4 Sums of Independent Random Variables

***Standing Assumptions*:** Assume throughout this section that $(\Omega, \mathscr{F}, P)$ is a fixed probability space and that $X_1, X_2, X_3, \dots$ are *independent* real-valued random variables on $(\Omega, \mathscr{F}, P)$. Let $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}$ be a nested sequence of $\sigma-$algebras[4] such that for every $n \geq 1$,

(F1) the random variable X_n is $\mathscr{F}_n-$measurable; and
(F2) the $\sigma-$algebras $\sigma(X_n, X_{n+1}, X_{n+2}, \dots)$ and $\mathscr{F}_{n-1}$ are independent.

For instance, we could take $\mathscr{F}_0 = \{\emptyset, \Omega\}$ and $\mathscr{F}_n = \sigma(X_1, X_2, \dots, X_n)$. For each $n = 0, 1, 2, \dots$ set

$$S_n = \sum_{j=1}^{n} X_j.$$

Lemma 4.1. *If Y, Z are independent, nonnegative random variables, then*

$$E(YZ) = (EY)(EZ). \tag{4.1}$$

Similarly, if X, Y are independent random variables with finite first moments, then the equality (4.1) *holds.*

Proof. If $Y = \mathbf{1}_F$ and $Z = \mathbf{1}_G$ are independent indicator variables then the equation (4.1) follows by definition of independence. Consequently, by linearity of expectation, (4.1) holds for any two independent *simple* random variables. To see that the result holds in general, observe that if Y and Z are independent nonnegative random variables then there exist sequences Y_n, Z_n of nonnegative simple random variables such that

$$\begin{gathered} 0 \leq Y_1 \leq Y_2 \leq \cdots \quad \text{and} \quad \lim_{n\to\infty} Y_n = Y; \\ 0 \leq Z_1 \leq Z_2 \leq \cdots \quad \text{and} \quad \lim_{n\to\infty} Z_n = Z; \quad \text{and} \\ Y_n, Z_n \text{ are independent.} \end{gathered}$$

(Exercise: Why?) Clearly, the sequence $Y_n Z_n$ converges monotonically to YZ. Hence, the monotone convergence theorem implies that

$$\begin{aligned} EY &= \lim_{n\to\infty} EY_n; \\ EZ &= \lim_{n\to\infty} EZ_n; \text{ and} \\ EYZ &= \lim_{n\to\infty} E(Y_n Z_n). \end{aligned}$$

Since $E(Y_n Z_n) = EY_n EZ_n$ for each $n = 1, 2, \dots$, it must be that $E(YZ) = EYEZ$. $\square$

[4] Probabilists often use the term *filtration* for a nested sequence of $\sigma-$algebras. If $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots$ is a filtration and $X_1, X_2, \dots$ a sequence of random variables such that each X_n is $\mathscr{F}_n-$measurable, then the sequence $(X_n)_{n\geq 1}$ is said to be *adapted* to the filtration.

Remark 4.2. It follows that if both $Y, Z \in L^1$ then their product $YZ \in L^1$. If not for the hypothesis that Y, Z are independent this would not be true. (See *Hölder's inequality*, sec. 3.5).

4.1 Stopping Times and the Wald Identities

Lemma 4.3. *Let T be a random variable taking values in the set $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$ of nonnegative integers. Then*

$$ET = \sum_{n=1}^{\infty} P\{T \geq n\}$$

Proof. Use the monotone convergence theorem and the fact that $T = \sum_{n=1}^{\infty} \mathbf{1}\{n \leq T\}$. □

Definition 4.4. A *stopping time* (relative to the *filtration* $(\mathscr{F}_n)_{n\geq 0}$) is a random variable T taking values in $\mathbb{Z}_+ \cup \{\infty\}$ such that for every $n \geq 0$ the event $\{T = n\}$ is an element of $\mathscr{F}_n$.

Example 4.5. (a) Let $B \in \mathscr{B}$ be a Borel set, and define τ_B to be the smallest $n \geq 0$ such that $S_n \in B$, or $\tau_B = \infty$ on the event that there is no such n. Then τ_B is a stopping time relative to any filtration $(\mathscr{F}_{n\,n\geq 0})$ with respect to which the sequence $(X_n)_{n\geq 1}$ is adapted. (b) Fix an integer $m \geq 0$, and let $\tau_{B,m}$ be the smallest $n \geq m$ such that $S_n \in B$, or $\tau_{B,m} = \infty$ on the event that there is no such n. Then $\tau_{B,m}$ is a stopping time. (c) Fix an integer $m \geq 0$, and let $\tau \equiv m$. Then τ is a stopping time.

Remark 4.6. If T is a stopping time then for any integer $m \geq 1$

(a) the event $\{T \geq m\} = \{T \leq m-1\}^c = (\cup_{n\leq m-1}\{T = n\})^c$ is in $\mathscr{F}_{m-1}$; and
(b) the random variable $T \wedge m$ is a stopping time.

