Moment conditions for a sequence with negative drift to be uniformly bounded in L^r

Robin Pemantle[1] and Jeffrey S. Rosenthal[2]

(August 27, 1998.)

ABSTRACT:

Suppose a sequence of random variables $\{X_n\}$ has negative drift when above a certain threshold and has increments bounded in L^p. When $p > 2$ this implies that $\mathbf{E}X_n$ is bounded above by a constant independent of n and the particular sequence $\{X_n\}$. When $p \leq 2$ there are counterexamples showing this does not hold. In general, increments bounded in L^p lead to a uniform L^r bound on X_n^+ for any $r < p - 1$, but not for $r \geq p - 1$. These results are motivated by questions about stability of queueing networks.

Keywords: L^p, p^{th} moments, supermartingale, martingale, linear boundary, Lyapunov function, stochastic adversary, queueing networks.

Subject classification: Primary: 60G07; Secondary: 60F25.

---

[1]Department of Mathematics, University of Wisconsin–Madison, Van Vleck Hall, 480 Lincoln Drive, Madison, WI 53706. Internet: pemantle@math.wisc.edu. Research supported in part by NSF grants DMS-9300191 and DMS-9803249.

[2]Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: jeff@utstat.toronto.edu. Research supported in part by NSERC of Canada.

1. Introduction.

Let $X_0, X_1, \ldots$ be a sequence of real-valued random variables. We wish to find a condition, along the lines of behaving like a supermartingale when sufficiently large, that will guarantee $\sup_n \mathbf{E}X_n < \infty$. In particular, we do not wish to assume any special properties of the increments such as independence, r-dependence, Markov property, symmetry, discreteness or nondiscreteness. Under what conditions can we guarantee that $\sup_n \mathbf{E}(X_n) < \infty$?

Specifically, we suppose that for some $a > 0$ and some J we have

$$\mathbf{E}(X_{n+1} - X_n \,|\, X_0, \ldots, X_n) \leq -a \quad \text{on the event} \quad \{X_n > J\} \tag{C1}$$

for all n. That is, the process has *negative drift* whenever it is above the point J. This condition alone says nothing about possible large jumps out of the interval $(-\infty, J)$, so we also assume that for some $p \geq 1$ and some $V < \infty$ we have

$$\mathbf{E}(|X_{n+1} - X_n|^p \,|\, X_0, \ldots, X_n) \leq V \tag{C2}$$

for all n. That is, the process has *increments with bounded* p^{th} *moments*. Conditions (C1) and (C2) are meant to characterize the behavior of sequences attracted to a basin, for example a nonnegative Lyapunov function of a Markov chain, which always decreases in expectation except when it is already small. Do these two conditions together imply that $\sup_n \mathbf{E}(X_n) < \infty$?

When $p = 1$ the answer is no. Although an honest supermartingale has $\mathbf{E}X_n \leq \mathbf{E}X_0$ for all n, a process behaving like a supermartingale above J may have $\mathbf{E}X_n$ unbounded. More surprisingly, the answer is still no for $p = 2$. However, if one assumes (C1) and (C2) with $p > 2$, then necessarily $\sup_n \mathbf{E}X_n < \infty$. More generally, we show (Theorem 1) that (C1) and (C2) imply that $\sup_n \mathbf{E}\left((X_n^+)^r\right) < \infty$ whenever $r < p - 1$, and that this bound on r in terms of p (or p in terms of r) is sharp. Furthermore, our bounds are may be explicitly computed and depend only on the parameters a, J, p, V, and r.

Our results are motivated by questions about queueing networks. Specifically, several authors (Borodin et al., 1996, 1998; Andrews et al., 1996) consider network loads under the influence of a *stochastic adversary*. Here X_n is the load of the network at time n. The adversary may add new packets to the network in virtually any manner, subject only to a load condition which leads to (C1) plus a moment condition such as (C2). (The load condition corresponds to the statement that, once the network is operating at full capacity, it processes packets more quickly on average than the adversary can add them.) The network is considered to be *stable* if the expected load remains bounded, i.e. if $\sup_n \mathbf{E}(X_n) < \infty$. In this context, our Corollary 2 may be interpreted as saying that a queueing network in the presence of a stochastic adversary is guaranteed to be stable, provided it satisfies the load condition (C1), and also the moment condition (C2) for some $p > 2$. On the other hand, if $p \leq 2$ then there is no such guarantee.

We note that there has been some previous work on related questions. For example, Hajek (1982) investigates bounds on hitting times for random sequences having bounded *exponential* moments. Also, Hastad et al. (1996) consider bounds on $\sup_n \mathbf{E}(X_n)$ for time-homogeneous *Markovian* systems, under hypotheses somewhat related to ours.

