LECTURE NOTES ON DONSKER'S THEOREM

DAVAR KHOSHNEVISAN

ABSTRACT. Some course notes on Donsker's theorem. These are for Math 7880-1 ("Topics in Probability"), taught at the Deparment of Mathematics at the University of Utah during the Spring semester of 2005.

They are constantly being updated and corrected. Read them at your own risk.

The authors thanks the National Science Foundation for their continued support.

1. Introduction

Let $\{X_n\}_{n=1}^{\infty}$ denote i.i.d. random variables, all taking values in $\mathbf{R}$. Define

$$S_n := X_1 + \cdots + X_n \qquad {}^{\forall} n \geq 1.$$

Recall the classical central limit theorem:

Theorem 1.1 (CLT). *If* $\mathrm{E}[X_1] = \mu$ *and* $\mathrm{Var}(X_1) := \sigma^2 \in (0,\infty)$, *then*

$$\frac{S_n - n\mu}{\sqrt{n}} \Rightarrow N(0,\sigma^2),$$

where $\Rightarrow$ *means weak convergence (or convergence in distribution), and* $N(m,v)$ *denotes the normal distribution with mean* $m \in \mathbf{R}$ *and variance* $v > 0$.

This is a most rudimentary example of an "invariance principle." Here we have a limit theorem where the limiting distribution depends on the approximating sequence only through μ and σ^2.

Here is an application in classical statistical theory:

Example 1.2. Suppose we have a population (e.g., heart rates) whose mean, μ, is unknown to us. In order to learn about this μ, we can take a large independent sample $X_1, \ldots, X_n$ from the said population, and construct the sample average $\bar{X}_n := (S_n/n)$. By the strong law of large numbers, $\bar{X}_n \approx \mu$. In order to find a more quantitative estimate we appeal to the CLT; it asserts that

$$\sqrt{n}\left(\bar{X}_n - \mu\right) \Rightarrow N(0,\sigma^2).$$

One can then use the preceding to derive "approximate confidence bounds" for μ. For instance, if n—the sample size—is large, then the CLT implies that

$$\mathrm{P}\left\{\mu = \bar{X}_n \pm \frac{2}{\sigma\sqrt{n}}\right\} \approx 0.90.$$

This relies on the fact that $\mathrm{P}\{N(0,1) \in [-2,2]\} \approx 0.90$, which you can find in a number of statistical tables. We have just found that, when the sample size is large, we are approximately 90% certain that μ is to within $2/\sqrt{\sigma^2 n}$ of the sample average.

The preceding example is quite elementary in nature. But it is a powerful tool in applied work. The reason for this is the said invariance property: We do not need to know much about the distribution of the sample to say something about the limit. Being in $L^2(\mathrm{P})$ suffices!

Now suppose you are drawing samples as time passes, and you wish to know if the mean of the underlying population has changed over time.

Then, it turns out that we need to consider the sample under the assumption that $X_1,\dots,X_n$ all have the same distribution. In that case, we compute

$$M_n := \max_{1\le j\le n}(S_j - j\mu) \qquad {}^{\forall} n\ge 1.$$

Is there a CLT for M_n? It turns out that the answer is a resounding "yes," and involves an invariance principle.

The same phenomenon holds for all sorts of other random variables that we can construct by applying nice "functionals" to $\{S_1,\dots,S_n\}$. Donsker's theorem asserts that something far deeper happens.

First, let us make the usual simplification that, without loss of generality, $\mathrm{E}[X_1]=0$ and $\mathrm{Var}(X_1)=1$. Else, replace the X_i's everywhere by $X_i' := (X_i-\mu)/\sigma$. Keeping this in mind, we can define for all $\omega\in\Omega$, $n\ge 1$, and $t\in(0,1]$,

$$\mathscr{S}_n(t,\omega) := \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[S_{i-1}(\omega) + n\left(t-\frac{i-1}{n}\right)X_i(\omega)\right]\mathbf{1}_{\left(\frac{i-1}{n},\frac{i}{n}\right]}(t). \tag{1.1}$$

Also define $\mathscr{S}_n(0,\omega):=0$. As usual, we do not write the dependence on ω. In this way, we see that $\mathscr{S}_n := \{\mathscr{S}_n(t);\, t\in[0,1]\}$ is a "random continuous function." This deserves to be made more precise. But before we do that, we should recognize that $\mathscr{S}_n$ is merely the linear interpolation of the normalized random walk $\{S_1/\sqrt{n},\dots,S_n/\sqrt{n}\}$, and is parametrized by $[0,1]$. Thus, for instance, by the CLT, $\mathscr{S}_n(1)\Rightarrow N(0,1)$.

Let $C[0,1]$ denote the collection of all continuous functions $f:[0,1]\to\mathbf{R}$, and metrize it with $d(f,g):=\sup_{x\in[0,1]}|f(x)-g(x)|$ for $f,g\in C[0,1]$.

Exercise 1.3. Check that $(C[0,1],d)$ is a complete, separable, metric vector space.

Now consider the measure space $(C[0,1],\mathscr{B})$, where $\mathscr{B} := \mathscr{B}(C[0,1])$ is the Borel σ-algebra on $C[0,1]$. Then, each $\mathscr{S}_n$ is now a random variable with values in $(C[0,1],\mathscr{C})$. Let P denote also the induced probability measure on the said measure space. [This is cheating a little bit, but no great harm will come of it.]