Proposition 4.7. *(Strong Markov Property) Let $X_1, X_2, \dots$ be independent, identically distributed random variables and let τ be a* finite *stopping time (i.e., a stopping time such that $P\{\tau < \infty\} = 1$). Then the random variables $X_{\tau+1}, X_{\tau+2}, \dots$ are independent, identically distributed and have the same joint distribution as do the random variables $X_1, X_2, \dots$, that is, for any integer $m \geq 1$ and Borel sets $B_1, B_2, \dots, B_m$,*

$$P\{X_{\tau+j} \in B_j \ \forall\, j \leq m\} = P\{X_j \in B_j \ \forall\, j \leq m\}.$$

Furthermore, the random variables $X_{\tau+1}, X_{\tau+2}, \dots$ are "conditionally independent of everything that has happened up to time τ", that is, for any integers $m, n \geq 0$ and Borel sets $B_1, B_2, \dots, B_{m+n}$,

$$\begin{aligned} &P\{\tau = m \text{ and } X_j \in B_j \ \ \forall\, j \leq m+n\} \\ &\qquad\qquad = P\{\tau = m \text{ and } X_j \in B_j \ \ \forall\, j \leq m\} P\{X_j \in B_j \ \forall\, m < j \leq m+n\}. \end{aligned}$$

Proof. Routine. □

Theorem 4.8. *(Wald's First Identity) Assume that the random variables X_i are independent, identically distributed and have finite first moment, and let T be a stopping time such that $ET < \infty$. Then S_T has finite first moment and*

$$ES_T = (EX_1)(ET). \tag{4.2}$$

Proof for Bounded Stopping Times. Assume first that $T \le m$. Then clearly, $|S_T| \le \sum_{i=1}^{m} |X_i|$, so the random variable S_T has finite first moment. Since T is a stopping time, for every $n \ge 1$ the event $\{T \ge n\} = \{T > n-1\}$ is in $\mathscr{F}_{n-1}$, and therefore is independent of X_n. Consequently,

$$\begin{aligned}
ES_T &= E\sum_{i=1}^{m} X_i \mathbf{1}\{T \ge i\} \\
&= E\sum_{i=1}^{m} X_i \mathbf{1}\{T > i-1\} \\
&= \sum_{i=1}^{m} EX_i \mathbf{1}\{T > i-1\} \\
&= \sum_{i=1}^{m} (EX_i)(E\mathbf{1}\{T > i-1\}) \\
&= EX_1 \sum_{i=1}^{m} P\{T \ge i\} \\
&= EX_1 ET.
\end{aligned}$$

□

Proof for Stopping Times with Finite Expectations. This is an exercise in the use of the monotone convergence theorem for expectations. We will first consider the case where the random variables X_i are nonnegative, and then we will deduce the general case by linearity of expectations.

Since the theorem is true for *bounded* stopping times, we know that for every $m < \infty$,

$$ES_{T\wedge m} = EX_1 E(T \wedge m). \tag{4.3}$$

As m increases the random variables $T \wedge m$ increase, and eventually stabilize at T, so by the monotone convergence theorem, $E(T \wedge m) \to ET$. Furthermore, if the random variables X_i are nonnegative then the partial sums S_k increase (or at any rate do not decrease) as k increases, and consequently so do the random variables

$$S_{T\wedge m} = \sum_{i=1}^{T\wedge m} X_i.$$

Clearly, $\lim_{m\to\infty} S_{T\wedge m} = S_T$, so by the monotone convergence theorem,

$$\lim_{m\to\infty} ES_{T\wedge m} = ES_T.$$

Thus, the left side of (4.3) converges to ES_T as $m \to \infty$, and so we conclude that the identity (4.2) holds when the summands X_i are nonnegative.

Finally, consider the general case, where the increments X_i satisfy $E|X_i| < \infty$ but are not necessarily nonnegative. Decomposing each increment X_i into its positive and negative parts gives

$$S_T = \sum_{k=1}^{T} X_k^+ - \sum_{k=1}^{T} X_k^- \quad \text{and}$$
$$|S_T| \le \sum_{k=1}^{T} X_k^+ + \sum_{k=1}^{T} X_k^-.$$

We have proved that the Wald identity (4.2) holds when the increments are nonnegative, so we have

$$E\left(\sum_{k=1}^{T} X_k^+\right) = EX_1^+ ET \quad \text{and}$$
$$E\left(\sum_{k=1}^{T} X_k^-\right) = EX_1^- ET$$

Adding these shows that $E|S_T| < \infty$, and subtracting shows that $ES_T = ETEX_1$. □

Remark 4.9. There is a subtle point in the last argument that should not be missed. The random variable T was assumed to be a stopping time, and implicit in this assumption is the understanding that T is a stopping time relative to a filtration $\mathscr{F}_0 \subset \mathscr{F}_1 \subset \mathscr{F}_2 \subset$ such that the properties (F1)- (F2) hold for the sequence $X_1, X_2, \ldots$. But to use the Wald identity for the sequences X_n^+ and X_n^-, one must check that properties (F1)- (F2) hold for these sequences. EXERCISE: Verify that if (F1)- (F2) hold for the sequence $X_1, X_2, \ldots$ then they hold for the sequences X_n^+ and X_n^-.