Finally, we note that while the notion of "stability" considered here (namely, that $\sup_n \mathbf{E}(X_n) < \infty$) is different from that of Markov chain stability (see e.g. Meyn and Tweedie, 1993), there are some connections. For example, it is known (see Tweedie, 1983, Theorem 2) that if $\{X_n\}$ is an aperiodic Harris-recurrent time-homogeneous Markov chain having stationary distribution $\pi(\cdot)$, and if $m \equiv \int x\, \pi(dx) < \infty$, then for π-a.e. x, $\mathbf{E}_{\delta_x}(X_n) \to m$, and hence $\sup_n \mathbf{E}_{\delta_x}(X_n) < \infty$. In other words, for such a Markov chain, stability in our sense is implied by standard Markov chain stability. In fact, it is known (e.g. Tuominen and Tweedie, 1994) that when $\{X_n\}$ is a random walk with negative drift, reset to zero when it attempts to leave the nonnegative half-line and having square integrable increments, then $\mathbf{E}_{\delta_x}(X_n)$ will converge and hence be bounded. This shows that our (C1) and (C2) do represent a greater generality than the random walk context.

The paper is organized as follows. In Section 2 we state the main result, along with two extensions. (The extensions are reasonably straightforward, but we include them in order to provide readily referenceable results that don't assume more than is needed.) We also provide in Section 2 a simpler proof of the main theorem in the case where $p > 4$ and $r = 1$, since in this case the back-of-the-napkin computation works, and anyone not interested in the sharp moment condition need read no further. In Section 3, we give examples to show why (C2) is needed with $p - 1 > r$ and why it is important to have moment bounds for the negative part of the increments as well as the positive part. Proofs are given in Sections 4 and 5, with Section 4 containing a reduction to a result on martingales and Section 5 containing a proof of the martingale result.

2. Main results.

Throughout this paper, the filtration $\{\mathcal{F}_n\}$ refers to any filtration to which $\{X_n\}$ is adapted. We continue to use (C1) and (C2) for conditional expectations with respect to $\mathcal{F}_n$, slightly generalizing the notation of the introductory section.

Theorem 1 *Let X_n be random variables and suppose that there exist constants $a > 0$, J, $V < \infty$, and $p > 2$, such that $X_0 \leq J$, and for all n,*

$$\mathbf{E}(X_{n+1} - X_n \mid \mathcal{F}_n) \leq -a \quad \text{on the event} \quad \{X_n > J\} \tag{C1}$$

and

$$\mathbf{E}(|X_{n+1} - X_n|^p \mid X_0, \ldots, X_n) \leq V \tag{C2}$$

Then for any $r \in (0, p-1)$ there is a $c = c(p, a, V, J, r) > 0$ such that $\mathbf{E}(X_n^+)^r < c$ for all n.

Applying this theorem to the process $X_n' := X_n - (X_0 - J)^+$ in the case $r = 1$ immediately yields the corollary:

Corollary 2 *Under hypotheses (C1) and (C2) of Theorem 1, but without assuming $X_0 \leq J$, we have*

$$\mathbf{E}(X_n \mid \mathcal{F}_0) \leq c(p, a, V, J, 1) + (X_0 - J)^+.$$

We also state an extension allowing the negative part of the increments to avoid the moment condition in (C2):

Corollary 3 *The conclusion of Theorem 1 still holds when $X_{n+1} - X_n$ is replaced by $(X_{n+1} - X_n)\mathbf{1}_{X_{n+1}-X_n > Z_n}$ in conditions (C1) and (C2), and $Z_n \le -a$ is any sequence adapted to $\{\mathcal{F}_n\}$.*

The proof of Theorem 1 proceeds by decomposing according to the last time U before time n that $\{X_k\}$ was less than J. When $p > 4$, Markov's inequality, together with a crude L^p estimate on $X_n - X_U$, gives bounds on the tails of X_n sufficient to yield Theorem 1. We finish the section by giving this argument.

Assume the notation and hypotheses of Theorem 1. Fix a positive integer n. Let $U = \max\{k \le n;\ X_k \le J\}$. Let $\mu_i = \mathbf{E}\,(X_{i+1} - X_i \,|\, \mathcal{F}_i)$, so that $\mu_i \le -a$ on $\{X_i > J\}$. We may recenter (see the proof of Corollary 5 for details) to obtain

$$\mathbf{E}\,((X_{n+1} - X_n - \mu_n)^p \,|\, \mathcal{F}_n) \ \le\ V'$$

for some $V' < \infty$.

But then, for $t > J$, we have

$$\mathbf{P}(X_n \ge t) = \sum_{k=0}^{n-1} \mathbf{P}(X_n \ge t,\ U = k)$$

$$\le \sum_{k=0}^{n-1} \mathbf{P}(X_n - X_k \ge t - J,\ X_k \le J,\ X_i > J \text{ for } k < i < n)$$

$$\le \sum_{k=0}^{n-1} \mathbf{P}\Big((X_n - X_{n-1} - \mu_{n-1}) + \ldots + (X_{k+1} - X_k - \mu_k) \ge t - J - V^{1/p} + a(n-k-1),$$
$$X_k \le J,\ X_i > J \text{ for } k < i < n\Big)$$

[since $\mu_i \leq -a$ for $k < i < n$, and $\mu_k \leq V^{1/p}$]

$$\leq \sum_{k=0}^{n-1} \mathbf{E}\Big(\Big|(X_n - X_{n-1} - \mu_{n-1}) + \ldots + (X_{k+1} - X_k - \mu_k)\Big|^p\Big) \Big(t - J - V^{1/p} + a(n-k-1)\Big)^{-p}$$