The bulk of these notes is concerned with the following invariance principle.

Theorem 1.4 (Donsker)**.** *Suppose* $\{X_i\}_{i=1}^\infty$ *is an i.i.d. sequence with* $\mathrm{E}[X_1]=0$ *and* $\mathrm{Var}(X_1)=1$. *Then,* $\mathscr{S}_n\Rightarrow W$ *as* $n\to\infty$, *where* W *denotes Brownian motion. The latter is viewed as a random element of* $(C[0,1],\mathscr{B})$.

It may help to recall that this means that for all bounded, continuous functions $\Lambda: C[0,1]\to\mathbf{R}$,

$$\lim_{n\to\infty}\mathrm{E}\left[\Lambda\left(\mathscr{S}_n\right)\right] = \mathrm{E}\left[\Lambda(W)\right].$$

Of course, *bounded* means that there exists a finite number $\|\Lambda\|$ such that for all $f \in C[0,1]$, $|\Lambda(f)| \le \|\Lambda\|$.

Definition 1.5. A measurable function $\Lambda : C[0,1] \to \mathbf{R}$ is called a *functional.*

2. Some Applications

Before we prove Theorem 1.4, we should learn to apply it in a variety of settings. That is precisely what we do here in this section.

Example 2.1. Let $h : \mathbf{R} \to \mathbf{R}$ be a bounded, continuous function. For $f \in C[0,1]$ define $\Lambda(f) := h(f(1))$. It is easy to see that Λ is a continuous functional. That is, $\Lambda(f_n) \to \Lambda(f)$ whenever $d(f_n, f) \to 0$. For this special Λ, Donsker's theorem says that

$$\mathrm{E}\left[h\left(\frac{S_n}{\sqrt{n}}\right)\right] = \mathrm{E}\left[h\left(\mathscr{S}_n(1)\right)\right] \to \mathrm{E}\left[h(W(1))\right] = \mathrm{E}\left[h(N(0,1))\right].$$

This is the CLT in disguise.

Example 2.2. Let $h : \mathbf{R} \to \mathbf{R}$ be a bounded, continuous function. For $f \in C[0,1]$ define $\Lambda(f) := h(\sup_{t\in[0,1]} f(t))$. Then, Λ is a bounded, continuous functional. Donsker's theorem says the following about this choice of Λ: As $n \to \infty$,

$$\mathrm{E}\left[h\left(\sup_{t\in[0,1]} \mathscr{S}_n(t)\right)\right] = \mathrm{E}\left[\Lambda\left(\mathscr{S}_n\right)\right] \to \mathrm{E}\left[\Lambda(W)\right] = \mathrm{E}\left[h\left(\sup_{t\in[0,1]} W(t)\right)\right].$$

By the reflection principle, $\sup_{t\in[0,1]} W(t)$ has the same distribution as $|N(0,1)|$. Therefore, we have proved that $\sup_{t\in[0,1]} \mathscr{S}_n(t) \Rightarrow |N(0,1)|$, where now $\Rightarrow$ denotes weak convergence in $\mathbf{R}$ (not $C[0,1]$). By convexity,

$$\sup_{t\in[0,1]} \mathscr{S}_n(t) = \max_{0\le j\le n} S_j/\sqrt{n}.$$

(Hash this out!) Therefore, we have proved that for all $x \ge 0$,

$$\lim_{n\to\infty} \mathrm{P}\left\{\max_{1\le j\le n} S_j \le x\sqrt{n}\right\} = \sqrt{\frac{2}{\pi}} \int_0^x \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz. \tag{2.1}$$

(Why?)

Example 2.3. Continue with Example 2.2, and note that for all $x \le 0$,

$$\lim_{n\to\infty} \mathrm{P}\left\{\min_{1\le j\le n} S_j \le x\sqrt{n}\right\} = \sqrt{\frac{2}{\pi}} \int_{-\infty}^x \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz.$$

(Why?) In fact, we can let $\alpha, \beta, \gamma \in \mathbf{R}$ be fixed, and note that

$$\alpha \min_{1\le j\le n} \frac{S_j}{\sqrt{n}} + \beta \max_{1\le j\le n} \frac{S_j}{\sqrt{n}} + \gamma \frac{S_n}{\sqrt{n}} \Rightarrow \alpha \inf_{t\in[0,1]} W(t) + \beta \sup_{t\in[0,1]} W(t) + \gamma \frac{S_n}{\sqrt{n}}.$$

(Why?) This is a statement about characteristic functions: It asserts that the characteristic function of the random vector $(\min_{j\le n} S_j, \max_{j\le n} S_j, S_n)/\sqrt{n}$ converges to that of $(\inf_{[0,1]} W, \sup_{[0,1]} W, W(1))$. Therefore, by the convergence theorem for three-dimensional Fourier transforms,

$$\left(\min_{1\le j\le n} \frac{S_j}{\sqrt{n}}\ ,\ \max_{1\le j\le n} \frac{S_j}{\sqrt{n}}\ ,\ \frac{S_n}{\sqrt{n}}\right) \Rightarrow \left(\inf_{t\in[0,1]} W(t)\ ,\ \sup_{t\in[0,1]} W(t)\ ,\ W(1)\right),$$

