Published in final edited form as:

Stat Probab Lett. 2007 August ; 77(14): 1486–1489. doi:10.1016/j.spl.2007.02.013.

Inequalities on the Overshoot beyond a Boundary for Independent Summands with Differing Distributions

John L. Spouge¹

National Center for Biotechnology Information National Library of Medicine

Abstract

Let $\{S_n : n \geq 0\}$ ($S_0 = 0$) denote the successive sums of independent non-negative random variates, of possibly differing distributions. Define: (1) the number $N(b) = \inf\{n \geq 0 : S_n > b\}$ of sums in the interval $[0, b]$; and (2) the overshoot $R_b = S_{N(b)} - b$. This paper bounds the tail $\mathbb{P}\{R_b > c\}$ and the moments $\mathbb{E}R_b^k$.

Keywords

Renewal theory; excess over the boundary; Lorden's inequality

1 Introduction and Statement of Results

As usual, define $\mathbb{N}_0 := \mathbb{N} \cup \{0\} := \{0, 1, 2, \dots\}$ (“:=” represents a definition). Let $(X_n \geq 0 : n \in \mathbb{N})$ be a sequence of independent non-negative random variates, not necessarily identically distributed. Denote sums of the $\{X_n\}$ by $S_n := \sum_{i=0}^n X_i$, where $S_0 := X_0 := 0$, and assume that $\lim_{n \rightarrow \infty} S_n = \infty$ with probability 1. For any set $A \subseteq \mathbb{N}$, define the number $N_m A := \#\{n \geq m : S_n \in A\}$ of sums following epoch m and falling in the set A . In particular, define

$$N_m(b) := N_m[0, b] = \inf\{n \geq m : S_n > b\}, \quad (1)$$

where $\inf \emptyset := \infty$. For brevity, we sometimes drop the subscript $m = 0$ on $N(b) := N_0(b)$. This paper presents inequalities on the distribution of the overshoot $R_b := S_{N(b)} - b$ beyond the boundary b .

The overshoot plays a central role in renewal theory (Asmussen 2003), where the $\{X_n\}$ are usually taken to be independent and identically distributed. In that case, let $X =_d X_1$ be the generic distributional representative of $\{X_n\}$. Elegant coupling methods (Lindvall 1992) yield inequalities on the tail $\sup_{b \geq 0} \mathbb{P}\{R_b > c\}$ and the moments $\sup_{b \geq 0} \mathbb{E}R_b^k$. The best bounds known are $\mathbb{E}R_b^k \leq (k+1)^{-1} (k+2) \mathbb{E}X^{k+1} / \mathbb{E}X$ (Chang 1994), with the improvement known as Lorden's inequality $\mathbb{E}R_b \leq \mathbb{E}X^2 / \mathbb{E}X$ for $k = 1$ (Lorden 1970). The coupling proofs rely heavily on the

¹National Center for Biotechnology Information National Library of Medicine Bethesda MD 20894 Phone: (301) 402-9310 Fax: (301) 480-2484 Email: spouge@ncbi.nlm.nih.gov

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

existence of a stationary distribution $\lim_{b \rightarrow \infty} \mathbb{P}\{R_b > c\}$ for the overshoot, however, so ready generalizations to non-identically distributed $\{X_n\}$ seem unavailable.

Lorden (1970) mentions the bound $\sup_{b \geq 0} \mathbb{P}\{R_b > c\} \leq \sup_{x \geq 0} \mathbb{P}\{X > x + c | X > x\}$, where the right side is 0 if $\mathbb{P}\{X > x\} = 0$. Without giving a specific page number, he ascribes the bound to Wald (1947). Wald's bound becomes particularly simple, if the failure rate $(d/dx)[-\log \mathbb{P}\{X > x\}]$ of X is monotone increasing for $x \geq 0$: because then $\log \mathbb{P}\{X > x\}$ is concave, i.e., $\mathbb{P}\{X > x + c | X > x\} \leq \mathbb{P}\{X > c | X > 0\}$, so $\sup_{b \geq 0} \mathbb{P}\{R_b > c\} \leq \mathbb{P}\{X > c | X > 0\}$, with equality if X is exponential. Brown (1980) also gives some related results.