[by Markov's inequality]

$$\leq \sum_{k=0}^{n-1} c_p V'(n-k)^{p/2} \Big(t - J - V^{1/p} + a(n-k-1)\Big)^{-p}$$

[by Lemma 7, which is a direct application of Burkholder's inequality]

$$\leq \sum_{\ell=0}^{\infty} c_p V'(\ell+1)^{p/2} \Big(t - J - V^{1/p} + a\ell\Big)^{-p}.$$

It then follows that

$$\begin{aligned} \mathbf{E}(X_n) &= \int_0^\infty dt\, \mathbf{P}(X_n \geq t) \\ \leq\ (J + V^{1/p} + 1) &+ \int_{J+V^{1/p}+1}^\infty dt \sum_{\ell=0}^{\infty} c_p V'(\ell+1)^{p/2} \Big(t - J - V^{1/p} + a\ell\Big)^{-p} \end{aligned}$$

This integral-of-sum does not depend on n. Furthermore, for $p > 4$ it is straightforward to check (by integrating first) that it is finite. This gives the result. □

3. Some counterexamples.

The following example is due to Madhu Sudan (personal communication via A. Borodin). Let $\{X_n\}$ be a time-inhomogeneous Markov chain such that $\mathbf{P}(X_2 = 0) = 2/3 = 1 - \mathbf{P}(X_2 = 2)$, and such that for $n \geq 2$,

$$\mathbf{P}(X_{n+1} = n+1 \,|\, X_n = n) = 1 - 2/n$$

$$\mathbf{P}(X_{n+1} = 0 \,|\, X_n = n) = 2/n$$

$$\mathbf{P}(X_{n+1} = n+1 \,|\, X_n = 0) = 1/n$$

$$\mathbf{P}(X_{n+1} = 0 \,|\, X_n = 0) = 1 - 1/n$$

These transition probabilities were chosen to ensure that

$$X_n \;=\; \begin{cases} 0, & \text{prob } 2/3 \\ n, & \text{prob } 1/3 \end{cases}$$

for all $n \geq 2$. Hence, $\mathbf{E}(X_n) = n/3$, so that $\sup_n \mathbf{E}(X_n) = \infty$.

On the other hand, it is easily verified that (C1) is satisfied with $a = 1$ and $J = 0$. Furthermore, (C2) is satisfied with $p = 1$ and $V = 3$. We conclude that condition (C1) alone, or combined with (C2) with $p = 1$, does not guarantee stability.

When (C2) holds with $p = 2$ it appears one has to do a little more to engineer a counterexample; specifically, we line up all the jumps out of $(-\infty, J)$ to amass at a fixed time M. Fix a large integer M, and define a time-inhomogeneous Markov chain by setting $X_0 = 0$, and, for $0 \leq n \leq M - 1$, letting

$$\mathbf{P}(X_{n+1} = X_n - 1 \,|\, X_n > 0) = 1$$

$$\mathbf{P}(X_{n+1} = 0 \,|\, X_n = 0) = 1 - (M - n)^{-2}$$

$$\mathbf{P}(X_{n+1} = 2(M - n) \,|\, X_n = 0) = (M - n)^{-2}$$

Then it is easily verified that (C1) is again satisfied with $a = 1$ and $J = 0$. Furthermore, (C2) is satisfied with $p = 2$ and $V = 4$.

On the other hand, setting $A = \exp\left(-\sum_{i=1}^{\infty} 1/i^2\right) > 0$, we compute that

$$\begin{aligned} \mathbf{E}(X_M) &= \sum_{k=0}^{M} (M - k + 1)\, \mathbf{P}(X_k > 0 \text{ and } X_j = 0 \text{ for } j < k) \\ &\geq \sum_{k=0}^{M} (M - k + 1) \left(A/(M - k + 1)^2\right) \\ &= \sum_{k=0}^{M} A/(M - k + 1) \;=\; \sum_{j=1}^{M+1} A/j \end{aligned}$$

which goes to infinity (like $A \log M$) as $M \to \infty$.

This shows that $\mathbf{E} X_n$ cannot be bounded in terms of a, J and V, and by "stringing together" such examples, for larger and larger choices of M, we can clearly make $\sup_n \mathbf{E}(X_n) = \infty$. We conclude that condition (C1), combined with (C2) with $p = 2$, still does not guarantee stability of $\{X_n\}$.

From the queueing theory perspective, it would be desirable, in condition (C2) of Theorem 1, to replace $|X_{n+1} - X_n|$ by $[X_{n+1} - X_n]^+$, i.e. to bound the p^{th} moments of just the *positive part* of the increments. Intuitively, this would correspond to allowing arbitrarily large *negative* increments, and bounding only the large *positive* increments. The problem with this is that the process is not sufficiently affected by its negative drift when this is all concentrated into a few unlikely large jumps. We give a counterexample to demonstrate this.