Theorem 4.10. *(Wald's Second Identity) Assume that the random variables X_i are independent, identically distributed with $E_{X_i} = 0$ and $\sigma^2 = EX_i^2 < \infty$. If T T is a stopping time such that $ET < \infty$ then*

$$ES_T^2 = \sigma^2 ET. \tag{4.4}$$

Proof. This is more delicate than the corresponding proof for Wald's First Identity. We do have *pointwise* convergence $S_{T\wedge m}^2 \to S_T^2$ as $m \to \infty$, so if we could first prove that the theorem is true for *bounded* stopping times then the Fatou Lemma and the monotone convergence theorem would imply that

$$ES_T^2 \le \lim_{m\to\infty} ES_{T\wedge m}^2 = \lim_{m\to\infty} \sigma^2 E(T \wedge m) = \sigma^2 ET.$$

The reverse inequality does not follow (at least in any obvious way) from the dominated convergence theorem, though, because the random variables $S_{T\wedge m}^2$ are not dominated by an integrable random variable. Thus, a different argument is needed. The key element of this argument will be the completeness of the metric space L^2 (with the metric induced by the L^2−norm).

First, observe that

$$S_{T\wedge m} = \sum_{k=1}^{m} X_k \mathbf{1}\{T \geq k\}.$$

Now let's calculate the *covariances* (i.e., L^2 inner product) of the summands. For any two integers $1 \leq m < n < \infty$,

$$E(X_m \mathbf{1}\{T \geq m\})(X_n \mathbf{1}\{T \geq n\}) = 0,$$

by Lemma 4.1, because the random variable X_n is independent of the three other random variables in the product. Hence, for any $0 \leq m < n < \infty$,

$$\begin{aligned}
E(S_{T\wedge n} - S_{T\wedge m})^2 &= E\left(\sum_{k=m+1}^{n} X_k \mathbf{1}\{T \geq k\}\right)^2 \\
&= \sum_{k=m+1}^{n} E X_k^2 \mathbf{1}\{T \geq k\} \\
&= \sigma^2 \sum_{k=m+1}^{n} P\{T \geq k\} \\
&= \sigma^2 ET \wedge n - \sigma^2 ET \wedge m.
\end{aligned}$$

Since $ET < \infty$, this implies (by the monotone convergence theorem) that the sequence $S_{T\wedge m}$ is Cauchy with respect to the L^2−norm. By the completeness of L^2, it follows that the sequence $S_{T\wedge m}$ *converges* in L^2−norm. But $S_{T\wedge m} \to S_T$ *pointwise*, so the only possible L^2−limit is S_T. Finally, since the random variables $S_{T\wedge m}$ converge in L^2 to S_T their L^2−norms also converge, and we conclude that

$$ES_T^2 = \lim_{m\to\infty} ES_{T\wedge m}^2 = \lim_{m\to\infty} \sigma^2 ET \wedge m = \sigma^2 ET.$$

□

Theorem 4.11. *(Wald's Third Identity) Assume that the random variables X_i are independent, identically distributed, nonnegative, and have expectation $EX_i = 1$. Then for any bounded stopping time T,*

$$E \prod_{i=1}^{T} X_i = 1. \tag{4.5}$$

Proof. Assume that T is a stopping time bounded by a nonnegative integer m. By Lemma 4.1, $E\prod_{i=k+1}^{m} X_i = 1$ for any two (nonrandom) integers $m \geq k \geq 0$. In addition, for each

$k < m$ the random variables $X_{k+1}, X_{k+2}, \dots, X_m$ are independent of $\mathbf{1}\{T = k\}$, and so by linearity of expectation

$$\begin{aligned}
E\prod_{i=1}^{T} X_i &= \sum_{k=0}^{m} E\prod_{i=1}^{T} X_i \mathbf{1}\{T = k\} \\
&= \sum_{k=0}^{m} E\prod_{i=1}^{k} X_i \mathbf{1}\{T = k\} \\
&= \sum_{k=0}^{m} E\prod_{i=1}^{k} X_i \mathbf{1}\{T = k\} E \prod_{i=k+1}^{m} X_i \\
&= \sum_{k=0}^{m} E\prod_{i=1}^{k} X_i \mathbf{1}\{T = k\} \prod_{i=k+1}^{m} X_i \\
&= \sum_{k=0}^{m} E\prod_{i=1}^{m} X_i \mathbf{1}\{T = k\} \\
&= E\prod_{i=1}^{m} X_i = 1
\end{aligned}$$

□

4.2 Nearest Neighbor Random Walks on $\mathbb{Z}$

Definition 4.12. The sequence $S_n = \sum_{i=1}^{n} X_i$ is said to be a *nearest neighbor random walk* (or a *p-q random walk*) on the integers if the random variables X_i are independent, identically distributed and have common distribution

$$P\{X_i = +1\} = 1 - P\{X_i = -1\} = p = 1 - q.$$

If $p = 1/2$ then S_n is called the *simple nearest neighbor random walk*. In general, if $p \neq 1/2$ then we shall assume that $0 < p < 1$ to avoid trivialities.

The Gambler's Ruin Problem. Fix two integers $A < 0 < B$. What is the probability that a $p - q$ random walk S_n (starting at the default initial state $S_0 = 0$) will visit B before A? This is the *gambler's ruin problem*. It is not difficult to see (or even to prove) that the random walk must, with probability one, exit the interval (A, B), by an argument that I will refer to as *Stein's trick*. Break time into successive blocks of length $A + B$. In any such block where all of the steps of the random walk are $+1$, the random walk must exit the interval (A, B), if it has not already done so. Since there are infinitely many time blocks, and since for each the probability of $A + B$ consecutive $+1$ steps is $p^{A+B} > 0$, the strong law of large numbers for Bernoulli random variables implies that with probability one there will eventually be a block of $A + B$ consecutive $+1$ steps.

