MAT 533, SPRING 2021, Stony Brook University

REAL ANALYSIS II

FOLLAND'S REAL ANALYSIS: CHAPTER 10
TOPICS IN PROBABILTY THEORY

Christopher Bishop

Chapter 10: Topics is Probabolity Theory

10.1 Basic concepts

10.2 The Law of Large Numbers

10.3 The Central Limit Theorem

Chapter 10.1: Basic Concepts

Analysts' Term	**Probabilists' term**
Measure space (X, M, μ), $\quad \mu(X) = 1.$	Sample space $(\Omega, \mathcal{B}, P)$
σ-algebra	σ-field
Measurable set	Event
Measurable real-valued function f	Random variable X
Integral of f, $\int f d\mu$	Expectation or mean of X, $E(X)$
$f \in L^p$	f has finite pth moment
Convergence in measure	Convergence in probability
Almost everywhere, a.e.	Almost surely, a.s.
Borel probability measure on $\mathbb{R}$	Distribution
Fourier transform of a measure	Characteristic function of a distribution
Characteristic function	Indicator function

Defn: The **variance** of a random variable (i.e., function) is

$$\sigma^2(X) = E[(X - E(X))^2],$$

and the standard deviation is

$$\sigma(X) = \sqrt{\sigma^2(X)}.$$

Defn: If $\phi : \Omega \to \Omega'$ is measurable map, and P is a measure on Ω, the **image measure** (or push forward measure) is defined by

$$P_\phi(E) = P(\phi^{-1}(E)).$$

Prop 10.1: With notation as above, if $f : \Omega' \to \mathbb{R}$ is a measurable function, then

$$\int_{\Omega'} f dP_\phi = \int_{\Omega} (f \circ \phi) dP$$

whenever either side is defined.

Defn: If X is a random variable on Ω then X is a measurable map from Ω to $\mathbb{R}$, so its image measure is a probability measure on $\mathbb{R}$, called its **distribution**.

Defn: Given X as above, the distribution function is

$$F(t) = P_X((-\infty, t]) = P)x \leq = t).$$

Defn: Two random variables with the same distribution function are called identically distributed.

Defn: For any finite sequence $\{X_1, ..., X_n\}$ of random variables on Ω consider this as a map $\Omega \mathbb{R}^n$. The image measure $P_{(X_1,...,X_n)}$ is called the joint distribution.

It is a general principle that all properties of random variables that are relevant to probability theory can be expressed in terms of their joint distributions.

$$E(X) = \int t dP_X(t)$$
$$\sigma^2(X) = \int (t - E(X))^2 dP_x(t)$$
$$E(X+Y) = \int (t+s) dP(X,Y)(t,s)$$

Given a Borel probability measure λ on $\mathbb{R}$ we define the mean and variance as

$$\overline{\lambda} = \int t d\lambda(t)$$
$$\sigma^2 = \int (t - \overline{\lambda})^2 d\lambda(t).$$

Defn: events E and F are independent if

$$P(E \cap F) = P(E)P(F).$$

If $P(E) > 0$, the conditional probability of F given E is

$$P_E(F) = P(E \cap F)/P(E).$$

Thus E and F are independent iff the probabilty of F is the same, whether or not E occurs.

Defn: a collection of events $\{E_\alpha\}$ is **independent** if

$$P_{(X_1,\dots,X_n)}(E_{\alpha_1} \cap \cdots \cap E_{\alpha_n}) = \prod_1^n P(E_{\alpha_j}),$$

for any distinct subset of the α's.

Defn: a collection of random variables $\{X_\alpha\}$ on Ω is **independent** if the events $X_\alpha \in B_\alpha$ are independent. This is equivalent to saying

$$P_{(X_1,\dots,X_n)} = \prod P_{X_i}.$$

Functions of independent variables are indepededent.

Prop. 10.2: Let $\{X_{nj} : 1 \leq j \leq J(n), 1 \leq n \leq N\}$ be independent random variables, and let $f_n : \mathbb{R}^{J(n)} \to \mathbb{R}$ be Borel measurable for $1 \leq n \leq N$. Then the random variables $Y_n = f_n(X_{n1}, .., X_{nJ(n)})$ $1 \leq n \leq N$, are independent.

An easy induction on (8.47) (definition of convolution of measures) shows that

$$\lambda_1 * \cdots * \lambda_n = \int \cdots \int \chi_E(t_1 + \cdots + t_n) d\lambda_1(t) d\lambda_n(t).$$

Prop. 10.4: If $\{X_j\}$ are independent random variables, then

$$P_{X_1 + \ldots X_n} = P_{X_1} * \cdots * P_{X_n}.$$

Prop. 10.5: Suppose that $\{X_j\}$ are independent random variables If $X_j \in L^1$ for all j, then

$$\prod X_j \in L^1, \qquad E(\prod X_j) = \prod E(X_j).$$

Cor. 10.6: If $\{X_j\}$ are independent random variables in L^2 then

$$\sigma^2(X_1 + \cdots + X_n) = \sum \sigma^(X_j).$$

In other words, independence implies orthogonal (but not conversely).

Proof. Let $Y_j = X_j - E(X_j)$. Then $\{Y_j\}$ are independent and have mean value zero so