Inequalities on the tail $\sup_{b \geq 0} \mathbb{P}\{R_b > c\}$ are useful outside classical renewal theory. Nonlinear renewal theory (Lai and Siegmund 1977; Lai and Siegmund 1979), where the $\{X_n\}$ might be neither independent nor identically distributed, sometimes requires the hypothesis $\lim_{c \rightarrow \infty} \sup_{b \geq 0} \mathbb{P}\{R_b > c\} = 0$ (Woodroffe 1990). Other applications, such as the statistical theory justifying the bootstrapping a database ((Green and Brenner 2002); see also the Appendix of (Schaffer et al. 2001)), require inequalities on $\sup_{b \geq 0} \mathbb{P}\{R_b > c\}$ for $\{X_n\}$ independent but not identically distributed. Accordingly, this paper presents overshoot inequalities for the case of independent but non-identically distributed $\{X_n\}$. With extra work, our methods can be applied in even more general circumstances.

To state our main result, let $X \leq_d X'$ denote stochastic inequality: $\mathbb{P}\{X \geq x\} \leq \mathbb{P}\{X' \geq x\}$ for all $x \in \mathbb{N}$ and let $X =_d X'$ denote equality of distribution for X and X' . Assume $Y_m \leq_d X_n \leq_d \tilde{Y}_m$ for all $n > m \geq 0$, where $(Y_n \geq 0; n \in \mathbb{N}_0)$ is stochastically increasing (i.e., $Y_n \leq_d Y_{n+1}$); and $(\tilde{Y}_n \geq 0; n \in \mathbb{N}_0)$, stochastically decreasing. For brevity, let $\underline{Y} := Y_0$ and $\tilde{Y} := \tilde{Y}_0$, and assume that $0 < \mathbb{E}\underline{Y} \leq \mathbb{E}\tilde{Y} < \infty$.

We note in passing that the bound Lorden attributes to Wald generalizes directly: for any $b, c \geq 0$ and $m \in \mathbb{N}_0$,

$$\mathbb{P}\{R_b > c\} \leq \mathbb{P}\{S_m > b + c\} + \sup_{y \geq 0} \frac{\mathbb{P}\{\tilde{Y}_m > y + c\}}{\mathbb{P}\{\underline{Y}_m > y\}}, \quad (2)$$

where the right side is 0 if $\mathbb{P}\{\underline{Y}_m > y\} = 0$. We prove Eq (2) later, but our main result is Theorem 1.1 below.

Let $\underline{N}_m A := \#\left\{n \geq 0; S_{m,n} \in A\right\}$, where $S_{m,n} := \sum_{i=0}^n \underline{Y}_{m,i}$, $\underline{Y}_{m,0} := 0$, and the $(\underline{Y}_{m,i}; i \in \mathbb{N})$ are independent variates with the same distribution as $\underline{Y}_m$. Abbreviate the average density of renewals as $\underline{\rho}_m(\delta) := \delta^{-1} \mathbb{E} \underline{N}_m [0, \delta)$, and $\underline{\rho}_0(\delta) := \underline{\rho}_m(\delta)$, and use the unambiguous abbreviation $x^{n+} := (x^+)^n$, where $x^+ := \max\{0, x\}$.

Theorem 1.1

Under the above set-up, for any $b, c \geq 0$, $\delta > 0$, and $m \in \mathbb{N}_0$,

$$\mathbb{P}\{R_b > c\} \leq \mathbb{P}\{S_m > b + c\} + \mathbb{E}\left[\tilde{Y}_m + \delta - c\right]^+ \underline{\rho}_m(\delta). \quad (3)$$

. From $\mathbb{E}R_b^k = \int_0^\infty kc^{k-1} \mathbb{P}\{R_b > c\} dc$ and $(k+1)^{-1} \mathbb{E}[\widehat{Y}_m + \delta]^{k+1} = \int_0^\infty kc^{k-1} \mathbb{E}[\widehat{Y}_m + \delta - c]^+ dc$, the integration of Eq (3) over $c \in [0, \infty)$ yields

$$\mathbb{E}R_b^k \leq \mathbb{E}[S_m - b]^{k+1} + (k+1)^{-1} \mathbb{E}[\widehat{Y}_m + \delta]^{k+1} \rho_m(\delta). \quad (4)$$

Note that the bounds in Eq (3) and (4) can be minimized over $\delta > 0$ and $m \in \mathbb{N}_0$.

Because $0 \leq [\widehat{Y} + \delta - c]^+ \leq \widehat{Y} + \delta$, we have $\lim_{c \rightarrow \infty} \mathbb{E}[\widehat{Y} + \delta - c]^+ = 0$ by dominated convergence. From Eq (3) with $m = 0$ and any $\delta > 0$, the conditions of Theorem 1.1 suffice (Woodroffe 1990) to demonstrate that $\lim_{c \rightarrow \infty} \sup_{b \geq 0} \mathbb{P}\{R_b > c\} = 0$.

