Random Harmonic Series

Byron Schmuland

1 Introduction.

The harmonic series is the first nontrivial divergent series we encounter. We learn that, although the individual terms $1/j$ converge to zero, together they accumulate so that their sum is infinite:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{j} + \cdots = \infty.$$

In contrast, we also learn that the alternating harmonic series converges; in fact,

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{j+1}\frac{1}{j} + \cdots = \ln 2.$$

Here the positive and negative terms partly cancel, allowing the series to converge.

To a probabilist, this alternating series suggests choosing plus and minus signs at random, by tossing a fair coin. Formally, let $(\varepsilon_j)_{j=1}^{\infty}$ be independent random variables with common distribution $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = 1/2$. Then, Kolmogorov's three series theorem [1, Theorem 22.8] or the martingale convergence theorem [1, Theorem 35.4] shows that the sequence $\sum_{j=1}^{n} \varepsilon_j/j$ converges almost surely. In this note, we investigate the distribution of the sum $X := \sum_{j=1}^{\infty} \varepsilon_j/j$.

2 Distribution of X.

Obviously, the distribution of X is symmetric about 0, so the mean $E(X)$ is zero. The second moment calculation

$$E(X^2) = \sum_{j=1}^{\infty} \frac{E(\varepsilon_j^2)}{j^2} = \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6},$$

in tandem with the Cauchy-Schwarz inequality $E(|X|) \le E(X^2)^{1/2}$, shows that the average absolute value $E(|X|)$ is no bigger than $\pi/\sqrt{6} = 1.28255$.

Exponential moments provide even more information. Simple properties of the exponential function give, for all $t \ge 0$,

$$\begin{aligned} E(\exp(tX)) &= \prod_{j=1}^{\infty} E(\exp(t\varepsilon_j/j)) = \prod_{j=1}^{\infty} \frac{\exp(t/j)+\exp(-t/j)}{2} \\ &\le \prod_{j=1}^{\infty} \exp(t^2/2j^2) = \exp(t^2\pi^2/12). \end{aligned}$$

For $x > 0$, Markov's inequality [1, (21.11)] tells us that

$$P(X > x) \le \inf_t \exp(t^2\pi^2/12 - tx) = \exp(-3x^2/\pi^2),$$

which shows that the probability of a very large sum is exceedingly small. On the other hand, we can show that it is never zero. Since $\sum_{j=1}^{\infty} \varepsilon_j/j$ converges almost surely, given any $\delta > 0$ we can choose N_1 so that

$$P(|\textstyle\sum_{j>N} \varepsilon_j/j| \le \delta/2) \ge 1/2 \qquad (1)$$

whenever $N \ge N_1$. Also, given any x in $\mathbb{R}$ we can select a nonrandom sequence $(e_j)_{j=1}^{\infty}$ of plus ones and minus ones so that $\sum_{j=1}^{\infty} e_j/j = x$. This is done by choosing plus signs until the partial sum exceeds x for the first time, then minus signs until the partial sum first becomes smaller than x, then iterating this procedure. Let N_2 be so big that $|\sum_{j=1}^{N} e_j/j - x| \le \delta/2$ for all $N \ge N_2$. Putting $N = \max\{N_1, N_2\}$, we have in view of (1) and the independence of the ε_j:

$$\begin{aligned} 0 < (1/2)^N(1/2) &\le P(\varepsilon_1 = e_1)\cdots P(\varepsilon_N = e_N)P(|\textstyle\sum_{j>N} \varepsilon_j/j| \le \delta/2) \\ &= P(\varepsilon_1 = e_1, \ldots, \varepsilon_N = e_N, |\textstyle\sum_{j>N} \varepsilon_j/j| \le \delta/2) \\ &\le P(|X - x| \le \delta). \end{aligned}$$

This shows that the distribution of X has full support on the real line, so there is no theoretical upper (or lower) bound on the random sum.

