Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER Stochastic Processes and their Applications 115 (2005) 579–591

stochastic processes and their applications

www.elsevier.com/locate/spa

The extremal behaviour over regenerative cycles for Markov additive processes with heavy tails

Niels Richard Hansen*, Anders Tolver Jensen

Department of Applied Mathematics and Statistics, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

Received 19 January 2003; received in revised form 22 July 2004; accepted 8 November 2004
Available online 30 November 2004

Abstract

We consider the reflection of an additive process with negative drift controlled by a Markov chain on a finite state space. We determine the tail behaviour of the distribution of the maximum over a regenerative cycle in the case with subexponential increments. Based on this, the asymptotic distribution of the running maximum is derived. Applications of the results to Markov modulated single server queueing systems are given.

Keywords: Extremes; Heavy tails; Markov additive processes; Markov chains; Markov modulated queues; Regeneration; Regular variation; Subexponential distributions

1. Introduction

Random walks and reflected random walks are important stochastic processes in several areas of applied probability. In dimensioning queueing systems with finite buffer capacity it is of great interest to study the probability of buffer overflow. Clearly, this is closely related to the study of the *actual waiting time* from the arrival of a customer until service starts, and for a stable *GI/G/1* queue this is equivalent to studying the distribution of a reflected random walk with negative drift, Asmussen

*Corresponding author. Tel.: +45 35 32 07 83; fax: +45 35 32 07 72.
E-mail address: richard@stat.ku.dk (N.R. Hansen).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2004.11.001

[4]. Another application of reflected random walks appears in biological sequence analysis. Here the best local similarity between two aligned sequences of random letters turns out to be precisely the maximum of a reflected random walk [15,9].

Iglehart [13] shows that the tail of the distribution of the maximal waiting time for a *GI/G/1* queue in the case of a light-tailed service time distribution is asymptotically exponential. Inspired for instance by problems from biological sequence analysis, Karlin and Dembo [15] derive similar results for a Markov controlled random walk with light-tailed increments. Independently, the corresponding result in the framework of Markov controlled queues was obtained by Asmussen and Perry [6].

The cycle maximum for reflected random walks with increment distributions that are subexponential was dealt with by Asmussen [3]—see also the paper by Heath et al. [12] for the regularly varying case. Similar results about the maximum over more general random intervals were treated by Foss and Zachary [11]. In Section 2 we extend the results of Asmussen [3] to the Markov controlled case, where the increments depend on an underlying finite state Markov chain. Assuming negative drift we show in Theorem 1 that if the (heaviest) increments are tail equivalent to a subexponential distribution H, the distribution of the maximum of the reflected Markov controlled additive process over a regenerative cycle is also tail equivalent to H. For technical reasons one needs H to belong to the slightly smaller class $\mathcal{S}^*$. In Corollary 2 we find that if H in addition is regularly varying the running maximum of the reflected process is asymptotically Fréchet distributed.

Finally, in Section 3, we deal with applications of the main result in the field of queueing theory. For a class of Markov modulated single-server queueing systems important processes such as the *residual workload* process may be sampled at random time points to obtain a reflected MAP. In this way we can use our result to study the asymptotic behaviour as time evolves of the tail of the maximal amount of work in the system.

2. Markov additive processes

2.1. Setup and main results

We consider a Markov chain $(J_n)_{n\geqslant 0}$ taking values in a finite state space E with transition probabilities governed by the transition matrix P. Furthermore, conditionally on $(J_n)_{n\geqslant 0}$, the process $(X_n)_{n\geqslant 1}$ is a sequence of independent random variables taking values in $\mathbb{R}$ such that the conditional distribution of X_n given $(J_n)_{n\geqslant 0}$ is $H_{J_{n-1}J_n}$ where H_{ij} for $i,j \in E$ is a matrix of probability measures on $\mathbb{R}$. These are called the increment distributions. We will from hereon use H_{ij} to denote both the probability measure and the distribution function. Let $\overline{H}_{ij}(x) = 1 - H_{ij}(x)$ denote the tail of the distribution function, put $F_{ij} = H_{ij}p_{ij}$ for $i,j \in E$ and $\overline{F}_{ij}(x) = \overline{H}_{ij}(x)p_{ij}$. We will throughout only consider the non-lattice case, that is H_{ij} is *not* concentrated on a grid:

$$\{d_j - d_i + nd \mid n \in \mathbb{Z}\},$$

for some constants $d, d_i > 0$ and $i, j \in E$, see Alsmeyer [2]. We will use $\mathbb{P}_i$ to denote the probability measure where $J_0 = i$ and $\mathbb{E}_i$ to denote expectation w.r.t. $\mathbb{P}_i$. Denote by $\mathbb{1}$ the column vector of ones.

