

---

# Explaining a Mysterious Maximal Inequality — and a Path to the Law of Large Numbers

---

J. Michael Steele

---

**Abstract.** In 1964 A. Garsia gave a stunningly brief proof of a useful maximal inequality of E. Hopf. The proof has become a textbook standard, but the inequality and its proof are widely regarded as mysterious. Here we suggest a straightforward first step analysis that may dispel some of the mystery. The development requires little more than the notion of a random variable, and, the inequality may be introduced as early as one likes in a graduate probability course. The benefit is that one gains access to a proof of the strong law of large numbers that is pleasantly free of technicalities or tricky ideas.

**1. EXPLORATION AND FIRST STEP ANALYSIS.** At first we consider an infinite sequence  $X, X_1, X_2, \dots$  of independent, identically distributed random variables that we assume to have a finite first moment, so in symbols  $E|X| < \infty$ . According to custom, we let  $S_k = X_1 + X_2 + \dots + X_k$ ; we think of the index  $k$  as time, and we call the sequence  $\{S_k, 1 \leq k < \infty\}$  a random walk (starting at  $S_0 = 0$ ). We also introduce the *maximal process*

$$M_n \stackrel{\text{def}}{=} \max(0, S_1, S_2, \dots, S_n) \quad \text{for } 1 \leq n < \infty, \quad (1)$$

and we emphasize two points: (1) we include zero as a maximand and (2) we do not take absolute values of the partial sums.

There are many good, nonmysterious reasons for being interested in  $M_n$ , but we leave those reasons aside for the moment. Our first goal is simply to see what one can say about  $M_n$ , if we just take one step at a time. The usual aim of such an exploration is to find a pleasing recurrence relation.

It would be nice if, after taking our first step, we were to have the identity

$$M_n(\omega) = X_1 + \max(0, X_2, X_2 + X_3, \dots, X_2 + X_3 + \dots + X_n), \quad (2)$$

but it is easy to find examples that show that this need not be true. Still, one can ask when it is true, and, if we ponder that possibility for a moment, it may not take long to guess that it is true for all  $\omega$  such that  $M_n(\omega) > 0$ .

This is a reasonable conjecture, and in two steps one can tease out a confirmation. If  $M_n(\omega) > 0$ , then random walk has a positive maximum at some time in the interval  $1 \leq k \leq n$ . If this maximum occurs at time  $k = 1$ , then  $M_n = X_1$ , the second summand of (2) is zero, and the identity holds. Alternatively, if the strictly positive maximum is attained at some  $1 < k \leq n$ , then the second summand of (2) is strictly positive. In this case, we can remove the leading zero from the set of maximands, and we see that the identity (2) again holds.

For  $\omega$  such that  $M_n(\omega) > 0$  we do have our recursion (2), and it expresses a certain symmetry which we can emphasize if we set

$$M_n^+ \stackrel{\text{def}}{=} \max(0, X_2, X_2 + X_3, \dots, X_2 + X_3 + \dots + X_{n+1}).$$

---

<http://dx.doi.org/10.4169/amer.math.monthly.122.5.490>  
MSC: Primary 60F15, Secondary 60G12

The motivation here—and one cause for the use of the word symmetry—is that for independent identically distributed sequences (or even for stationary sequences) we have “distributional invariance under the plus-operation”; specifically, we have the distributional equality

$$M_k^+ \stackrel{d}{=} M_k \quad \text{for all } 1 \leq k < \infty.$$

Now, if we use an indicator function to bundle up our assumption  $M_n(\omega) > 0$ , then from our discussion of (2) we have for all  $\omega$  that

$$\begin{aligned} M_n \mathbb{1}(M_n > 0) &= X_1 \mathbb{1}(M_n > 0) + M_{n-1}^+ \mathbb{1}(M_n > 0) \\ &\leq X_1 \mathbb{1}(M_n > 0) + M_n^+ \mathbb{1}(M_n > 0), \end{aligned}$$

where in the second line we used  $M_{n-1}^+ \leq M_n^+$ , which follows just from the definition of  $M_n^+$ . Next, we collect terms