$$E(Y_j Y_k) = E(Y_j)E(Y_k) = 0 \cdot 0 = 0,$$

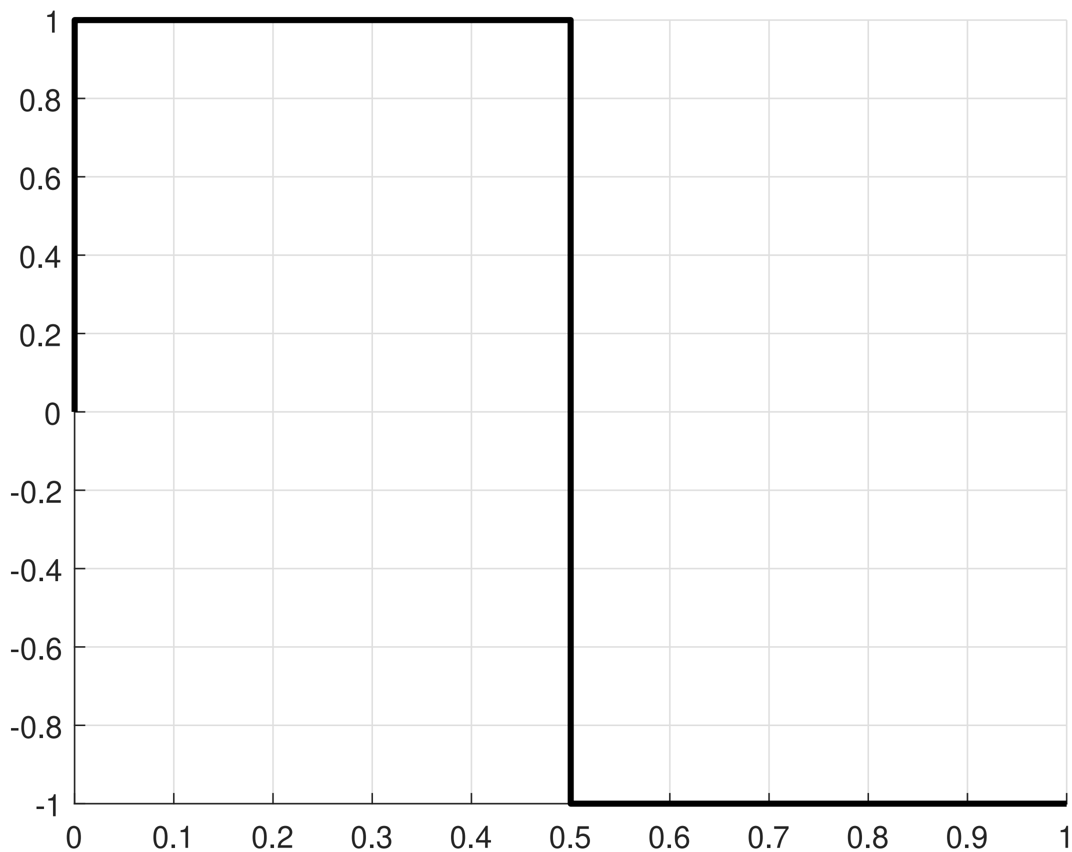
so they are orthogonal. The result follows from Pythagorean theorem (5.23). $\square$

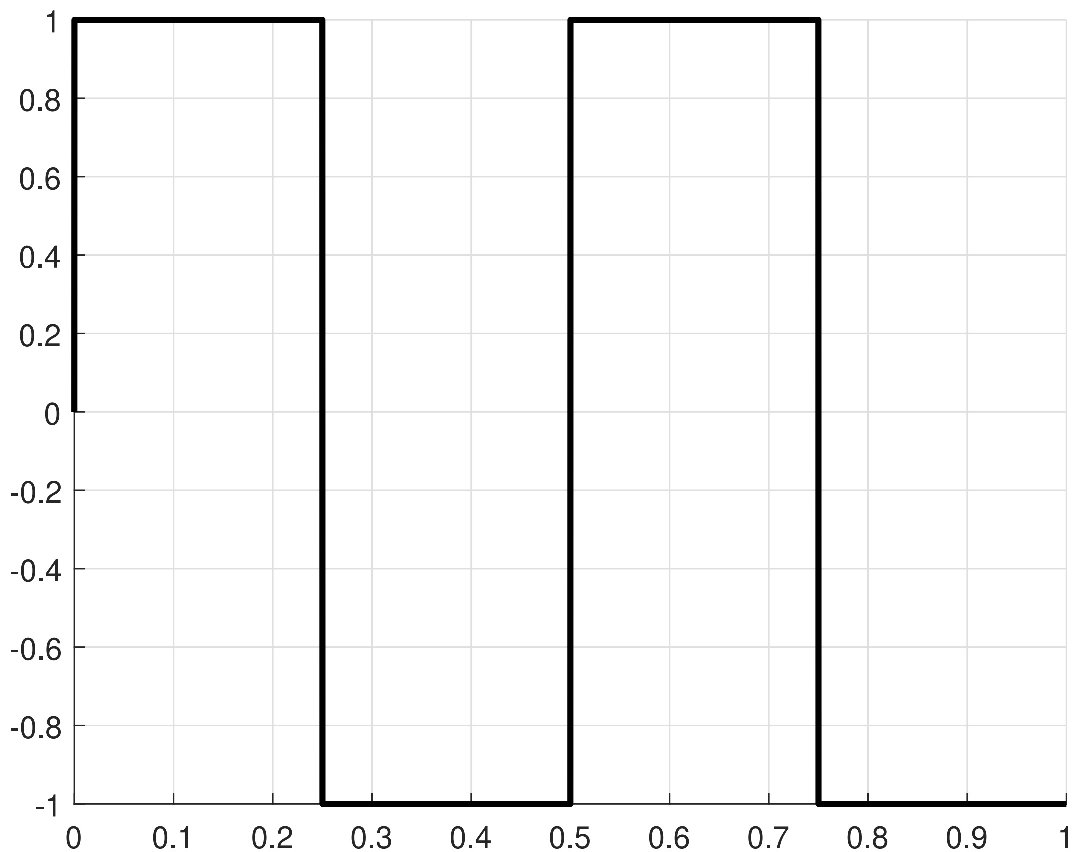
Independence is natually asociated to product spaces. Random variables are indenpendent if they depend on disjoint sets of coordinates.

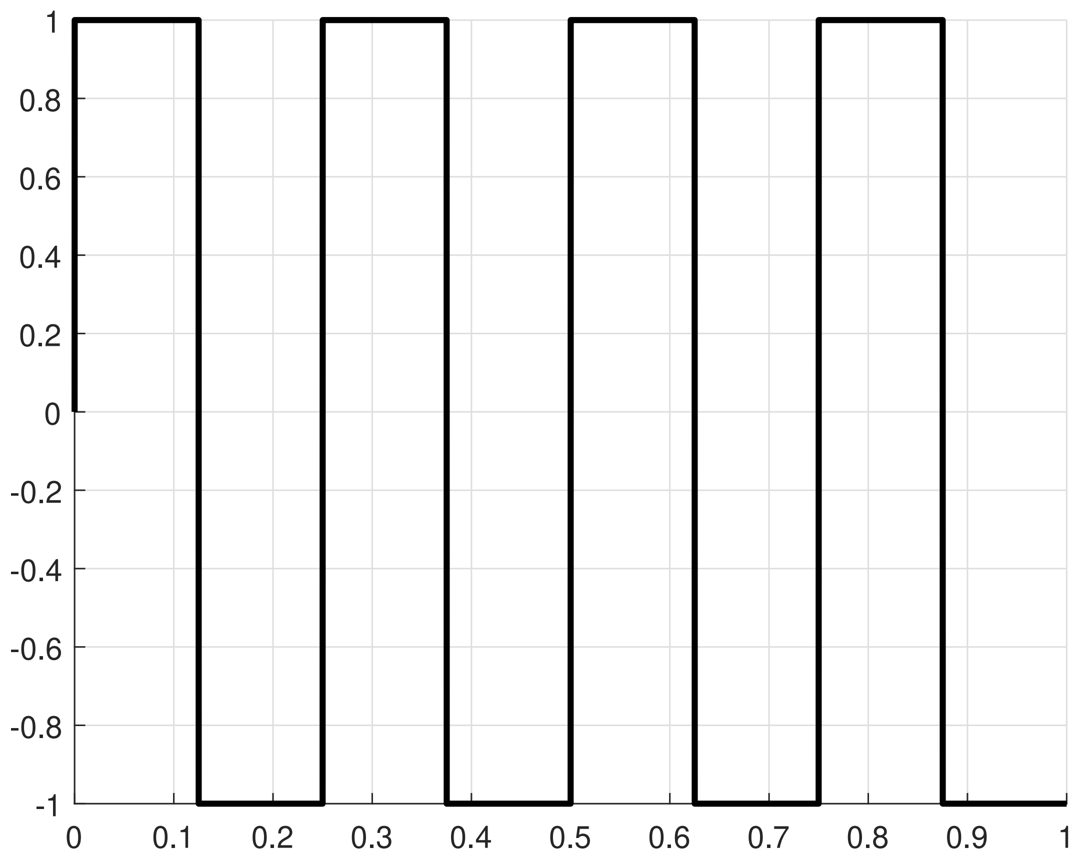
Example: Rademacher functions:

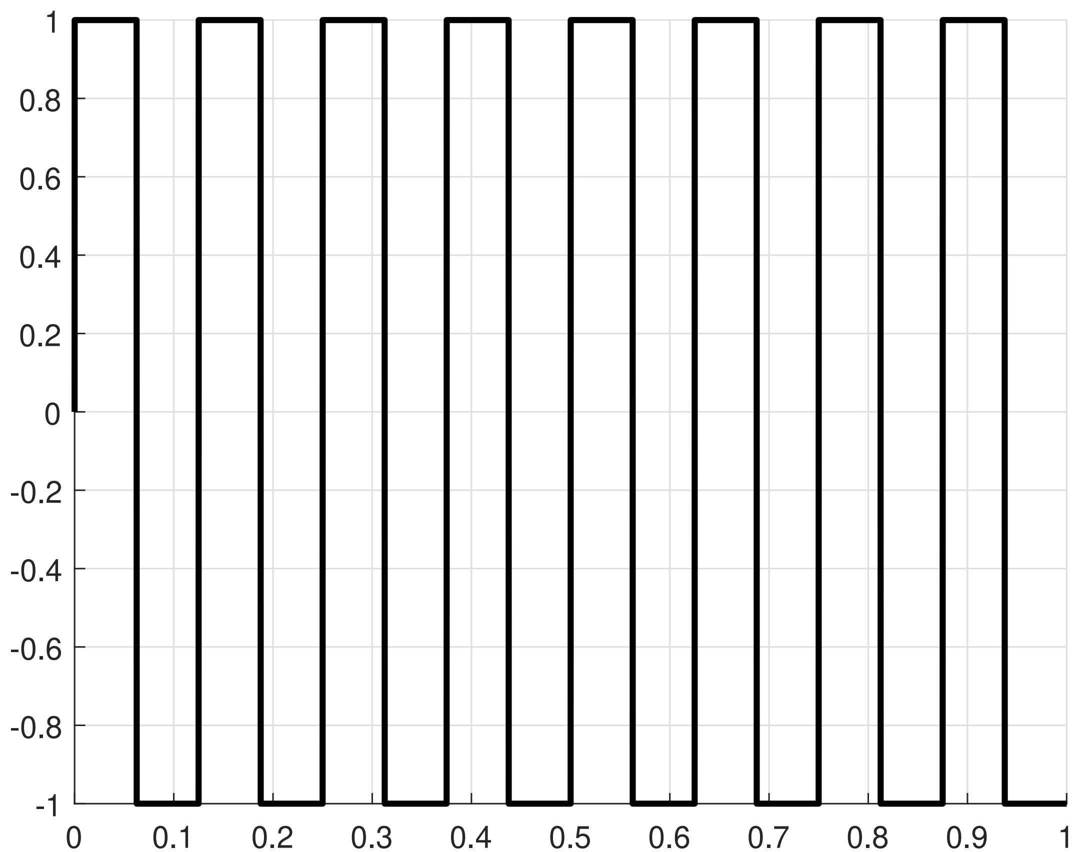
$$r_n(t) = \mathrm{sgn}(\sin(2^{\mathrm{n+1}}\pi \mathrm{t})).$$

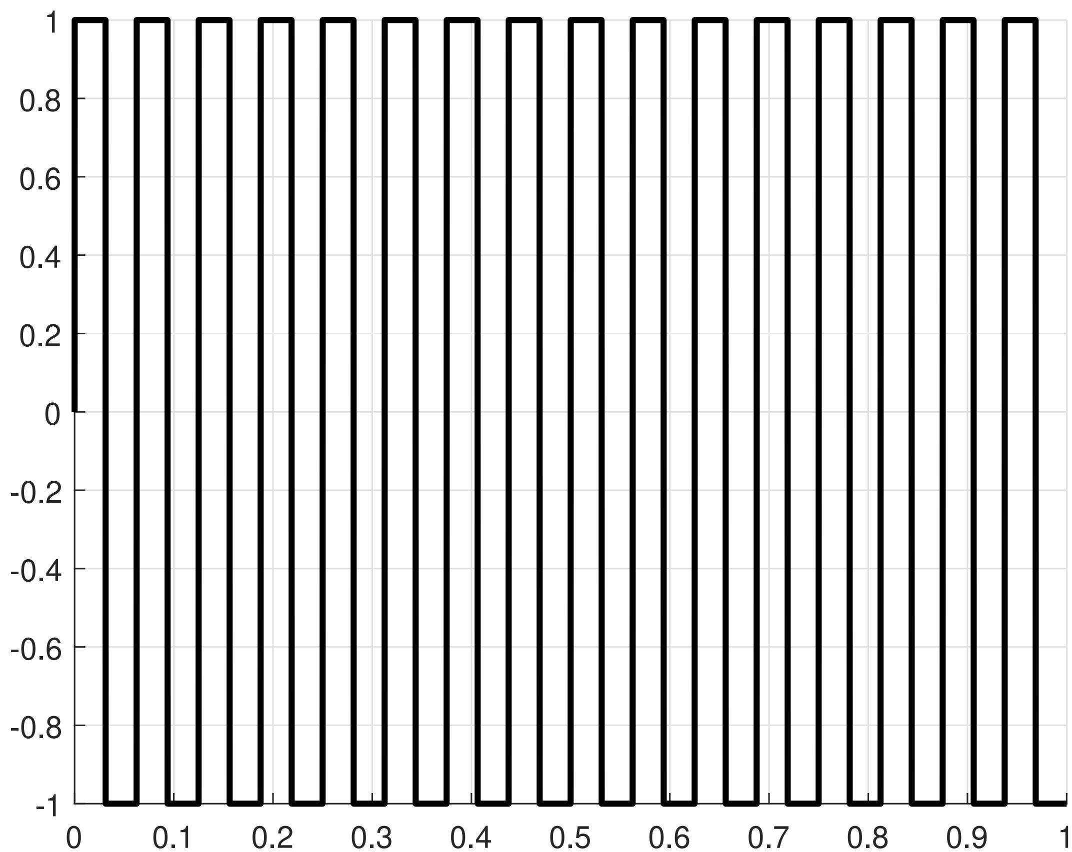
This is $+1$ if nth binary digit of x is 0 and -1 if nth binary digit is 1.


