Ergodic Theorems

The first two sections are based on the book by Breiman (1968, Chapter 6). (See also Steele, 2015.) The third section is based on an elegant short paper by Steele (1989).

The Breiman book is available in electronic form via the Yale Library at

`http://hdl.handle.net/10079/bibid/12786310` .

1 Strong laws of large numbers

Many weeks ago I stated, but did not prove, the following result.

<1> **Theorem.** *Let $X_1, X_2, \dots$ be a sequence of independent, identically distributed random variables defined on some $(\Omega, \mathcal{F}, \mathbb{P})$). Then*

$$n^{-1} \sum\nolimits_{i \le n} X_i(\omega) \to \mathbb{P}X_1 \qquad a.e.[\mathbb{P}].$$

This Theorem is actually a special case of a result for stationary sequences.

<2> **Theorem.** *Let $X_1, X_2, \dots$ be a stationary sequence of random variables defined on some $(\Omega, \mathcal{F}, \mathbb{P})$). If $\mathbb{P}|X_1| < \infty$ then*

$$n^{-1} \sum\nolimits_{i \le n} X_i(\omega) \to \mathbb{P}_{\mathcal{G}}X_1 \qquad a.e.[\mathbb{P}],$$

where $\mathcal{G}$ is a (soon to be specified) sub-sigma-field of $\sigma(X)$.

And this Theorem can be derived from an even more general result involving ***measure preserving transformations***.

<3> **Definition.** *Suppose* $(\omega, \mathcal{F}, \mathbb{P})$ *is a probability space and* $T : \Omega \to \Omega$ *is* $\mathcal{F}\backslash\mathcal{F}$-*measurable. The map* T *is said to be measure preserving if the image of* $\mathbb{P}$ *under* T *is* $\mathbb{P}$ *itself. That is,* $\mathbb{P}f(\omega) = \mathbb{P}f(T\omega)$ *at least for all* f *in* $\mathcal{M}^+(\Omega, \mathcal{F})$.

It is easy to manufacture stationary process from a measure preserving transformation (m.p.t.) T. Let f be any measurable, real-valued function on Ω define $X_1(\omega) = f(\omega)$ and $X_2(\omega) = f(T\omega)$,..., and $X_n(\omega) = f(T^{n-1}\omega)$, ... Then

$$\begin{aligned}
&\mathbb{P}g(X_{n+1}, \dots, X_{n+k}) \\
&\quad = \mathbb{P}g\left(f(T^n\omega), \dots, f(T^{n+k-1}(\omega)\right) \\
&\quad = \mathbb{P}g\left(f(T^0\omega), \dots, f(T^{k-1}(\omega)\right) \qquad \text{because } T^n \text{ is also m.p.} \\
&\quad = \mathbb{P}g(X_1, \dots, X_k)
\end{aligned}$$

Note the convention that T^0 is the identity map.

Associated with each m.p.t. T is an ***invariant sigma-field***:

$$\mathcal{I} = \{F \in \mathcal{F} : T^{-1}F = F\}.$$

An easy argument using limits of simple functions shows that an $\mathcal{F}$-measurable real valued function f is $\mathcal{I}$-measurable if and only $f(\omega) = f(T\omega)$ for all $\omega \in \Omega$. Such a function is said to be ***invariant under*** T. The m.p.t. T is said to be ***ergodic*** if $\mathcal{I}$ is trivial, that is, if $\mathbb{P}F$ is either zero or one for each F in $\mathcal{I}$. Equivalently, T is ergodic if and only if every invariant measurable function is constant.

<4> **Example.** Suppose T is a m.p.t. on $(\Omega, \mathcal{F}, \mathbb{P})$ and $f : \Omega \to \mathbb{R}$ is $\mathcal{F}\backslash\mathcal{B}(\mathbb{R})$-measurable. Define $S_n(\omega) = \sum_{0 \le i < n} f(T^i\omega)$ and $h(\omega) = \limsup_n S_n(\omega)/n$. Then h is invariant under T:

$$h(T\omega) = \limsup_n S_n(T\omega)/n = \limsup_n \frac{n+1}{n} \frac{S_{n+1}(\omega) - f(\omega)}{n+1},$$

which equals $h(\omega)$ because $(n+1)/n \to 1$ and $f(\omega)/(n+1) \to 0$ as $n \to \infty$.