where, now, $\Rightarrow$ denotes weak convergence in $\mathbf{R}^3$. Write this as $\boldsymbol{V}_n \Rightarrow \boldsymbol{V}$, where $\boldsymbol{V}_n$ and $\boldsymbol{V}$ are the (hopefully) obvious three-dimensional random variables described above. It follows that for any bounded, continuous $h : \mathbf{R}^3 \to \mathbf{R}$, $\lim_n \mathrm{E}[h(\boldsymbol{V}_n)] = \mathrm{E}[h(\boldsymbol{V})]$. Apply this to $h(x, y, z) := y-x$ to find that $\max_{1\le j\le n} |S_j|/\sqrt{n} \Rightarrow \sup_{t\in[0,1]} |W(t)|$. But the distribution of $\sup_{[0,1]} W$ is known (Khoshnevisan, 2005, Corollary 10.22, page 178). It follows from the formula for the said distribution function that

$$\lim_{n\to\infty} \mathrm{P}\left\{\max_{1\le j\le n} |S_j| \le x\sqrt{n}\right\} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp\left(-\frac{(2k+1)^2\pi^2}{8x^2}\right). \tag{2.2}$$

Exercise 2.4. Prove that for all bounded, continuous $f : \mathbf{R} \to \mathbf{R}$,

$$\frac{1}{n}\sum_{i=1}^{n} f\left(\frac{S_i}{\sqrt{i}}\right) \Rightarrow \int_0^1 f(W(s))\, ds.$$

Exercise 2.5 (Hard)**.** Exercise 2.4 can be improved. For example, prove

$$\frac{1}{n}\sum_{i=1}^{n} 1_{[0,\infty)}(S_i) \Rightarrow |\{0 \le s \le 1 :\ W(s) \ge 0\}|,$$

where $|\{\cdots\}|$ denotes Lebesgue measure. We will return to this exercise later on.

3. Proof of Donsker's Theorem

For all positive integers $k \ge 1$, and all continuous functions $f : [0,1] \to \mathbf{R}$ (i.e., all $f \in C[0,1]$), we can define a piecewise-linear function $(\pi_k f) \in C[0,1]$ as follows: $(\pi_k f)(0) := f(0)$, and for all $t \in (0,1]$,

$$(\pi_k f)(t) := \sum_{i=1}^{2^k} \left[f\left(\frac{i-1}{2^k}\right) + 2^k\left(t - \frac{i-1}{2^k}\right)\left\{ f\left(\frac{i}{2^k}\right) - f\left(\frac{i-1}{2^k}\right)\right\}\right] \mathbf{1}_{\left(\frac{i-1}{2^k}, \frac{i}{2^k}\right]}(t).$$

That is, $(\pi_k f)(i2^{-k}) = f(i2^{-k})$ for all $i = 0, \ldots, k$, and it interpolate between these values linearly. You should compare this to (1.1); for instance, check that $\pi_k \mathscr{S}_{2^n} = \mathscr{S}_{2^n}$ for all integers $k \ge 2^n$. Also note the projection property: $\pi_{k+1}\pi_k = \pi_{k+1}$; equivalently, if $i \le j$ then $\pi_i \pi_j = \pi_j$.

The following is an immediate consequence of the a.s.-continuity of Brownian motion, and does not merit a proof. But it is worth writing out explicitly.

Lemma 3.1. *With probability one,* $\lim_{k\to\infty} d(\pi_k W, W) = 0$.

The idea behind the proof of Donsker's theorem is this: We know that $\pi_k W \approx W$ a.s., and hence in distribution. Out task would be two-fold: On one hand, we prove that uniformly for all n (large) we can find k_0 such that for all $k \geq k_0$, $\pi_k \mathscr{S}_n \approx \mathscr{S}_n$. This is called "tightness." On the other hand, we will prove, using the ordinary CLT, that for each k fixed, $\pi_k \mathscr{S}_n \Rightarrow \pi_k W$. This called the "weak convergence of the fdd's." Assembling the pieces then gets us through the proof. Now we get on with actually proving things. The first step involves truncation to ensure that we can have as many moments as we would like.

3.1. **Truncation.** Define, for all $\nu > 0$ integers $i, n \geq 1$,

$$Y_i^{(\nu)} := X_i \mathbf{1}_{\{|X_i| \leq \nu\}}, \text{ and } X_i^{(\nu)} := (Y_i^{(\nu)} - \mathrm{E}[Y_1^{(\nu)}]).$$

Also define $S_n^{(\nu)}$ and $\mathscr{S}_n^{(\nu)}$ analogously.

Lemma 3.2. *For all* $\nu > 0$,

$$\sup_{n\geq 1} \left\| d\left(\mathscr{S}_n^{(\nu)}, \mathscr{S}_n\right)\right\|_2 \leq 2\sqrt{\mathrm{E}[X_1^2; |X_1| > \nu]}. \tag{3.1}$$

In particular, for all $\lambda > 0$,

$$\lim_{\nu\to\infty} \sup_n \mathrm{P}\left\{d\left(\mathscr{S}_n^{(\nu)}, \mathscr{S}_n\right) > \lambda\right\} = 0.$$

Proof. The second statement follows from the first, and Chebyshev's inequality. We will concentrate on the first assertion.