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

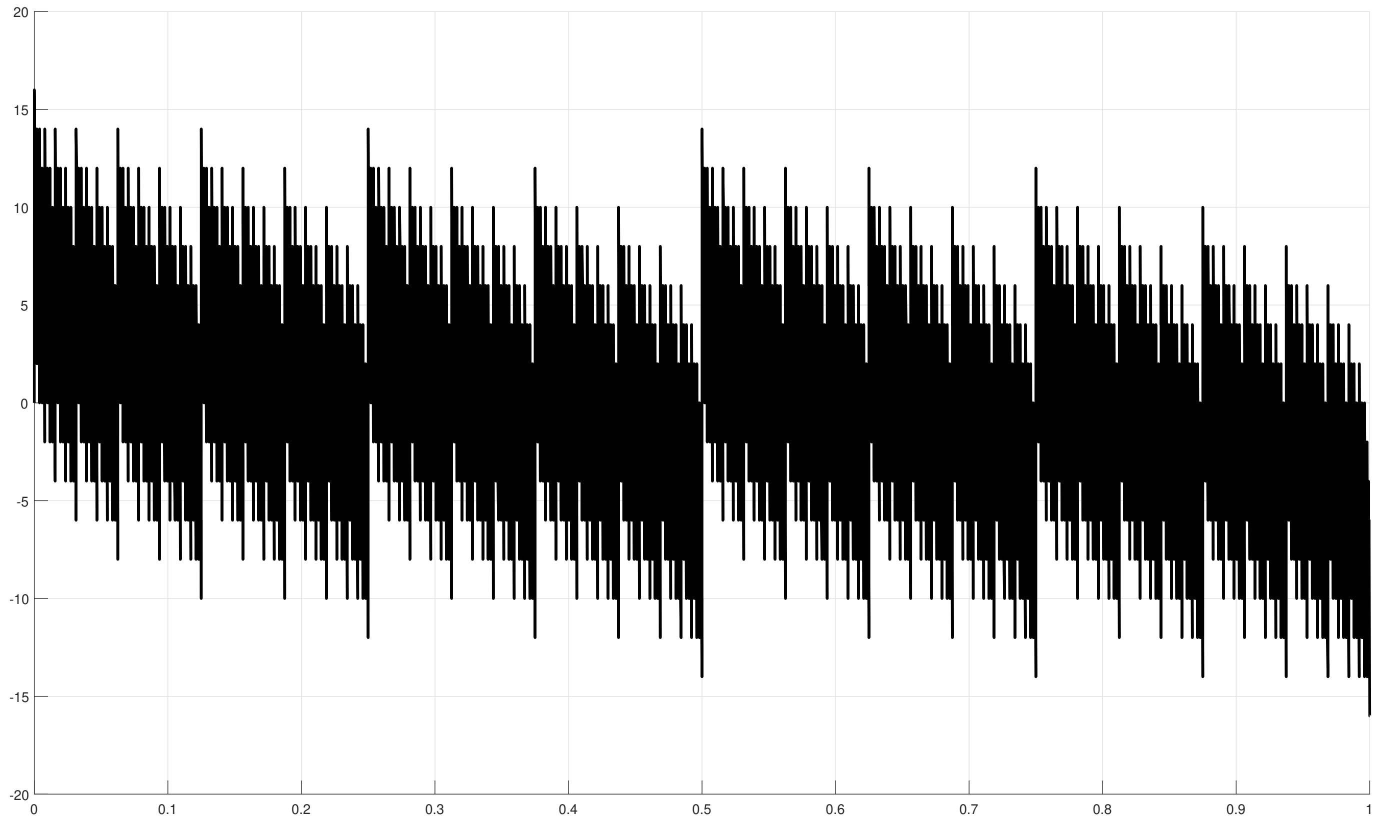

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

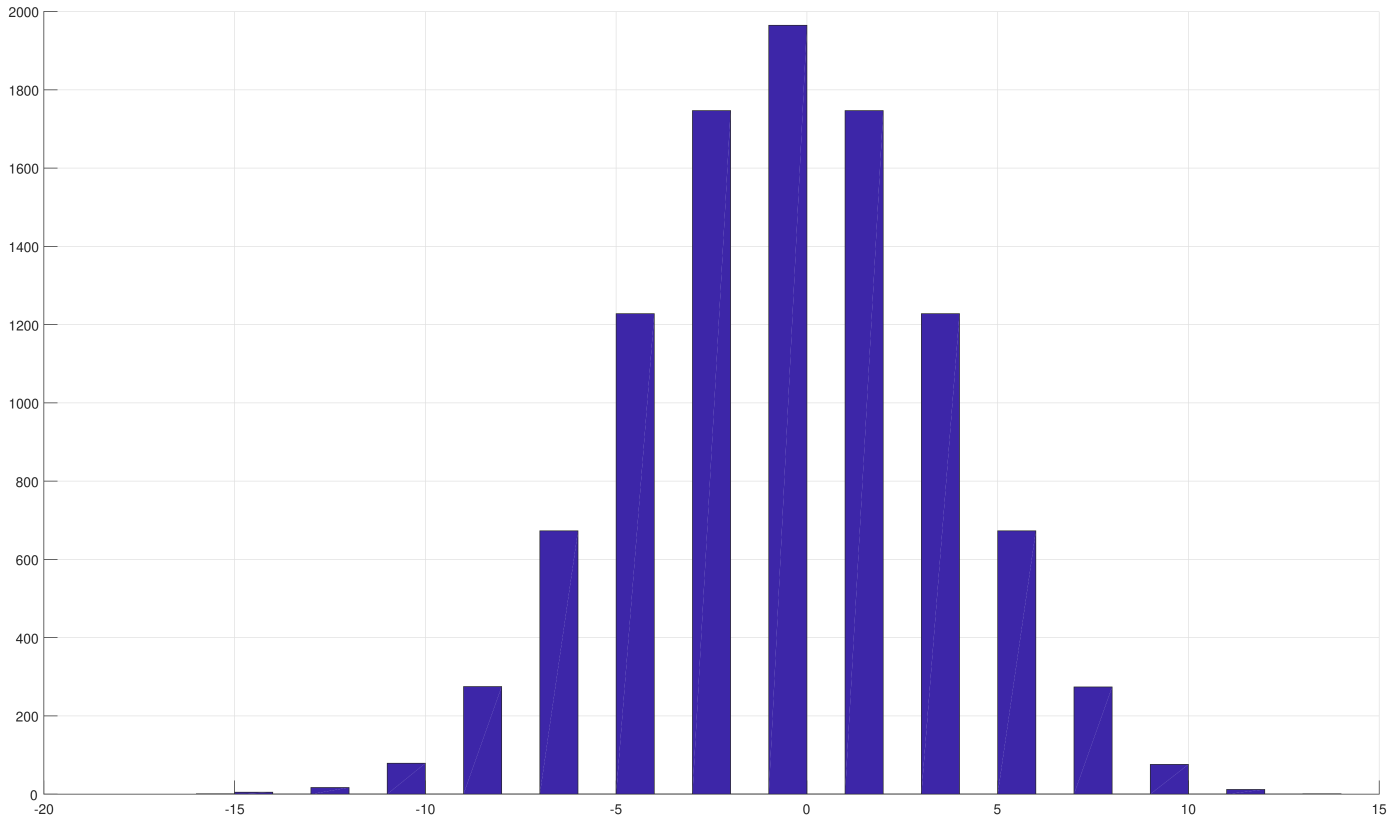

1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


20
15
10
5
0
-5
-10
-15
-20
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


2000
1800
1600
1400
1200
1000
800
600
400
200
0
-20
-15
-10
-5
0
5
10
15

Hans Rademacher (1892–1969)

Chapter 10.2: The Law of Large Numbers

If you generate a sequence of numbers independently at random, with probability p_i of choosing the i number then for almost ever sequence, the number i occurs with frequency p_i.

If we generate binary sequences by choosing 0 or 1 with equal probability $1/2$ then almost every sequence has an "equal" number of 0's and 1's.

Almost every real number has a binary expansion where 0's and 1's occur equally often.

The laws of large numbers make statements like this precise.

10.9 The Weak Law of Large Numbers: Let $\{X_j\}_1^\infty$ be a sequence of independent L^2 random variables with means $\{\mu_j\}$ and variances $\{\sigma_j^2\}$. If

$$\frac{1}{n^2}\sum_1^n \sigma_j^2 \to 0$$

as $n \to \infty$, then

$$S_n = \frac{1}{n}\sum_1^n (X_j - \mu_j) \to 0$$

in probability as $n \to \infty$.

Proof. S_n has mean 0 and variance

$$\frac{1}{n^2}\sum_1^n \sigma_j^2.$$

So by Chebyshev's inequality, for any $\epsilon > 0$

$$P(|S_n| > \epsilon) \le \frac{1}{(n\epsilon)^2}\sum_1^n \sigma_j^2 \to 0$$

as $n \to \infty$. $\square$

The limsup of a sequence of sets is the collection of points that are in infinitely many of the sets. Also

$$\limsup A_n = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n.$$

10.10 The Borel-Cantelli Lemma: Let $\{A_n\}$ be a sequence of events.

(a) If $\sum_1^\infty P(A_n) < \infty$, then $P(\limsup A_n) = O$.

(b) If A_n's are independent and $\sum_1^\infty P(A_n) = \infty$, then $P(\limsup A_n) = 1$.

Proof. We have

$$P(\limsup A_n) \leq \inf_k P(\cup_{n=k}^{\infty} A_n) \leq \inf_k \sum_{n \geq k} P(A_n).$$

The latter sum tends to zero if the $P(A_n)$ are summable. This proves (a).

To prove (b), suppose that $\sum P(A_n)$ diverges and the A_n's are independent. We must show

$$P((\limsup A_n)^c) = P\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n^c\right) = 0.$$

It is enough to show

