Chapter 3

Independent Sums

Independence and Convolution

One of the central ideas in probabilty is the notion of independence. In intuitive terms two events are independent if they have no influence on each other. The formal definition is

Definition 3.1. Two events A and B are independent if $P[A \cap B] = P[A].P[B]$

Exercise 3.1. If A and B are independent prove that so are A^c and B.

Definition 3.2. Two random variables X and Y are independent if the events $X \in A$ and $Y \in B$ are independent for any two Borel sets A and B on the line i.e.

$$P[X \in A, Y \in B] = P[X \in A].P[Y \in B]$$

for all Borel sets A and B.

There is a natural extension to a finite or even an infinite collection of random variables.

Definition 3.3. A finite collection collection $\{X_j : 1 \leq j \leq n\}$ of random variables are said to be independent if for any n Borel sets $A_1, \ldots, A_n$ on the line

$$P\Big[\cap_{1 \leq j \leq n} [X_j \in A_j]\Big] = \Pi_{1 \leq j \leq n} P[X_j \in A_j]$$

Definition 3.4. An infinite collection of random variables is said to independent if every finite subcollection is independent.

Lemma 3.1. *Two random variables X, Y defined on (Ω, Σ, P) are independent if and only if the measure induced on $\mathbf{R}^2$ by (X, Y), is the product measure $\alpha \times \beta$ where α and β are the distributions on $\mathbf{R}$ induced by X and Y respectively.*

Proof: Left as an exercise.

The important thing to note is that if X and Y are independent and one knows their distributions α and β, then their joint distribution is automatically determined as the product measure.

If X and Y are random variables having α and β for their distributions, the distribution of the sum $Z = X + Y$ is determined as follows. First we construct the product measure $\alpha \times \beta$ on $\mathbf{R} \times \mathbf{R}$ and then consider the induced distribution of the function $f(x, y) = x + y$. This distribution, called the convolution of α and β, is denoted by $\alpha * \beta$. An elementary calculation using Fubini's theorem provides the following identities.

$$(\alpha * \beta)(A) = \int \alpha(A - x)\, d\beta = \int \beta(A - x)\, d\alpha \tag{3.1}$$

In terms of characteristic function, we can express the characteristic function of the convolution as

$$\begin{aligned} \int \exp[\,i\,t\,x\,]d(\alpha * \beta) &= \int\int \exp[\,i\,t\,(x + y)\,]\,d\,\alpha\,d\,\beta \\ &= \int \exp[\,i\,t\,x\,]\,d\,\alpha \int \exp[\,i\,t\,x\,]\;d\,\beta \end{aligned}$$

or equivalently

$$\phi_{\alpha*\beta}(t) = \phi_\alpha(t)\phi_\beta(t) \tag{3.2}$$

which provides a direct way of calculating the distributions of sums of indepenedent random variables by the use of characteristic functions.

Exercise 3.2. If X and Y are independent show that for any two measurable functions f and g, $f(x)$ and $g(Y)$ are independent.

Exercise 3.3. Use Fubini's theorem to show that if X and Y are independent and if f and g are measurable functions with both $E[|f(X)|]$ and $E[|f(Y)|]$ finite then

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$$

Exercise 3.4. If X and Y are any two random variables then $E(X+Y) = E(X) + E(Y)$, and if X and Y are independent then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

where

$$\mathrm{Var}(X) = E[[X - E[X]]^2] = E[X^2] - [E[X]]^2$$

If $X_1, X_2, \cdots, X_n$ are n independent random variables, then the distribution of their sum $S_n = X_1 + X_2 + \cdots + X_n$ can be computed in terms of the distributions of the summands. If α_j is the distribution of X_j, then the distribution of μ_n of S_n is given by the convolution $\mu_n = \alpha_1 * \alpha_2 * \cdots * \alpha_n$ that can be calculated inductively by $\mu_{j+1} = \mu_j * \alpha_{j+1}$. In terms of their characteristic functions $\psi_n(t) = \phi_1(t)\phi_2(t)\cdots\phi_n(t)$. The first two moments of S_n are computed easily.

$$E(S_n) = E(X_1) + E(X_2) + \cdots E(X_n)$$

and

$$\begin{aligned} \mathrm{Var}(S_n) &= E[S_n - E(S_n)]^2 \\ &= \sum_j E[X_j - E(X_j)]^2 \\ &\quad +2 \sum_{1 \le i < j \le n} E[X_i - E(X_i)][X_j - E(X_j)] \end{aligned}$$

For $i \neq j$, because of independence

$$E[X_i - E(X_i)][X_j - E(X_j)] = E[X_i - E(X_i)]E[X_j - E(X_j)] = 0$$

and we get the formula

$$\mathrm{Var}(S_n) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) \tag{3.3}$$