□

<5> **Theorem.** *(Ergodic theorem) Let* T *be a m.p.t on* $(\Omega, \mathcal{F}, \mathbb{P})$ *and let* f *be a function in* $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. *Define* $S_n(\omega) = \sum_{0 \le i < n} f(T^i\omega)$ *and* $Z = \mathbb{P}_{\mathcal{I}}f$. *Then* $S_n/n \to Z$ *almost surely and* $\mathbb{P}|S_n/n - Z| \to 0$.

If the transformation T *is ergodic then* Z *can be taken as the constant* $\mathbb{P}f$.

The next Section contains the main ideas in the proof of the almost sure convergence part of the Ergodic theorem. Problem [2] shows that the $\mathcal{L}^1$-convergence is a simple Corollary.

2 Heuristics and proof for the Ergodic theorem

We can assume that Z is an invariant function. We need to show that $S_n(\omega)/n \to Z(\omega)$ a.e.[$\mathbb{P}$]. Equivalently, if $f_1(\omega) = f(\omega) - Z(\omega)$, we need to show that

$$n^{-1}\sum\nolimits_{0\le i<n} f_1(T^{i-1}\omega) \to 0 \quad a.e.[\mathbb{P}].$$

To save on notation it is cleaner to assert that, without loss of generality, $Z = 0$. That is, $\mathbb{P}_{\mathcal{I}}f = 0$ so that $\mathbb{P}(Df) = 0$ for each D in $\mathcal{I}$.

For a fixed $\epsilon > 0$, the following argument shows that the set

$$D = \{\omega : \limsup S_n(\omega)/n > \epsilon\}$$

has zero probability. If we cast out a sequence of negligible sets (for a sequence of ϵ's tending to zero) we then get $\limsup S_n/n \le 0$ a.e.[$\mathbb{P}$]. A similar argument with f replaced by $-f$ then gives $\limsup(-S_n)/n \le 0$ a.e.[$\mathbb{P}$], that is, $\liminf S_n/n \ge 0$ a.e.[$\mathbb{P}$], and we are done.

The argument to show that $\mathbb{P}D = 0$ is very elegant. The idea is to prove that

<6> $$\mathbb{P}(fD) \ge \epsilon \mathbb{P}D,$$

then note that $\mathbb{P}(fD) = 0$ because $D \in \mathcal{I}$.

Inequality <6> comes from an analogous inequality with D replaced by the set

<7> $$D_n = \{\omega \in D, H_n(\omega) > 0\} \qquad \text{where } H_n(\omega) = \max_{1\le i\le n}\left(S_i(\omega) - i\epsilon\right).$$

Note that $H_n(\omega)$ increases with n and the sets $\{H_n(\omega) > 0\}$ increase to the set

$$A = \{\omega : \sup_i S_i(\omega)/i > \epsilon\} \supset \{\omega : \limsup_i S_i(\omega)/i > \epsilon\} = D.$$

Thus

$$D_n = D\{H_n > 0\} \uparrow DA = D \qquad \text{as } n \to \infty.$$

The inequality $\mathbb{P}(fD_n) \geq \epsilon \mathbb{P} D_n$ will be a consequence of an integrated version of a pointwise inequality

<8> $$f(\omega) + H_n^+(T\omega) = \epsilon + H_{n+1}(\omega) \geq \epsilon + H_n(\omega) \qquad \text{for all } \omega.$$

Let me first show the integration steps and then explain where <8> comes from.