In [3, sec. 5.2], Kent E. Morrison also considers the distribution of the random variable X. His numerical integration suggests that X has a density of the form in Figure 1.

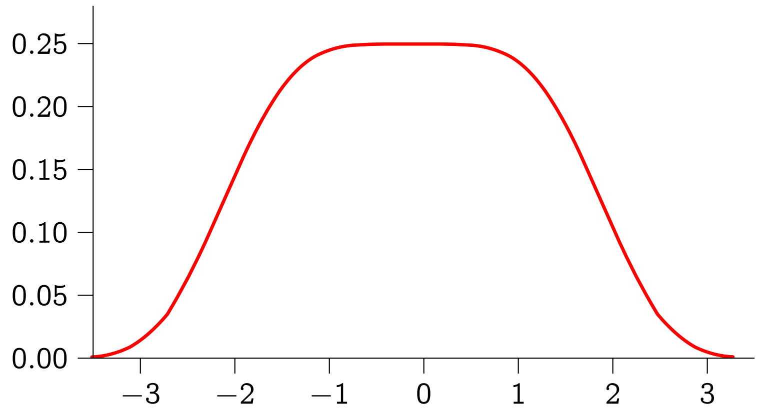

Figure 1. Density of $\sum_{j=1}^{\infty} \varepsilon_j/j$.

Looking at Figure 1, it is easy to believe that X has a smooth density with a flat top. Morrison [3, p. 723] notes that the value of the density at 0 is "suspiciously close to 1/4," and he also conjectures that its value at 2 is 1/8.

Unfortunately, in trying to justify such claims, the approach to X via the partial sums $\sum_{j=1}^{n} \varepsilon_j/j$ does not offer much of a foothold. These partial sums are discrete random variables and do not have densities. After a brief interlude on coin tossing, in section 4 we take an alternative approach to X and in section 5 settle Morrison's two conjectures. This was first done by Morrison himself in an unpublished paper [4] in 1998. In section 6, we explain his proof as well.

3 Binary digits and coin tossing.

An infinite sequence of fair coin tosses can be modelled by selecting a random number uniformly from the unit interval. This observation underlies much of Mark Kac's delightful monograph [2] but can also be found in many probability texts in connection with Borel's normal number theorem [1, sec. 1]. This model is based on the nonterminating dyadic expansion

$$\omega = \sum_{j=1}^{\infty} d_j(\omega)/2^j$$

of ω in $[0,1]$, that is, $(d_j(\omega))_{j=1}^{\infty}$ is the sequence of binary digits of ω. To avoid ambiguity we use the nonterminating expansion for $\omega > 0$. For instance, $1/2 = .0111\ldots$ rather than $1/2 = .1000\ldots$.

If we equip $[0,1]$ with Lebesgue measure, the point ω is said to be chosen *uniformly* from $[0,1]$, in the sense that

$$P(\alpha \leq \omega \leq \beta) = \beta - \alpha \tag{2}$$

for $0 \leq \alpha \leq \beta \leq 1$. Equation (2) shows that there is no location bias in selecting ω, informally, every ω in $[0,1]$ is equally likely to be chosen. It follows [1, (1.9)] that the random variables $(d_j)_{j=1}^{\infty}$ are independent and have common distribution

$$P(d_j = 0) = P(d_j = 1) = 1/2.$$

To recap, the binary digits of a uniformly chosen number from the unit interval act like a sequence of fair coin tosses.

The transformation $\omega \mapsto 1 - 2\omega$ preserves uniformity but changes the underlying interval to $[-1,1]$ and changes the coefficients from zeros and ones to plus ones and minus ones. Thus the classical model of fair coin tosses by a uniform random number implies the following proposition.

