Quenched phantom distribution functions for Markov chains

Adam Jakubowski*†, Patryk Truszczyński

†Nicolaus Copernicus University, Faculty of Mathematics and Computer Science, ul. Chopina 12/18, 87-100 Toruń, Poland

Abstract

It is known that random walk Metropolis algorithms with heavy-tailed target densities can model atypical (slow) growth of maxima, which in general is exhibited by processes with the extremal index zero. The asymptotics of maxima of such sequences can be analyzed in terms of continuous phantom distribution functions. We show that in a large class of positive Harris recurrent Markov chains (containing the above Metropolis chains) a phantom distribution function can be recovered by starting “at the point” rather than from the stationary distribution.

Keywords: stochastic extremes, Markov chains, phantom distribution function, relative extremal index, random walk Metropolis algorithm, coupling, Harris chains

2010 MSC: 60G70, 60J05, 60F05

1. Introduction and the statement of results

A stationary sequence \( \{X_n\} \) of random variables with marginal distribution function \( F(x) = \mathbb{P}(X_j \leq x) \) and partial maxima \( M_n = \max_{0 \leq j \leq n-1} X_j \) admits a phantom distribution function \( G \) if

\[
\sup_{x \in \mathbb{R}} \left| \mathbb{P}(M_n \leq x) - G^n(x) \right| \to 0, \quad \text{as} \quad n \to \infty.
\]

Clearly, \( G \) is not uniquely determined, since for (1) to hold only the behavior of \( G(x) \) near the right end \( G_* = \sup\{x; G(x) < 1\} \) is of importance. It was observed in [7, Proposition 1] that under natural conditions (regularity), if \( H \) is any other phantom distribution function for \( \{X_n\} \), it must be tail equivalent to \( G \), i.e.

\[
\lim_{x \to G_*} \frac{1 - H(x)}{1 - G(x)} = 1.
\]

*Corresponding author

Email addresses: adjakubo@mat.umk.pl (Adam Jakubowski), 502121@doktorant.umk.pl (Patryk Truszczyński)

Preprint submitted to Statistics and Probability Letters October 5, 2018
Phantom distribution functions were introduced by O’Brien in [14], as an extension of the notion of Leadbetter’s extremal index. The latter corresponds to \( G(x) \) of the form \( F^\theta(x) \), for some \( \theta \in (0,1] \) (see [11]). Contrary to extremal indices, the existence of phantom distribution functions is a quite common phenomenon for weakly dependent stationary sequences. For example, phantom distribution functions can be almost explicitly constructed for Markov chains with regenerative structure [16, Theorem 3.1]. Moreover, it was proved in the recent paper [7] that any \( \alpha \)-mixing stationary sequence with continuous marginals \( F \) admits a continuous phantom distribution function.

It follows that the asymptotic behavior of maxima of stationary sequences with no extremal index or with the extremal index zero can still be analyzed using phantom distribution functions. Recall that \( \{X_n\} \) has the extremal index \( \theta = 0 \) if

\[
\mathbb{P}(M_n \leq u_n(\tau)) \rightarrow 1,
\]

for every sequence \( u_n(\tau) \) such that \( n(1 - F(u_n(\tau))) \rightarrow \tau \) (see [11]). This means that maxima \( M_n \) increase slower than maxima of i.i.d. random variables with marginal’s tails comparable to \( 1 - F(x) \), and so asymptotic properties of \( M_n \) cannot be expressed in terms of \( F \). Such a situation appears, for example, when Lindley’s process has subexponential innovations (see [1]) or when the continuous target distribution of the random walk Metropolis algorithm has heavy tails (see [15]).