$$P(\cap_{n=k}^{\infty} A_n^c) = 0$$

for all k. Using independence and $1 - t \leq e^t$ we get

$$P(\cap_{n=k}^{K} A_n^c) = \prod_{k}^{K}(1 - P(A_n)) \leq \prod_{k}^{K} \exp(-P(A_n)) = \exp(-\sum_{k}^{K} P(A_n))$$

which tends to zero as $K \to \infty$. $\square$

Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics

by

DENNIS SULLIVAN

I.H.E.S., Bures-sur-Yvette, France

Contents

Disjoint spheres, approximation by imaginary quadratic numbers, and the logarithm law for geodesics by Dennis Sullivan.

§1. Abstract Borel-Cantelli

If $A_1, A_2, \ldots$ is a sequence of subsets of a probability space X and $A_\infty = \{x : x \in A_i$ for infinitely many $i\}$ we want to compare the conditions

(i) $\Sigma_i |A_i| = \infty$, $|A_i| =$ measure A_i

and

(ii) $|A_\infty| > 0$.

The first proposition is very standard. We recall the proof to establish notation.

PROPOSITION 1. *If* $|A_\infty| > 0$, *then* $\Sigma_i |A_i| = \infty$.

Proof. Let $\varphi_N(x) =$ sum of the characteristic functions of A_i for $i \leq N$. Then by definition $A_\infty = \{x\colon \lim_{N\to\infty} \varphi_N(x) = \infty\}$. Since the φ_N are monotone increasing

$$\lim \int \varphi_N = \int \lim \varphi_N$$

by the Lebesgue monotone convergence theorem. One side is $\Sigma_i |A_i|$ and the other is infinity if $|A_\infty| > 0$.

PROPOSITION 2 (quasi-independent Borel-Cantelli). *Suppose* $\Sigma_i |A_i| = \infty$ *and there is a constant* $c \geq 1$ *so that for* $i<j$, $|A_i \cap A_j| \leq c|A_i| \cdot |A_j|$. *Then* $|A_\infty|>0$, *in fact there is a set of positive measure* $\bar{A}$ *so that for* $x \in \bar{A}$,

$$\limsup_N \frac{\operatorname{card}\{i: x \in A_i, i \leq N\}}{|A_1|+\ldots+|A_N|} > 0.$$

Proof. Consider φ_N as in Proposition 1. Let $|\varphi_N|_2, |\varphi_N|_1$ denote $(\int \varphi_N^2)^{1/2}$ and $\int \varphi_N$ respectively. By Schwarz $|\varphi_N|_1 \leq |\varphi_N|_2$. Conversely, using our hypothesis,

$$\begin{aligned}
\int \varphi_N^2 &= \sum_{i \leq j \leq N} |A_i \cap A_j| \\
&= \sum_{i \leq N} |A_i| + \sum_{i<j \leq N} |A_i \cap A_j| \\
&\leq \sum_{i \leq N} |A_i| + c \sum_{i<j \leq N} |A_i||A_j| \quad \text{(by quasi-independence)} \\
&\leq c \sum_{i \leq j \leq N} |A_i||A_j| \\
&= c\left(\int \varphi_N\right)^2.
\end{aligned}$$

Thus $|\varphi_N|_2 \leq \sqrt{c}\, |\varphi_N|_1$.