Proposition 4.13. *Let S_n be a simple nearest neighbor random walk on $\mathbb{Z}$, and for any integers $A < 0 < B$ let $T = T_{A,B}$ be the first time n such that $S_n = A$ or B. Then*

$$P\{S_T = B\} = 1 - P\{S_T = A\} = \frac{|A|}{|A| + B} \quad \textit{and} \tag{4.6}$$

$$ET = |AB|. \tag{4.7}$$

Proof. Wald 1 and 2. To see that $ET < \infty$, observe that T is dominated by $(|A| + B)$ times a geometric random variable, by Stein's trick. □

Corollary 4.14. *Let S_n be a simple nearest neighbor random walk on $\mathbb{Z}$. For any integer $a \neq 0$ define τ_a to be the smallest integer n such that $S_n = a$, or $\tau_a = \infty$ if there is no such n. Then*

$$P\{\tau_a < \infty\} = 1 \quad \textit{and} \quad E\tau_a = \infty. \tag{4.8}$$

Proof. Without loss of generality assume that $a > 0$. Clearly, $\tau_a < \infty$ on the event that $T_{A,a} < \infty$ and $S_{T_{A,a}} = a$, so for any $A > -\infty$,

$$P\{\tau_a < \infty\} \geq \frac{|A|}{a + |A|}$$

It follows that $P\{\tau_a < \infty\} = 1$. Furthermore, $\tau_a \geq T_{A,a}$, so for any $A > -\infty$

$$E\tau_a \geq |A|a.$$

□

Proposition 4.15. *Let S_n be the $p - q$ nearest neighbor random walk on $\mathbb{Z}$, and for any integers $A < 0 < B$ let $T = T_{A,B}$ be the first time n such that $S_n = A$ or B. Then*

$$P\{S_T = B\} = 1 - P\{S_T = A\} = \frac{1 - (q/p)^A}{(q/p)^B - (q/p)^A}. \tag{4.9}$$

Proof. The random variable T is almost surely finite, by Stein's trick, and so $T \wedge m \uparrow T$ and $S_{T\wedge m} \to S_T$ as $m \to \infty$. Observe that $E(q/p)^{X_i} = 1$, so Wald's third identity implies that for each $m = 1, 2, \ldots$,

$$E\left(\frac{q}{p}\right)^{S_{T\wedge m}} = E\prod_{i=1}^{T\wedge m}(q/p)^{X_i} = 1.$$

Now the random variables $(q/p)^{S_{T\wedge m}}$ are uniformly bounded, because up until time T the random walk stays between A and B; consequently, the dominated convergence theorem implies that

$$E\left(\frac{q}{p}\right)^{S_T} = 1.$$

Thus,

$$\left(\frac{q}{p}\right)^B P\{S_T = B\} + \left(\frac{q}{p}\right)^A P\{S_T = A\} = 1;$$

since $P\{S_T = A\} = 1 - P\{S_T = B\}$, the equality (4.9) follows. □

Corollary 4.16. *Let S_n be the $p-q$ nearest neighbor random walk on $\mathbb{Z}$ with $q < \frac{1}{2} < p$, and for any integer $a \neq 0$ define τ_a to be the smallest integer n such that $S_n = a$, or $\tau_a = \infty$ if there is no such n. Then*

$$P\{\tau_a < \infty\} = 1 \quad \textit{if } a \geq 1,$$
$$P\{\tau_a < \infty\} = (q/p)^{|a|} \quad \textit{if } a \leq -1.$$

Exercise 4.17. For $p-q$ nearest neighbor random walk on $\mathbb{Z}$, calculate $ET_{A,B}$.

First-Passage Time Distribution for Simple Random Walk. Let S_n be simple random walk with initial state $S_0 = 0$, and let $\tau = \tau(1)$ be the first passage time to the level 1, as in Corollary 4.14. We will now deduce the complete distribution of the random variable τ, by using Wald's third identity to calculate the probability generating function Es^τ. For this, we need the moment generating function of ξ_1:

$$\varphi(\theta) = Ee^{\theta\xi_1} = \frac{1}{2}(e^\theta + e^{-\theta}) = \cosh\theta.$$

Recall that the function $\cosh\theta$ is even, and it is strictly increasing on the half-line $\theta \in [0,\infty)$; consequently, for every $y > 1$ the equation $\cosh\theta = y$ has two real solutions $\pm\theta$. Fix $0 < s < 1$, and set $s = 1/\varphi(\theta)$; then by solving a quadratic equation (exercise) you find that for $\theta > 0$,

$$e^{-\theta} = \frac{1 \pm \sqrt{1-s^2}}{s}.$$

Because $e^{-\theta} < 1$ for $\theta > 0$, the relevant root is

$$e^{-\theta} = \frac{1 - \sqrt{1-s^2}}{s}.$$

Now let's use the third Wald identity. Since this only applies directly to *bounded* stopping times, we'll use it on $\tau \wedge n$ and then hope for the best in letting $n \to \infty$. The identity gives

$$E\left(\frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}}\right) = 1.$$

We will argue below that if $\theta > 0$ then it is permissible to take $n \to \infty$ in this identity. Suppose for the moment that it is; then since $S_\tau \equiv 1$, the limiting form of the identity will read, after the substitution $s = 1/\varphi(\theta)$,

$$e^\theta Es^\tau = 1.$$

Using the formula for $e^{-\theta}$ obtained above, we conclude that

$$\boxed{Es^{\tau} = \frac{1-\sqrt{1-s^2}}{s}} \tag{4.10}$$

To justify letting $n \to \infty$ above, we use the dominated convergence theorem. First, since $\tau < \infty$ (at least with probability one),

$$\lim_{n\to\infty} \frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}} = \frac{\exp\{\theta S_{\tau}\}}{\varphi(\theta)^{\tau}}.$$