We will focus on the latter example. Let us recall basic definitions. Let \( \{Z_j\} \) be an i.i.d. sequence with the marginal distribution function \( H \) given by the proposal density \( h \), which is symmetric about 0, and let \( \{U_j\} \) be an i.i.d. sequence distributed uniformly on \([0,1]\), independent of \( \{Z_j\} \). Choose and fix the target probability density \( f(x) \) (heavy-tailed in our case). Then the random walk Metropolis algorithm is the Markov chain given by the recursive equation

\[
X_{j+1} = X_j + Z_{j+1} \mathbb{1}\{U_{j+1} \leq \psi(X_j, X_j + Z_{j+1})\}, \tag{2}
\]

where \( \psi(x, y) \) is defined as

\[
\psi(x, y) = \begin{cases} 
\min \{f(y)/f(x), 1\} & \text{if } f(x) > 0, \\
1 & \text{if } f(x) = 0. 
\end{cases} \tag{3}
\]

Random walk Metropolis algorithms have been designed for the purposes of simulation and therefore can be efficiently used in modeling of atypically slow growth of partial maxima of stationary sequences. As observed in [7, Remark 7] such a flexible family of models allows classifying stationary processes by relation of having a relative extremal index in the sense of [8]. To be more precise, given a “model” stationary sequence \( \{X_n\} \) we can consider a class of stationary sequences \( \{X'_n\} \) satisfying for some \( \theta \in (0, +\infty) \)

\[
\sup_{x \in \mathbb{R}} |\mathbb{P}(M_n \leq x) - \mathbb{P}^\theta(M'_n \leq x)| \rightarrow 0, \quad \text{as } n \rightarrow \infty, \tag{4}
\]

where \( M'_n = \max_{1 \leq j \leq n-1} X'_j \). Writing \( \{X_n\} \sim_\theta \{X'_n\} \) if (4) holds, we see that \( \{X_n\} \sim_\theta \{X'_n\} \) implies \( \{X'_n\} \sim_{1/\theta} \{X_n\} \) and that \( \mathbb{P}(M_n \leq v_n) \rightarrow \alpha \in (0,1) \)
implies $P(M_n' \leq v_n) \to \alpha^{1/\theta}$, what clearly relates asymptotic quantiles of both sequences.

For weakly dependent sequences relation (4) means that there exists a phantom distribution function $G$ for $\{X_n\}$ such that $G^{1/\theta}$ is a phantom distribution function for $\{X_n'\}$ (see the proof of [8, Theorem 1.5]). It follows that the idea of comparison to models with atypical growth of maxima requires efficient methods of finding phantom distribution functions for random walk Metropolis algorithms with heavy-tailed proposals. We refer to the discussion in [10] for peculiarities related to the choice of the proposal density of the Monte Carlo Markov Chain algorithms.

Here we want to address another issue: whether we can start our random walk Metropolis algorithms “at the point” rather than at the target density. The question is not trivial, since starting “at the point” we loose stationarity.

We adopt formula (1) as the definition of a phantom distribution function of a non-stationary sequence (see [9]). For the terminology related to general space Markov chains $\{Y_n\}$ we refer to the classic source [12]. In particular, $P_s(\cdot)$ means the probability conditional on $\{Y_0 = s\}$, while $P_{\pi}(\cdot)$ means that $Y_0$ is distributed according to $\pi$.

**Theorem 1.** Let $\{Y_n\}$ be a positive Harris and aperiodic chain taking values in $(\mathcal{S}, \mathcal{S})$ and with a stationary distribution $\pi$. Let $f : (\mathcal{S}, \mathcal{S}) \to (\mathbb{R}^1, \mathcal{B}^1)$ be a measurable function.

Let us define

$$X_n = f(Y_n), \quad n = 0, 1, 2, \ldots, \quad M_n = \max_{0 \leq j \leq n-1} X_j, \quad n = 1, 2, \ldots.$$  

If $\{X_n\}$ admits a continuous phantom distribution function $G$ under some initial distribution $\lambda$, i.e. if we have

$$\sup_{x \in \mathbb{R}^1} \left| P_{\lambda}(M_n \leq x) - G^n(x) \right| \to 0, \text{ as } n \to \infty,$$

then $G$ is also a continuous phantom distribution function for the stationary (under $\pi$) sequence $\{X_n\}$.