Put $S_0 = 0$ and for $n \geqslant 1$

$$S_n = \sum_{k=1}^{n} X_k, \tag{1}$$

together with $W_0 = 0$ and recursively for $n \geqslant 1$

$$W_n = (W_{n-1} + X_n)^+. \tag{2}$$

We call $(J_n, S_n)_{n \geqslant 0}$ a Markov Additive Process—from hereon abbreviated MAP—and $(J_n, W_n)_{n \geqslant 0}$ is the corresponding *reflected* MAP. Both processes are Markov processes on $E \times \mathbb{R}$ and $E \times [0, \infty)$, respectively.

We are interested in the case where the H_{ij}'s have subexponential tails. Hence let $\mathscr{S}$ denote the set of subexponential distributions on $[0, \infty)$ and let $\mathscr{S}^*$ denote the smaller set of distributions, G, on $[0, \infty)$ with finite expectation defined by the requirement

$$\lim_{x \to \infty} \int_0^x \overline{G}(y) \frac{\overline{G}(x-y)}{\overline{G}(x)} \, \mathrm{d}y = 2 \int_0^\infty \overline{G}(y) \, \mathrm{d}y,$$

see [16].

Assume for the rest of this paper that there exists $H \in \mathscr{S}^*$ such that for all $i, j \in E$

$$\lim_{x \to \infty} \frac{\overline{F}_{ij}(x)}{\overline{H}(x)} = \gamma_{ij} \tag{3}$$

for some $\gamma_{ij} \geqslant 0$ and at least one $\gamma_{ij} > 0$. Let Γ denote the matrix $(\gamma_{ij})_{i,j \in E}$. Using that the convergence $\overline{H}(x-y)/\overline{H}(x) \to 1$ is uniform in y on compact sets [8, Lemma 1.3.5] we get that

$$\lim_{x \to \infty} \frac{\overline{F}_{ij}(x-y)}{\overline{H}(x)} = \gamma_{ij} \tag{4}$$

uniformly in y on compact sets. Assume that P is irreducible with invariant probability measure π and put

$$\mu_{ij} = \int y F_{ij}(\mathrm{d}y) = \int y H_{ij}(\mathrm{d}y) p_{ij}, \tag{5}$$

$$\mu = \sum_{i,j \in E} \pi_i \mu_{ij}. \tag{6}$$

We will finally assume that $\mu < 0$, so the additive process $(S_n)_{n \geqslant 0}$ tends to $-\infty$ by ergodicity of $(J_n, X_n)_{n \geqslant 0}$. Let

$$\tau_-^n = \inf\{k > \tau_-^{n-1} \mid S_k \leqslant S_{\tau_-^{n-1}}\}, \quad (\tau_-^0 = 0)$$

be the successive descending ladder times, which are finite almost surely due to the negative drift. Then $(J_{\tau_-^n})_{n\geqslant 0}$ is a Harris recurrent Markov chain [2],[1] in particular there exists $i_0 \in E$ such that for all $i \in E$

$$\mathbb{P}_i(J_{\tau_-^n} = i_0 \text{ for some } n) = 1.$$

Consequently the stopping time

$$\sigma = \inf\{n \geqslant 1 \mid W_n = 0,\ J_n = i_0\} \tag{7}$$

is finite $\mathbb{P}_i$-almost surely for all $i \in E$. The time σ is a regeneration time for $(J_n, W_n)_{n\geqslant 0}$. We want to study the maximum of $(W_n)_{n\geqslant 0}$ up to time σ;

$$\mathscr{M}_\sigma = \max_{0\leqslant n\leqslant \sigma} W_n. \tag{8}$$

Theorem 1. *Under assumptions* (3) *and* $\mu < 0$

$$\mathbb{P}_{i_0}(\mathscr{M}_\sigma > x) \sim \overline{H}(x)\mathbb{E}_{i_0}(\sigma)\,\pi\Gamma\mathbb{1} \tag{9}$$

for $x \to \infty$.

The qualitative content of Theorem 1 is that the distribution of $\mathscr{M}_\sigma$ is tail-equivalent to the heaviest tail of the increment distributions. In addition Theorem 1 provides an explicit and intuitive representation of the constant of proportionality as the mean cycle length times a weighted average of the tail indices Γ.

Condition (3) may seem technical and only tied up with the techniques of the proof. However, Foss and Zachary [11, Theorem 1(ii)] state that for the corresponding result to hold in the random walk case the condition is also necessary. Since the ordinary random walk is covered by Theorem 1 the condition is certainly also necessary for this theorem to be valid in general.

The result obtained in Theorem 1 should be compared to the (tail of the) time invariant distribution of W_n, which is known to be tail equivalent to the integrated tail of H when e.g. $H \in \mathscr{S}^*$ [14]. Thus in the heavy-tailed framework the distribution of the cycle maximum and the invariant distribution of W_n are not in general tail equivalent in contrast to the light tailed case. In this case, as remarked by Asmussen [3] for the reflected random walk, the process $(W_n)_{n\geqslant 0}$ has extreme value index 0, and the analysis of the asymptotic behaviour of the running maximum of W_n cannot be related directly to the tail of the invariant distribution. However, it is possible to derive the asymptotic behaviour on the basis of the tail of the distribution of $\mathscr{M}_\sigma$ taking advantage of the regenerative structure of the reflected MAP.