Multiply both sides of inequality <8> by the indicator function of D_n:

$$fD_n + H_n^+(T\omega)\{\omega \in D\}\{H_n(\omega) > 0\} \geq \epsilon D_n + H_n(\omega)\{\omega \in D\}\{H_n(\omega) > 0\}.$$

On the right-hand side the indicator $\{H_n > 0\}$ converts the H_n to an H_n^+; the left-hand side gets bigger if we just discard the $\{H_n > 0\}$; and the invariance of D lets us replace $\{\omega \in D\}$ by $\{T\omega \in D\}$ on the left-hand side. Those modifications leave us with

<9> $$fD_n + H_n^+(T\omega)\{T\omega \in D\} \geq \epsilon D_n + H_n^+(\omega)\{\omega \in D\}.$$

Notice that we have a function $g(\omega) := H_n^+(\omega)\{\omega \in D\}$ on the right-hand side paired with $g(T\omega)$ on the left-hand side. The measure preserving property of T ensures that $\mathbb{P}g(T\omega) = \mathbb{P}g(\omega)$. If we take expected values of both sides of inequality <9> the g contributions cancel, leaving the desired inequality $\mathbb{P}(fD_n) \geq \epsilon \mathbb{P} D_n$. Dominated Convergence as $n \to \infty$ then gives <6>.

Finally , for the pointwise inequality <8> note that

$$f(\omega) + S_i(T\omega) = f(\omega) + f(T\omega) + \cdots + f(T^i\omega) = S_{i+1}(\omega)$$

so that

$$\begin{aligned} &f(\omega) + H_n^+(T\omega) \\ &\quad = \max\left(f(\omega) + 0, f(\omega) + S_1(T\omega) - \epsilon, \ldots, f(\omega) + S_n(T\omega) - n\epsilon\right) \\ &\quad = \max\left(S_1(\omega), S_2(\omega) - \epsilon, \ldots, S_{n+1}(\omega) - n\epsilon\right) \\ &\quad = \epsilon + H_{n+1}(\omega). \end{aligned}$$

Very neat.

3 Stationary processes (Theorem <2>)

Remember that a sequence of random variables $\{X_i : i \in \mathbb{N}\}$ is stationary if for each $k \in \mathbb{N}$ the random vector $(X_1, \dots, X_k)$ has the same distribution as the random vector $(X_{n+1}, \dots, X_{n+k})$ for each $n \in \mathbb{N}$. Equivalently, for every k and every g in $\mathcal{M}^+(\mathbb{R}^k, \mathcal{B}(\mathbb{R})^k)$,

$$\mathbb{P}g(X_{n+1}, \dots, X_{n+k}) = \mathbb{P}g(X_1, \dots, X_k) \qquad \text{for every } n.$$

In particular, $\mathbb{P}|X_i|$ is the same for all i.

As you know from HW2.3, the sequence defines a map

$$X(\omega) = (X_1(\omega), X_2(\omega), \dots)$$

from Ω into $\mathbb{R}^{\mathbb{N}}$. The map is $\mathcal{F}\backslash\mathcal{B}$-measurable for $\mathcal{B} = \mathcal{B}(\mathbb{R})^{\mathbb{N}}$, the product sigma-field on $\mathbb{R}^{\mathbb{N}}$. The distribution of X is a probability measure P on $\mathcal{B}$. If the sequence is stationary then P is preserved by the shift map

$$T(x_1, x_2, \dots) = (x_2, x_3, \dots).$$

That is, $Pf(Tx) = Pf(x)$ at least for each f in $\mathcal{M}^+(\mathbb{R}^{\mathbb{N}}, \mathcal{B})$.

If $\mathbb{P}|X_1| < \infty$ then the function $f(x) = x_1$ belongs to $\mathcal{L}^1(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, P)$. Define $\mathfrak{S}(x) = f(x) + \dots + f(T^{n-1}x) = \sum_{i \le n} x_i$. Theorem <5>, for the probability space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}, P)$, implies that

$$\mathfrak{S}(x)/n \to Z(x) = P_{\mathcal{J}} f \qquad \text{both almost surely and in } \mathcal{L}^1(P).$$

Put another way, the almost sure convergence means that the set

$$E = \{x \in \mathbb{R}^{\mathbb{N}} : \mathfrak{S}(x)/n \to Z(x)\}$$

has P-measure 1.