Conversely, if $\{X_n\}$ admits a continuous phantom distribution function $G$ under $\pi$, then there exists a set $S_0 \in \mathcal{S}$ satisfying $\pi(S_0) = 0$ and such that relation (5) holds for every initial distribution $\lambda$ with the property that $\lambda(S_0) = 0$.

**Theorem 2.** In assumptions of Theorem 1 if $\pi \circ f^{-1}$ (i.e. the marginal law of $\{X_n\}$ under the stationary distribution $\pi$) is continuous and unbounded above, then there exists a continuous distribution function $G$ such that for each $s \in \mathcal{S}$

$$\sup_{x \in \mathbb{R}^1} \left| P_s(M_n \leq x) - G^n(x) \right| \to 0, \text{ as } n \to \infty.$$  

The proofs are given in the next section. We will need a complement to [9, Corollary 5], which might be of independent interest.
Recall that a distribution function $G$ is regular (in the sense of O'Brien), if
\begin{equation}
G(G_-) = 1 \quad \text{and} \quad \lim_{x \to G_-} \frac{1 - G(x-)}{1 - G(x)} = 1.
\end{equation}

Due to an observation made long time ago by Theorem 2, $G$ is regular if, and only if, for some $\gamma \in (0, 1)$ there exists a sequence $\{v_n = v_n(\gamma)\}$ such that
\[
G^n(v_n) \to \gamma.
\]

Notice that if $G$ is regular then the sequence $\{v_n(\gamma)\}$ exists for every $\gamma \in (0, 1)$ and that $\{v_n(\gamma)\}$ can always be chosen non-decreasing.

**Theorem 3.** Let $X_0, X_1, X_2, \ldots$ be an arbitrary sequence of random variables with partial maxima $M_n = \max_{0 \leq j \leq n-1} X_j$. Then the following conditions are equivalent.

(i) $\{X_j\}$ admits a continuous phantom distribution function.

(ii) $\{X_j\}$ admits a regular phantom distribution function.

(iii) There exists $\beta > 0$ and a non-decreasing sequence of levels $\{v_n = v_n(\beta)\}$ such that
\[
P(M_{[nt]} \leq v_n(\beta)) \to \exp(-\beta t), \quad t \in D,
\]
where $D \subset \mathbb{R}^+$ is dense.

(iv) For every $\beta > 0$ there exists a non-decreasing sequence of levels $\{v_n = v_n(\beta)\}$ such that (7) holds for every $t > 0$ (i.e. for $D = \mathbb{R}^+$).

**Remark 1.** If (7) holds for some dense subset $D$ of $\mathbb{R}^+$, then it is satisfied uniformly in $t \in \mathbb{R}^+$.

**Remark 2.** We have suggested one possible motivation for considering quenched (i.e. started “at the point”) phantom distribution function. But such a notion is interesting by itself, as in the context of the central limit theorem or the functional central limit theorem. It must be stressed that passing from a “usual” limit theorem to its quenched form need not be automatic (see [17]).

According to our knowledge this paper is the first that addresses “quenched” questions in the extreme value limit theory.

2. Proofs

2.1. **Proof of Theorem 3**

Statement (iv) trivially implies (iii), so let us assume that (iii) holds, i.e. (7) is satisfied for some $\beta > 0$ and some non-decreasing sequence of levels $\{v_n(\beta)\}$.
Applying Corollary 5 in [9] we obtain a phantom distribution function $G$ for \( \{X_j\} \) given by the formula

\[
G(x) = \begin{cases} 
0, & \text{if } x < v_1(\beta), \\
\exp(-\beta)^{1/n}, & \text{if } v_n(\beta) \leq x < v_{n+1}(\beta), \\
1, & \text{if } x \geq \sup_n v_n(\beta).
\end{cases}
\]  

(8)

$G$ of the above form is regular, hence (ii) holds. By [7, Remark 1, p. 704] there exists a continuous distribution function $H$ such that

\[
\sup_{x \in \mathbb{R}} \left| G^n(x) - H^n(x) \right| \to 0, \quad \text{as } n \to \infty.
\]

Clearly, $H(x)$ is a continuous phantom distribution function for $\{X_j\}$. So (i) also holds. (The interested reader may find an explicit formula for $H$ on p. 704 in [7]).