If we let $\mathscr{M}_n = \max_{0\leqslant k\leqslant n} W_n$, we obtain the following corollary as a consequence of the regenerative structure of $(J_n, W_n)_{n\geqslant 0}$ together with Proposition VI.4.10 in Asmussen [4].

---

[1] The paper by Alsmeyer gives a result for a general state space E. In the case considered here with E finite, a simple proof can be given.

Corollary 2. *If $\overline{H}$ is regularly varying at infinity with exponent $-\alpha$, then for any $i \in E$*

$$\mathbb{P}_i\left(\frac{\mathcal{M}_n}{b_n} \leqslant x\right) \to \exp(-x^{-\alpha}), \quad n \to \infty \tag{10}$$

with b_n satisfying

$$n\overline{H}(b_n) \to \frac{1}{\pi\Gamma\mathbb{1}}, \quad n \to \infty.$$

One could for instance choose $b_n = H^{\leftarrow}(1 - 1/(n\pi\Gamma\mathbb{1}))$.

Note that $H^{\leftarrow}$ denotes the *generalised inverse* of H defined as

$$H^{\leftarrow}(t) = \inf\{x \in \mathbb{R} \mid H(x) \geqslant t\}.$$

The proof of Theorem 1 is divided into a number of lemmas. The idea in the proof is to use the ''one big jump'' heuristic for subexponential distributions, thus an extreme value for the reflected MAP occurs due to one extreme X_n-value. We split the extreme event $(\mathcal{M}_\sigma > x)$ into the event where the jump to a level above x happens from an intermediate level in $[x_0, x]$, $x_0 < x$, and the event where the jump happens from a level below x_0. Then we show that the probability of the first event is asymptotically negligible and that the probability of the last event has the desired asymptotic behaviour. To deal with the former we use some non-trivial downcrossing results due to Asmussen [3] in the random walk set-up—see also Foss and Zachary [10]. This argument is developed in Lemmas 3–5 of Section 2.2. In Lemma 6 we give the asymptotics when jumps occur from levels below some x_0 and Lemma 7 shows, using the downcrossing results, that the other probability vanishes asymptotically.

2.2. Proofs

For a matrix, A, of σ-finite measures, A_{ij}, on $\mathbb{R}$ let $||A||$ denote the matrix whose ij'th element is the total mass $A_{ij}(\mathbb{R})$. Given another matrix, B, of σ-finite measures on $\mathbb{R}$ we define the convolution product of A and B by

$$(A * B)_{ij} = \sum_{k \in E} A_{ik} * B_{kj}.$$

Let R be the matrix of occupation measures for the MAP, i.e. let

$$R_{ij}(D) = \sum_{n=0}^{\infty} \mathbb{P}_i(J_n = j, S_n \in D).$$

It is easily seen that $R = \sum_{n=0}^{\infty} F^{*n}$.

Denote by (J_n^*, S_n^*) the time reversal of (J_n, S_n), which is a MAP with transition probabilities $F_{ij}^* = \pi_j F_{ji}/\pi_i$. Let G_- be the matrix of descending ladder height distributions for (J_n, S_n), let G_+^* be the matrix of ascending ladder height distributions for (J_n^*, S_n^*) and define the matrix ${}^{\#}G_+$ by ${}^{\#}G_{+,ij} = \pi_i G_{+,ji}^*/\pi_j$.

Define the corresponding matrices of renewal measures by ${}^{\#}U_{+} = \sum_{n=0}^{\infty}({}^{\#}G_{+})^{*n}$ and $U_{-} = \sum_{n=0}^{\infty} G_{-}^{*n}$, and let

$$R_{-,ij}(D) = \mathbb{E}_i \sum_{n=0}^{\tau_{-}-1} 1_{(J_n=j, S_n \in D)}$$

be the occupation measure up to time $\tau_{-} = \tau_{-}^{1} = \inf\{n \geqslant 1 \mid S_n \leqslant 0\}$. From Asmussen [4, Proposition XI.2.13] it follows that ${}^{\#}U_{+} = R_{-}$ and by the negative drift assumption it follows from [4, Proposition XI.2.14] that the spectral radius of $||{}^{\#}G_{+}||$ is strictly less than one. Consequently ${}^{\#}U_{+}$ is a matrix of finite measures and we conclude that in particular

$$\mathbb{E}_i(\tau_{-}) = \sum_{j \in E} R_{-,ij}(\mathbb{R}_{+}) < \infty.$$

It is clear that the process $(\tau_{-}^{n}, J_{\tau_{-}^{n}})_{n \geqslant 0}$ is a Markov renewal process and due to Harris recurrence of $(J_{\tau_{-}^{n}})_{n \geqslant 0}$ it follows that

$$\mathbb{E}_{i_0}(\sigma) = \frac{\sum_{i \in E} v_i \mathbb{E}_i(\tau_{-})}{v_{i_0}} < \infty,$$

where $(v_i)_{i \in E}$ is the invariant distribution of $(J_{\tau_{-}^{n}})_{n \geqslant 0}$, cf. [4, Proposition VII.4.2]. This is important since we need to consider a regenerative process with a cycle length distribution that has finite mean.