Fix $0 < \epsilon < 1$, and consider the following *time-homogeneous* Markov chain $\{X_n\}$. Let $X_0 = 0$, and for $n \geq 0$, let

$$\begin{aligned} \mathbf{P}(X_{n+1} = 1 \mid X_n = 0) &= 1 \\ \mathbf{P}(X_{n+1} = x+1 \mid X_n = x > 0) &= 1 - (1+\epsilon)/(x+1) \\ \mathbf{P}(X_{n+1} = 0 \mid X_n = x > 0) &= (1+\epsilon)/(x+1) \end{aligned}$$

Then (C1) is satisfied with $J = 0$ and $a = \epsilon$. Also, $[X_{n+1} - X_n]^+ \leq 1$, so (C2) would indeed hold (for any $p > 0$, and with $V = 1$) if we replaced $|X_{n+1} - X_n|$ by $[X_{n+1} - X_n]^+$.

On the other hand, it is straightforward to see that $\mathcal{L}(X_n)$ converges weakly to a stationary distribution $\pi(n)$, which is such that $\pi(n) \sim C n^{-1-\epsilon}$ as $n \to \infty$. In particular, $\sum_n n\, \pi(n) = \infty$. It follows that $\mathbf{E}(X_n) \to \infty$, i.e. that $\{X_n\}$ is *not* stable in this case. We conclude that Theorem 1 does *not* continue to hold if we consider only the positive part of $X_{n+1} - X_n$ in condition (C2).

Remark. This counter-example only works when $a \leq V^{1/p}$. In the case where $X_n \geq 0$

for all n, this appears to be an extremal counterexample, leading to the following open question:

> Does Theorem 1 continues to hold for sufficiently large a if we assume $X_n \geq 0$ and replace $|X_{n+1} - X_n|$ by $[X_{n+1} - X_n]^+$ in (C2) ?

Despite this counter-example, the hypotheses of Theorem 1 may indeed be weakened to allow some large negative increments. However, both condition (C1) *and* condition (C2) must be identically modified so that negative drift is still manifested. This is the motivation for having stated Corollary 3 as an extension to the main theorem.

4. Reduction to a martingale question.

We will derive Theorem 1 and Corollary 3 from the following martingale result.

Theorem 4 *Let $\{M_n : n = 0, 1, 2, \ldots\}$ be a sequence adapted to a filtration $\{\mathcal{F}_n\}$ and let Δ_n denote $M_{n+1} - M_n$. Suppose that the sequence started at M_1 is a martingale (i.e., $\mathbf{E}(\Delta_n \,|\, F_n) = 0$ for $n \geq 1$), and that $M_0 \leq 0$. Suppose further that for some $p > 2$ and $b > 0$ we have*

$$\mathbf{E}(|\Delta_n|^p \,|\, \mathcal{F}_n) \leq b \tag{1}$$

for all n including $n = 0$. Let $\tau = \inf\{n > 0 : M_n \leq n\}$. Then for any $r \in (0, p)$ there is a constant $C = C(b, p, r)$ such that

$$\mathbf{E}\left((M_t^+)^r \mathbf{1}_{\tau > t}\right) \leq C t^{r-p}. \tag{2}$$

We defer the proof of Theorem 4 until the following section. In the remainder of this section, we assume Theorem 4, and derive Theorem 1 and Corollary 3 as consequences.

Corollary 5 *Let $\{Y_n\}$ be adapted to $\{\mathcal{F}_n\}$ with $Y_0 \leq 0$. Suppose $\mathbf{E}(|\Delta_n'|^p \,|\, \mathcal{F}_n) \leq B$ for all n and $\mathbf{E}(\Delta_n' \,|\, \mathcal{F}_n) \leq 0$ for all $1 \leq n < \sigma$, where $\Delta_n' = Y_{n+1} - Y_n$ and $\sigma = \inf\{n > 0 : Y_n \leq n\}$. Then for $0 < r < p$ there is a constant $K = K(B, p, r)$ such that*

$$\mathbf{E}\left((Y_t^+)^r \mathbf{1}_{\sigma > t}\right) \leq K t^{r-p}.$$

Proof: An easy fact useful here and later is that $z^+ \leq 1 + |z|^p$ and hence

$$\mathbf{E}|Z|^p \leq b \;\Rightarrow\; \mathbf{E}Z^+ \leq 1 + b. \tag{3}$$

Recall (see e.g. Durrett 1996, p. 237) that the supermartingale $\{Y_{n\wedge\sigma} : n \geq 1\}$ may be decomposed as $Y_{n\wedge\sigma} = M_n - A_n$ where $\{M_n : n \geq 1\}$ is a martingale and $\{A_n : n \geq 1\}$ is an increasing predictable process with $A_1 = 0$. Let μ_n denote $\mathbf{E}(\Delta_n' \,|\, \mathcal{F}_n)$. Then the increments $\Delta_n := M_{n+1} - M_n$ satisfy

$$\mathbf{E}(|\Delta_n|^p \,|\, \mathcal{F}_n) = \mathbf{E}(|\Delta_n' - \mu_n|^p \,|\, \mathcal{F}_n) \leq 2^p \mathbf{E}(|\Delta_n'|^p + |\mu_n|^p \,|\, \mathcal{F}_n) \leq 2^p(B + (1+B)^p).$$