Weak Law of Large Numbers

Let us look at the distribution of the number of succesees in n independent trials, with the probability of success in a single trial being equal to p.

$$P\{S_n = r\} = \binom{n}{r} p^r (1-p)^{n-r}$$

and

$$\begin{aligned}
P\{|S_n - np| \geq n\delta\} &= \sum_{|r-np|\geq n\delta} \binom{n}{r} p^r (1-p)^{n-r} \\
&\leq \frac{1}{n^2\delta^2} \sum_{|r-np|\geq n\delta} (r-np)^2 \binom{n}{r} p^r (1-p)^{n-r} \\
&\leq \frac{1}{n^2\delta^2} \sum_{1\leq r\leq n} (r-np)^2 \binom{n}{r} p^r (1-p)^{n-r} \\
&= \frac{1}{n^2\delta^2} E[S_n - np]^2 = \frac{1}{n^2\delta^2} \mathrm{Var}(S_n) \\
&= \frac{1}{n^2\delta^2} np(1-p) \\
&= \frac{p(1-p)}{n\delta^2}
\end{aligned}$$

We have used the fact that $S_n = X_1 + X_2 + \cdots + X_n$ where the X_i are independent and have the simple distribution $P\{X_i = 1\} = p$ and $P\{X_i = 0\} = 1 - p$. Therefore $E(S_n) = np$ and $\mathrm{Var}(S_n) = n\mathrm{Var}(X_1) = np(1-p)$

It follows now that

$$\lim_{n\to\infty} P\{|S_n - np| \geq n\delta\} = \lim_{n\to\infty} P\{|\frac{S_n}{n} - p| \geq \delta\} = 0$$

or the average $\frac{S_n}{n}$ converges to p in probability. This is seen easily to be equivalent to the statement that the distribution of $\frac{S_n}{n}$ converges to the distribution degenerate at p.

The above argument works for any sequence of independent and identically distributed random variables. If we assume that $E(X_i) = m$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$, then $E(\frac{S_n}{n}) = m$ and $\mathrm{Var}(\frac{S_n}{n}) = \frac{\sigma^2}{n}$. Chebychev's inequality states that for any random variable X

$$\begin{aligned} P\{|X - E[X]| \geq \delta\} &= \int_{|X-E[X]|\geq\delta} dP \\ &\leq \frac{1}{\delta^2}\int_{|X-E[X]|\geq\delta}[X - E[X]]^2 dP \\ &= \frac{1}{\delta^2}\int [X - E[X]]^2 dP \\ &= \frac{1}{\delta^2}\mathrm{Var}(X) \end{aligned} \tag{3.4}$$

This can be used to prove the weak law of large numbers for the general case of independent identically distributed random variables with finite second moments.

Theorem 3.2. *If $X_1, X_2 \ldots, X_n, \ldots$ is a sequence of independent identically distributed random variables with $E[X_j] \equiv m$ and $\mathrm{VarX_j} \equiv \sigma^2$ then for*

$$S_n = X_1 + X_2 + \cdots + X_n$$

we have

$$\lim_{n\to\infty} P\Big[|\frac{S_n}{n} - m| \geq \delta = 0$$

for any $\delta > 0$.

Proof: Use Chebychev's inequality to estimate

$$P\Big[|\frac{S_n}{n} - m| \geq \delta\Big] \leq \frac{1}{\delta^2}\mathrm{Var}(\frac{S_n}{n}) = \frac{\sigma^2}{n\delta^2}$$

Actually it is enough to assume that $E|X_i| < \infty$ and the existence of the second moment is not needed. We will provide two proofs of the statement

Theorem 3.3. *If $X_1, X_2, \cdots X_n$ are independent and identically distributed with a finite first moment and $E(X_i) = m$, then $\frac{X_1+X_2+\cdots+X_n}{n}$ converges to m in probability as $n \to \infty$.*

Proof 1. Let C be a large constant and let us define X_i^C as the truncated random variable $X_i^C = X_i$ if $|X_i| \leq C$ and $X_i^C = 0$ otherwise.