Now consider $\psi_N(x)=\varphi_N(x)/|\varphi_N|_1$ and choose a weak limit ψ in the ball of radius $\sqrt{c}$ of square integrable funcitons. Since $(\psi_N, 1)\to(\psi, 1)$ we have $|\psi|_1=1$.

Similarly, ψ is non-negative, so ψ is positive on a set of positive measure. If $A=$support ψ, then for all $x\in A$, $\lim_N \varphi_N(x)=\infty$, because $\lim_N |\varphi_N|_1\to\infty$. Thus $A_\infty\supset A$ has positive measure.

Now we show there is a set $\bar{A}$ of positive measure in A_∞ so that if $x\in\bar{A}$

$$\limsup_N \frac{\varphi_N(x)}{|\varphi_N|_1}>0.$$

If for subset of positive measure in A the lim sup is zero, then for a further subset of positive measure the ratios are $\leqslant 1$ for N sufficiently large. By dominated convergence $\int\psi=0$ on this subset. This is a contradiction. Thus $\bar{A}$ may be taken to have full measure in A.

10.11 Kolmogorov's Inequality: Let $X_1, ..., X_n$ be independent random variables with mean 0 and variances $\sigma_1^2, \ldots, \sigma_n^2$, and let $S_k k = X_1 + ... + X_k$. For any $\epsilon > 0$,

$$P(\max_{1\leq k\leq n} |S_k \geq \epsilon) \leq \epsilon^{-2} \sum_1^n \sigma_k^2.$$

Proof. Let A_k be the set where $|S_j| < \epsilon$ for $j < k$ and $|S_k| \geq \epsilon$ (first time $|S_k|$ exceeds ϵ). These sets are disjoint and the union is $\{\max j \leq k|S_k| \geq \epsilon$. Hence

$$P(\max |S_k| \geq \epsilon) = \sum_1^n P(A_k) \leq \epsilon^{-2} \sum_1^n E(\chi_{A_k} S_k^2)$$

since $S_k^2 \geq \epsilon^2$ on A_k.

On the other hand,

$$\begin{aligned}
E(S_n^2) &\geq \sum_1^n E(\chi_{A_k} S_n^2) \\
&= \sum_1^n E(\chi_{A_k}[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2]) \\
&\geq \sum_1^n E(\chi_{A_k} S_k^2) + 2\sum_1^n E(\chi_{A_k} S_k(S_n - S_k) + (S_n - S_k)^2].
\end{aligned}$$

We claim each term in the second sum is zero.

If the claim holds, then

$$E(S_n^2) \;\geq\; \sum_1^n E(\chi_{A_k} S_k^2)$$

so

$$P(\max|S_n| \geq \epsilon) \leq \epsilon^{-2} E(S_n^2) = \epsilon^{-2} \sum_1^n \sigma_k^2$$

by Cor 10.6since the X_k's have mean value zero.

But χ_{A_k} only depends on the values of $X_1, \ldots, X_k$, whereas $S_n - S_k$ only depends on the value of $X_{k+1}, \ldots, X_n$. Since there are disjoint sets of independent variables, these two quantities are also independent. Thus

$$E(\chi_{A_k} S_k(S_n - S_k)) = E(\chi_{A_k} S_k) \cdot E(S_n - S_k) = E(\chi_{A_k} S_k) \cdot 0 = 0. \quad \square$$

10.12 Kolmogorov's Strong Law of Large numbers If $\{X_n\}_1^\infty$ are independent L^2 random variables with means $\{\mu_n\}$ and variances $\{\sigma_n^2\}$ so that

$$\sum_1^\infty n^{-2}\sigma_n^2 < \infty,$$

then

$$S_n \frac{1}{n}\sum_1^n (X_j - \mu_j) \to 0,$$

almost surely as $n \to \infty$.

Proof. Given $\epsilon > 0$, for $k \in \mathbb{N}$ let A_k be the set where $|S_n|/n \geq \epsilon$ for some $n \in [2^{k-1}, 2^k]$. Then on A_k we have $|S_n| \geq \epsilon 2^{k-1}$ for some $n < 2^k$, so by Kolmogorov's inequality,

$$P(A_k) \leq (\epsilon 2^{k-1})^{-2} \sum_1^{2^k} \sigma_n^2.$$

Therefore

$$\begin{aligned} P(A_k) &\le \frac{4}{\epsilon^2} \sum k = 1^{\infty} \sum_{m}^{2^k} 2^{-2k} \sigma_n^2 \\ &\le \frac{4}{\epsilon^2} \sum n = 1^{\infty} \sum_{k \ge \log_2 b} 2^{-2k} \sigma_n^2 \\ &\le \frac{8}{\epsilon^2} \sum n = 1^{\infty} n^{-2} \sigma_n^2 \end{aligned}$$

Thus $\sum P(A_k) < \infty$, so by Borel-Cantelli $P(\limsup A_k) = 0$ (almost every pooint is only in finitely many A_k). But $\limsup A_k$ is the set where $|S_n| \ge \epsilon n$ for infinitely many n, so

$$P(\limsup \frac{1}{n}|S_n| < \epsilon) = 1,$$

for any $\epsilon > 0$. Taking $\epsilon \to 0$ through a countable sequence proves the result $\square$

10.13 Khinchine's Strong Law of Large Numbers: If $\{X_n\}$ is a sequence of independent identically distributed L^1 random variables with mean μ, then

$$S_n = \frac{1}{n}\sum_{1}^{n} X_j \to \mu$$

almost surely as $n \to \infty$.