$$(M_n - M_n^+) \mathbb{1}(M_n > 0) \leq X_1 \mathbb{1}(M_n > 0),$$

and we also note by the non-negativity of  $M_n^+$  that we have the trivial bound

$$(M_n - M_n^+) \mathbb{1}(M_n \leq 0) \leq 0.$$

Summing the last two equations gives us a key inequality

$$M_n - M_n^+ \leq X_1 \mathbb{1}(M_n > 0). \tag{3}$$

One should note a curious fact of this derivation. So far, we have not used *any* properties of the random sequence  $\{X_1, X_2, \dots\}$ , so the inequality (3) holds without restriction. What we have is simply a statement about real numbers, and, as far as pure logic goes, the language of random variables was unnecessary here. Nevertheless, the relation (3) does emerge naturally when one is guided by the imagery of random walk and the traditional reasoning of first step analysis.

When we take the expectation in (3) we have our main result. It may look special and modest, but it is general and powerful.

**Lemma 1 (Garsia’s  $L^1$  Maximal Inequality).** *If  $\{X_1, X_2, \dots, X_n\}$  are integrable random variables such that  $EM_n = EM_n^+$ , then we have*

$$0 \leq E[X_1 \mathbb{1}(M_n > 0)]. \tag{4}$$

The condition  $EM_n = EM_n^+$  certainly holds if the random variables  $\{X_1, X_2, \dots\}$  are independent and identically distributed, but one does not need so much. Stationarity of  $\{X_1, X_2, \dots\}$  would be enough.

In fact, the lemma stated here is rather different from the Hopf maximal inequality that is given in Garsia [2], but, except for the motivation and exposition, the logic of the proof is unchanged. It is curious—but perhaps only curious—that in the present formulation one sees that the exchangeability of  $\{X_1, X_2, \dots, X_{n+1}\}$  would suffice for the maximal inequality (4) to hold.

**2. PASSAGE TO A MORE CONVENTIONAL FORM.** The inequality (4) does not have the shape of more familiar maximal inequalities like those of Lévy or Kolmogorov, but it is easy to pass from (4) to an inequality with a more conventional appearance. If we take a constant  $\lambda > 0$  and apply inequality (4) to the random variables  $X'_i = X_i - \lambda$ , then we have

$$\begin{aligned} 0 &\leq E \left[ (X_1 - \lambda) \mathbb{1} \left[ \max_{0 \leq k \leq n} \{0, S_1 - \lambda, S_2 - 2\lambda, \dots, S_n - n\lambda\} > 0 \right] \right] \\ &= E \left[ (X_1 - \lambda) \mathbb{1} \left[ \max_{1 \leq k \leq n} \{S_1, S_2/2, \dots, S_k/k, \dots, S_n/n\} > \lambda \right] \right]. \end{aligned}$$

By linearity this implies

$$P \left[ \max_{1 \leq k \leq n} \{S_1, \dots, S_k/k, \dots, S_n/n\} > \lambda \right] \leq \frac{1}{\lambda} E(|X_1|),$$

which is a maximal inequality of the classical form; specifically, it is a weak-type  $L^1$  maximal inequality. If we let  $n \rightarrow \infty$ , then we have an even nicer version

$$P \left[ \max_{1 \leq k < \infty} \{S_k/k\} \geq \lambda \right] \leq \frac{1}{\lambda} E(|X_1|). \quad (5)$$

**3. PROOF OF THE STRONG LAW OF LARGE NUMBERS.** The SLLN for bounded random variables is something that one can take as given; it is often proved in the earliest days of a graduate course in probability. It is most pleasingly obtained as an immediate consequence of Hoeffding's inequality and the Borel–Cantelli lemma, but it is more commonly proved by applying Markov's inequality to the fourth power of  $S_n/n$ .

The SLLN for bounded random variables is a baby theorem, and for a genuine adult strength SLLN one wants to reduce the moment assumption to the logical minimum. Here we take  $\{X_1, X_2, \dots\}$  to be a sequence of independent, identically distributed random variables such that  $E(|X_i|) < \infty$  and  $E(X_i) = 0$ . Our goal is to show that  $S_n/n$  converges to zero with probability one.