Wiener–Hopf theory for MAPs [4, Theorem XI.2.12] gives the following factorisation identity:

$$I - F = \left(I - {}^{\#}G_{+}\right) * (I - G_{-})$$

which for the occupation measure amounts to

$$R(D) = U_{-} * {}^{\#}U_{+}(D) \tag{11}$$

valid for $D \subseteq \mathbb{R}$ a bounded set.

We will need the following non-lattice version of the Markov renewal theorem, cf. Alsmeyer [1].

Lemma 3. *If $\mu < 0$ the matrix $||G_{-}||$ is stochastic and it holds that*

$$R_{ij}((z, z+y]) \to y \frac{\pi_j}{|\mu|}, \quad z \to -\infty$$

for all $y > 0$.

For $x > 0$ let

$$N_{\sigma}(x) = \sum_{n=0}^{\sigma-1} 1_{(W_n > x, W_{n+1} \leqslant x)}$$

be the number of downcrossings of level x before time σ. Let

$$\rho(x) = \inf\{n \geqslant 1 \mid S_n \leqslant -x\}$$

be the time of the first downcrossing of the MAP of level $-x$, and if also $y>0$ denote by

$$N(x,y) = \sum_{n=0}^{\rho(x+y)-1} 1_{(S_n>-x, S_{n+1}\leqslant -x)}$$

the number of downcrossings of level $-x$ *before* the barrier $-(x+y)$ is reached. Note that $N(x,y) \nearrow N(x) = \sum_{n=0}^{\infty} 1(S_n > -x, S_{n+1} \leqslant -x)$ for $y \to \infty$, where $N(x)$ is the *total* number of downcrossings of level $-x$. Let in the following m_- be the matrix $(m_{-,ij})_{i,j\in E}$ given by

$$m_{-,ij} = \int_0^\infty F_{ij}(-z)\,\mathrm{d}z.$$

Lemma 4. *When $\mu<0$ it holds for all $i \in E$ that*

$$\mathbb{E}_i N(x) \to \frac{\pi m_- \mathbb{1}}{|\mu|}, \quad x \to \infty$$

and the convergence

$$\mathbb{E}_i N(x,y) \nearrow \mathbb{E}_i N(x), \quad y \to \infty$$

is uniform in x.

Proof. By conditioning on the value of (J_n, S_n) for $n \geqslant 0$ we get using Lemma 3 that for all $i \in E$

$$\begin{aligned}
\mathbb{E}_i N(x) &= \sum_{j,k\in E} \int_{-x}^{\infty} F_{jk}(-(x+y)) R_{ij}(\mathrm{d}y) \\
&= \sum_{j,k\in E} \int_{0}^{\infty} F_{jk}(-z) R_{ij}(\mathrm{d}z - x) \\
&\to \sum_{j,k\in E} \frac{\pi_j}{|\mu|} \int_{0}^{\infty} F_{jk}(-z)\,\mathrm{d}z, \quad x \to \infty \\
&= \frac{\pi m_- \mathbb{1}}{|\mu|}.
\end{aligned}$$

For the second result, use the strong Markov property to obtain

$$\begin{aligned}
&0 \leqslant \mathbb{E}_i N(x) - \mathbb{E}_i N(x,y) \\
&= \mathbb{E}_i \sum_{n=\rho(x+y)}^{\infty} 1_{(S_n>-x, S_{n+1}\leqslant -x)} \\
&\leqslant \max_j \mathbb{P}_j(\mathscr{M} > y) \sup_z \mathbb{E}_j(N(z)),
\end{aligned}$$

where $\mathscr{M} = \max_{n\geqslant 0} S_n$. Since $\mathscr{M} < \infty$ almost surely the result follows if $\sup_z \mathbb{E}_j(N(z)) < \infty$. By the negative drift $U_{-,ij}$ is finite on compact sets, hence from (11) we conclude that $R_{jk}([x,\infty)) < \infty$ for all x. Lemma 3 then implies that there exists α such that $R_{jk}([x, x+1)) \leqslant \alpha$ for all x, hence

$$\begin{aligned}\mathbb{E}_j(N(z)) &= \sum_{k,l\in E}\int_0^\infty F_{kl}(-x)R_{jk}(\mathrm{d}x - z)\\ &\leqslant \sum_{k,l\in E}\sum_{n=0}^\infty F_{kl}(-n)R_{jk}([n-x, n-x+1))\\ &\leqslant \alpha\sum_{k,l\in E}\sum_{n=0}^\infty F_{kl}(-n).\end{aligned}$$