It remains to prove (i) $\Rightarrow$ (iv). So let us suppose that $\{X_j\}$ admits a continuous phantom distribution function, i.e. (1) holds for some continuous $G$. Choose $\beta > 0$ and define

\[
v_n(\beta) = \inf \{ x : G^n(x) = \exp(-\beta) \}.
\]

(notice that such $v_n(\beta)$ exists by the continuity of $G$). Then for any $t > 0$

\[
G^{[nt]}(v_n(\beta)) = \exp(-\beta)^{[nt]/n} \to \exp(-t\beta), \quad \text{as } n \to \infty,
\]

and

\[
\mathbb{P}(M_{[nt]} \leq v_n(\beta)) = G^{[nt]}(v_n) + o(1) \to \exp(-t\beta), \quad \text{as } n \to \infty.
\]

2.2. Proof of Theorem 1

Let $\{Y_j\}$ be an aperiodic positive Harris chain and let $\pi$ be its unique stationary initial distribution. Choose and fix an initial distribution $\lambda$. By [2, Proposition 3.13, p. 205] there exist a coupling of $\{X_n\}$ under $\pi$ and $\lambda$ with some a.s. finite coupling time $\tau$. Recall, that this means that on some probability space one can define two stochastic processes $\{Y'_j\}$ and $\{Y''_j\}$ such that $\{Y'_j\}$ has the same distribution as $\{Y_j\}$ under the initial distribution $\pi$, $\{Y''_j\}$ has the same distribution as $\{Y_j\}$ under the initial distribution $\lambda$ and

\[Y'_t(\omega) = Y''_t(\omega), \text{ whenever } t \geq \tau(\omega).
\]

Let $X'_t = f(Y'_t)$ and $X''_t = f(Y''_t)$ and let $\{M'_n\}$ and $\{M''_n\}$ be the partial maximum processes for $\{X'_t\}$ and $\{X''_t\}$, respectively.

Suppose that (5) holds for some continuous $G$. By Theorem 3 there exists $\beta > 0$ and a non-decreasing sequence of levels $v_n \nearrow G_*$ such that in the present notation

\[
\mathbb{P}_\lambda(M_{[nt]} \leq v_n) = \mathbb{P}(M''_{[nt]} \leq v_n) \to \exp(-\beta t), \quad t > 0.
\]

(9)
We claim that
\[
\mathbb{P}_\lambda(X_n < G_*) = \mathbb{P}(X_n' < G_*) = 1, \quad n = 0, 1, 2, \ldots. \tag{10}
\]
Indeed, if \(\mathbb{P}(X_n' \geq G_*) = \delta > 0\) for some \(n_0\), then for \(n\) such that \(nt \geq n_0\) we have
\[
\mathbb{P}(M'_n \leq v_n) \leq \mathbb{P}(X_n' < G_*) = 1 - \delta,
\]
while by (4)
\[
\mathbb{P}(M'_n \leq v_n) \to \exp(-\beta t) > 1 - \delta,
\]
if \(t < -\ln(1 - \delta)/\beta\).