Proposition 1. *If $(\varepsilon_j)_{j=1}^{\infty}$ are independent random variables with common distribution $P(\varepsilon_j = 1) = P(\varepsilon_j = -1) = 1/2$, then the sum $\sum_{j=1}^{\infty} \varepsilon_j/2^j$ has a uniform distribution on $[-1,1]$.*

4 Regrouping the series.

Proposition 1 of the previous section shows that a sequence of discrete random variables can sum to a continuous random variable with a well-known density. We exploit this result by rewriting our sum $\sum_{j=1}^{\infty} \varepsilon_j/j$ as follows:

$$\begin{aligned} X &= \frac{\varepsilon_1}{1} + \frac{\varepsilon_2}{2} + \frac{\varepsilon_4}{4} + \cdots =: U_0 \\ &+ \frac{\varepsilon_3}{3} + \frac{\varepsilon_6}{6} + \frac{\varepsilon_{12}}{12} + \cdots =: U_1 \\ &+ \frac{\varepsilon_5}{5} + \frac{\varepsilon_{10}}{10} + \frac{\varepsilon_{20}}{20} + \cdots =: U_2 \\ &+ \cdots \end{aligned}$$

For every $j \geq 0$, Proposition 1 implies that

$$U_j = \frac{2}{2j+1} \left(\sum_{i=1}^{\infty} \varepsilon_{2^{(i-1)}(2j+1)} / 2^i \right)$$

has a uniform distribution on $[-2/(2j+1), 2/(2j+1)]$. Since the U_j are defined using distinct ε variables, they are independent as well. That is, we can write $X = U_0 + U_1 + U_2 + \cdots$ as the sum of independent uniform random variables.

This new series warrants a closer look, since regrouping a conditionally convergent series can give a different sum. For example, the alternating harmonic series has value

$$x = 1 - 1/2 + 1/3 - 1/4 + \cdots = \ln 2,$$

but regrouping the series we get, for every $j \geq 0$,

$$u_j = (2j+1)^{-1}(1 - 1/2 - 1/4 - 1/8 - \cdots) = 0.$$

For the alternating sequence of plus and minus signs, we have $x \neq \sum_{j=0}^{\infty} u_j$. Luckily, this turns out to be a rare exception.

It is not hard to see that you can group a finite number of U_j without changing the sum X, for instance,

$$X = U_0 + \frac{\varepsilon_3}{3} + \frac{\varepsilon_5}{5} + \frac{\varepsilon_6}{6} + \frac{\varepsilon_7}{7} + \frac{\varepsilon_9}{9} + \frac{\varepsilon_{10}}{10} + \cdots,$$

and

$$X = (U_0 + U_1) + \frac{\varepsilon_5}{5} + \frac{\varepsilon_7}{7} + \frac{\varepsilon_9}{9} + \frac{\varepsilon_{10}}{10} + \cdots,$$

are both legitimate equations. That this grouping leaves the sum intact has nothing to do with randomness, it works for any sequence of plus and minus signs. So for any $n \geq 0$ we can write

$$X = (U_0 + U_1 + \cdots + U_n) + \sum_{j \in \Lambda_n} \frac{\varepsilon_j}{j},$$

where Λ_n is the collection of indices not used in $U_0 + U_1 + \cdots + U_n$. The mean square difference satisfies

$$E[(X - \sum_{j=0}^{n} U_j)^2] = \sum_{j \in \Lambda_n} \frac{1}{j^2} \leq \sum_{j=2n+1}^{\infty} \frac{1}{j^2} \to 0,$$

so that $\sum_{j=0}^{n} U_j \to X$ in mean square. On the other hand, the martingale convergence theorem shows that $\sum_{j=0}^{n} U_j \to \sum_{j=0}^{\infty} U_j$ almost surely. Both mean square convergence and almost sure convergence imply convergence in probability, where we have almost surely unique limits. That is, $X = \sum_{j=0}^{\infty} U_j$ almost surely.