Hence, by the DCT, it will follow that interchange of limit and expectation is allowable provided the integrands are dominated by an integrable random variable. For this, examine the numerator and the denominator separately. Since $\theta > 0$, the random variable $e^{\theta S_{\tau\wedge n}}$ cannot be larger than e^{θ}, because on the one hand, $S_\tau = 1$, and on the other, if $\tau > n$ then $S_n \le 0$ and so $e^{S_{\tau\wedge n}} \le 1$. The denominator is even easier: since $\varphi(\theta) = \cosh\theta \ge 1$, it is always the case that $\varphi(\theta)^{\tau\wedge n} \ge 1$. Thus,

$$\frac{\exp\{\theta S_{\tau\wedge n}\}}{\varphi(\theta)^{\tau\wedge n}} \le e^{\theta},$$

and so the integrands are uniformly bounded.

The exact distribution of the first-passage time $\tau = \tau(1)$ can be recovered from the generating function (4.10) with the aid of *Newton's binomial formula*, according to which

$$\sqrt{1-s^2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-s^2)^n \quad \text{for all } |s| < 1. \tag{4.11}$$

From equation (4.10) we now deduce that

$$Es^{\tau} = \sum_{n=1}^{\infty} s^n P\{\tau = n\} = \sum_{n=1}^{\infty} (-1)^n \binom{1/2}{n} s^{2n-1}.$$

Matching coefficients, we obtain

Proposition 4.18. $P\{\tau = 2n-1\} = (-1)^n \binom{1/2}{n}$ *and* $P\{\tau = 2n\} = 0$.

Exercise 4.19. Verify that $P\{\tau = 2n-1\} = 2^{-2n+1}(2n-1)^{-1}\binom{2n-1}{n}$. This implies that

$$P\{\tau = 2n-1\} = \frac{P\{S_{2n-1} = 1\}}{2n-1} \tag{4.12}$$

Exercise 4.20. Show that $P\{\tau = 2n-1\} \sim C/n^{3/2}$ for some constant C, and identify C. (Thus, the density of τ obeys a *power law* with exponent $3/2$.)

Exercise 4.21. (a) Show that the generating function $F(s) = Es^{\tau}$ given by equation (4.10) satisfies the relation

$$1 - F(s) \sim \sqrt{2}\sqrt{1-s} \qquad \text{as } s \to 1-. \tag{4.13}$$

(b) The random variable $\tau(m) = \min\{n : S_n = m\}$ is the sum of m independent copies of $\tau = \tau(1)$, and so its probability generating function is the nth power of $F(s)$. Use this fact and the result of part (a) to show that for every real number $\lambda > 0$,

$$\lim_{m\to\infty} E\exp\{-\lambda\tau(m)/m^2\} = e^{-\sqrt{2\lambda}} \tag{4.14}$$

Remark 4.22. The function $\varphi(\lambda) = \exp\{-\sqrt{2\lambda}\}$ is the Laplace transform of a probability density called the *one-sided stable law of exponent* $1/2$. This is the distribution of the first-passage time to the level 1 for the *Wiener process* (also called *Brownian motion*). In effect, the result of Exercise 4.21 (b) implies that the random variables $\tau(m)/m^2$ converge in distribution to the stable law of exponent $1/2$.

4.3 L^2–Maximal Inequality and Convergence of Random Series

Assume in this section that $X_1, X_2, \dots$ are independent – but not necessarily identically distributed – random variables with

$$EX_i = 0 \quad \text{and} \quad EX_i^2 := \sigma_i^2 < \infty.$$

Set $S_n = \sum_{i=1}^n X_i$ and $S_0 = 0$. The next proposition is an extension of Wald's second identity to sums of non-identically distributed random variables.

Proposition 4.23. *For any bounded stopping time T,*

$$ES_T^2 = E\sum_{i=1}^{T} \sigma_i^2.$$

Proof. HW. □

Corollary 4.24. *(L^2 Maximal Inequality) For any scalar $\alpha > 0$ and any integer $m \geq 0$,*

$$P\{\max_{n\le m} |S_n| \geq \alpha\} \leq \alpha^{-2}\sum_{i=1}^{m}\sigma_i^2 \quad \textit{and therefore}$$

$$P\{\sup_{n\ge 1} |S_n| \geq \alpha\} \leq \alpha^{-2}\sum_{i=1}^{\infty}\sigma_i^2.$$

Theorem 4.25. *If $\sum_{n=1}^{\infty}\sigma_n^2 < \infty$ then the random variables S_n converge in L^2–norm and almost surely as $n \to \infty$ to a limit S_∞ with expectation $ES_\infty = 0$.*