Let $Y_i^C = X_i - X_i^C$ so that $X_i = X_i^C + Y_i^C$. Then

$$\begin{aligned}\frac{1}{n}\sum_{1\le i\le n} X_i &= \frac{1}{n}\sum_{1\le i\le n} X_i^C + \frac{1}{n}\sum_{1\le i\le n} Y_i^C \\ &= \xi_n^C + \eta_n^C\end{aligned}$$

If we denote by $a_C = E(X_i^C)$ and $b_C = E(Y_i^C)$ we always have $m = a_C + b_C$. Consider the quantity

$$\begin{aligned}\delta_n &= E[|\frac{1}{n}\sum_{1\le i\le n} X_i - m|] \\ &= E[|\xi_n^C + \eta_n^C - m|] \\ &\le E[|\xi_n^C - a_C|] + E[|\eta_n^C - b_C|] \\ &\le \Big[E[|\xi_n^C - a_C|^2]\Big]^{\frac{1}{2}} + 2E[|Y_i^C|]\end{aligned}$$

Taking the limsup as $n \to \infty$, by the earlier argument for each fixed cutoff level C the first term goes to zero. Therefore

$$\limsup_{n\to\infty} \delta_n \le 2E[|Y_i^C|]$$

If we now let the cutoff level C to go to infinity, by the integrability of X_i, $E|Y_i^C| \to 0$ as $C \to \infty$ and we are done. The final step of eastablishing that for any sequence Y_n of random variables $E[|Y_n|] \to 0$ implies that $Y_n \to 0$ in probability, is left as an exercise.

Proof 2. We can use characteristic functions. If we denote the characteristic function of X_i by $\phi(t)$, then the characteristic function of $\frac{1}{n}\sum_{1\le i\le n} X_i$ is given by $\psi_n(t) = [\phi(\frac{t}{n})]^n$. The existence of the first moment assures us that $\phi(t)$ is differentiable at $t = 0$ with a derivative equal to im where $m = E(X_i)$. Therefore

$$\phi(\frac{t}{n}) = 1 + \frac{i\,m\,t}{n} + o(\frac{1}{n})$$

and this implies

$$\lim_{n\to\infty} \psi_n(t) = \exp[\,i\,m\,t\,]$$

and therefore the distribution of $\frac{S_n}{n}$ tends to the degenerate distribution at the point m. The weak law of large numbers is thereby established.

Exercise 3.5. If the underlying distribution is a Cauchy distribution with density $\frac{1}{\pi(1+x^2)}$ and characteristic function $\phi(t) = e^{-|t|}$, prove that the weak law does not hold.

Exercise 3.6. The weak law may hold sometimes even if the mean does not exist. If we dampen the tails of the Cauchy ever so slightly with a density $f(x) = \frac{c}{(1+x^2)\log(1+x^2)}$, show that the weak law of large numbers holds.

Exercise 3.7. In the case of the Binomial distribution with $p = \frac{1}{2}$, use Stirling's formula

$$n! \simeq \sqrt{2\pi}\ e^{-n}\ n^{n+\frac{1}{2}}$$

to estimate the probability

$$\sum_{r \geq nx} \binom{n}{r} \frac{1}{2^n}$$

and show that it decays geometrically in n. Can you calculate the geometric ratio

$$\rho(x) = \lim_{n\to\infty} \left[\sum_{r \geq nx} \binom{n}{r} \frac{1}{2^n} \right]^{\frac{1}{n}}$$

explicitly as a function of x for $x > \frac{1}{2}$?

Strong Limit Theorems

The weak law of large numbers is really a result concerning the behavior of

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

where $X_1, X_2, \cdots, X_n, \ldots$ is a sequence of independent and identically distributed random variables on some probability space $(\Omega, \mathcal{B}, P)$. Under the assumption that X_i are integrable with an integral equal to m, the weak law asserts that as $n \to \infty$ $\frac{S_n}{n} \to m$ in Probability. Since almost everywhere convergence is generally stronger than convergence in Probabilty one may ask if

$$P\Big[\omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = m\Big] = 1$$

This is called the Strong Law of Large Numbers. Strong laws are statements that hold for almost all ω.

Let us look at functions of the form $f_n = \chi_{A_n}$. It is easy to verify that $f_n \to 0$ in probability if and only if $P(A_n) \to 0$. On the other hand

Lemma 3.4. (Borel-Cantelli). *If*

$$\sum_n P(A_n) < \infty$$

then

$$P[\omega : \lim_{n\to\infty} \chi_{A_n}(\omega) = 0] = 1.$$

If the events A_n are mutually independent the converse is also true.

Remark: Note that the complementary event

$$[\omega : \limsup_{n\to\infty} \chi_{A_n}(\omega) = 1]$$

is the same as $\cap_{n=1}^{\infty} \cup_{j=n}^{\infty} A_j$, or the event that infinitely many of the events $\{A_j\}$ occcur.

Proof. By monotone convergence theorem if $\sum_n P(A_n) < \infty$, then $\sum_n \chi_{A_n}(\omega) = S(\omega)$ is finite almost everywhere and

$$E(S(\omega)) = \sum_n P(A_n) < \infty.$$

If an infinite series has a finite sum then the n-th term must go to 0, thereby proving the direct part. To prove the converse we need to show that if $\sum_n P(A_n) = \infty$, then $\lim_{m\to\infty} P(\cup_{n=m}^{\infty} A_n) > 0$. We can use independence and the continuity of probability under monotone limits, to calculate for every m,

$$\begin{aligned} P(\cup_{n=m}^{\infty} A_n) &= 1 - P(\cap_{n=m}^{\infty} A_n^c) \\ &= 1 - \prod_{n=m}^{\infty} (1 - P(A_n)) = 1 - 0 = 1 \end{aligned}$$

and we are done.