5 Densities and characteristic functions.

The smoothness of any random variable Y is related to the decay at infinity of its *characteristic function* ϕ_Y, defined by $\phi_Y(t) = E(\exp(itY))$. For instance, if ϕ_Y is absolutely integrable over the line, then Y has a continuous density function given by the inversion formula [1, (26.20)]:

$$f_Y(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx)\phi_Y(t)\, dt. \tag{3}$$

In addition, if

$$\int_{-\infty}^{\infty} |t^n \phi_Y(t)|\, dt < \infty,$$

then the density f_Y is n times continuously differentiable.

For each $j \geq 0$, let f_j and ϕ_j denote the density and the characteristic function of U_j, respectively:

$$\begin{aligned} f_j(x) &= \begin{cases} (2j+1)/4 & \text{if } -2/(2j+1) \leq x \leq 2/(2j+1), \\ 0 & \text{otherwise;} \end{cases} \\ \phi_j(t) &= \frac{\sin(2t/(2j+1))}{2t/(2j+1)}. \end{aligned}$$

The density g_n of the partial sum $U_0 + U_1 + \cdots + U_n$ is the convolution product $g_n = f_0 * f_1 * \cdots * f_n$, while the characteristic function ψ_n of the partial sum is the product $\psi_n = \phi_0\phi_1 \cdots \phi_n$.

Since $U_0 + U_1 + \cdots + U_n \to X$, the characteristic functions ψ_n converge [1, Theorem 26.3] to the characteristic function ψ of X, namely,

$$\psi(t) = \prod_{j=0}^{\infty} \frac{\sin(2t/2j+1)}{2t/(2j+1)}.$$

The powers of t in the denominator show that ψ has very strong decay at infinity. Bounding $|\psi(t)|$ using the first $n+2$ factors gives

$$|t^n \psi(t)| \leq \frac{1 \cdot 3 \cdots (2(n+1)+1)}{2^{n+1}} t^{-2}.$$

This shows that $t \mapsto |t^n \psi(t)|$ is integrable over $\mathbb{R}$, so ψ has n continuous derivatives in $(-\infty, \infty)$. This is true for all n, ensuring that X has a smooth density function g.

The inversion formula (3) gives

$$\begin{aligned} |g_n(x) - g(x)| &= \frac{1}{2\pi} \left| \int \exp(-itx)(\psi_n(t) - \psi(t))\, dt \right| \\ &\leq \frac{1}{2\pi} \int |\psi_n(t) - \psi(t)|\, dt. \end{aligned}$$

Since $\psi_n(t) \to \psi(t)$, $|\psi_n(t)| \leq |\psi_1(t)|$, and $|\psi_1|$ is integrable, the dominated convergence theorem shows that g_n converges to g uniformly on $\mathbb{R}$.

The densities g_n can be calculated explicitly using the convolution scheme

$$g_0(x) = 1_{[-2,2]}(x)/4, \quad g_n(x) = \int g_{n-1}(x-y) f_n(y)\, dy \tag{4}$$

for $n \geq 1$, so we now have the tools to study the limit density g. Note that the property of being symmetric about 0 and nonincreasing on $[0, \infty)$ is closed under convolution. Therefore the functions g_n all share this property (see Figures 2–4), as do the functions h_n used in proving Proposition 3. This observation lies at the heart of both our proofs.

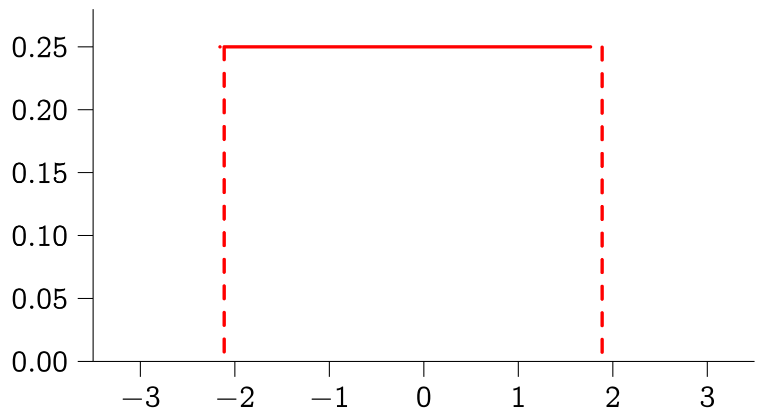

Figure 2. Density g_0 of U_0.