Evidently,

$$\begin{aligned}\mathrm{E}\left[\left|S_n - S_n^{(\nu)}\right|^2\right] &= \mathrm{Var}\left(\sum_{i=1}^n X_i \mathbf{1}_{\{|X_i|>\nu\}}\right)\\ &= n\mathrm{Var}(X_1; |X_1| > \nu) \leq n\mathrm{E}[X_1^2; |X_1| > \nu].\end{aligned}$$

Now $\{S_n - S_n^{(\nu)}\}_{n=1}^\infty$ is a mean-zero, P-integrable random walk. Therefore, it is a martingale. By Doob's L^2 inequality,[1]

$$\mathrm{E}\left[\max_{1\leq j\leq n}\left|S_j - S_j^{(\nu)}\right|^2\right] \leq 4\mathrm{E}\left[\left|S_n - S_n^{(\nu)}\right|^2\right] \leq 4n\mathrm{E}[X_1^2; |X_1| > \nu].$$

[1]Recall that Doob's L^2 inequality states the following: If M is an L^2-martingale, then $\mathrm{E}[\max_{j\leq n} M_j^2] \leq 4\mathrm{E}[M_n^2]$.

This proves the lemma because $\max_{1\le j\le n}|S_j - S_j^{(\nu)}| = \sqrt{n}\sup_{t\in[0,1]}|\mathscr{S}_n(t) - \mathscr{S}_n^{(\nu)}(t)|$, thanks to piecewise linearity. □

Lemma 3.3. *For each* $\lambda > 0$,

$$\lim_{k\to\infty}\sup_{n\ge 1} \mathrm{P}\left\{d\left(\pi_k \mathscr{S}_n^{(\nu)}, \mathscr{S}_n^{(\nu)}\right) > \lambda\right\} = 0.$$

Proof. Thanks to piecewise linearity, if $n > 2^k$ then

$$d\left(\pi_k \mathscr{S}_n^{(\nu)}, \mathscr{S}_n^{(\nu)}\right) \le \frac{1}{\sqrt{n}} \max_{\substack{1\le j\le 2^k \\ i\in[(j-1)\lfloor n2^{-k}\rfloor,\, j\lfloor n2^{-k}\rfloor]}} \left|S^{(\nu)}_{j\lfloor n2^{-k}\rfloor} - S_i^{(\nu)}\right|.$$

Therefore,

$$\begin{aligned}\mathrm{P}\left\{d\left(\pi_k \mathscr{S}_n^{(\nu)}, \mathscr{S}_n^{(\nu)}\right) > \lambda\right\} &\le \sum_{j=1}^{2^k} \mathrm{P}\left\{\max_{(j-1)n2^{-k}\le i\le jn2^{-k}} \left|S^{(\nu)}_{j\lfloor n2^{-k}\rfloor} - S_i^{(\nu)}\right| \ge \lambda\sqrt{n}\right\}\\ &\le 2^k \mathrm{P}\left\{\max_{1\le i\le n2^{-k}} \left|S_i^{(\nu)}\right| \ge \lambda\sqrt{n}\right\},\end{aligned}$$

because $S_{a+b}^{(\nu)} - S_b^{(\nu)}$ has the same distribution as $S_a^{(\nu)}$. Now, $\{S_i^{(\nu)}\}_{i=1}^{\infty}$ is a mean-zero random walk whose increments are bounded by 2ν (why not ν?). According to Azuma–Hoeffding inequality (Khoshnevisan, 2005, Exercise 8.45, p. 143),

$$\mathrm{P}\left\{\max_{1\le i\le m}\left|S_i^{(\nu)}\right| \ge t\right\} \le 2\exp\left(-\frac{t^2}{8\nu^2 m}\right) \qquad \forall\, t > 0,\ m\ge 1.$$

Therefore,

$$\mathrm{P}\left\{d\left(\pi_k \mathscr{S}_n^{(\nu)}, \mathscr{S}_n^{(\nu)}\right) > \lambda\right\} \le 2^{k+1}\exp\left(-\frac{\lambda^2 n}{8\nu^2[n2^{-k}]}\right) \le 2^{k+1}\exp\left(-\frac{\lambda^2 2^k}{8\nu^2}\right).$$

The lemma follows. □

3.2. **Weak Convergence of Finite-Dimensional Distributions.**

Definition 3.4. Let $\{Y(t)\}_{t\in T}$ denote a stochastic process indexed by some set T. Then the *finite-dimensional distributions* of the process Y are the totality of distributions of vectors of the form $(Y(t_1),\ldots,Y(t_m))$ as $t_1,\ldots,t_m$ vary over T.

Define

$$\sigma_\nu^2 := \mathrm{Var}\left(X_1^{(\nu)}\right) = \mathrm{Var}\left(X_1; |X_1|\le\nu\right). \tag{3.2}$$

According to the CLT, for any fixed $s\in(0,1)$,

$$\frac{S^{(\nu)}_{\lfloor ns\rfloor}}{\sqrt{n}} \Rightarrow \sigma_\nu^2 N(0,s),$$

where ⇒ denotes weak convergence in $\mathbf{R}$. Because $W(s)$ has the same distribution as $N(0,s)$, the preceding is equivalent to the following:

$$\frac{S^{(\nu)}_{\lfloor ns\rfloor}}{\sqrt{n}} \Rightarrow \sigma^2_\nu W(s).$$