Now let me can pull the result back to Ω. First note that

$$S_n(\omega) = \sum\nolimits_{i \le n} X_i(\omega) = \mathfrak{S}_n(X\omega).$$

Define

$$\mathcal{G} = \{X^{-1}J : J \in \mathcal{J}\},$$

a sub-sigma-field of $\mathcal{F}$. By a result proved in class near the start of the semester,

$$\mathcal{M}^+(\Omega, \mathcal{G}) = \{h \circ X : h \in \mathcal{M}^+(\mathbb{R}^{\mathbb{N}}, \mathcal{J})\}.$$

This representation shows that $Z(X\omega)$ is a version of $\mathbb{P}_{\mathcal{G}}X_1$: for each $h \in \mathcal{M}^+(\mathbb{R}^{\mathbb{N}}, \mathcal{I})$

$$\mathbb{P}h(X\omega)X_1(\omega) = Ph(x)f(x) = Ph(x)Z(x) = \mathbb{P}\mathbb{P}h(X\omega)Z(X\omega).$$

Also $\mathbb{P}\{\omega : X\omega \in E\} = PE = 1$ and, for $\omega \in X^{-1}E$,

$$S_n(\omega)/n \to Z(X\omega),$$

which is the assertion made by Theorem <2>.

Remark. You might wonder whether it is really necessary to pass to the image measure before invoking the Ergodic theorem. Is it possible to define a measure preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$ then invke the Ergodic theorem for that transformation? See Doob (1953, Section X.1) for discussion of this question.

4 The strong law of large numbers (Theorem <1>)

A sequence of iid random variables is clearly stationary. If we can show that the invariance sigma-field $\mathcal{I}$ on $\mathbb{R}^{\mathbb{N}}$, as defined in Section 3, is trivial then the sigma-field $\mathcal{G}$ on Ω will also be trivial. It will then follow that $\mathbb{P}_{\mathcal{G}}X_1 = \mathbb{P}X_1$, as needed for the SLLN.

Triviality of $\mathcal{I}$ is equivalent to the assertion that every function f in $\mathcal{M}_{\text{bdd}}(\mathbb{R}^{\mathbb{N}}, \mathcal{I})$ is constant.

Consider such an f. Without loss of generality suppose $0 \le f(x) \le 1$ for all $x \in \mathbb{R}^{\mathbb{N}}$. Invariance tells us that

$$f(x) = f(x_{k+1}, x_{k+2}, \dots) \qquad \text{for every } k.$$

HW7.3 tells us that for each $\epsilon > 0$ there exists a k and a $g_k \in \mathcal{M}(\mathbb{R}^k, \mathcal{B}(\mathbb{R})^k)$ for which

$$P|f(x) - g_k(x_1, \dots, x_k)| < \epsilon.$$

This approximation might appear counterintuitive because the previous display tells us that f is independent of g_k. In fact that is the reason that f must be degenerate at its expected value $c = Pf$.

Without loss of generality we may assume that $0 \le g_k \le 1$, so that

$$\epsilon > P|f - g_k| \ge P|f - g_k|^2 = P|f - c - (g_k - c)|^2$$

However we also have

$$P|f - c - (g_k - c)|^2 = P|f - c|^2 - 2P(f - c)(g_k - c) + P|g_k - c|^2.$$

Independence kills the cross-product term, leaving a sum of two nonnegative terms. We can conclude that $P|f - c|^2 < \epsilon$ for every $\epsilon > 0$.

Remark. The same method can be used to prove the Kolmogorov zero-one law, UGMTP Example 4.12.

5 Problems

[1] Suppose g is an $\mathcal{F}$-measurable real-valued function on Ω.

