Law of Large Numbers *Let $\{f_n\}$, $n = 1, 2\ldots$ be a sequence of orthogonal functions on a probability space $(X, d\nu)$ and suppose $E(f^2) = \int |f|^2 d\nu \leq 1$. 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 $\sum_n |g_n|^2 < \infty$ a.e. $(d\nu)$ and hence $g_n \to 0$ a.e. $(d\nu)$.

Using this, it is easy to verify the law of large numbers (LLN) for $n \to \infty$ along the sequence of squares. Namely,

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

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

$$\int \frac{1}{n^2}|S_{n^2}|^2 d\nu \leq \frac{1}{n^2}.$$

Since the right hand side is summable, the observation above implies $g_n \to 0$ a.e. $(d\nu)$. This is the same as $\frac{1}{n^2}S_{n^2} \to 0$, a.e..

To deal with limit over all the integers take $m^2 \leq n < (m+1)^2$ and set $m(n) = \lfloor\sqrt{n}\rfloor$. 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. Put

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

Then since each $m = m(n)$ is associated to at most $2m + 1$ different n's we get

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

so by the initial observation, $g_n \to 0$ a.e. with respect to ν. This implies $\frac{1}{m(n)^2}S_n \to 0$ a.e., which in turn implies $\frac{1}{n}S_n \to 0$ a.e., which is what we wanted. $\square$

This version is sometimes called the strong law of large numbers because it gives a.e. convergence, as opposed to the weak version which only says that $\frac{1}{n}S_n$ converges to 0 in L^2.

As a remark we should note that 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. This means that for any n and any collection of measurable sets $\{A_1, \ldots, A_n\}$ we have

$$\nu(x \in X : f_j(x) \in A_j, j = 1, \ldots, n\}) = \prod_{j=1}^{n} \nu(\{x \in X : f_j(x) \in A_j\}).$$

Roughly, this says that knowing the values of any of the f_j's at x does not give us any information about the values of the remaining functions there.

By 1915 Hausdorff had proved that if $\{f_n\} \in L^2(\nu) \cap L^1(\nu)$ are independent, orthonormal (orthogonal and have L^2 norm 1) and satisfy $\int f_n d\nu = 0$ 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-Wintree in 1941 says

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