Proof. The summands X_i are uncorrelated (that is, orthogonal in L^2) by Lemma 4.1. Consequently, the L^2- distance between S_n and S_{n+m} is

$$\|S_{n+m} - S_n\|_2^2 = \sum_{i=n+1}^{n+m} \sigma_i^2.$$

Since $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$, it follows that the sequence S_n is Cauchy in L^2, and hence by the completeness of L^2 there exists a random variable $S_\infty \in L^2$ such that

$$\lim_{n\to\infty} E|S_\infty - S_n|^2 = 0.$$

To prove that $S_n \to S_\infty$ almost surely, it suffices to show that for every $\varepsilon > 0$ there exists $n_\varepsilon < \infty$ such that if $n \geq n_\varepsilon$ then

$$P\{|S_\infty - S_n| > \varepsilon \text{ for some } n \geq n_\varepsilon\} \leq \varepsilon.$$

This follows from the Maximal Inequality, which implies that for any $m < \infty$,

$$P\{|S_m - S_n| > \varepsilon/2 \text{ for some } n \geq m\} \leq \frac{4}{\varepsilon^2} \sum_{n=m}^{\infty} \sigma_n^2.$$

Finally, since $S_n \to S_\infty$ in L^2, the random variables S_n are uniformly integrable. Since $S_n \to S_\infty$ almost surely, it follows that $ES_n \to ES_\infty$. But by hypothesis, $ES_n = 0$. □

Example 4.26. Let $X_1, X_2, \ldots$ be independent, identically distributed Rademacher$-1/2$, that is, $P\{X_i = +1\} = P\{X_i = -1\} = 1/2$. Then the random series

$$\sum_{n=1}^{\infty} \frac{X_n}{n}$$

converges almost surely and in L^2. The series does *not* converge absolutely.

4.4 Kolmogorov's Strong Law of Large Numbers

Proposition 4.27. *(Kronecker's Lemma) Let a_n be an* increasing *sequence of positive numbers such that* $\lim_{n\to\infty} a_n = \infty$*, and let x_k be a sequence of real numbers such that the series $\sum_{n=1}^{\infty}(x_n/a_n)$ converges (not necessarily absolutely). Then*

$$\lim_{m\to\infty} \frac{1}{a_m} \sum_{n=1}^{m} x_n = 0. \tag{4.15}$$

Proof. This is an exercise in *summation by parts*, a technique that is frequently of use in dealing with sequences of sums. The idea is to represent the summands x_i of interest as differences of successive terms: in this case,

$$x_n = a_n(s_n - s_{n+1}) \quad \text{where} \quad s_n = \sum_{i=n}^{\infty} \frac{x_i}{a_i}.$$

The hypothesis ensures that the series defining s_n converge, and also imply that $\lim_{n\to\infty} s_n = 0$. Now write

$$\begin{aligned}\frac{1}{a_m}\sum_{n=1}^{m} x_n &= \frac{1}{a_m}\sum_{n=1}^{m} a_n(s_n - s_{n+1}) \\ &= \frac{1}{a_m}\sum_{n=2}^{m}(a_n - a_{n-1})s_n + \frac{a_1}{a_m}s_1 - s_{m+1}.\end{aligned}$$

It is clear that the last two terms converge to 0 as $m \to \infty$, because $a_m \to \infty$. Therefore, to prove the proposition it suffices to show that $a_m^{-1}\sum_{n=2}^{m}(a_n - a_{n-1})s_n$ converges to 0.

Fix $\varepsilon > 0$, and choose $K - K(\varepsilon)$ so large that $|s_n| < \varepsilon$ for all $n \geq K$. Write

$$a_m^{-1}\sum_{n=2}^{m}(a_n - a_{n-1})s_n = a_m^{-1}\sum_{n=2}^{K}(a_n - a_{n-1})s_n + a_m^{-1}\sum_{n=K+1}^{m}(a_n - a_{n-1})s_n = f_m + g_m.$$

Since $a_m \to \infty$ and since the sum $\sum_{n=2}^{K}(a_n - a_{n-1})s_n$ does not change as m increases, we have $\lim_{m\to\infty} f_m = 0$. On the other hand, since the sequence a_n is nondecreasing and since $|s_n| < \varepsilon$ for all of the indices $K < n \leq m$,

$$\begin{aligned}|g_m| &\leq a_m^{-1}\sum_{n=K+1}^{m}(a_n - a_{n-1})|s_n| \\ &\leq a_m^{-1}\sum_{n=K+1}^{m}(a_n - a_{n-1})\varepsilon \\ &= a_m^{-1}\varepsilon\sum_{n=K+1}^{m}(a_n - a_{n-1}) = \varepsilon\left(1 - \frac{a_K}{a_m}\right) \leq \varepsilon.\end{aligned}$$

Finally, since $\varepsilon > 0$ is arbitrary, (4.15) follows. □

Theorem 4.28. *(L^2-Strong Law of Large Numbers) Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables with mean $EX_n = 0$ and finite variance $\sigma^2 = EX_n^2 < \infty$, and let $S_n = \sum_{i=1}^{n} X_i$. Then with probability one,*

$$\lim_{n\to\infty} S_n/n = 0. \tag{4.16}$$

Proof. Theorem 4.25 implies that the series $\sum_{n=1}^{\infty}(X_n/n)$ converges almost surely, because the variances are summable. Kronecker's Lemma implies that on the event that the series $\sum_{n=1}^{\infty}(X_n/n)$ converges, the averages (4.17) converge to 0. □

In fact, the hypothesis that the summands have finite variance is extraneous: only finiteness of the first moment is needed. This is *Kolmogorov's Strong Law Of Large Numbers.*