Applying Theorem 4 to $\{M_n\}$ with $b = 2^p(B + (1+B)^p)$ and $M_0 := Y_0$ yields

$$\mathbf{E}\left((M_t^+)^r \mathbf{1}_{\tau > t}\right) \leq C t^{r-p}. \tag{4}$$

When $\sigma > t$ it follows that $M_n \geq n + A_n$ for $1 \leq n \leq t$ and hence that $\tau > t$. Also, when $\sigma > t$, we know that $M_t = Y_t + A_t \geq Y_t$ and therefore that

$$Y_t^+ \mathbf{1}_{\sigma > t} \leq M_t^+ \mathbf{1}_{\tau > t}.$$

The conclusion of the lemma now follows from (4), with $K = C(2^p(B + (1+B)^p)), p, r)$. □

The above argument uses no properties of the process A_n, other than its being nonincreasing and adapted. In particular, it need not be predictable. If the increments Δ_n' can be decomposed into the sum of two parts, one satisfying the hypotheses of the corollary and one nonincreasing and adapted, then the second of these can be absorbed into the process $\{A_n\}$ and the result will still hold. Without loss of generality, the second piece can be taken to be $\Delta_n' \mathbf{1}_{\Delta_n' \leq Z_n}$ for some adapted nonpositive $\{Z_n\}$. In other words, the moment condition need not apply to the negative tail of the increment, as long as the mean is still nonpositive when the negative tail is excluded. We state this more precisely as the following corollary.

Corollary 6 *Let $\{Y_n\}$ be adapted to $\{\mathcal{F}_n\}$ with $Y_0 \le 0$. Let $\{Z_n\}$ be any adapted nonpositive sequence. Suppose $\mathbf{E}(|\Delta'_n|^p \mathbf{1}_{\Delta'_n > Z_n} \,|\, \mathcal{F}_n) \le B$ for all n and $\mathbf{E}(\Delta'_n \mathbf{1}_{\Delta'_n > Z_n} \,|\, \mathcal{F}_n) \le 0$ for all $1 \le n < \sigma$, where $\Delta'_n = Y_{n+1} - Y_n$ and $\sigma = \inf\{n > 0 : Y_n \le n\}$. Then for $0 < r < p$,*

$$\mathbf{E}\left((Y_t^+)^r \mathbf{1}_{\sigma > t}\right) \le K t^{r-p}.$$

□

We now use these corollaries to derive Theorem 1 and Corollary 3.

Proof of Theorem 1 from Corollary 5, and of Corollary 3 from Corollary 6: First assume that $a = 1$ and $J = 0$. Given $\{X_n\}$ as in the hypotheses of the theorem, fix an $N \ge 1$; we will compute an upper bound for $\mathbf{E}(X_N^+)^r$ that does not depend on N. Let $U := \max\{k \le N : X_k \le 0\}$ denote the last time up to N that X takes a nonpositive value. Decompose according to the value of U:

$$\mathbf{E}(X_N^+)^r = \sum_{k=0}^{N-1} \mathbf{E}\left((X_N^+)^r \mathbf{1}_{U=k}\right).$$

To evaluate the summand, define for any $k < N$ a process $\{Y_n^{(k)}\}$ by $Y_n^{(k)} = (X_{k+n} + n)\mathbf{1}_{X_k \le 0}$. In other words, if $X_k > 0$ the process $\{Y_N^{(k)}\}$ is constant at zero, and otherwise it is the process $\{X_n\}$ shifted by k and compensated by adding 1 each time step. We apply Corollary 5 to the process $\{Y_n^{(k)}\}$. Hypothesis (C1) of Theorem 1, together with the fact that $X_{k+j} > 0$ for $0 < j < \sigma^{(k)}$, imply that $\mathbf{E}(\Delta'_n \,|\, \mathcal{F}_n) \le 0$ when $1 \le n \le \sigma^{(k)}$. Also, $\mathbf{E}(|\Delta'_n|^p \,|\, \mathcal{F}_n) \le \mathbf{E}(|1 + X_{n+1} - X_n|^p \,|\, \mathcal{F}_n) \le B := 2^p(1+V)$. The conclusion is that

$$\mathbf{E}\left([(Y_{N-k}^{(k)})^+]^r \mathbf{1}_{\sigma^{(k)} > N-k}\right) \le K t^{r-p}.$$

But for each k,

$$X_N^+ \mathbf{1}_{U=k} \le Y_{N-k}^{(k)} \mathbf{1}_{\sigma^{(k)} > N-k}$$

and it follows that

$$\mathbf{E}\left((X_N^+)^r \mathbf{1}_{U=k}\right) \le K(N-k)^{r-p}.$$

Now sum to get

$$\mathbf{E}(X_N^+)^r \le \sum_{k=0}^{N-1} K(N-k)^{r-p} \le K\zeta(p-r).$$

This completes the case $a = 1, J = 0$.