The right-hand side is finite and independent of z, hence we see that $\sup_z \mathbb{E}_j(N(z)) < \infty$. $\square$

Lemma 5. *Under assumptions* (3) *and* $\mu < 0$

$$\lim_{x\to\infty}\frac{\mathbb{E}_{i_0}N_\sigma(x)}{\overline{H}(x)} = \frac{\mathbb{E}_{i_0}(\sigma)}{|\mu|}\,\pi\Gamma\mathbb{1}\pi m_-\mathbb{1}.$$

Proof. By Asmussen [4, Proposition XI.2.11] the reflected MAP (J_n, W_n) has an invariant distribution λ which coincides with the distribution of $(J_0^*, \mathscr{M}^*)$ where $\mathscr{M}^* = \max_{n\geqslant 0} S_n^*$ and J_0^* has distribution π. We find that

$$\lim_{x\to\infty}\frac{\overline{F}_{ij}^*(x)}{\overline{H}(x)} = \lim_{x\to\infty}\frac{\pi_j\overline{F}_{ji}(x)/\pi_i}{\overline{H}(x)} = \pi_j\gamma_{ji}/\pi_i.$$

With $\hat{H}(x) = \int_x^\infty \overline{H}(y)\,\mathrm{d}y$ the integrated tail of H we get from Jelenković and Lazar [14, Theorem 4], using that $H \in \mathscr{S}^*$ implies $\hat{H} \in \mathscr{S}$, that

$$\frac{\lambda(i,(x;\infty))}{\hat{H}(x)} = \frac{\mathbb{P}_\pi(J_0^* = i, \mathscr{M}^* > x)}{\hat{H}(x)} \to \frac{\pi_i}{|\mu|}\,\pi\Gamma\mathbb{1}, \quad x\to\infty.$$

Using that $H \in \mathscr{S}^*$ Corollary 1 of Asmussen et al. [5] implies that for all $i,j \in E$

$$\frac{\int_x^\infty F_{ij}(x-y)\lambda(i,\mathrm{d}y)}{\overline{H}(x)} \to \frac{\pi\Gamma\mathbb{1}}{|\mu|}\pi_i\int_0^\infty F_{ij}(-z)\,\mathrm{d}z, \quad x\to\infty. \tag{12}$$

Since σ is a regeneration time for (J_n, W_n), its time-invariant distribution may be represented as

$$\lambda(\{i\}\times[0;t]) = \frac{1}{\mathbb{E}_{i_0}(\sigma)}\mathbb{E}_{i_0}\left(\sum_{n=0}^{\sigma-1} 1_{(J_n=i, W_n\leqslant t)}\right),$$

hence

$$\mathbb{E}_{i_0}N_\sigma(x) = \mathbb{E}_{i_0}(\sigma)\sum_{i,j\in E}\int_x^\infty F_{ij}(x-y)\lambda(i,\mathrm{d}y),$$

Summing over i,j in (12) we conclude that

$$\frac{\mathbb{E}_{i_0}N_\sigma(x)}{\overline{H}(x)} \to \frac{\mathbb{E}_{i_0}(\sigma)}{|\mu|}\,\pi\Gamma\mathbb{1}\,\pi m_-\mathbb{1}, \quad x\to\infty. \quad \square$$

Note that $\mathbb{E}_{i_0}(\sigma)$ can be expressed explicitly in terms of the invariant distribution for $(J_n, W_n)_{n\geqslant 0}$,

$$\mathbb{E}_{i_0}(\sigma) = \frac{1}{\pi_{i_0}\mathbb{P}_{i_0}(\mathscr{M}^* = 0)} = \frac{1}{\pi_{i_0}\left(1 - \sum_{j\in E}||G_+^*(i_0,j)||\right)}.$$

Put

$$\tau(x) = \inf\{n\geqslant 1 \mid W_n > x\},$$

so $(\mathscr{M}_\sigma > x) = (\tau(x) < \sigma)$ and let for $x_0 < x$, $y_0 \geqslant 0$ and $j \in E$,

$$A(x, x_0, y_0) = (\tau(x) < \sigma,\, W_{\tau(x)} > x + y_0,\, W_{\tau(x)-1} < x_0).$$

That is, $A(x, x_0, y_0)$ is the event that the W_n-process will exceed x before time σ, and when doing so the W_n-process jumps from a value below x_0 to a value above $x + y_0$.

Lemma 6. *Under assumptions* (3) *and* $\mu < 0$

$$\lim_{x_0\to\infty}\lim_{x\to\infty}\frac{\mathbb{P}_{i_0}(A(x, x_0, y_0))}{\overline{H}(x)} = \mathbb{E}_{i_0}(\sigma)\,\pi\Gamma\mathbb{1} \tag{13}$$

for all $y_0 \geqslant 0$.

