpha|\mathcal{H}_\alpha)| \le p_\alpha\psi(l_n) = P(n^{-1}A_1 \log 2 \in D)\psi(l_n)$$

almost surely. Therefore, we obtain

$$b_3 = \sum_{\alpha\in\mathcal{I}} E[|E(X_\alpha - p_\alpha|\mathcal{H}_\alpha)|]$$
$$\le nP(n^{-1}A_1 \log 2 \in D)\psi(l_n)$$
$$\le n\frac{1}{nu}C\theta^{-l_n}$$
$$= \frac{C}{u}\frac{l_n}{n},$$

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

where we used (1.6) and the last step follows from the choice of $l_n$ as given in (1.8). The above upper bound, along with (2.8) and (2.9), yields (2.6) thanks to Theorem 2.1.

We now move on to estimating the second term in (2.5). We first use Taylor's theorem to obtain the inequality $|\log(1+x) - x| \leq x^2/2$, which can be rewritten as

$$\left|\frac{\log(1+x)}{x} - 1\right| \leq \frac{x}{2} \tag{2.10}$$

for all $x > 0$. Using this inequality, we shall now bound the second-order term of the convergence in (1.5).

To this end, note that

$$\begin{aligned}|nP(n^{-1}A_1 \log 2 \in D) - u^{-1}| &= \left|\frac{n \log(1 + \lceil nu/\log 2\rceil^{-1})}{\log 2} - \frac{1}{u}\right| \\ &\leq \frac{n}{\log 2}\left|\log\left(1 + \frac{1}{\lceil nu/\log 2\rceil}\right) - \log\left(1 + \frac{1}{nu/\log 2}\right)\right| \\ &\quad + \frac{1}{u}\left|\frac{\log(1 + n^{-1}u^{-1}\log 2)}{n^{-1}u^{-1}\log 2} - 1\right|.\end{aligned}$$

By virtue of (2.10), the second term above is bounded by $(\log 2)/2u^2n$. On the other hand, using the mean value theorem, we can estimate the first term as follows:

$$\frac{n}{\log 2}\left|\log\left(1 + \frac{1}{\lceil nu/\log 2\rceil}\right) - \log\left(1 + \frac{1}{nu/\log 2}\right)\right| \leq \frac{n}{\log 2}\frac{(\log 2)^2}{n^2u^2} = \frac{\log 2}{u^2n}.$$

Therefore, by [**8**, Lemma (8)], it follows that

$$d_{\mathrm{TV}}(\tilde{\mathcal{E}}_n^u, \mathcal{E}_*^u) \leq |nP(n^{-1}A_1 \log 2 \in D) - u^{-1}| \leq \frac{3\log 2}{2u^2}\frac{1}{n}. \tag{2.11}$$

The above inequality, (2.6) and (2.5) imply that there exists a constant $\kappa \in (0, \infty)$ such that, for all $u > 0$ and for all $n \geq 1$,

$$d_{\mathrm{TV}}(\mathcal{E}_n^u, \mathcal{E}_*^u) \leq \kappa \max\left\{\frac{1}{u}, \frac{1}{u^2}\right\}\frac{l_n}{n},$$

from which Theorem 1.1 follows.

*Remark 2.2.* We would like to mention here an alternative approach pointed out to us by an anonymous referee. Smith [**18**, Theorem 1] gives a similar Chen–Stein type upper bound in the more general set-up of non-stationary processes. It is possible to use this result to give a bound on $d_{\mathrm{TV}}(\mathcal{E}_n^u, \tilde{\mathcal{E}}_n^u)$ in our work, leaving the estimation of $d_{\mathrm{TV}}(\tilde{\mathcal{E}}_n^u, \mathcal{E}_*^u)$ (based on hard analysis) as it is. This will involve (in the notation of [**18**]) coming up with the function $g(n, r)$, the subsets $I_{nrk}$ and $I^*_{nrk}$ ($\subseteq I_{nrk}$), the latter being very similar to a neighbourhood of dependence, and verifying Condition $D'$ of [**18**]. We think that this will be more involved than the estimation of the terms $b_1$ and $b_2$ of our paper. On the other hand, Condition $D$ of [**18**] will follow directly from the exponential mixing property (1.7) of our paper, and this verification will be shorter than bounding the term $b_3$ in our work. Overall, we feel that an application of [**18**, Theorem 1], instead of [**2**, Theorem 2], will perhaps result in an argument of similar length. However, we have not compared the rates obtained by these two results in our set-up.

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

2.3. *New proof of (1.9).* By [**11**, Theorem 4.7], in order to establish (1.9), it is enough to show the following results.

(i) For all $u, v \in (0, \infty]$ with $u < v$,

$$E(Q_n((u, v])) \longrightarrow E(Q_*((u, v])) = \nu((u, v]) = u^{-1} - v^{-1} \tag{2.12}$$

as $n \to \infty$. Of course, we follow the convention $\infty^{-1} = 0$.

(ii) Whenever $0 < u_1 < v_1 < u_2 < v_2 < \cdots < u_k < v_k \le \infty$,

$$\begin{aligned} P(Q_n((u_1, v_1]) = 0, Q_n((u_2, v_2]) = 0, \ldots, Q_n((u_k, v_k]) = 0) \\ \to P(Q_*((u_1, v_1]) = 0, Q_*((u_2, v_2]) = 0, \ldots, Q_*((u_k, v_k]) = 0) \\ = \prod_{i=1}^{k} e^{-\nu((u_i, v_i])} = \exp\left\{ -\nu\left( \bigcup_{i=1}^{k} (u_i, v_i] \right) \right\} \end{aligned} \tag{2.13}$$

as $n \to \infty$.