Once again, ⇒ denotes weak convergence in $\mathbf{R}$ [$s \in (0,1)$ is fixed here.] Now choose and fix $0 \le s_1 < s_2 < \ldots < s_m \le 1$, and note that $\{S^{(\nu)}_{\lfloor ns_{i+1}\rfloor} - S^{(\nu)}_{\lfloor ns_i\rfloor}\}_{i=1}^{m-1}$ is an independent sequence. Moreover, $S^{(\nu)}_{s_{i+1}} - S^{(\nu)}_{s_i}$ has the same distribution as $S^{(\nu)}_{s_{i+1}-s_i}$. Because W also has independent increments, this verifies that as $n \to \infty$,

$$\left(\frac{S^{(\nu)}_{\lfloor ns_{i+1}\rfloor} - S^{(\nu)}_{\lfloor ns_i\rfloor}}{\sqrt{n}}\right)_{1\le i\le m} \Rightarrow \sigma^2_\nu \left(W(s_{i+1}) - W(s_i)\right)_{1\le i\le m}, \tag{3.3}$$

where ⇒ denotes weak convergence in $\mathbf{R}^m$, and $(x_i)_{1\le i\le m}$ designates the m-vector whose ith coordinate is x_i.

Exercise 3.5. If $(X^1_n,\ldots,X^m_n)$ converges weakly in $\mathbf{R}^m$ to $(X^1,\ldots,X^m)$, then $(X^1_n, X^1_n+X^2_n,\ldots,X^1_n+\cdots+X^m_n)$ converges weakly in $\mathbf{R}^m$ to $(X^1, X^1+X^2,\ldots,X^1+\cdots+X^m)$.

Apply this exercise together with (3.3) to find the following.

Lemma 3.6. *As $n\to\infty$, the finite-dimensional distributions of the stochastic process $\left\{n^{-1/2}S^{(\nu)}_{\lfloor ns\rfloor}\right\}_{s\in[0,1]}$ converge weakly to those of $\sigma^2_\nu W$.*

Let $\mathscr{C}_k$ denote the collection of all bounded, continuous functionals $\Lambda : C[0,1]\to\mathbf{R}$ such that $\Lambda(f)$ depends only on the values $f(i2^{-k})$ ($i = 0,\ldots,2^k$). More precisely, if $f(i2^{-k}) = g(i2^{-k})$ for all $i = 0,\ldots,2^k$ then $\Lambda(f) = \Lambda(g)$ for all $\Lambda\in\mathscr{C}_k$. Note that whenever $\Lambda : C[0,1]\to\mathbf{R}$ is a bounded, continuous functional, then $\Lambda\circ\pi_k\in\mathscr{C}_k$. Therefore, Lemma 3.6, and the definition of weak convergence on $\mathbf{R}^{1+2^k}$, together imply that for each $\nu>0$ and $k\ge 1$ fixed,

$$\pi_k\mathscr{S}^{(\nu)}_n \Rightarrow \sigma^2_\nu \pi_k W, \tag{3.4}$$

where ⇒ defines weak convergence in $C[0,1]$.

3.3. **Proof of Donsker's Theorem (Theorem 1.4).** Let $\Lambda : C[0,1]\to\mathbf{R}$ denote a bounded, continuous functional once more. We write

$$\begin{aligned}\left|\mathrm{E}\left[\Lambda\left(\mathscr{S}^{(\nu)}_n\right)\right] - \mathrm{E}\left[\Lambda(\sigma^2_\nu W)\right]\right| &\le \left|\mathrm{E}\left[\Lambda\left(\pi_k\mathscr{S}^{(\nu)}_n\right)\right] - \mathrm{E}\left[\Lambda(\sigma^2_\nu\pi_k W)\right]\right|\\ &\quad+\left|\mathrm{E}\left[\Lambda\left(\pi_k\mathscr{S}^{(\nu)}_n\right)\right] - \mathrm{E}\left[\Lambda\left(\mathscr{S}^{(\nu)}_n\right)\right]\right|\\ &\quad+\left|\mathrm{E}\left[\Lambda(\sigma^2_\nu\pi_k W)\right] - \mathrm{E}\left[\Lambda(\sigma^2_\nu W)\right]\right|\\ &:= T_1+T_2+T_3.\end{aligned}$$

We established in (3.4) that $\lim_{n\to\infty} T_1 = 0$ for each $k \ge 1$ fixed. Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \left| \mathrm{E}\left[\Lambda\left(\mathscr{S}_n^{(\nu)}\right)\right] - \mathrm{E}\left[\Lambda(\sigma_\nu^2 W)\right]\right| \\
&\le \limsup_{k\to\infty} \sup_n \left| \mathrm{E}\left[\Lambda\left(\pi_k \mathscr{S}_n^{(\nu)}\right)\right] - \mathrm{E}\left[\Lambda\left(\mathscr{S}_n^{(\nu)}\right)\right]\right| \\
&\quad + \limsup_{k\to\infty} \left|\mathrm{E}\left[\Lambda(\pi_k W)\right] - \mathrm{E}\left[\Lambda(W)\right]\right|.
\end{aligned}$$

Thanks to Lemma 3.1 and the dominated convergence theorem, the final term vanishes. We aim to prove that

$$\lim_{k\to\infty} \sup_n \left\| \Lambda\left(\pi_k \mathscr{S}_n^{(\nu)}\right) - \Lambda\left(\mathscr{S}_n^{(\nu)}\right)\right\|_1 = 0. \tag{3.5}$$

Now we use a result from measure theory that I will leave to you as an exercise.