Proof. Put $\sigma(x) = \sigma \wedge \tau(x)$. Then

$$\begin{aligned}\mathbb{P}_{i_0}(A(x, x_0, y_0)) &= \sum_{n=1}^{\infty} \mathbb{P}_{i_0}(W_n > x + y_0,\, W_{n-1} < x_0,\, \sigma(x) \geqslant n)\\ &= \sum_{i,j\in E}\sum_{n=1}^{\infty} \mathbb{P}_{i_0}(W_n > x + y_0,\, W_{n-1} < x_0,\, J_{n-1} = i, J_n = j, \sigma(x) \geqslant n)\\ &= \sum_{i,j\in E}\sum_{n=1}^{\infty} \int_0^{x_0} \overline{F}_{ij}(x + y_0 - y)\mathbb{P}_{i_0}(W_{n-1} \in \mathrm{d}y | J_{n-1} = i, \sigma(x) \geqslant n)\\ &\quad \times \mathbb{P}_{i_0}(J_{n-1} = i, \sigma(x) \geqslant n).\end{aligned}$$

Using that the convergence in (4) is uniform for $y \in [0, x_0]$ and that $\sigma(x) \nearrow \sigma$ for $x \to \infty$ we get that

$$\frac{\mathbb{P}_{i_0}(A(x, x_0, y_0))}{\overline{H}(x)} \to \sum_{i,j\in E}\sum_{n=1}^{\infty} \gamma_{ij}\mathbb{P}_{i_0}(W_{n-1} < x_0, J_{n-1} = i, \sigma \geqslant n).$$

Letting $x_0 \to \infty$, the result follows by using that

$$\sum_{n=1}^{\infty} \mathbb{P}_{i_0}(J_{n-1} = i, \sigma \geqslant n) = \sum_{n=0}^{\sigma-1} \mathbb{P}_{i_0}(J_n = i) = \pi_i\mathbb{E}_{i_0}(\sigma). \quad \square$$

For $x_0 < x$ let $A(x, x_0) = A(x, x_0, 0) = (\tau(x) < \sigma, W_{\tau(x)-1} < x_0)$ and put $B(x, x_0) = (\tau(x) < \sigma, W_{\tau(x)-1} \geqslant x_0)$.

Lemma 7. *Under assumptions* (3) *and* $\mu<0$

$$\lim_{x_0\to\infty}\limsup_{x\to\infty}\frac{\mathbb{P}_{i_0}(B(x,x_0))}{\overline{H}(x)}=0. \tag{14}$$

Proof. Put $\kappa=(\pi m_-\mathbb{1})/|\mu|$, let $\varepsilon>0$ be given, and choose x and y_0 according to Lemma 4 so that

$$\mathbb{E}_j(N(y-x,x))\geqslant\kappa-\varepsilon$$

for $y\geqslant x+y_0$ and all $j\in E$. On the event $(\tau(x)<\sigma)$, the number of times W_n crosses level x from above after time $\tau(x)$ and before time σ is larger than the number of times W_n crosses level x from above after time $\tau(x)$ and before hitting zero. Hence the strong Markov property of $(J_n,W_n)_{n\geqslant 0}$ gives

$$\begin{aligned}
&\mathbb{E}_{i_0}(N_\sigma(x);A(x,x_0))\\
&\quad\geqslant\mathbb{E}_{i_0}(\mathbb{E}_{J_{\tau(x)}}(N(W_{\tau(x)}-x,x));A(x,x_0))\\
&\quad=\sum_{j\in E}\int_x^\infty\mathbb{E}_j(N(y-x,x))\mathbb{P}_{i_0}(W_{\tau(x)}\in\mathrm{d}y,J_{\tau(x)}=j,A(x,x_0))\\
&\quad\geqslant\sum_{j\in E}\int_{x+y_0}^\infty\mathbb{E}_j(N(y-x,x))\mathbb{P}_{i_0}(W_{\tau(x)}\in\mathrm{d}y,J_{\tau(x)}=j,A(x,x_0))\\
&\quad\geqslant(\kappa-\varepsilon)\mathbb{P}_{i_0}(A(x,x_0,y_0)).
\end{aligned}$$

Using Lemma 6

$$\lim_{x_0\to\infty}\liminf_{x\to\infty}\frac{\mathbb{E}_{i_0}(N_\sigma(x);A(x,x_0))}{\overline{H}(x)}\geqslant\mathbb{E}_{i_0}(\sigma)(\kappa-\varepsilon)\,\pi\Gamma\mathbb{1}.$$

Since $\varepsilon>0$ was arbitrary and $1_{B(x,x_0)}\leqslant N_\sigma(x)1_{B(x,x_0)}=N_\sigma(x)-N_\sigma(x)1_{A(x,x_0)}$ we get from Lemma 5 that

$$\lim_{x_0\to\infty}\limsup_{x\to\infty}\frac{\mathbb{P}_{i_0}(B(x,x_0))}{\overline{H}(x)}\leqslant\mathbb{E}_{i_0}(\sigma)\pi\Gamma\mathbb{1}\left(\frac{\pi m_-\mathbb{1}}{|\mu|}-\kappa\right)=0.\qquad\square$$