Proof. Replacing X_n by $X_n - \mu$, we may assume that X_n has mean zero. Let λ be the common distribution of the X_j's; we are thus assuming that

$$\int |t| d\lambda(t) < \infty, \qquad \int t d\lambda(t) = 0.$$

Let $Y_j = X_j$ on the set where $|X_j| \leq j$ and zero elsewhere. Then

$$\begin{aligned} \sum_1^\infty P(Y_j \neq X_j) &= P(|X_j| > j) \\ &= \sum_1^\infty \lambda(t : |t| > j) \\ &= \sum_{j=1}^\infty \sum_{k=j}^\infty \lambda(t : k < |t| \leq k+1). \end{aligned}$$

Since

$$\sum_{j=1}^\infty \sum_{k=j}^\infty = \sum_{k=1}^\infty \sum_{j=1}^k,$$

interchanging the order of summation gives

$$
\begin{aligned}
\sum_{1}^{\infty} P(Y_j \neq X_j) &= \sum_{k=1}^{\infty} k\lambda(t : k < |t| \leq k+1) \\
&\leq \sum_{k=1}^{\infty} |t| d\lambda(t) \\
&< \infty.
\end{aligned}
$$

By the Borel-Cantelli lemma, we almost surely have $X_j = Y_j$ for j sufficiently large, so it suffices to show almost surely that

$$
\frac{1}{n} \sum_{1}^{n} Y_j \to 0.
$$

We have

$$\sigma_n^2(Y_n) \leq Y(Y_n^2) = \int_{|t|\leq n} t^2 d\lambda(t),$$

and hence

$$\begin{aligned}
\sum_1^\infty n^{-2}\sigma^2(Y_n) &\leq \sum_{n=1}^\infty \sum_{j=1}^n n^{-2} \int_{j-1<|t|\leq j} t^2 d\lambda(t) \\
&\leq \sum_{n=1}^\infty \sum_{j=1}^n j \cdot n^{-2} \int_{j-1<|t|\leq j} |t| d\lambda(t).
\end{aligned}$$

Reversing the order of summation, and using

$$\sum_{n=j}^\infty n^{-2} \leq 2/j,$$

we get

$$\sum_{n=1}^\infty n^{-2}\sigma^2(Y_n) \leq 2\sum_{n=1}^\infty \int_{j-1<|t|\leq j} |t| d\lambda(t) = 2\int_{-\infty}^\infty |t| d\lambda(t) < \infty.$$

By Kolmogorov's SLLN (10.12) if $\mu_j = E(Y_j)$ we have $n^{-1}\sum_1^n (Y_j - \mu_j) \to 0$ almost surely. By the dominated convergence theorem

$$\mu_j = \int_{|t|\le j} t d\lambda(t) \to \int_{-\infty}^{\infty} t d\lambda(t) = 0.$$

It follows (Exercise 12) that $n^{-1}\sum_1^n \mu_j \to 0$ also. Hence

$$\frac{1}{n}\sum_1^n Y_j \to 0$$

almost surely, as desired. $\square$

Here is another version of the SLLN that does not require independence, only orthogonality. This is very usely in many cases. However, we have to assume the variances are bounded, which is more than was assumed in Kolmogorov's SLLN.

This is Theorem 1.5.2 from Fractals in Probabilty and Analysis by Bishop and Peres.

Strong Law of Large Numbers: Let $(X, d\nu)$ be a probability space and $\{f_n\}$, $n = 1, 2 \ldots$ a sequence of orthogonal functions in $L^2(X, d\nu)$. Suppose $E(f_n^2) = \int |f_n|^2 \, d\nu \leq 1$, for all n. Then

$$\frac{1}{n} S_n = \frac{1}{n} \sum_{k=1}^{n} f_k \to 0,$$

a.e. (with respect to ν) as $n \to \infty$.

Proof. We begin with the simple observation that if $\{g_n\}$ is a sequence of functions on a probability space $(X, d\nu)$ such that

$$\sum_n \int |g_n|^2 \, d\nu < \infty,$$

then $\sum_n |g_n|^2 < \infty$ ν-a.e. and hence $g_n \to 0$ ν-a.e.

Using this, it is easy to verify the Strong Law of Large Numbers (LLN) for $n \to \infty$ along the sequence of squares. Specifically, since the functions $\{f_n\}$ are orthogonal,

$$\int \left(\frac{1}{n} S_n\right)^2 \, d\nu = \frac{1}{n^2} \int |S_n|^2 d\nu = \frac{1}{n^2} \sum_{k=1}^{n} \int |f_k|^2 \, d\nu \leq \frac{1}{n}.$$

Thus if we set $g_n = \frac{1}{n^2} S_{n^2}$, we have

$$\int g_n^2 \, d\nu \leq \frac{1}{n^2}.$$

Since the right hand side is summable, the observation made above implies that $g_n = n^{-2} S_{n^2} \to 0$ ν-a.e.

To handle the limit over all positive integers, suppose that $m^2 \leq n < (m+1)^2$. Then

$$\begin{aligned}
\int |\frac{1}{m^2} S_n - \frac{1}{m^2} S_{m^2}|^2 \, d\nu &= \frac{1}{m^4} \int |\sum_{k=m^2+1}^{n} f_k|^2 \, d\nu \\
&= \frac{1}{m^4} \int \sum_{k=m^2+1}^{n} |f_k|^2 \, d\nu \\
&\leq \frac{2}{m^3},
\end{aligned}$$

since the sum has at most $2m$ terms, each of size at most 1. Set $m(n) = \lfloor \sqrt{n} \rfloor$ and

$$h_n = \frac{S_n}{m(n)^2} - \frac{S_{m(n)^2}}{m(n)^2}.$$

Now each integer m equals $m(n)$ for at most $2m+1$ different choices of n. Therefore,

$$\sum_{n=1}^{\infty} \int |h_n|^2 \, d\mu \leq \sum_{n=1}^{\infty} \frac{2}{m(n)^3} \leq \sum_{m} (2m+1) \frac{2}{m^3} < \infty,$$

so by the initial observation, $h_n \to 0$ a.e. with respect to ν. This yields that

$$\frac{1}{m(n)^2} S_n \to 0 \text{ a.e.}$$

which, in turn, implies that $\frac{1}{n} S_n \to 0$ a.e., as claimed. $\square$

In the strong law of large numbers better estimates for the decay of S_n are possible if we assume that the functions $\{f_n\}$ are independent with respect to the measure ν.