Exercise 3.7. Suppose $\{Z_n\}_{n=1}^\infty$ is a sequence of bounded, real-valued random variables. If $Z_n \to Z$ in probability, then $f(Z_n)$ converges in $L^1(\mathrm{P})$ to $f(Z)$ for all bounded, continuous functions $f : \mathbf{R} \to \mathbf{R}$.

This and Lemma 3.3 together imply (3.5). This, in turn, proves that

$$\lim_{n\to\infty} \left| \mathrm{E}\left[\Lambda\left(\mathscr{S}_n^{(\nu)}\right)\right] - \mathrm{E}\left[\Lambda(\sigma_\nu^2 W)\right]\right| = 0.$$

That is, we have established that $\mathscr{S}_n^{(\nu)} \Rightarrow \sigma_\nu^2 W$, where $\Rightarrow$ denotes weak convergence in $C[0,1]$. But $\lim_{\nu\to\infty} \sigma_\nu^2 = \mathrm{E}[X_1^2] = 1$. Therefore,

$$\lim_{\nu\to\infty} \mathrm{E}\left[\Lambda\left(\sigma_\nu^2 W\right)\right] = \mathrm{E}[\Lambda(W)].$$

Also, another round of the preceding exercise tells us that

$$\lim_{\nu\to\infty} \sup_n \left| \mathrm{E}\left[\Lambda\left(\mathscr{S}_n^{(\nu)}\right)\right] - \mathrm{E}\left[\Lambda\left(\mathscr{S}_n\right)\right]\right| = 0;$$

consult Lemma 3.2. This proves Donsker's theorem.

4. The Arcsine Law

According to Exercise 2.5, if $\{X_n\}_{n=1}^\infty$ are i.i.d., mean-zero, variance-one random variables, and $S_n := X_1 + \cdots + X_n$, then

$$\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{S_i>0\}} \Rightarrow \int_0^1 \mathbf{1}_{\{W(s)>0\}}\, ds.$$

We now find the distribution of the latter random variable. To do so, it is enough to consider the random walk $\{S_n\}_{n=1}^\infty$ of choice, find the distribution of $\sum_{i=1}^n \mathbf{1}_{\{S_i>0\}}$ for that walk, normalize and take limits.

Theorem 4.1 (Paul Lévy)**.** *For all* $a \in (0,1)$,

$$\mathrm{P}\left\{\int_0^1 \mathbf{1}_{\{W(s)>0\}}\,ds \le a\right\} = \frac{2}{\pi}\arcsin\left(\sqrt{a}\right).$$

Consequently, for any mean-zero, variance-one random walk $\{S_n\}_{n=1}^\infty$,

$$\lim_{n\to\infty} \mathrm{P}\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{S_i>0\}} \le a\right\} = \frac{2}{\pi}\arcsin\left(\sqrt{a}\right) \qquad \forall a \in (0,1).$$

We will do this by working with the special case where $X_1 = \pm 1$ with probability 1/2 each. The resulting random walk $\{S_n\}_{n=1}^\infty$ is called the *simple symmetric walk*. It goes also by the title of the Bernoulli walk, Rademacher walk, etc.

4.1. **Some Combinatorics for the Simple Walk.** We delve a bit into the combinatorial structure of simple walks. Feller (1968, Chapter 3) is an excellent reference.

Definition 4.2. By a *path of length n* (more aptly, lattice path) we mean a collection of points $(k, s_0), \ldots, (k+n, s_n)$ where $s_i \in \mathbf{Z}$ and $|s_{i+1} - s_i| = 1$. In this case, we say that the path *goes from* (k, s_0) to $(k+n, s_n)$.

At any step of its construction, a lattice path can go up or down. Therefore, there are 2^n paths of length n. Consequently, with probability 2^{-n}, the first n steps of our simple walk $\{S_k\}_{k=0}^\infty$ are equal to a given path of length n that starts from $(0,0)$, where $S_0 := 0$. That is, all paths that start from $(0,0)$ and have length n are equally likely. Therefore, if Π is a property of paths of length n,

$$\begin{aligned}&\mathrm{P}\left\{\{S_i\}_{i=0}^n \in \Pi\right\}\\ &= \frac{\#\text{ of paths of length } n \text{ that start from } (0,0) \text{ and are in } \Pi}{2^n}.\end{aligned} \tag{4.1}$$

Thus, any probabilistic problem for the simple walk has a combinatorial variant, and vice versa.

Let $N_{n,x}$ denote the number of paths that go from $(0,0)$ to (n,x). An elementary computation shows that

$$N_{n,x} = \begin{cases} \binom{n}{(n+x)/2}, & \text{if } n+x \text{ is even},\\ 0, & \text{otherwise.}\end{cases} \tag{4.2}$$

Another important fact is the 'reflection principle' of Désiré André (1887).

Theorem 4.3 (The Reflection Principle)**.** *Suppose* $n, x, y > 0$ *and* $k \ge 0$ *are integers. Let* M *denote the number of paths that go from* (k,x) *to* $(k+n, y)$ *and hit zero at some point. Then* M *is equal to the number of paths that go from* $(0,-x)$ *to* (n,y)*. That is,* $M = N_{n,x+y}$.