For the general case, let $X_n' = (X_n - J)/a$. This process is covered by the analysis of the $a = 1, J = 0$ case above, with V/a^p in place of V. We conclude that $\mathbf{E}(X_n')^r \le c(p, 1, V/a^2, 0, r)$, and hence that $\mathbf{E}X_n \le c(p, a, V, J, r) := J + a^r c(p, 1, V/a^p, 0, r)$.

The proof of Corollary 3 from Corollary 6 is virtually identical. □

5. The proof of Theorem 4.

We now concern ourselves with the proof of Theorem 4. We begin with two lemmas.

Lemma 7 *Let $\{M_n\}$ be a martingale with $M_0 = 0$, and with increments bounded in L^p:*

$$\mathbf{E}(|M_n - M_{n-1}|^p \,|\, \mathcal{F}_{n-1}) \le L.$$

Then there is c_p such that $\mathbf{E}|M_n|^p \le c_p L n^{p/2}$.

Proof: Burkholder's inequality (see Stout 1974, Theorem 3.3.6, or Burkholder 1966) tells us that there is a constant c_p for which

$$\mathbf{E}|M_n|^p \le c_p \mathbf{E}\left(\sum_{k=1}^{n}(M_k - M_{k-1})^2\right)^{p/2}.$$

(In fact, for $p > 1$ we may take $c_p = 18p^{3/2}(p-1)^{-1/2}$, cf. Chow and Teicher, 1997.) For any $Z_1, \ldots, Z_n$, Hölder's inequality gives

$$\mathbf{E}|Z_1 + \cdots + Z_n|^{p/2} \le n^{p/2} \max_{1\le k\le n} \mathbf{E}|Z_k|^{p/2}.$$

Set $Z_k = (M_k - M_{k-1})^2$ and observe that $\mathbf{E}Z_k^{p/2} \le \mathbf{E}((\mathbf{E}(Z_k \,|\, \mathcal{F}_{k-1})^{p/2}) \le \mathbf{E}(\mathbf{E}|M_k - M_{k-1}|^p \,|\, \mathcal{F}_{k-1}) \le L$, so the conclusion of the lemma follows. □

Lemma 8 *Assume the notation and hypotheses of Theorem 4. For $x > 0$, let $S_x = \inf\{k : M_k \geq x\}$ be the time to hit value x or greater. Then there is a $C' = C'(b,p)$ such that*

$$\mathbf{P}(\tau > S_x) \leq \frac{C'}{x^{p/2}}.$$

Proof: Fix $x \geq 1$ and bound in two ways the quantity $\mathbf{E}\,|M_{\tau \wedge S_x}|^p$. First, since $\{M_{\tau \wedge S_x \wedge n} : n \geq 1\}$ is a martingale, $|x|^p$ is convex, and $\tau \wedge S_x \geq 1$ is a stopping time bounded above by x, we have

$$\mathbf{E}\,|M_{\tau \wedge S_x}|^p \leq \mathbf{E}\,|M_x|^p. \tag{5}$$

Using Lemma 7 gives $\mathbf{E}|M_x - M_1|^p \leq c_p b x^{p/2}$, and since $\mathbf{E}|M_1|^p \leq b$, this yields

$$\mathbf{E}|M_x|^p = ||M_x||_p^p \leq (||M_1||_p + ||M_x - M_1||_p)^p \leq \left(b^{1/p} + (c_p b x^{p/2})^{1/p}\right)^p \leq c'(b,p) x^{p/2} \tag{6}$$

with $c'(b,p) := c_p b(1 + c_p^{-1})^p$. On the other hand, on the event $\{\tau > S_x\}$ we have $M_{\tau \wedge S_x} = M_{S_x} \geq x$, so that

$$x^p \mathbf{P}(\tau > S_x) \leq \mathbf{E}|M_{\tau \wedge S_x}|^p,$$

and combining this with (5) and (6) gives

$$\mathbf{P}(\tau > S_x) \leq x^{-p} c'(p,b) x^{p/2}$$

which proves the result for $x \geq 1$. Finally, for $x < 1$ we use $\mathbf{P}(\tau > S_x) \leq 1$, so the lemma follows with $C'(p,b) := \max(1, c'(p,b))$. $\square$

Proof of Theorem 4: Let $T = \inf\{k \geq 0 : \Delta_k \geq t/4\}$ be the time of the first large jump. Since $\tau > t$ implies $S_x < x$ for all $x \leq t$, we can write

$$\mathbf{E}\left((M_t^+)^r \mathbf{1}_{\tau > t}\right) = \mathbf{E}\left((M_t^+)^r \mathbf{1}_G\right) + \mathbf{E}\left((M_t^+)^r \mathbf{1}_H\right), \tag{7}$$

where $G = \{T \geq S_{t/2} < t < \tau\}$ and $H = \{T < S_{t/2} < t < \tau\}$.