The recipe has just three ingredients: truncation, the SLLN for bounded random variables, and the  $L^1$  maximal inequality. We first fix  $\varepsilon > 0$ , and then for a constant  $K > 0$  we consider the truncation representation

$$X_i = X_i \mathbb{1}(|X_i| \leq K) + X_i \mathbb{1}(|X_i| > K) \stackrel{\text{def}}{=} X'_i(K) + X''_i(K).$$

By the SLLN for bounded random variables, we have with probability one that

$$\limsup_{n \rightarrow \infty} \frac{1}{n} S_n = \limsup_{n \rightarrow \infty} \left\{ \frac{1}{n} S'_n(K) + \frac{1}{n} S''_n(K) \right\} \leq E[X'_1(K)] + \sup_{1 \leq n < \infty} |S''_n(K)/n|.$$

Since  $EX_1 = 0$ , the dominated convergence theorem tells us there is an  $K_0$  such that for all  $K \geq K_0$  we have  $|EX'_1(K)| \leq \varepsilon$ . Hence for all  $K \geq K_0$  the  $L^1$  maximal inequality (5) gives us

$$P \left( \limsup_{n \rightarrow \infty} \frac{1}{n} S_n > 2\varepsilon \right) \leq \frac{1}{\varepsilon} E[|X''_1(K)|].$$

If we let  $K \rightarrow \infty$ , then the upper bound goes to zero, so when we recall that  $\varepsilon > 0$  is arbitrary, we have

$$\limsup_{n \rightarrow \infty} \frac{1}{n} S_n \leq 0 \quad \text{with probability one.}$$

Finally, by applying the same argument to the sequence  $\{-X_i : 1 \leq i < \infty\}$ , one gets a corresponding lower bound on the limit infimum. The two bounds complete the proof of the strong law of large numbers.

**4. CONNECTIONS AND COMMENTS.** In comparison to the most often quoted proofs of the SLLN, the proof via an  $L^1$  maximal inequality has two notable benefits: (a) one only needs a naive, fixed-level truncation, and (b) the finish via the maximal inequality needs no further trickery.

Kolmogorov’s proof uses a maximal inequality that is of comparable complexity to (5), but the finish is not nearly so quick. Typically, one uses a “moving truncation” of the form  $X_k = X_k \mathbb{1}(|X_i| \leq k) + X_k \mathbb{1}(|X_k| > k)$ , a clever moment calculation, a Borel–Cantelli argument, and, for the *coup de grâce*, a fact about real sequences such as Kronecker’s lemma.

The argument introduced by Etemadi [1] is shorter than Kolmogorov’s, and it also covers new ground. Still, it has its own subtleties. First, a sagacious restriction to non-negative random variables makes it possible to restrict attention to proving that one has almost sure convergence to the mean when one moves along subsequences of the form  $n_k = \lfloor (1 + \delta)^k \rfloor$ ,  $\delta > 0$ . After introducing an exponential moving truncation one is then left with “just” a moment calculation and a Borel–Cantelli argument. Nevertheless, the moment calculation is a clever one which one could easily miss, or fail to remember.

The adult strength SLLN is a special case of the ergodic theorem, but this is a lesson that tends to come toward the end of a traditional course in probability theory, if at all. Here we completely sidestepped the language of ergodic theory and kept close to the concepts that are already familiar to beginners.

The lemma of Hopf [3] that motivated Garsia [2] had its origins in a proof due to Yosida and Kakutani [6] of the classical Birkhoff ergodic theorem, for which there are now many proofs. If one were to deal openly with ergodic theory when first introducing the SLLN, then one could use the streamlined arguments of Katznelson and Weiss [4] or Kean and Petersen [5] to prove the SLLN. This is certainly a sensible possibility, but the path would probably be a bit more slippery for beginners than the path that proceeds through Lemma 1 and the  $L^1$  maximal inequality (5).