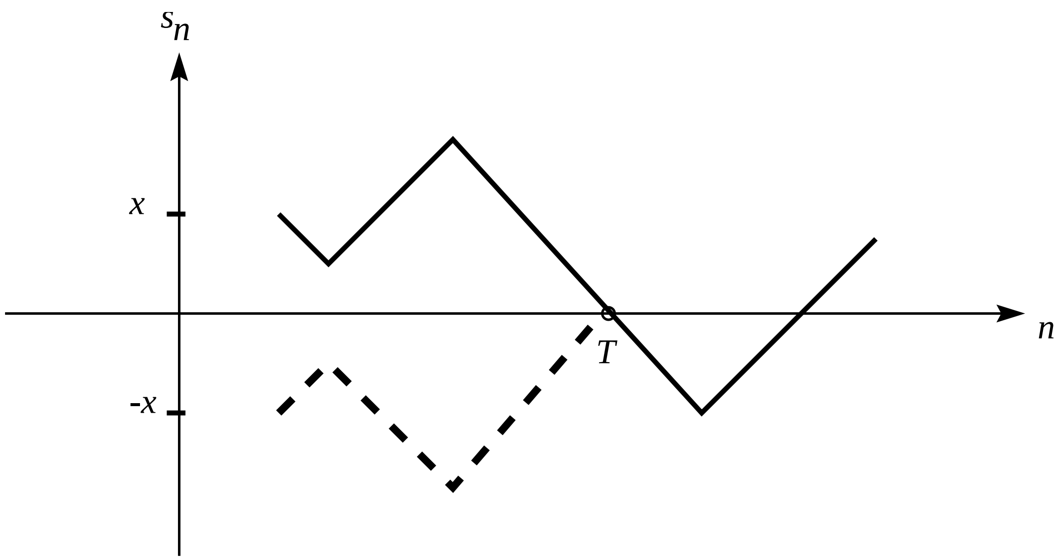

FIGURE 4.1. A reflected lattice path

Proof. Let T denote the first instant when a given path from (k,x) to $(k+n,y)$ crosses zero. Reflect the pre-T portion of this path to obtain a path that goes from $(k,-x)$ to $(k+n,y)$ (Figure 4.1). This map is an invertible operation, therefore M is equal to the number of paths that go from $(k,-x)$ to $(k+n,y)$. It is easy to see that M does not depend on k. □

The following is the key to our forthcoming analysis.

Theorem 4.4 (The Ballot Problem)**.** *Let $n,x>0$ be integers. Then the number of paths that go from $(0,0)$ to (n,x) and $s_1,\ldots,s_n>0$ is $(x/n)N_{n,x}$.*

Define T_0 to be the first time the simple walk crosses $y=0$; i.e.,

$$T_0 := \inf\{k\geq 1:\ S_k=0\} \qquad (\inf\varnothing := \infty). \tag{4.3}$$

Then, the ballot theorem is saying the following (check!):

$$\mathrm{P}\{T_0>n \mid S_n=x\} = \frac{x}{n} \qquad \forall x=1,\ldots,n.$$

Proof of the Ballot Theorem. Let M denote the number of paths that go from $(0,0)$ to (n,x) and $s_1,\ldots,s_n>0$. All paths in question have the property that they go from $(1,1)$ to (n,x). Therefore, we might as well assume that $x\leq n$, whence

$$\begin{aligned} M &= \#\left[\text{paths from } (1,1) \text{ to } (n,x)\right] \\ &\quad - \#\left[\text{paths from } (1,1) \text{ to } (n,x) \text{ and cross } y=0 \text{ at some intervening time}\right] \\ &= N_{n-1,x-1} - N_{n-1,x+1}. \end{aligned}$$

We have applied the reflection principle in the very last step. If $n+x$ is odd then $N_{n-1,x-1} = N_{n-1,x+1} = N_{n,x} = 0$, and the result follows. On the other hand, if $n+x$ is even, then by (4.2),

$$\begin{aligned} N_{n-1,x-1} - N_{n-1,x+1} &= \binom{n-1}{\frac{n+x}{2}-1} - \binom{n-1}{\frac{n+x}{2}} \\ &= \frac{(n-1)!}{\left(\frac{n+x}{2}-1\right)!\left(\frac{n-x}{2}\right)!} - \frac{(n-1)!}{\left(\frac{n+x}{2}\right)!\left(\frac{n-x}{2}-1\right)!} \\ &= \frac{N_{n,x}}{n}\left\{\left(\frac{n+x}{2}\right) - \left(\frac{n-x}{2}\right)\right\}, \end{aligned}$$

whence the result follows. □

Define

$$\ell_{2n} := \frac{1}{n+1}\binom{2n}{n} \qquad \forall n \geq 0. \tag{4.4}$$

Theorem 4.5. *ℓ_{2n-2} is the number of paths of length $2n$ which have $s_i > 0$ for all $i \leq 2n-1$ and $s_{2n} = 0$.*

Proof. Let M denote the number of all paths of length $2n$ which have $s_i > 0$ for all $i \leq 2n-1$ and $s_{2n} = 0$. Evidently, M is the number of all paths that go from $(0,0)$ to $(2n-1,1)$ and do not cross $y = 0$. By the ballot theorem,

$$M = \frac{1}{2n-1} N_{2n-1,1} = \frac{1}{2n-1}\binom{2n-1}{n},$$

which is equal to ℓ_{2n-2}. □

Theorem 4.6. *ℓ_{2n} is the number of paths which have $s_i \geq 0$ for all $i \leq 2n-1$ and $s_{2n} = 0$.*

Proof. If M denotes the number of paths of interest, then M is the number of paths that go from $(0,0)$ to $(2n+1,1)$ such that $s_i > 0$ for all $i \leq 2n+1$; cf. Figure 4.2. [Simply shift the axes to the dashed ones, and add a piece that goes from $(0,0)$ to $(1,1)$ in the new coordinate system.] Theorem 4.5 does the rest. □

Define $f_0 := 0$ and $u_0 = 1$. Then, for all $n \geq 1$ define

$$u_{2n} := \frac{\binom{2n}{n}}{2^{2n}}, \quad f_{2n} := \frac{u_{2n-2}}{2n}.$$

The following result computes various probabilities of interest.