Proof of Theorem 1. From the identity

$$\mathbb{P}_{i_0}(\mathscr{M}_\sigma>x)=\mathbb{P}_{i_0}(A(x,x_0))+\mathbb{P}_{i_0}(B(x,x_0))$$

result (9) follows immediately from Lemmas 6 and 7. $\square$

3. Examples

In this section we discuss possible applications of Theorem 1. With focus on a Markov-modulated queueing system with a single server we explain how the investigation of certain performance characteristics are related to the study of a reflected MAP. For concrete examples of single server queues operating in Markovian environments we refer to the work by Dudin and Klimenok [7].

3.1. Markov-modulated queues

Consider a single server queueing system where the server handles incoming work at unit rate. Denote by $\tilde{S}_t$ the total amount of work that has arrived up to time t. In terms of the *netput process*, $S_t = \tilde{S}_t - t$, the *residual workload* present in the queue at time t may be expressed as

$$V_t = S_t - \inf_{s \leqslant t} S_s.$$

It is valuable to study the behaviour of V_t since many interesting problems concerning, e.g. buffer overflow and the actual waiting time of "customers" arriving to the queue can be given formulations in terms of the process $(V_t)_{t \geqslant 0}$, Asmussen [4, Chapter IV.1].

Suppose that we can find sampling times ρ_n, $n \geqslant 0$, and a Markov chain $(J_n)_{n \geqslant 0}$ such that $(S_{\rho_n}, J_n)_{n \geqslant 0}$ is a MAP. Then the inequality

$$W_n \leqslant V_{\rho_n}$$

involving the reflected MAP $(W_n, J_n)_{n \geqslant 0}$, holds, but in general equality does not hold. However, under the additional assumption that $t \to S_t$ is monotone on each interval $[\rho_n, \rho_{n+1}]$ it can easily be verified that $W_n = V_{\rho_n}$ and

$$M_t := \sup_{s \leqslant t} V_s = \max_n V_{\rho_n \wedge t} \left(= \max\{ \max_{n: \rho_n \leqslant t} W_n, V_t \} \right).$$

Occasionally it is convenient to consider instead sampling times ρ_n, $n \geqslant 0$, such that $(S_{\rho_n -}, J_n)_{n \geqslant 0}$ is a MAP. Also in this case the monotonicity assumption above implies that $W_n = V_{\rho_n -}$ and hence

$$M_n := \max_{1 \leqslant k \leqslant n} V_{\rho_k -} = \max_{1 \leqslant k \leqslant n} W_k.$$

These identities relate the study of M_t and M_n to the study of the running maximum of a reflected MAP. Moreover, if the increments of the MAP can be shown to be heavy-tailed the asymptotics of M_t and M_n as $t \to \infty$ resp. $n \to \infty$ can be derived from Theorem 1. If the increments have regularly varying tails, say, the asymptotics is given by Corollary 2.

Example 8 (*Markov-modulated single-server queue*). Suppose that customers arrive to a single-server queue at times $(\rho_n)_{n \geqslant 0}$ and denote by B_n the service time of customer n. If $X_n = B_n - (\rho_{n+1} - \rho_n)$ are conditionally independent given a finite-state Markov chain $(J_n)_{n \geqslant 0}$ such that X_n given $(J_n)_{n \geqslant 0}$ depends only on (J_{n-1}, J_n) then $(S_{\rho_n -}, J_n)_{n \geqslant 0}$ is a MAP. Since obviously $t \to S_t$ is decreasing on $[\rho_n, \rho_{n+1})$ the actual waiting time, $V_{\rho_n -}$, of customer n equals the value, W_n, of the reflected MAP with the increments $(X_n)_{n \geqslant 0}$.

Example 9 (*Markov-controlled fluid-models*). Instead of assuming that work arrives in jumps we can assume that work flows continuously into to the queue with a stochastic rate, λ_n, that changes at the times $(\rho_n)_{n \geqslant 0}$. If the increments $X_n = (\lambda_n - 1)(\rho_{n+1} - \rho_n)$ are conditionally independent given a finite-state Markov chain $(J_n)_{n \geqslant 0}$ such that X_n given $(J_n)_{n \geqslant 0}$ depends only on (J_{n-1}, J_n) then $(S_{\rho_n}, J_n)_{n \geqslant 0}$

is a MAP. Since $t \to S_t$ is linear, hence monotone, on each $[\rho_n, \rho_{n+1}]$ we conclude that

$$M_t = \max_n V_{\rho_n \wedge t}$$

with $(V_{\rho_n})_{n \geqslant 0} = (W_n)_{n \geqslant 0}$ the reflected MAP with increments $(X_n)_{n \geqslant 0}$.