---

#### REFERENCES

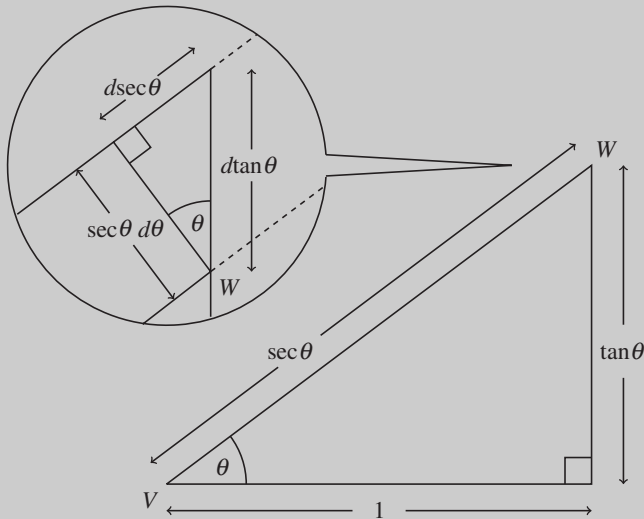
1. N. Etemadi, An elementary proof of the strong law of large numbers, *Z. Wahrsch. Verw. Gebiete* **55** no. 1 (1981) 119–122.
2. A. M. Garsia, A simple proof of E. Hopf’s maximal ergodic theorem, *J. Math. Mech.* **14** (1965) 381–382.
3. E. Hopf, The general temporally discrete Markoff process, *J. Rational Mech. Anal.* **3** (1954) 13–45.
4. Y. Katznelson, B. Weiss, A simple proof of some ergodic theorems, *Israel J. Math.* **42** no. 4 (1982) 291–296.
5. M. Keane, K. Petersen, Easy and nearly simultaneous proofs of the ergodic theorem and maximal ergodic theorem, in *Dynamics & Stochastics*. Vol. 48, IMS Lecture Notes Monogr. Ser., Edited by D. Denteneer, F. den Hollander, and E. Verbitskiy, Inst. Math. Statist., Beachwood, OH, 2006. 248–251.
6. K. Yosida, S. Kakutani, Birkhoff’s ergodic theorem and the maximal ergodic theorem, *Proc. Imp. Acad., Tokyo* **15** (1939) 165–168.

**J. MICHAEL STEELE** received his Ph.D. in mathematics from Stanford University. He has taught at U.B.C., Stanford, CMU, Princeton, and the Wharton School of the University of Pennsylvania. He enjoys mathematical inequalities and sequential decision making, and he pays the rent with stochastic calculus and financial time series.

*Department of Statistics, Wharton School, University of Pennsylvania, JM Huntsman Hall 447, Philadelphia, PA, 19104*  
*steele@wharton.upenn.edu*

### Using Infinitesimals to Differentiate Secant and Tangent

For this figure, let  $d\theta$  be an infinitely small increment of the angle  $\theta$  at vertex  $V$ . Under Keisler's microscope, which has infinite magnification [1], we see corresponding infinitely small changes  $d \tan \theta$  and  $d \sec \theta$  in the lengths of the side opposite  $\theta$  and the hypotenuse. The two rays of the infinitely small angle appear parallel under the microscope, and the tiny arc of the circle centered at  $V$  passing through  $W$  appears as a straight line. The length of the arc is the radius times the radian measure of the angle; thus, it is  $\sec \theta d\theta$ .



The big triangle and the infinitely small triangle are similar. Thus,

$$\frac{d \sec \theta}{\sec \theta d \theta} = \tan \theta, \quad \text{and so} \quad \frac{d}{d \theta} \sec \theta = \sec \theta \tan \theta$$

and

$$\frac{d \tan \theta}{\sec \theta d \theta} = \sec \theta, \quad \text{and so} \quad \frac{d}{d \theta} \tan \theta = \sec^2 \theta.$$

The derivatives of sine and cosine follow as corollaries.

#### REFERENCE

1. H. J. Keisler, *Elementary Calculus: An Infinitesimal Approach*, Online Edition, <http://www.math.wisc.edu/~keisler/calc.html>, 2000.

—Submitted by Michael Hardy, St. Cloud State University, St. Cloud, MN, USA

<http://dx.doi.org/10.4169/amer.math.monthly.122.5.494>  
 MSC: Primary 33B10