With $m = 0$, Eqs (3) and (4) show that $\mathbb{P}\{R_b > c\} \leq \mathbb{E}[\widehat{Y} + \delta - c]^+ \rho(\delta)$ and

$\mathbb{E}R_b^k \leq (k+1)^{-1} \mathbb{E}[\widehat{Y} + \delta]^{k+1} \rho(\delta)$ independently of $b \geq 0$. The bounds are particularly useful in theoretical contexts, because the best bounds known for identically distributed random variates $\{X_n\}$ (Chang 1994) contain moments of the same order.

If required, the usual methods yield explicit bounds for the renewal density $\rho_m(\delta)$. Coupling, e.g., shows that if $p_m(\delta) := \mathbb{P}\{Y_m \geq \delta\}$, then Bernoulli $(p_m(\delta)) \leq_d Y_m$, so $N_m[0, \delta] \leq_d$ Geometric $(p_m(\delta))$ (the right side being the waiting time for the sum of independent Bernoulli $(p_m(\delta))$ variates to leave $[0, 1)$). Thus, $\mathbb{E}N_m[0, \delta] \leq 1/p_m(\delta)$ so $\rho_m(\delta) \geq \delta^{-1}/p_m(\delta)$. Thus, from Eq (4) for $m = 0$ and $k = 1$,

$$\sup_{b \geq 0} \mathbb{E}R_b \leq \frac{1}{2} \inf_{\delta > 0} \frac{\mathbb{E}[\widehat{Y} + \delta]^2}{\delta \mathbb{P}\{\widehat{Y} \geq \delta\}}. \quad (5)$$

Although Eq (5) does not reduce to equality in any special case, it compares reasonably well with Lorden's inequality $\sup_{b \geq 0} \mathbb{E}R_b \leq \mathbb{E}X^2/\mathbb{E}X$, which hypothesizes variates with equal distributions.

In passing, note that the moment generating function $\phi_m(\theta) := \mathbb{E} \exp(\theta Y_m)$ and a Chernoff-type inequality with $\theta < 0$ also yield a bound for the renewal density $\rho_m(\delta)$:

$$\begin{aligned} \mathbb{E}N_m[0, \delta] &= \sum_{n=0}^{\infty} \mathbb{P}\left\{S_{m,n} < \delta\right\} \leq \sum_{n=0}^{\infty} \mathbb{E}\left[\exp\left\{\theta\left(S_{m,n} - \delta\right)\right\}; S_{m,n} < \delta\right] \\ &\leq \sum_{n=0}^{\infty} \mathbb{E}\left[\exp\left\{\theta\left(S_{m,n} - \delta\right)\right\}\right] \leq e^{-\theta\delta} \sum_{n=0}^{\infty} \phi_m^n(\theta) = e^{-\theta\delta} / \{1 - \phi_m(\theta)\}. \end{aligned} \quad (6)$$

The next section proves Theorem 1.1 and concludes with a brief proof of the inequality Eq (2).

2 Proof of Theorem 1.1

Let us bound the tail of the overshoot distribution:

$$\begin{aligned}
\mathbb{P}\{R_b > c\} &= \int_{[0,b]} \sum_{n \geq 1} \mathbb{P}\{S_{n-1} = b - dy; X_n > y + c\} \\
&\leq \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \sum_{n > m} \mathbb{P}\{S_{n-1} = b - dy; X_n > y + c\} \\
&= \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \sum_{n > m} \mathbb{P}\{S_{n-1} = b - dy\} \mathbb{P}\{X_n > y + c\} \\
&\leq \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \sup_{n > m} \mathbb{P}\{X_n > y + c\} \sum_{n \geq m} \mathbb{P}\{S_n = b - dy\} \\
&\leq \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \mathbb{P}\{\widehat{Y}_m > y + c\} \mathbb{E}N_m\{b - dy\}.
\end{aligned} \tag{7}$$

To continue from Eq (7), let $j = \lfloor b/\delta \rfloor$, so $[0, b] \subset \bigcup_{i=0}^j [i\delta, (i+1)\delta]$. Substitute the inclusion into the domain of integration, noting that $\mathbb{P}\{\widehat{Y}_m > y + c\} \leq \mathbb{P}\{\widehat{Y}_m > i\delta + c\}$ for $y \in [i\delta, (i+1)\delta]$:

$$\begin{aligned}
\mathbb{P}\{R_b > c\} &\leq \mathbb{P}\{S_m > b + c\} + \sum_{i=0}^j \mathbb{P}\{\widehat{Y}_m > i\delta + c\} \mathbb{E}N_m(b - (i+1)\delta, b - i\delta] \\
&\leq \mathbb{P}\{S_m > b + c\} + \mathbb{E}\left[\widehat{Y}_m + \delta - c\right]^+ \sup_{t \in \mathbb{R}} \delta^{-1} \mathbb{E}N_m(t, t + \delta],
\end{aligned} \tag{8}$$