So assume (10). By the convergence in total variation of marginals to the stationary distribution (see e.g. [12, Theorem 13.3.3, p. 328]) we have also for \(m = 0, 1, 2, \ldots\)
\[
\mathbb{P}(X_m' < G_*) = \mathbb{P}_\pi(X_1 < G_*) = \lim_{n \to \infty} \mathbb{P}_\lambda(X_n < G_*) = 1. \tag{11}
\]
Since the coupling time \(\tau\) is a.s. finite, we have
\[
\mathbb{P}\left(\max_{0 \leq j \leq \tau-1} X'_j < G_*\right) = \sum_{k=1}^\infty \mathbb{P}\left(\max_{0 \leq j \leq k-1} X'_j < G_*, \tau = k\right) = \sum_{k=1}^\infty \mathbb{P}(\tau = k) = 1,
\]
hence for every \(t > 0\)
\[
1 \geq \mathbb{P}\left(\max_{0 \leq j \leq \lfloor nt \rfloor \wedge \tau-1} X'_j \leq v_n\right) \geq \mathbb{P}\left(\max_{0 \leq j \leq \tau-1} X'_j \leq v_n\right) \geq \mathbb{P}(\tau = k) = 1.
\]
In a similar way we obtain that \(\mathbb{P}\left(\max_{0 \leq j \leq \tau-1} X''_j < G_*\right) = 1\) and for every \(t > 0\)
\[
\mathbb{P}\left(\max_{0 \leq j \leq \lfloor nt \rfloor \wedge \tau-1} X''_j \leq v_n\right) \to 1.
\]
Therefore we can write
\[
\mathbb{P}_\pi(M_{[nt]} \leq v_n) = \mathbb{P}(M'_{[nt]} \leq v_n)
\]
\[
= \mathbb{P}\left(\max_{0 \leq j \leq \lfloor nt \rfloor \wedge \tau-1} X'_j \leq v_n, \max_{\lfloor nt \rfloor \wedge \tau \leq j \leq \lfloor nt \rfloor-1} X'_j \leq v_n\right)
\]
\[
= \mathbb{P}\left(\max_{\lfloor nt \rfloor \wedge \tau \leq j \leq \lfloor nt \rfloor-1} X'_j \leq v_n\right) + o(1) \tag{12}
\]
\[
= \mathbb{P}\left(\max_{\lfloor nt \rfloor \wedge \tau \leq j \leq \lfloor nt \rfloor-1} X''_j \leq v_n\right) + o(1)
\]
\[
= \mathbb{P}\left(\max_{0 \leq j \leq \lfloor nt \rfloor \wedge \tau-1} X''_j \leq v_n, \max_{\lfloor nt \rfloor \wedge \tau \leq j \leq \lfloor nt \rfloor-1} X''_j \leq v_n\right) + o(1)
\]
\[
= \mathbb{P}(M''_{[nt]} \leq v_n) + o(1) \to \exp(-\beta t), \quad t > 0.
\]
Applying Theorem 3 we obtain that \(G\) is a phantom distribution function for \(\{X_n\}\) under \(\pi\).
To prove the other part of Theorem 1 let us suppose \( \{X_n\} \) admits a continuous phantom distribution function \( G \) under \( \pi \). Then by Theorem 3
\[
\mathbb{P}_\pi(M_{[nt]} \leq v_n) = \mathbb{P}(M'_{[nt]} \leq v_n) \rightarrow \exp(-\beta t), \quad t > 0,
\]
for some \( \beta > 0 \) and a non-decreasing sequence of levels \( v_n \) \( \nearrow \) \( G_* \). Similarly as for (10) we deduce that
\[
\mathbb{P}_\pi(X_1 < G_*) = \mathbb{P}(X'_{n} < G_*) = 1, \quad n = 0, 1, 2, \ldots.
\]

Let us define “the bad set”
\[
B_0 = \{ y : f(y) \geq G_* \}.
\] (13)

Since \( \pi(B_0) = 0 \) and the chain \( \{X_n\} \) is \( \pi \)-irreducible (see [12, Theorem 10.4.9, p. 246]) we obtain that also the set of states that communicate with \( B_0 \) has \( \pi \)-measure zero (see [12, Proposition 4.2.2, p. 83]):
\[
\pi(\{ y : P^n(y, B_0) > 0 \text{ for some } n \geq 1 \}) = 0.
\]

So let us set
\[
S_0 = B_0 \cup \{ y : P^n(y, B_0) > 0 \text{ for some } n \geq 1 \},
\]
and assume that an initial distribution \( \lambda \) has the property that
\[
\lambda(S_0) = 0.
\]

Then we have
\[
\mathbb{P}_\lambda(X_n < G_*) = \mathbb{P}(X''_{n} < G_*) = 1, \quad n = 0, 1, 2, \ldots.
\]

This means that the two crucial relations (10) and (11) are satisfied and so we may repeat step by step the reasoning in (12), but in reverse order. The theorem follows.