Theorem 4.29. *(Kolmogorov) Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables with finite first moment $E|X_1| < \infty$ and mean $EX_n = \mu$, and let $S_n = \sum_{i=1}^n X_i$. Then with probability one,*

$$\lim_{n\to\infty} S_n/n = \mu. \tag{4.17}$$

Lemma 4.30. *Let $X_1, X_2, \ldots$ be identically distributed random variables with finite first moment $E|X_1| < \infty$ and mean $EX_n = 0$. Then for each $\varepsilon > 0$*

$$P\{|X_n| \geq \varepsilon n \ \textit{ infinitely often}\} = 0.$$

Proof. By Borel-Cantelli it suffices to show that $\sum_{n=1}^\infty P\{|X_n| \geq \varepsilon n\} < \infty$. Since the random variables are identically distributed, it suffices to show that $\sum_{n=1}^\infty P\{|X_1| \geq \varepsilon n\} < \infty$. But $|X_1|/\varepsilon := Y$ has finite first moment $EY = E|X_1|/\varepsilon$, and hence so does $[Y]$ (where $[\cdot]$ denotes the greatest integer function). Since Y takes values in the set of nonnegative integers,

$$EY = \sum_{n=1}^\infty P\{Y \geq n\} = \sum_{n=1}^\infty P\{Y|X_1| \geq \varepsilon n\}.$$

□

Proof of Theorem 4.29. Without loss of generality, we may assume that $\mu = 0$. For each $n = 1, 2, \ldots$ define Y_n by truncating X_n at the levels $\pm n$, that is, $Y_n = X_n \mathbf{1}\{|X_n| \leq n\}$, and let $S_n^Y = \sum_{i=1}^n Y_i$. By Lemma 4.30, with probability one $Y_n = X_n$ except for at most finitely many indices n. Consequently, to prove that $S_n/n \to 0$ almost surely it suffices to show that $S_n^Y/n \to 0$ almost surely.

The random variables $Y_1, Y_2, \ldots$ are independent but no longer identically distributed, and furthermore the expectations EY_n need not $= 0$. Nevertheless,

$$EY_n = EX_n \mathbf{1}\{|X_n| \leq n\} = EX_1 \mathbf{1}\{|X_1| \leq n\} \longrightarrow 0$$

by the dominated convergence theorem (since $E|X_1| < \infty$). Therefore, the averages $n^{-1}\sum_{i=1}^n EY_i$ converge to 0 as $n \to \infty$. Thus, to prove that $S_n^Y/n \to 0$ almost surely, it suffices to prove that with probability 1,

$$\frac{1}{n}\sum_{i=1}^n (Y_i - EY_i) \longrightarrow 0.$$

By Kronecker's Lemma, it now suffices to show that with probability one the sequence $\sum_{i=1}^n (Y_i - EY_i)/i$ converges to a finite limit, and for this it suffices, by the Khintchine-Kolmogorov theorem, to prove that $\sum_{n=1}^\infty \mathrm{Var}(Y_n/n) < \infty$. Finally, since $\mathrm{Var}(Y_n) = E(Y_n - EY_n)^2 \leq EY_n^2$, it suffices to show that

$$\sum_{n=1}^\infty n^{-2} EY_n^2 < \infty.$$

Here we go:

$$\begin{aligned}\sum_{n=1}^{\infty} n^{-2}EY_n^2 &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} n^{-2}EX_1^2\mathbf{1}\{k-1<|X_1|\le k\}\\ &= \sum_{k=1}^{\infty} EX_1^2\mathbf{1}\{k-1<|X_1|\le k\}\sum_{n=k}^{\infty} n^{-2}\\ &\le 2\sum_{k=1}^{\infty} EX_1^2\mathbf{1}\{k-1<|X_1|\le k\}k^{-1}\\ &\le 2\sum_{k=1}^{\infty} k^2P\{k-1<|X_1|\le k\}k^{-1}\\ &= 2\sum_{k=1}^{\infty} kP\{k-1<|X_1|\le k\}\\ &\le 2(E|X_1|+1)<\infty.\end{aligned}$$

Here we have used the fact that $\sum_{n=k}^{\infty} n^{-2}\le\int_{k-1}^{\infty} t^{-2}\,dt=(k-1)^{-1}\le 2k^{-1}$, and (of course) the hypothesis that the first moment of $|X_1|$ is finite.

□

Definition 4.31. A sequence $X_1, X_2, \dots$ of random variables is said to be *$m-$dependent* for some integer $m\ge 1$ if for every $n\ge 1$ the $\sigma-$algebras $\sigma(X_i)_{i\le n}$ and $\sigma(X_i)_{i\ge n+m+1}$ are independent.

Exercise 4.32. If $X_1, X_2, \dots$ are $m-$dependent then for each i the random variables

$$X_i, X_{i+m+1}, X_{i+2m+2}, \dots$$

are independent.

Corollary 4.33. *If $X_1, X_2, \dots$ are $m-$dependent random variables all with the same distribution, and if $E|X_1|<\infty$ and $EX_i=\mu$ then with probability one,*

$$\lim \frac{S_n}{n}=\mu.$$

4.5 The Kesten-Spitzer-Whitman Theorem

Next, we will use Kolmogorov's Strong Law of Large Numbers to derive a deep and interesting theorem about the behavior of random walks on the integer lattices $\mathbb{Z}^d$. A *random walk* on $\mathbb{Z}^d$ is just the sequence $S_n=\sum_{k=1}^n X_k$ of partial sums of a sequence $X_1, X_2, \dots$ of independent, identically distributed random *vectors* taking values in $\mathbb{Z}$; these random vectors X_k are called the *steps* of the random walk, and their common distribution is the *step distribution*. For example, the simple nearest neighbor random walk on $\mathbb{Z}$ has step distribution

$$P\{X_k = \pm e_i\} = \frac{1}{4}$$

where e_1 and e_2 are the standard unit vectors in $\mathbb{R}$.