(i) Show that $h(\omega) := \limsup n^{-1} \sum_{0 \le i < n} g(T^i\omega)/n$ is $\mathcal{I}$-measurable.

(ii) Deduce that $D := \{\omega : \limsup S_n(\omega)/n > \epsilon\}$ is $\mathcal{I}$-measurable, for each $\epsilon > 0$.

(iii) Suppose $g = g \circ T$ almost surely. That is, $\mathbb{P}\{\omega : g(\omega) = g(T\omega)\} = 1$. Show that $g = h$ almost surely.

[2] Show that the convergence in Theorem <5> also holds in the $\mathcal{L}^1$ sense. Hint: Split f_0 into a sum of $f_1 = f_0\{|f_0| \le C\}$ and $f_2 = f_0\{|f_0| > C\}$, with corresponding decompositions $S_n = S_{n,1} + S_{n,2}$ and $\mathbb{P}_{\mathcal{I}} f_0 = h_1 + h_2$ where $h_i = \mathbb{P}_{\mathcal{I}} f_i$. Choose the constant C large enough that $\mathbb{P}|f_2| < \epsilon$. Show that $\mathbb{P}|S_{n,2}| < \epsilon$ and $\mathbb{P}|h_2| < \epsilon$. There is no need for Egoroff's theorem.

[3] Suppose $\{g_n : n \in \mathbb{N}\}$ and $\{h_n : n \in \mathbb{N}\}$ are subadditive non-negative processes. Define $f_n = g_n h_n$. Show that $\{f_n : n \in \mathbb{N}\}$ is subadditive.

[4] Suppose f is an $\mathcal{F}\mathcal{B}(\mathbb{R})$-measurable function on $\Omega = \mathbb{R}^{\mathbb{N}}$ for which $f(\omega) \le f(T\omega)$ for all ω, as in the proof of Theorem <10>.

(i) For each real α show that $B_\alpha := \{\omega : f(\omega) > \alpha\}$ is a subset of $\{\omega : f(T\omega) > \alpha\} = T^{-1}B_\alpha$.

(ii) Deduce from the fact that T preserves measures that $B_\alpha = T^{-1}B_\alpha$ almost surely.

(iii) Cast out a sequence of negligible sets to deduce that $f(\omega) = f(T\omega)$ almost surely.

(iv) Deduce via Problem [1](iii) that there exists an invariant function h that equals f almost surely.

References

Breiman, L. (1968). *Probability.* Reading, Massachusets: Addison-Wesley.

Doob, J. L. (1953). *Stochastic Processes.* New York: Wiley.

Kingman, J. (1976). Subadditive processes. In *Ecole d'Eté de Probabilités de Saint-Flour V-1975*, pp. 167–223. Springer.

Kingman, J. F. (1968). The ergodic theory of subadditive stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological) 30*(3), 499–510.

Kingman, J. F. C. (1973). Subadditive ergodic theory. *The Annals of Probability 1*(6), 883–899.

Steele, J. M. (1989). Kingman's subadditive ergodic theorem. *Annales de l'Institut Henri Poincaré—Probabilités et Statistiques 25*(1), 93–98.

Steele, J. M. (2004). *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities.* Cambridge University Press.

Steele, J. M. (2015). Explaining a mysterious maximal inequality — and a path to the law of large numbers. *The American Mathematical Monthly 122*(5), 490–494.

Ignore this Section

6 The subadditive ergodic theorem

Kingman (1968) proved a useful extension of the ergodic theorem, which he later (Kingman, 1973, 1976) discussed in more leisurely fashion. J. Michael Steel simplified the proof in a gem of a paper Steele (1989), which you should all read to see how a probability grand master explains technical ideas. If you are impressed by this paper, you should also look at the book Steele (2004). It would be a much greater pleasure to read the literature if everyone wrote as well as JMC.