because $\mathbb{P}\{\widehat{Y}_m > i\delta + c\} \leq \mathbb{P}\{\widehat{Y}_m > y - \delta + c\}$ for $y \in [i\delta, (i+1)\delta]$, and

$$\delta \sum_{i=0}^{\infty} \mathbb{P}\{\widehat{Y}_m > i\delta + c\} \leq \int_{y=0}^{\infty} \mathbb{P}\{\widehat{Y}_m > y - \delta + c\} dy = \mathbb{E}\left[\widehat{Y}_m + \delta - c\right]^+. \tag{9}$$

Eq (8) motivates a bound on $\mathbb{E}N_m(t, t + \delta]$. A minor variation on a standard coupling argument (e.g., (Lindvall 1992)) provides the bound $\mathbb{E}N_m(t, t + \delta] \leq \mathbb{E}N_m[0, \delta]$. To sketch the argument,

construct a probability space containing sequences $\left\{ \underset{\sim}{Y}'_{m,n} : n \in \mathbb{N}_0 \right\}$ and $\left\{ \underset{\sim}{Y}''_{m,n} : n \in \mathbb{N}_0 \right\}$, where $\underset{\sim}{Y}'_{m,n} = \underset{\sim}{Y}''_{m,n} = \underset{\sim}{Y}_{m,n}$ for $n \in \mathbb{N}_0$ and $\underset{\sim}{Y}'_{m, N(t)+n} = \underset{\sim}{Y}''_{m,n}$ for $n \in \mathbb{N}$. In an obvious notation, $N'_m(t, t + \delta] \leq N''_m[0, \delta]$ on this space, so Eq (3) follows. This concludes the proof of Theorem 1.1.

A brief demonstration of the inequality Eq (2) follows. Start as in Eq (7):

$$\begin{aligned}
\mathbb{P}\{R_b > c\} &\leq \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \sum_{n > m} \mathbb{P}\{S_{n-1} = b - dy\} \mathbb{P}\{X_n > y + c\} \\
&= \mathbb{P}\{S_m > b + c\} + \int_{[0,b]} \sum_{n > m} \mathbb{P}\{S_{n-1} = b - dy; X_n > y\} \mathbb{P}\{X_n > y + c | X_n > y\} \\
&\leq \mathbb{P}\{S_m > b + c\} + \sup_{n > m, y \geq 0} \mathbb{P}\{X_n > y + c | X_n > y\} \sum_{n > m} \int_{[0,b]} \mathbb{P}\{S_{n-1} = b - dy; X_n > y\}.
\end{aligned} \tag{10}$$

Because

$$\sum_{n>m} \int_{[0,b]} \mathbb{P}\{S_{n-1}=b-dy; Y_n>y\} = \sum_{n>m} \mathbb{P}\{N(b)=n\} \leq 1, \quad (11)$$

Eq (2) is proved.

Acknowledgments

This research was supported by the Intramural Research Program of the National Library of Medicine, NIH.

References

- Asmussen, S. Applied Probability and Queues. Springer-Verlag; New York: 2003.
- Brown M. Bounds, inequalities, and monotonicity properties for some specialized renewal processes. *Annals of Probability* 1980;8:227–240.
- Chang JT. Inequalities for the overshoot. *Ann Appl Probab* 1994;4:1223–1233.
- Green RE, Brenner SE. Bootstrapping and normalization for enhanced evaluations of pairwise sequence comparison. *Proceedings of the IEEE* 2002;90:1834–1847.
- Lai TL, Siegmund D. Nonlinear Renewal Theory with Applications to Sequential-Analysis. 1. *Annals of Statistics* 1977;5:946–954.
- Lai TL, Siegmund D. Non-Linear Renewal Theory with Applications to Sequential-Analysis. *Annals of Statistics* 1979;7:60–76.
- Lindvall, T. Lectures on the Coupling Method. Wiley and Sons; New York: 1992.
- Lorden G. On the excess over the boundary. *Annals of Mathematical Statistics* 1970;41:520–527.
- Schaffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV, Altschul SF. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Research* 2001;29:2994–3005. [PubMed: 11452024]
- Wald, A. Sequential Analysis. Wiley; New York: 1947.
- Woodroffe, M. *Annals of Probability*. Vol. 18. On the nonlinear renewal theorem; 1990. p. 1790-1805.