Theorem 4.34. *(Kesten-Spitzer-Whitman) Let S_n be a random walk on $\mathbb{Z}^d$. For each $n = 0, 1, 2, \dots$ define R_n to be the number of distinct sites visited by the random walk in its first n steps, that is,*

$$R_n := \text{cardinality}\{S_0, S_1, \dots, S_n\}. \tag{4.18}$$

Then

$$\frac{R_n}{n} \longrightarrow P\{\textit{no return to } S_0\} \quad \textit{a.s.} \tag{4.19}$$

I will only prove the weaker statement that R_n/n converges to $P\{\text{no return}\}$ *in probability*. Even the weaker statement has quite a lot of information in it, though, as the next corollary shows.

Corollary 4.35. *Let S_n be a random walk on $\mathbb{Z} = \mathbb{Z}^1$ whose step distribution has finite first moment and mean* 0*. Then*

$$P\{\textit{no return to } 0\} = 0.$$

Proof. Since the increments $X_n = S_n - S_{n-1}$ have finite first moment and mean zero, Kolmogorov's SLLN implies that $S_n/n \to 0$ almost surely. This in turn implies that for every $\varepsilon > 0$, eventually $|S_n| \le n\varepsilon$, and so the number of distinct sites visited by time n (at least for large n) cannot be much larger than the total number of integers between $-n\varepsilon$ and $+n\varepsilon$. Thus, for sufficiently large n,

$$R_n \le 4\varepsilon n.$$

Since $\varepsilon > 0$ is arbitrary, it follows that $\lim R_n/n = 0$ almost surely. The KSW theorem does the rest. □

Proof of the KSW Theorem. To calculate R_n, run through the first $n+1$ states S_j of the random walk and for each count $+1$ if S_j is not revisited by time n, that is,

$$R_n = \sum_{j=0}^{n} \mathbf{1}\{S_j \text{ not revisited before time } n\}.$$

The event that S_j is not revisited by time n contains the event that S_j is never revisited at all; consequently,

$$R_n \ge \sum_{j=0}^{n} \mathbf{1}\{S_j \text{ never revisited}\} = \sum_{j=0}^{n} \mathbf{1}\{S_j \ne S_{m+j} \text{ for any } m \ge 1\}.$$

This clearly implies that

$$ER_n/n \geq P\{\text{no return}\}. \tag{4.20}$$

We can also obtain a simple *upper* bound for R_n by similar reasoning. For this, consider again the event that S_j is not revisited by time n. Fix $M \geq 1$. If $j \leq n - M$, then this event is contained in the event that S_j is not revisited in the next M steps. Thus,

$$R_n \leq \sum_{j=0}^{n-M} \mathbf{1}\{S_j \neq S_{j+i} \text{ for any } 1 \leq i \leq M\} + M. \tag{4.21}$$

The random variable $Y_j^M := \mathbf{1}\{S_j \neq S_{j+i} \text{ for any } 1 \leq i \leq M\}$ is a Bernoulli random variable that depends only on the increments $X_{j+1}, X_{j+2}, \ldots, X_{j+M}$ of the underlying random walk. Since these increments are independent and identically distributed, it follows that for any M the sequence $\{Y_j^M\}_{j\geq 1}$ is an $M-$dependent sequence of identically distributed Bernoulli random variables, and so the strong law of large numbers applies: in particular, with probability one,

$$\lim_{n\to\infty} n^{-1} \sum_{j=1}^{n} Y_j^M = EY_1^M = P\{S_i \neq 0 \text{ for any } i \leq M\}.$$

Consequently, by (4.21), for every $M \geq 1$, with probability one,

$$\limsup_{n\to\infty} \frac{R_n}{n} \leq P\{S_i \neq 0 \text{ for any } i \leq M\}.$$

The dominated convergence theorem implies that the probabilities on the right converge (down) to $P\{\text{no return}\}$, so this proves that with probability one

$$\limsup_{n\to\infty} \frac{R_n}{n} \leq P\{\text{no return}\}.$$

So here is what we have proved: (a) the random variables R_n/n have limsup no larger than $P\{\text{no return}\}$, and (b) have expectations no *smaller* than $P\{\text{no return}\}$. Since $R_n/n \leq 1$, this implies, by the next exercise, that in fact

$$R_n/n \xrightarrow{P} P\{\text{no return}\}.$$

Exercise 4.36. Let Z_n be a sequence of uniformly bounded random variables (that is, there exists a constant $C < \infty$ such that $|Z_n| \leq C$ for every n) such that $\limsup Z_n \leq \alpha$ almost surely and $EZ_n \geq \alpha$. Prove that $Z_n \to \alpha$ in probability.

□

Exercise 4.37. Use the Kesten-Spitzer-Whitman theorem to calculate $P\{\text{no return to } 0\}$ for $p - q$ nearest-neighbor random walk on $\mathbb{Z}$ when $p > q$.