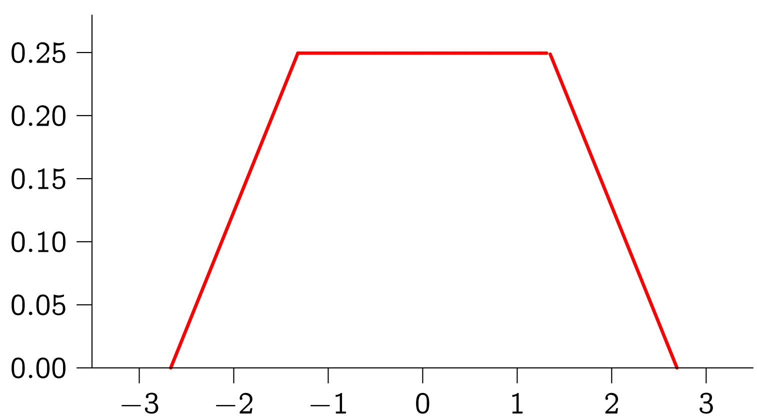

Figure 3. Density g_1 of $U_0 + U_1$.

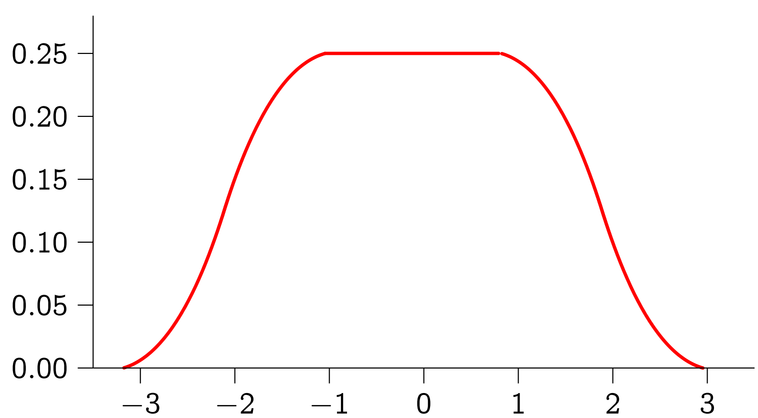

Figure 4. Density g_2 of $U_0 + U_1 + U_2$.

Proposition 2. *The value $g(0)$ is strictly less than 1/4.*

Proof. Each of the densities g_n takes its maximum value at $x = 0$. Using (4) we calculate

$$g_n(0) = \int g_{n-1}(0 - y) f_n(y)\, dy \le g_{n-1}(0) \int f_n(y)\, dy = g_{n-1}(0),$$

so that $g_n(0) \downarrow g(0)$. We note also that $g_n(x) \le 1/4$ for all x in $\mathbb{R}$ and $n \ge 0$. Therefore, for $n \ge 1$, $g_n(0)$ can equal 1/4 only if $g_{n-1}(x) = 1/4$ for all x in the support of the density f_n. We will demonstrate that this happens only when $n \le 6$.

We show by induction:

$$g_n(x) < 1/4 \quad \text{if and only if} \quad |x| > 2 - \sum_{j=1}^{n} 2/(2j+1). \tag{5}$$

Direct inspection shows that this is true for $n = 0$, when, as usual, an "empty sum" means 0. Suppose that (5) is true for g_{n-1}. Convolution gives

$$g_n(x) = \frac{2n+1}{4} \int_{x-2/(2n+1)}^{x+2/(2n+1)} g_{n-1}(y)\, dy.$$

If $|x| > 2 - \sum_{j=1}^{n} 2/(2j+1)$, then the interval

$$[x - 2/(2n+1), x + 2/(2n+1)] \cap \{y : g_{n-1}(y) < 1/4\}$$

is nonempty and $g_n(x) < 1/4$; otherwise the interval is empty and $g_n(x) = 1/4$. This completes the induction proof.