2.3. Proof of Theorem 2

By virtue of [12, Theorem 13.3.3] one obtains \( \alpha \)-mixing of \( \{X_n\} \) under stationary distribution \( \pi \) (for more details see e.g. [4] or [6]).

If \( \pi \circ f^{-1} \) is continuous then \( \{X_n\} \) admits a continuous phantom distribution function \( G \) by [7, Theorem 6]. Moreover, since \( \pi \circ f^{-1} \) is unbounded above, \( G_* = +\infty \) and the bad set \( B_0 \) given by (13) is empty. Hence our Theorem 1 states that \( G \) is a phantom distribution function for \( \{X_n\} \) under any initial distribution \( \lambda \). In particular for every \( s \in S \)
\[
\sup_{x \in \mathbb{R}^1} \left| \mathbb{P}_n(M_n \leq x) - G^n(x) \right| \rightarrow 0, \quad \text{as } n \rightarrow \infty.
\]
References

[1] Asmussen, S. (1998). Subexponential asymptotics for stochastic processes: extremal behavior, stationary distributions and first passage probabilities. *Ann. Appl. Probab.*, 8:354–374.

[2] Asmussen, S. (2003). *Applied Probability and Queues*. Springer, New York.

[3] Barrera, D., Peligrad, C., and Peligrad, M. (2016). Macroscopic models for long-range dependent network traffic. *Stochastic Process. Appl.*, 126:1885–1900.

[4] Bradley, R. (2007). *Introduction to strong mixing conditions*, volume 2. Kendrick Press, Heber City UT.

[5] Derriennic, Y. and Lin, M. (2003). The central limit theorem for markov chains started at a point. *Probab. Theory Relat. Fields*, 125:73–76.

[6] Doukhan, P. (1994). *Mixing: Properties and Examples*, volume 85 of *Lecture Notes in Statist.* Springer, New York.

[7] Doukhan, P., Jakubowski, A., and Lang, G. (2015). Phantom distribution functions for some stationary sequences. *Extremes*, 18:697–725.

[8] Jakubowski, A. (1991). Relative extremal index of two stationary processes. *Stochastic Process. Appl.*, 37:281–297.

[9] Jakubowski, A. (1993). An asymptotic independent representation in limit theorems for maxima of nonstationary random sequences. *Ann. Probab.*, 21:819–830.

[10] Jarner, S. and Roberts, G. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.*, 34:781–815.

[11] Leadbetter, M. R. (1983). Extremes and local dependence in stationary sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 65:291–306.

[12] Meyn, S. and Tweedie, R. (2009). *Markov Chains and Stochastic Stability*. 2ND Ed. Cambridge University Press, Cambridge.

[13] O’Brien, G. L. (1974). Limit theorems for the maximum term of a stationary process. *Ann. Probab.*, 2:54–545.

[14] O’Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Ann. Probab.*, 15:281–292.

[15] Roberts, G. O., Rosenthal, J., Segers, J., and Sousa, B. (2006). Extremal indices, geometric ergodicity of Markov chains and MCMC. *Extremes*, 9:213–229.
[16] Rootzén, H. (1988). Maxima and exceedances of stationary Markov chains. *Adv. Appl. Probab.*, 20:371-390.

[17] Volný, D. and Woodroofe, M. (2010). An example of non-quenched convergence in the conditional central limit theorem for partial sums of linear processes. In Berkes, I., Bradley, R., Dehling, H., Peligrad, M., and Tichy, R., editors, *Dependence in Probability, Analysis and Number Theory*, pages 317–322. Kendrick Press, Heber City UT.