By 1915 Hausdorff had proved that if $\{f_n\}$ are independent and satisfy $\int f_n d\nu = 0$ and $\int f_n^2\, d\nu = 1$, then

$$\lim_{N\to\infty} \frac{1}{N^{\frac{1}{2}+\epsilon}} \sum_{n=0}^{N} f_n(x) = 0 \text{ for a.e. } x$$

and for every $\epsilon > 0$. After that Hardy–Littlewood, and independently Khinchin, proved

$$\lim_{N\to\infty} \frac{1}{\sqrt{N \log N}} \sum_{n=0}^{N} f_n(x) = 0 \text{ for a.e. } x.$$

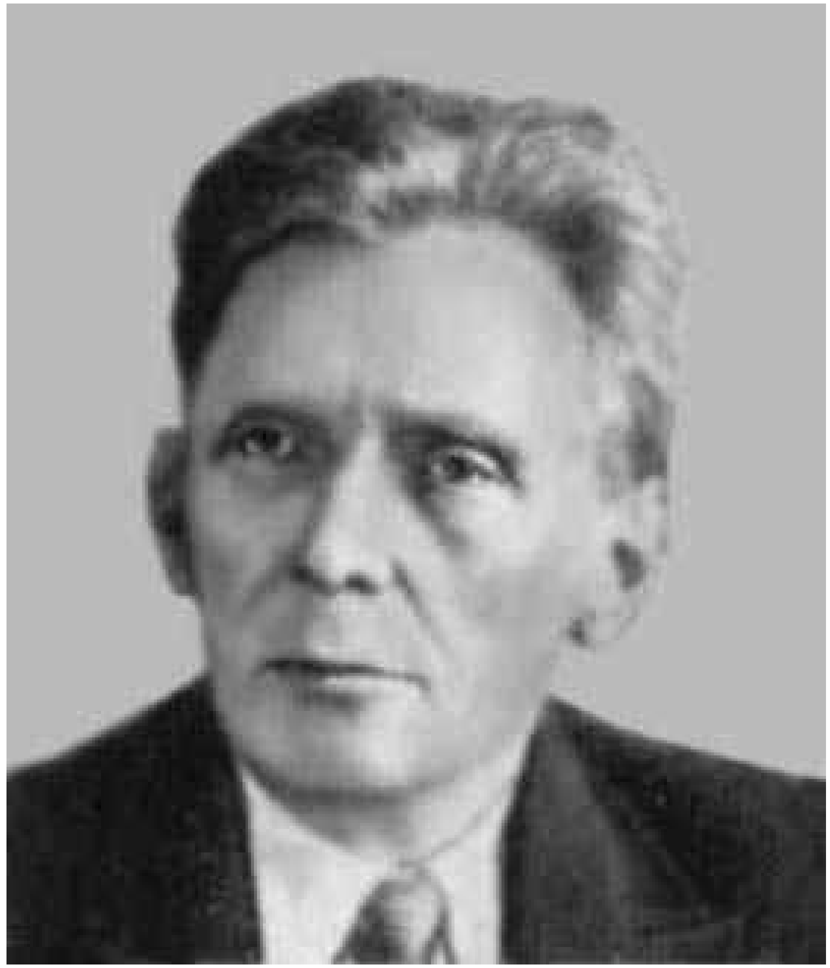
The "final" result, found by Khinchin for a special case in 1928 and proved in general by Hartman–Wintner says

$$\limsup_{N\to\infty} \frac{1}{\sqrt{2N\log\log N}} \sum_{n=0}^{N} f_n(x) = 1 \text{ for a.e. } x.$$

If we only assume orthogonality, bounded variances cannot be improved to the weaker condition in Kolmorgorov's SLLN. See

ON THE STRONG LAW OF LARGE NUMBERS FOR PAIRWISE INDEPENDENT RANDOM VARIABLES by S. Csorgo, K. Tandori, and V. Totik

Andrey Nikolaevich Kolmogorov (1903–1987



Aleksandr Yakovlevich Khinchin (1894–1959)

Chapter 10.3: The Central Limit Theorem

Defn: The probability measure $\nu_\mu^{\sigma^2}$ on $\mathbb{R}$ defined by

$$d\nu_\mu^{\sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2} dt$$

is called the **normal distribution** or **Gaussian distribution** with mean μ and variance σ^2. When $\mu = 0$ and $\sigma = 1$ this is called the **standard normal distribution**.

Theorem 10.14: Let λ be a Borel probability measure on $\mathbb{R}$ such that

$$\int t^2 d\lambda(t) = 1, \qquad \int t d\lambda(t) = 0.$$

(The finiteness of the first integral implies the existence of the second.) For $n \in \mathbb{N}$ let $\lambda^{*n} = \lambda * \cdots * \lambda$ (n factors) and define the measure λ_n by

$$\lambda_n(E) = \lambda^{*n}(\sqrt{n}E).$$

where $\sqrt{n}aE = \{\sqrt{n}x : x \in E\}$. Then $\lambda_n \to \nu_0^1$ vaguely (in weak* topology).

Proof. The hypotheses on the measure imply that its Fourier transform is of class C^2 and satisfies

$$\hat{\lambda}(0) = 1, \qquad (\hat{\lambda})'(0) = 0, \qquad (\hat{\lambda})''(0) = -4\pi^2.$$

(differentiate the integral twice in Thm 8.22d). Thus

$$\hat{\lambda}(\xi) = 1 - 2\pi^2\xi^2 + o(\xi^2).$$

Moreover,

$$(\lambda^{*n})^\wedge = (\hat{\lambda})^n$$

so

$$\hat{\lambda}_n(\xi) = [\hat{\lambda}(\xi/\sqrt{n})]^n = [1 - \frac{2\pi^2\xi^2}{n} + o\left((\frac{\xi^2}{n})^2\right)]^n.$$

Since $\log 1 + x = x + o(x)$ near zero,

$$\log \hat{\lambda}_n(\xi) = n\log[[1 - \frac{2\pi^2\xi^2}{n} + o\left((\frac{\xi^2}{n})^2\right)]^2 = -2\pi^2\xi^2 + no(\xi^2/n)$$

which tends to $-2\pi^2\xi^2$ as $n \to \infty$. In other words, $\hat{\lambda}_n(\xi) \to \exp(-2\pi^2\xi^2)$ so the conclusion follows from Propositions 8.24 (the Fourier transform of the Gauss kernel) and 8.50 (pointwise convergence of the Fourier transform of a measure implies vague convergence). $\square$

10.15 The Central Limit Theorem: Let $\{X_j\}$ be a sequence of independent identically distributed L^2 random variables with mean μ and variance σ^2. As $n \to \infty$ the distribution of $(\sigma\sqrt{n})^{-1}\sum_1^n(X_j - \mu)$ converges vaguely to the standard normnal ν_0^1, and for all $a \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(\frac{1}{\sigma\sqrt{n}}\sum_1^n(x_j - \mu) \leq a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^a e^{-t^2/2}dt.$$

Proof. Replacing X_j by $\sigma^{-1}(X_j -)$, we may assume that $\mu = 0$ and $\sigma = 1$. If λ is the common distribution of the X_j's, then λ satisfies the hypotheses of Theorem 10.14, and in the notation used there, λ_n is the distribution of $n^{-1/2}\sum_1^n X_j$. The first assertion thus follows immediately, and the seeond one is equlvalent to it by Proposition 7.19 (characterization of vague convergence in terms of pointwise convergence of distribution functions). $\square$