To bound the first term, abbreviate $S := S_{t/2}$ and begin by observing that $M_S \leq 3t/4$ on G, since the level $t/2$ or higher has just been obtained and the increment was no more than $t/4$. Thus

$$\mathbf{E}\left((M_t^+)^r \mathbf{1}_G\right) \leq \mathbf{P}(\tau > S) \mathbf{E}((M_t^+)^r \mathbf{1}_{\tau > t} \,|\, \mathcal{F}_S).$$

The first factor may be bounded via Lemma 8:

$$\mathbf{P}(\tau > S) \le \frac{2^{p/2}C'}{t^{p/2}}. \tag{8}$$

The second factor is bounded using the formula

$$\mathbf{E}\left(Z^r \mathbf{1}_{Z>u}\right) = u^r \mathbf{P}(Z > u) + \int_u^\infty r y^{r-1} \mathbf{P}(Z > y)\, dy. \tag{9}$$

By Lemma 7 conditionally on $\mathcal{F}_S$, $\mathbf{E}(|M_t - M_S|^p \mid F_S) \le c_p b(t-S)^{p/2} \le c_p b t^{p/2}$. Hence by Markov's inequality, $\mathbf{P}(M_t - M_S \ge y \mid \mathcal{F}_S) \le c_p b t^{p/2}/y^p$. Therefore,

$$\begin{aligned}
\mathbf{E}((M_t^+)^r \mathbf{1}_{\tau>t} \mid \mathcal{F}_S) &\le \mathbf{E}((M_t^+)^r \mathbf{1}_{M_t>t} \mid \mathcal{F}_S) \\
&= t^r \mathbf{P}(M_t \ge t \mid \mathcal{F}_S) + \int_t^\infty r y^{r-1} \mathbf{P}(M_t \ge y \mid \mathcal{F}_S)\, dy \\
&\le t^r \mathbf{P}(M_t - M_S \ge t/4 \mid \mathcal{F}_S) + \int_{t/4}^\infty r y^{r-1} \mathbf{P}(M_t - M_S \ge y \mid \mathcal{F}_S)\, dy \\
&\le c_p b 4^p t^{r-p/2} + \int_{t/4}^\infty r y^{r-1} c_b p t^{p/2} y^{-p}\, dy \\
&\le c_p b(4^p + \frac{r}{p-r} 4^{p-r}) t^{r-p/2} \\
&\le c_2(b,p,r) t^{r-p/2},
\end{aligned}$$

where $c_2(b,p,r) := c_p b(4^p + 4^{p-r} r/(p-r))$. Combining with (8) gives

$$\mathbf{E}\left((M_t^+)^r \mathbf{1}_G\right) \le 2^{p/2} c_2 C' t^{r-p}. \tag{10}$$

We will bound the second term by decomposing according to the value of T. A preliminary computation is to bound the quantity $\mathbf{E}((M_t^+)^r \mathbf{1}_{T=k, M_t>t} \mid \mathcal{F}_k)$. Break this into three pieces: the part up to time k, the jump at time k, and the part from time $k+1$ to time t. For any $0 < r < p-1$, $|x+y+z|^r \le 3^r(|x|^r + |y|^r + |z|^r)$ (use convexity when $r \ge 1$ and sublinearity when $r \le 1$). Hence

$$(M_t^+)^r \le 3^r \left[(M_k^+)^r + (\Delta_k^+)^r + ((M_t - M_{k+1})^+)^r\right].$$

The event $\{T = k\}$ implies $M_k \le 3t/4$, and is also in the initital σ-field of the martingale $\{M_n - M_{k+1} : n \ge k+1\}$. Therefore, when we condition on $\mathcal{F}_k$, we get

$$\begin{aligned}\mathbf{E}((M_t^+)^r \mathbf{1}_{T=k, M_t > t} \,|\, \mathcal{F}_k) \;\; &\le \;\; 3^r \left[\left(\frac{3t}{4}\right)^r \mathbf{P}(\Delta_k \ge t/4 \,|\, \mathcal{F}_k) \right. \\ &\quad + \mathbf{E}((\Delta_k^+)^r \mathbf{1}_{\Delta_k \ge t/4} \,|\, \mathcal{F}_k) \\ &\quad \left. + \, \mathbf{E}(|M_t - M_{k+1}|^r \mathbf{1}_{T=k, M_t - M_{k+1} \ge t/4} \,|\, \mathcal{F}_k) \right]\end{aligned}$$

The moment condition $\mathbf{E}(|\Delta_k|^p \,|\, \mathcal{F}_k) \le b$ implies that $\mathbf{P}(\Delta_k \ge y) \le b y^{-p}$, hence the first of these contributions is at most

$$3^r (\frac{3t}{4})^r b (\frac{t}{4})^{-p}.$$

Using (9) again, we bound the second of the three contributions by

$$3^r (\frac{t}{4})^r \mathbf{P}(\Delta_k \ge \frac{t}{4} \,|\, \mathcal{F}_k) + 3^r \int_{t/4}^{\infty} r y^{r-1} \mathbf{P}(\Delta_k \ge y) \, dy$$

which is at most

$$3^r b (\frac{t}{4})^{r-p} + 3^r \frac{r}{p-r} (\frac{t}{4})^{r-p}.$$

Lemma 7 implies $\mathbf{E}(|M_t - M_{k+1}|^r \,|\, \mathcal{F}_{k+1}) \le c_p b t^{r/2}$, while $\mathbf{1}_{T=k} \in \mathcal{F}_{k+1}$ and has conditional expectation at most $b(t/4)^{-p}$ given $\mathcal{F}_k$. Therefore the third contribution is bounded by

$$3^r b (\frac{t}{4})^{-p} c_p b t^{r/2}.$$

Summing the three contributions gives

$$\mathbf{E}((M_t^+)^r \mathbf{1}_{T=k, M_t > t} \,|\, \mathcal{F}_k) \le c_3 t^{r-p} \tag{11}$$

where $c_3 = 3^r 4^p b (c_p b + \frac{p}{p-r} + 3^r)$.