Theorem 4.7. *The following are valid for all $n \geq 1$:*

(1) $\mathrm{P}\{S_{2n} = 0\} = u_{2n}$;

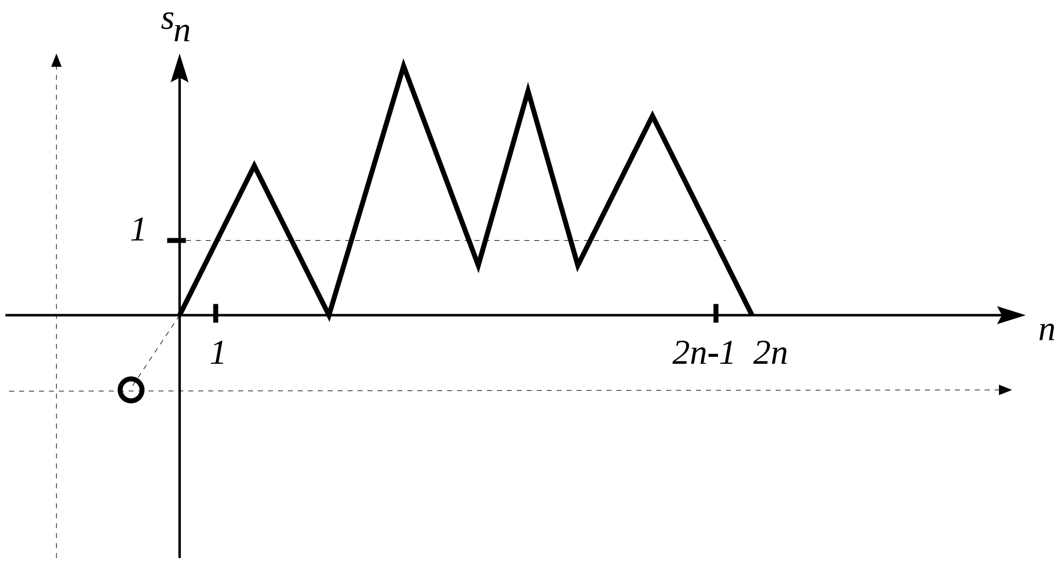

FIGURE 4.2. Adding a Piece to the Path

(2) $\mathrm{P}\{T_0 = 2n\} = f_{2n}$;
(3) $\mathrm{P}\{S_i \ge 0 \text{ for all } i \le 2n-3, S_{2n-2} = 0,\ S_{2n-1} = -1\} = f_{2n}$;
(4) $\mathrm{P}\{T_0 > 2n\} = u_{2n}$;
(5) $\mathrm{P}\{S_i \ge 0 \text{ for all } i \le 2n\} = u_{2n}$.

Proof. Because there are $N_{2n,0} = 2^{2n} u_{2n}$ paths that end up at zero at time $2n$, (1) follows. To prove (2) note that the number of paths that hit zero, for the first time, at time $2n$ is exactly twice the number of paths which satisfy $s_i > 0$ $(i \le 2n-1)$ and $s_{2n} = 0$. Theorem 4.5 then implies (2). Similarly, Theorem 4.6 implies (5).

By (2),

$$\mathrm{P}\{T_0 > 2n\} = 1 - \sum_{j=1}^{n} \mathrm{P}\{T_0 = 2j\} = 1 - \sum_{j=1}^{n} f_{2j}.$$

But it is easy to check that $f_{2n} = u_{2n-2} - u_{2n}$ for all $n \ge 1$. Because $u_0 = 1$, this means that

$$\mathrm{P}\{T_0 > 2n\} = 1 - \Big[(1-u_2) + (u_2 - u_4) + \cdots + (u_{2n-2} - u_{2n})\Big] = u_{2n}.$$

This proves (4). It remains to verify (3). But this follows from Theorem 4.6: the number of paths that satisfy $s_i \ge 0$ $(i \le 2n-3)$, $s_{2n-2} = 0$ and $s_{2n-1} = -1$ is the same as the number of paths that satisfy $s_i \ge 0$ $(i \le 2n-2)$ and $s_{2n-2} = 0$, which is ℓ_{2n-2}. □

REFERENCES

Feller, William. 1968. *An Introduction to Probability Theory and Its Applications. Vol. I*, Third edition, John Wiley & Sons Inc., New York.

Khoshnevisan, Davar. 2005. *Probability*, University of Utah, Salt Lake City, UT, lecture notes.

DEPARTMENT OF MATHEMATICS, THE UNIVERSITY OF UTAH, 155 S. 1400 E. SALT LAKE CITY, UT 84112–0090

E-mail address: davar@math.utah.edu

URL: http://www.math.utah.edu/~davar