<10> **Theorem.** *Suppose T is a measure-preserving transformation on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $\{g_n : n \in \mathbb{N}\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is a sequence of integrable functions with the subadditivity property*

$$g_{i+j}(\omega) \le g_i(\omega) + g_j(T^i\omega) \qquad \text{for all } i, j \in \mathbb{N} \text{ and all } \omega.$$

Then there exists an invariant, integrable function h, possibly taking the value $-\infty$, for which $g_n(\omega)/n \to h(\omega)$ almost surely.

<11> **Example.** If f is integrable, the process

$$g_n(\omega) = f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega)$$

is additive (a special case of subadditivity) because

$$g_{k+\ell}(\omega) = g_k(\omega) + \sum\nolimits_{k \le i < k+\ell} f(T^i\omega) = g_k(\omega) + g_\ell(T^k\omega).$$

□

The previous Example suggests (correctly) that the Ergodic Theorem is a special case of Theorem <5>. It is amusing that the more general theorem can be proved using the weaker theorem.

Proof (of Theorem <10>) First a simple inequality:

$$\begin{aligned} g_n(\omega) &\le g_{n-1}(\omega) + g_1(T^{n-1}\omega) \\ &\le g_{n-2}(\omega) + g_1(T^{n-2}\omega) + g_1(T^{n-1}\omega) \\ &\dots \\ &\le g_1(\omega) + g_2(\omega) + \dots + g_n(T^{n-1}\omega). \end{aligned}$$

The process defined by

$$f_n(\omega) = g_n(\omega) - n - \sum\nolimits_{i=0}^{n-1} g_1(T^i\omega)$$

is also subadditive and $f_n(\omega)/n \le -1$. Moreover, for each n,

$$f_{n+1}(\omega) \le f_n(\omega) + f_1(T\omega) < f_n(\omega) \qquad \text{because } f_1 \le -1.$$

Thus

<12> $$0 > f_1(\omega) > f_2(\omega) > \dots > f_n(\omega) > \dots$$

If we can show that $f_n(\omega)/n$ converges almost surely to an invariant function then the result for g_n/n follows via the Ergodic Theorem <5> with $f_0 = g_1$.

Define $f(\omega) = \liminf f_n(\omega)/n$. If we take the lim inf of both sides of the pointwise inequality

$$f_{n+1}(\omega)/n \le \left(f_1(\omega) + f_n(T\omega)\right)/n$$

we get $f(\omega) \le f(T\omega)$ for each ω. By Problem [4], there exists an invariant function h for which function $f = h$ almost surely. Of course we may assume $h(\omega) \le -1$ everywhere.

Consider any invariant function γ for which

$$h(\omega) < \gamma(\omega) < 0 \qquad \text{for all } \omega.$$

(For example, $\gamma(\omega) = \epsilon + \max(-M, h(\omega))$.) The next part of the proof will show that $\limsup f_n(\omega)/n \le \gamma(\omega)$ almost surely on sets with probability arbitrarily close to 1. By taking a sequence of γ_i's that decrease pointwise to h and sets with probabilities converging rapidly enough to 1, we can then conclude that $f_n(\omega)/n \to h(\omega)$ almost surely.

Next comes a surprising use of the Ergodic Theorem. By construction, $f_n(\omega)/n < \gamma(\omega)$ infinitely often, for each ω. The sets

$$A_N := \{\omega : \min_{i\le N} f_i(\omega)/i < \gamma(\omega)\} \uparrow \Omega \qquad \text{for each } \omega.$$

By the conditional expectation version of Monotone Converge, $\mathbb{P}_{\mathcal{J}} A_N \uparrow 1$ almost surely as N goes to infinity.

For an arbitrarily small $\delta > 0$ define $B_N := \{\omega : \mathbb{P}_{\mathcal{J}} A_N > 1 - \delta\}$. With δ fixed, choose N large enough to make $\mathbb{P} B_N > 1 - \delta$.