Now we bound the second term, by decomposing according to the value of T.

$$\begin{aligned}\mathbf{E}\left((M_t^+)^r \mathbf{1}_H\right) \;\; &= \;\; \sum_{k=0}^{\lfloor t/2 \rfloor} \mathbf{E}\left((M_t^+)^r \mathbf{1}_H \mathbf{1}_{T=k}\right) \\ &= \;\; \sum_{k=0}^{\lfloor t/2 \rfloor} \mathbf{E}\left[\mathbf{E}((M_t^+)^r \mathbf{1}_H \mathbf{1}_{T=k} \,|\, \mathcal{F}_k)\right]. \end{aligned} \tag{12}$$

The event $\{\tau > k\}$ is in $\mathcal{F}_k$ and contains the event $H \cap \{T = k\}$, so we have

$$\mathbf{E}((M_t^+)^r \mathbf{1}_H \mathbf{1}_{T=k} \mid F_k) \leq \mathbf{1}_{\tau>k}\, \mathbf{E}((M_t^+)^r \mathbf{1}_{T=k} \mid \mathcal{F}_k)$$

and hence

$$\mathbf{E}((M_t^+)^r \mathbf{1}_H) \leq \sum_{k=0}^{\lfloor t/2 \rfloor} \mathbf{P}(\tau > k)\, \mathbf{E}\left[\mathbf{E}((M_t^+)^r)\mathbf{1}_{T=k} \mid \mathcal{F}_k)\right].$$

Plugging in the upper bound (11) and using Lemma 8 gives

$$\mathbf{E}((M_t^+)^r \mathbf{1}_H) \leq \sum_{k=0}^{\lfloor t/2 \rfloor} C' k^{-p/2} c_3 t^{r-p}$$

and summing yields a bound of

$$\mathbf{E}((M_t^+)^r \mathbf{1}_H) \leq c_4 t^{r-p} \tag{13}$$

for the second term, where $c_4 := C' c_3 \zeta(p/2)$. By (7), the two bounds (10) and (13) together imply the conclusion of Theorem 4. $\square$

This completes the proof of Theorem 4, and hence also the proof of Theorem 1 and Corollary 3.

Acknowledgements. We are grateful to Allan Borodin for bringing this problem to our attention. We thank Jim Fill, Leslie Goldberg, Jim Kuelbs, Tom Kurtz, Jeremy Quastel, Tom Salisbury, and Richard Tweedie for very helpful comments.

References

[1] Andrews, M., Awerbuch, B., Fernández, A., Kleinberg, J., Leighton, F.T., and Liu, Z. (1996). Universal stability results for greedy contention-resolution protocols. *Proc. 37th Ann. IEEE Sympos. on Foundations of Computer Science*, 380–389.

[2] Borodin, A., Kleinberg, J., Raghavan, P., Sudan, M., and Williamson, D.P. (1996). Adversarial queueing theory. *Proc. 28th Ann. ACM Sympos. on Theory of Computing*, 376–385.

[3] Borodin, A., Kleinberg, J., Raghavan, P., Sudan, M., and Williamson, D.P. (1998). Adversarial queueing theory. In preparation.

[4] Burkholder, D. (1966). Martingale transforms. *Ann. Math. Stat.* **37** 1497–1504.

[5] Chow, Y.S. and Teicher, H. (1997). *Probability theory, independence, interchangeability, Martingales, 3rd edition.* Springer: New York.

[6] Durrett, R. (1996). *Probability: theory and examples, 2nd edition.* Duxbury Press: Belmont, CA.

[7] Hajek, B. (1982). Hitting-time and occupation-time bounds implied by drift analysis with applications. *Adv. Appl. Prob.* **14**, 502–525.

[8] Hastad, J., Leighton, T., and Rogoff, B. (1996). Analysis of backoff protocols for multiple access channels, *SIAM J. Comput.* **25**(4), 740–774.

[9] Meyn, S.P. and Tweedie, R.L. (1993). *Markov chains and stochastic stability.* Springer-Verlag: London.

[10] Stout, W. (1974). *Almost sure convergence.* Academic Press: New York.

[11] Tweedie, R.L. and Tuominen, P. (1994), Subgeometric Rates of Convergence of f-ergodic Markov Chains. *Ann. Appl. Prob.* **26**, 775–798.

[12] Tweedie, R.L. (1983). The existence of moments for stationary Markov chains. *J. Appl. Prob.* **20**, 191–196.