In both examples the residual workload process $(V_t)_{t \geqslant 0}$ can be regarded as controlled by an underlying *semi*-Markov process with the durations between jumps being $(\rho_{n+1} - \rho_n)_{n \geqslant 0}$. The increments in the second example can then be heavy tailed if either the rate, λ_n, or the durations, $\rho_{n+1} - \rho_n$, (when $\lambda_n > 1$) are heavy tailed. It could be interesting to be able to handle a semi-Markov modulated queue where the arrival intensity is > 1 during time periods that are heavy tailed. In particular, we have taken an interest in the residual workload of a queue where the arrival process is a doubly stochastic Poisson process with an intensity that depends on the value of a finite state semi-Markov chain. The presence of heavy tails is then introduced by letting the duration distributions of the semi-Markov chain be heavy-tailed so that small positive increments can aggregate over long periods of time. Though such a model does not satisfy the monotonicity assumption presented above we expect that results about e.g. the workload process can be obtained along the same lines as in the preceding examples. Unfortunately a direct translation of Theorem 1 is not possible.

4. Concluding remarks

Theorem 1 has intrinsic value as a generalisation of the corresponding result for random walks with heavy tailed increments. In addition, it allows us to analyse certain continuous time processes which contain an embedded MAP obtained by sampling at random time points. The Achilles' heel is the path regularity conditions that the processes must satisfy in between the sampling points. Unfortunately the monotonicity assumption suggested in Section 3 rules out several interesting examples as described above. It is an ongoing work to extend our results so that we can handle more general continuous time processes, e.g. semi-Markov modulated queues.

Acknowledgements

We would like to thank Søren Asmussen for giving us access to the revised version of *Applied Probability and Queues* and for drawing our attention to the paper by Foss and Zachary [10]. We would also like to thank MaPhySto for organising and especially Gennady Samorodnitsky for giving the excellent concentrated advanced course "Long Range Dependence, Heavy Tails and Rare Events" in the spring of 2002, which greatly inspired this work. Finally we thank the referees for valuable suggestions.

References

[1] G. Alsmeyer, On the Markov renewal theorem, Stochastic Proc. Appl. 50 (1994) 37–56.
[2] G. Alsmeyer, The ladder variables of a Markov random walk, Probab. Math. Statist. 20 (1) (2000) 151–168.
[3] S. Asmussen, Subexponential asymptotics for stochastic processes: extremal behavior, stationary distributions and first passage probabilities, Ann. Appl. Probab. 8 (2) (1998) 354–374.
[4] S. Asmussen, Applied Probability and Queues, in: Applications of Mathematics, Stochastic Modelling and Applied Probability, vol. 51, Springer, New York, 2003.
[5] S. Asmussen, V. Kalashnikov, D. Konstantinides, C. Klüppelberg, G. Tsitsiashvili, A local limit theorem for random walk maxima with heavy tails, Statist. Probab. Lett. 56 (4) (2002) 399–404.
[6] S. Asmussen, D. Perry, On cycle maxima, first passage problems and extreme value theory for queues, Comm. Statist. Stochastic Models 8 (3) (1992) 421–458.
[7] A. Dudin, V. Klimenok, Multi-dimensional quasitoeplitx Markov chains, J. Appl. Math. Stochastic Anal. 12 (4) (1998) 393–415.
[8] P. Embrechts, C. Klüppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, in: Applications of Mathematics, Stochastic Modelling and Applied Probability, vol. 33, Springer, Berlin, 1997.
[9] W.J. Ewens, G.R. Grant, Statistical methods in bioinformatics: an introduction, in: Statistics for Biology and Health, Springer, New York, 2001.
[10] S. Foss, S. Zachary, Asymptotics for the maximum of a modulated random walk with heavy-tailed increments, Anal. Methods Appl. Probab. 207 (2002) 37–52.
[11] S. Foss, S. Zachary, The maximum on a random time interval of a random walk with long-tailed increments and negative drift, Ann. Appl. Probab. 13 (1) (2003) 37–53.
[12] D. Heath, S. Resnick, G. Samorodnitsky, Patterns of buffer overflow in a class of queues with long memory in the input stream, Ann. Appl. Probab. 7 (4) (1997) 1021–1057.
[13] D.L. Iglehart, Extreme values in the $GI/G/1$ queue, Ann. Math. Statist. 43 (1972) 627–635.
[14] P.R. Jelenković, A.A. Lazar, Subexponential asymptotics of a Markov-modulated random walk with queueing applications, J. Appl. Probab. 35 (2) (1998) 325–347.
[15] S. Karlin, A. Dembo, Limit distributions of maximal segmental score among Markov-dependent partial sums, Adv. Appl. Probab. 24 (1) (1992) 113–140.
[16] C. Klüppelberg, Subexponential distributions and integrated tails, J. Appl. Probab. 25 (1) (1988) 132–141.