Since

$$2 - \sum_{j=1}^{7} 2/(2j+1) < 0 < 2 - \sum_{j=1}^{6} 2/(2j+1),$$

we see that $g_7(0) < 1/4 = g_6(0)$, and conclude that $g(0) \le g_7(0) < 1/4$. $\square$

The next result is proved with similar ideas, but depends on the symmetry (or lack thereof) of g_n in a neighborhood of 2. For example, $g_1(2)$ is exactly 1/8, and since $x \mapsto g_1(2 - x) - g_1(2)$ is an odd function for x near 0, the convolution of g_1 with a symmetric uniform distribution over a small interval will not change its value at 2. In this way we see that $g_2(2)$ is also equal to 1/8. Eventually though, the neighborhood of "oddness" is smaller than the support of the next convolution, and from that point on, $g_n(2)$ begins to decrease strictly.

Proposition 3. *The value $g(2)$ is strictly less than 1/8.*

Proof. For $n \geq 0$ define h_n by $h_n(x) = g_n(2+x) + g_n(2-x)$. Then

$$h_0(x) = 1_{[-4,4]}(x)/4 + 1_{[0]}(x)/4$$

and $h_n = h_{n-1} * f_n$ for $n \geq 1$. The functions h_n are symmetric and non-increasing on $[0, \infty)$ so, as in the proof of Proposition 2, we find that h_n takes its maximum value at $x = 0$ and that $h_n(0) = 2g_n(2) \downarrow 2g(2)$.

As in the proof of Proposition 2, induction shows that $h_n(x) < 1/4$ if and only if $|x| > 4 - \sum_{j=1}^{n} 2/(2j+1)$. Since

$$4 - \sum_{j=1}^{56} 2/(2j+1) < 0 < 4 - \sum_{j=1}^{55} 2/(2j+1),$$

we see that $2g_{57}(2) = h_{57}(0) < 1/4 = h_{56}(0) = 2g_{56}(2)$, and conclude that $g(2) \leq g_{57}(2) < 1/8$. $\square$

6 Morrison's proof.

For comparison, let's look at Morrison's proof of Propositions 2 and 3. These are found on pages 13 and 14, at the end of section 5, of [4].

In effect, Morrison decomposes X into U_0 plus a remainder $R := \sum_{j=1}^{\infty} U_j$. Then $g = g_0 * r$, where r is the density of R, so

$$g(x) = \int r(y) g_0(x-y)\, dy = (1/4)P(x-2 < R < x+2). \tag{6}$$

The argument in section 2 shows that R has support on the whole real line, so that $P(-2 < R < 2) < 1$ and hence $g(0) < 1/4$. Similarly