By linearity of expectation, in order to establish (2.12), it is enough to do so with $u \in (0, \infty)$ and $v = \infty$. This special case follows using stationarity of the $A_i$ and (1.5) as shown below:

$$\begin{aligned} E(Q_n((u, \infty])) &= E\left[ \sum_{i=1}^{n} \delta_{n^{-1} A_i \log 2}((u, \infty]) \right] \\ &= n P(n^{-1} A_1 \log 2 > u) \\ &\to \nu((u, \infty]), \end{aligned}$$

as $n \to \infty$). This proves (2.12).

On the other hand, verification of (2.13) will need a tiny detour of the proof of Theorem 1.1 (as carried out in [**4**] in the context of Gaussian free fields) and Theorem 2.1 will again play a significant role in the proof. To this end, fix $0 < u_1 < v_1 < u_2 < v_2 < \cdots < u_k < v_k \le \infty$ and set

$$D = \bigcup_{i=1}^{k} (u_i, v_i]. \tag{2.14}$$

Note that, by (1.4) and [**15**, Proposition 3.12], it follows that

$$n P(n^{-1} A_1 \log 2 \in D) \to \nu(D) \tag{2.15}$$

as $n \to \infty$ for $D$ as in (2.14). Therefore by changing the definition of $D$ from (2.7) to (2.14) in the proof of Theorem 1.1 and using (2.15), it is easy to show that

$$d_{\mathrm{TV}}(Q_n(D), Q_*(D)) \to 0$$

as $n \to \infty$. In particular, $P(Q_n(D) = 0) \to P(Q_*(D) = 0) = e^{-\nu(D)}$, which is a restatement of (2.13). This completes the proof of (1.9) based on the Chen–Stein method of [**2**].

*Acknowledgements.* This work was initiated during a visit by M.K. and P.R. to the Tata Institute of Fundamental Research, Mumbai, and a significant portion of the work was carried out when the authors were at the International Centre for Theoretical

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press

Sciences, Bangalore, for the programme 'Probabilistic Methods in Negative Curvature' (ICTS/pmnc2019/03). We thank both institutes for their hospitality and the lovely working conditions. We would also like to acknowledge an anonymous reviewer and an executive editor for their careful reading of the paper. Their detailed comments have significantly improved our work (especially Remarks 1.3 and 2.2). A.G. gratefully acknowledges support from a grant from the Indo-French Centre for the Promotion of Advanced Research, a Swarnajayanti Fellowship from Department of Science and Technology, Government of India, and a MATRICS grant from the Science and Engineering Research Board. P.R. acknowledges the support from a MATRICS grant from the Science and Engineering Research Board and a Swarnajayanti Fellowship from Department of Science and Technology, Government of India.

## References

[**1**] J. Aaronson. *An Introduction to Infinite Ergodic Theory*. American Mathematical Society, Providence, RI, 1997.
[**2**] R. Arratia, L. Goldstein and L. Gordon. Two moments suffice for Poisson approximations: the Chen–Stein method. *Ann. Probab.* **17** (1989), 9–25.
[**3**] Y. Chang and J. Ma. Some distribution results of the Oppenheim continued fractions. *Monatsh. Math.* **184**(3) (2017), 379–399.
[**4**] A. Chiarini, A. Cipriani and R. S. Hazra. A note on the extremal process of the supercritical Gaussian free field. *Electron. Commun. Probab.* **20** (2015), paper no. 74.
[**5**] R. Davis. Stable limits for partial sums of dependent random variables. *Ann. Probab.* **11**(2) (1983), 262–269.
[**6**] R. Davis and T. Hsing. Point processes for partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**(2) (1995), 879–917.
[**7**] W. Doeblin. Remarques sur la théorie métrique des fractions continues. *Compos. Math.* **7** (1940), 353–371.
[**8**] D. Freedman. The Poisson approximation for dependent events. *Ann. Probab.* **2** (1974), 256–269.
[**9**] J. Galambos. The distribution of the largest coefficient in continued fraction expansions. *Quart. J. Math.* **23**(2) (1972), 147–151.
[**10**] M. Iosifescu. A Poisson law for $\psi$-mixing sequences establishing the truth of a Doeblin's statement. *Rev. Roumaine Math. Pures Appl.* **22** (1977), 1441–1447.
[**11**] O. Kallenberg. *Random Measures*, 3rd edn. Akademie-Verlag, Berlin, 1983.
[**12**] A. Khintchine. *Continued Fractions*. University of Chicago Press, Chicago, 1964.
[**13**] H. Nakada and R. Natsui. On the metrical theory of continued fraction mixing fibred systems and its application to Jacobi–Perron algorithm. *Monatsh. Math.* **138** (2003), 267–288.
[**14**] W. Philipp. A conjecture of Erdös on continued fractions. *Acta Arith.* **28**(4) (1976), 379–386.
[**15**] S. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, New York, 1987.
[**16**] S. Resnick. *Heavy-Tail Phenomena*. Springer, New York, 2007.
[**17**] S. Resnick and L. de Haan. Second-order regular variation and rates of convergence in extreme-value theory. *Ann. Probab.* **24**(1) (1989), 97–124.
[**18**] R. L. Smith. Extreme value theory for dependent sequences via the Stein–Chen method of Poisson approximation. *Stochastic Process. Appl.* **30**(2) (1988), 317–327.
[**19**] M. Tyran-Kamińska. Weak convergence to Lévy stable processes in dynamical systems. *Stoch. Dyn.* **10**(2) (2010), 263–289.

https://doi.org/10.1017/etds.2019.64 Published online by Cambridge University Press