For almost all $\omega \in B_N$, and all n large enough,

$$n^{-1} \sum\nolimits_{0 \le i < n} \{T^i \omega \in A_N\} > 1 - \delta.$$

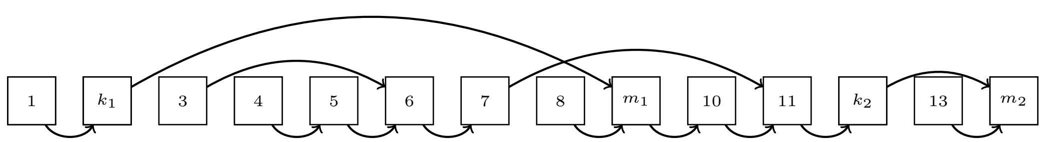
Consider any such ω and a suitably large n (with $n > N$.) Call an integer k in the range $1 \le k \le n - N$ 'good' if $T^k \omega \in A_N$. By definition, for good k there must be an integer ℓ (depending on $\omega, k, n, \delta, \gamma$) with $1 \le \ell \le N$ for which

<13> $$f_\ell(T^k \omega)/\ell < \gamma(T^k \omega) = \gamma(\omega),$$

the final equality by the invariance of γ. Draw an arrow pointing from the integer k to the integer $m = k + \ell$. Note that $m \le n$ because $k \le n - N$ and $\ell \le N$.

Call all the other integers in $[n] := \{i \in \mathbb{N} : 1 \le i \le n\}$ 'bad'. Make the arrow from a bad i point to its successor, $i + 1$.

In the following picture, the arrows for the good integers are on top and the short arrows for the bad integers are on the bottom.



The picture is slightly misleading. If n is large enough there should be at least $(n - N)(1 - \delta)$ good integers, which means there are at most

$$n - (n - N)(1 - \delta) \le N + n\delta$$

bad integers.

Think of the elements of $[n]$ as a one-dimensional version of a Snakes and Ladders board, without the snakes. Play a non-random version of the game by starting at 1 and following the arrows. For example, for the picture the sequence would be

$$1, 2 = k_1, 9 = m_1 = k_1 + \ell_1, 10, 11, 12 = k_2, 14 = m_2 = k_2 + \ell_2, \ldots$$

Suppose we visit good sites $k_1, k_2, \ldots, k_r$ before reaching n. By construction, $k_i < m_i = k_i + \ell_i \le k_{i+1}$ for each i.

By the decreasing property <12> and the bound <13> for each good integer,

$$\begin{aligned} f_n(\omega) &\le f_{m_r}(\omega) \\ &\le f_{k_r}(\omega) + \gamma(\omega)\ell_r \\ &\le f_{m_{r-1}}(\omega) + \gamma(\omega)\ell_r \\ &\le f_{k_{r-1}}(\omega) + \gamma(\omega)\ell_{r-1} + \gamma(\omega)\ell_r \\ &\le \dots \\ &\le f_{k_1}(\omega) + \gamma(\omega)\ell_1 + \dots + \gamma(\omega)\ell_r \\ &\le \gamma(\omega)\sum\nolimits_{i=1}^{r} \ell_i. \end{aligned}$$

The path from 1 to n passes through r good integers and s bad integers, which implies

$$(s \times 1) + \ell_1 + \dots + \ell_r = n - 1,$$

so that

$$\ell_1 + \dots + \ell_r \ge n - 1 - n + (n - N)(1 - \delta) \ge n(1 - \delta) - 1 - N(1 - \delta).$$

The fact that $\gamma(\omega) < 0$ then gives us

$$f_n(\omega)/n \le \gamma(\omega)(1 - \delta - o(1)) \qquad \text{as } n \to \infty.$$

Thus $\limsup f_n(\omega)/n \le \gamma(\omega)(1 - \delta)$ for almost all ω in B_N, for arbitrarily small $\delta > 0$ and $\gamma > h$. That is not quite what I promised but it is good enough for you to prove that $\limsup f_n(\omega)/n \le h(\omega)$ almost surely.

□

7 An application

What would be a good example? Maybe translate the example used by Steele (1989) into a USLLN.