$$g(2) = (1/4)P(0 < R < 4) = (1/8)P(-4 < R < 4) < 1/8.$$

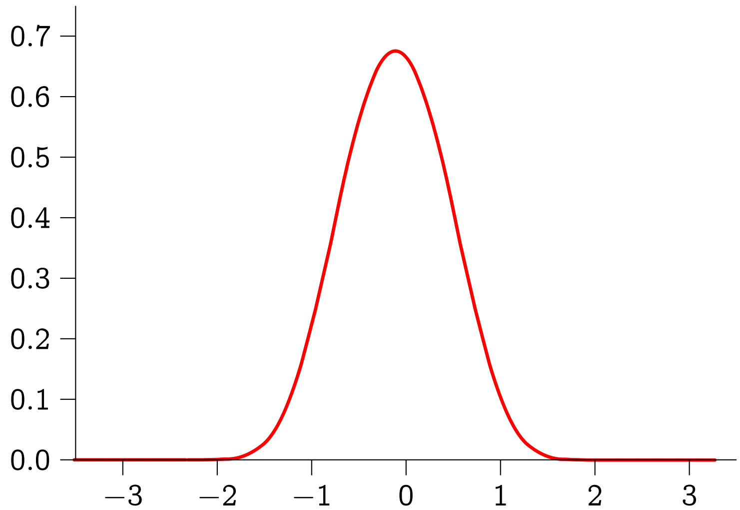

Figure 5. Density of R.

Although R has full support on the real line, the density in Figure 5 shows that both $P(-2 < R < 2)$ and $P(-4 < R < 4)$ are nearly equal to 1. This explains why $g(0)$ is so close to $1/4$ and $g(2)$ so close to $1/8$.

7 Numerical results.

In approximating X by the partial sum $U_0 + U_1 + \cdots + U_n$ we replace the tail $\sum_{j=n+1}^{\infty} U_j$ with zero. Of course, the tail is not exactly zero; in fact, a glance at Figure 5 hints that, for $n = 0$, the tail is close to a normal random variable. Indeed, it is easy to pursue this hint and rigorously prove a central limit theorem:

$$\frac{\sum_{j=n+1}^{\infty} U_j}{\sigma_n} \Rightarrow Z,$$

where $\sigma_n^2 = \text{Var}\left(\sum_{j=n+1}^{\infty} U_j\right) = (4/3) \sum_{j=n+1}^{\infty} (2j+1)^{-2}$ and Z is a standard normal random variable. That is, the tail is close to a normal random variable with variance σ_n^2. This suggests using the approximation $U_0 + U_1 + \cdots + U_n + \sigma_n Z \approx X$, which practice shows to be superior to $U_0 + U_1 + \cdots + U_n \approx X$. For instance, even with $n = 0$, this gives a density function already impressively close to the limit. To ten decimal places, this density has value .2499150393 at $x = 0$, and .1250000000 at $x = 2$ (see Figure 6).

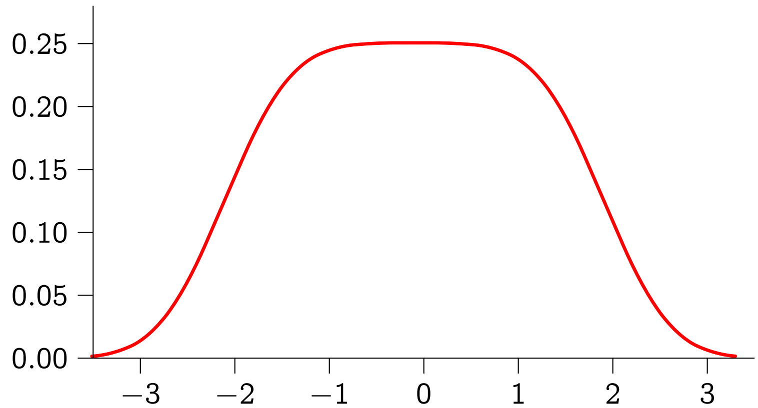

Figure 6. Density of $U_0 + \sigma_0 Z$.

In terms of the characteristic functions, using normal tails means approximating $\psi(t)$ by $\psi_n(t)\exp(-\sigma_n^2 t^2/2)$ rather than $\psi_n(t)$. For the symmetric random variable X, the inversion formula (3) gives

$$g(x) = \frac{1}{\pi}\int_0^\infty \cos(xt)\psi(t)\,dt \approx \frac{1}{\pi}\int_0^\infty \cos(xt)\psi_n(t)\exp(-\sigma_n^2 t^2/2)\,dt.$$

With $n = 150$, we integrated from $t = 0$ to $t = 15$ using a Riemann sum with $dt = 0.02$ and the midpoints of the subintervals for the points of evaluation. We determined that this is accurate to ten decimal places, giving $g(0) = .2499943958$ and $g(2) = .1250000000$.

8 Other random sums.

Replacing the tail of a series by an appropriate normal random variable is a good way of investigating other random sums. For example, the random sum $\sum_{j=1}^\infty \varepsilon_j/j^2$ has the smooth density pictured in Figure 7. The lumps in the distribution come from the first three choices of random sign, while the remaining part of the random sum is essentially determined by a normal tail random variable.

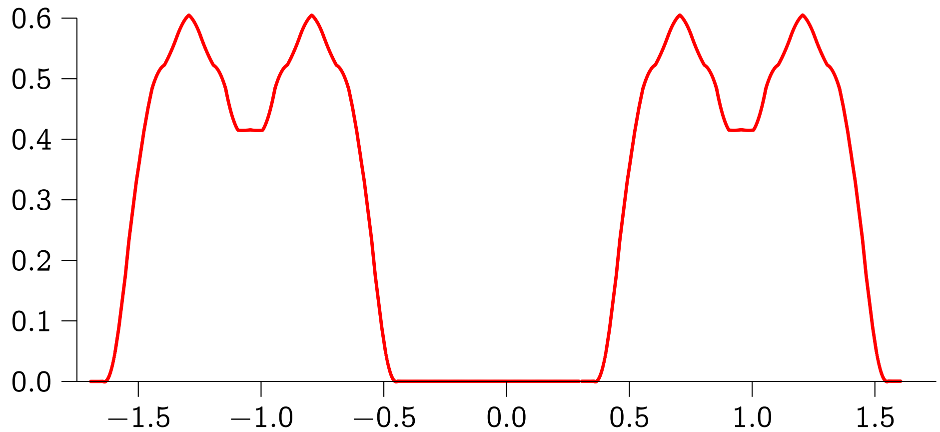

Figure 7. Density of $\sum_{j=1}^{\infty} \varepsilon_j/j^2$.

We conclude with some exercises and further food for thought.

1. Find a nonrandom sequence $(e_j)_{j=1}^{\infty}$ of plus ones and minus ones with $\sum_{j=1}^{n} 1_{[e_j=1]}/n \to 1/2$, but $\sum_{j=1}^{\infty} e_j/j = \infty$. Balancing the plus and minus signs does not guarantee convergence.

2. Prove that $\sum_{j=1}^{\infty} \varepsilon_j/j^2$ has a smooth density. The regrouping trick doesn't work here.

3. Use Morrison's formula (6) to show that $g''(0) = (1/2)r'(2)$. Argue that r is strictly decreasing on $(0, \infty)$ and therefore $g''(0) < 0$. The density g does not have a flat top.

4. Investigate the distribution of the random sum $\sum_{j=1}^{\infty} \varepsilon_j/3^j$.

References

[1] P. Billingsley, *Probability and Measure, 2nd ed.* John Wiley & Sons, New York, 1986.

[2] M. Kac, *Statistical Independence in Probability, Analysis and Number Theory*, Carus Mathematical Monographs, no. 12, Mathematical Association of America, Washington, D.C., 1959.

[3] K. E. Morrison, Cosine products, Fourier transforms, and random sums, *Amer. Math. Monthly* 102 (1995) 716–724.

[4] K. E. Morrison, The final resting place of a fatigued random walker, unpublished manuscript, 1998.

BYRON SCHMULAND teaches probability and statistics at the University of Alberta. His research is mainly about Markov processes, but he likes working on all kinds of mathematical problems. In his leisure, he struggles through language classes in Korean and Mandarin Chinese, and enjoys listening to contemporary surf music.

Department of Mathematical and Statistical Sciences
University of Alberta
Edmonton, Alberta,
Canada T6G 2G1
schmu@stat.ualberta.ca