Another digression that we want to make into measure theory at this point is to discuss Kolmogorov consistency theorem. How do we know that there are probability spaces that admit a sequence of independent identically distributed random variables with specified distributions? By the construction of product measures that we outlined earlier we can construct a measure on $\mathbf{R}^n$ for every n which is the joint distribution of the first n random variables. Let us denote by P_n this probabilty measure on $\mathbf{R}^n$. They are consistent in the sense that if we project in the natural way from $\mathbf{R}^{n+1} \to \mathbf{R}^n$, P_{n+1} projects to P_n. Such a family is called a consistent family of finite dimensional distributions. We look at the space $\Omega = \mathbf{R}^\infty$ consisting of all real sequences $\omega = \{x_n : n \geq 1\}$ with a natural σ-field Σ generated by the field $\mathcal{F}$ of finite dimensional cylinder sets of the form $B = \{\omega : (x_1, \cdots, x_n) \in A\}$ where A varies over Borel sets in $\mathbf{R}^n$ and varies over positive integers.

Theorem 3.5. (Kolmogorov Connsistency Theorem). *Given a consistent family of finite dimensional distributions P_n, there exists a unique P on (Ω, Σ) such that for every n, under the natural projection $\pi_n(\omega) = (x_1, \cdots, x_n)$, the induced measure $P\pi_n^{-1} = P_n$ on $\mathbf{R}^n$.*

Proof. The consistency is just what is required to be able to define P on $\mathcal{F}$ by

$$P(B) = P_n(A).$$

Once we have P defined on the field $\mathcal{F}$, we have to prove the countable additivity of P on $\mathcal{F}$. The rest is then routine. Let $B_n \in \mathcal{F}$ and $B_n \downarrow \Phi$, the empty set. If possible let $P(B_n) \geq \delta$ for all n and for some $\delta > 0$. Then $B_n = \pi_{k_n}^{-1} A_{k_n}$ for some k_n and without loss of generality we assume that $k_n = n$, so that $B_n = \pi_n^{-1} A_n$ for some Borel set $A_n \subset \mathbf{R}^n$. According to Exercise 3.8 below, we can find a closed bounded subset $K_n \subset A_n$ such that

$$P_n(A_n - K_n) \leq \frac{\delta}{2^{n+1}}$$

and define $C_n = \pi_n^{-1} K_n$ and $D_n = \cap_{j=1}^n C_j = \pi_n^{-1} F_n$ for some closed bounded set $F_n \subset K_n \subset \mathbf{R}^n$. Then

$$P(D_n) \geq \delta - \sum_{j=1}^{n} \frac{\delta}{2^{j+1}} \geq \frac{\delta}{2}.$$

$D_n \subset B_n$, $D_n \downarrow \Phi$ and each D_n is nonempty. If we take $\omega^{(n)} = \{x_j^n : j \geq 1\}$ to be an arbitrary point from D_n, by our construction $(x_1^n, \cdots x_m^n) \in F_m$ for $n \geq m$. We can definitely choose a subsequence (diagonlization) such that $x_j^{n_k}$ converges for each j producing a limit $\omega = (x_1, \cdots, x_m, \cdots)$ and, for every m, we will have $(x_1, \cdots, x_m) \in F_m$. This implies that $\omega \in D_m$ for every m, contradicting $D_n \downarrow \Phi$. We are done.

Exercise 3.8. We have used the fact that given any borel set $A \subset \mathbf{R}^n$, and a probability measure α on $\mathbf{R}^n$, for any $\epsilon > o$, there exists a closed bounded subset $K_\epsilon \subset A$ such that $\alpha(A - K_\epsilon) \leq \epsilon$. Prove it by showing that the class of sets $\mathcal{A}$ with the above property is a monotone class that contains finite disjoint unioms of measurable rectangles and therefore contains the Borel σ-field. To prove the last fact, establish it first for $n = 1$. To handle $n = 1$, repeat the same argument starting from finite disjoint unions of right-closed left-open intevals. Use the countable additivity to verify this directly.

The following is a strong version of the Law of Large Numbers.

Theorem 3.6. If $X_1, \cdots, X_n \cdots$ is a sequence of independent identically distributed random variables with $E|X_i|^4 = C < \infty$, then

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = E(X_1)$$

with probability 1.

Proof. We can assume without loss of generality that $E(X_i) = 0$. Just take $Y_i = X_i - E(X_i)$. A simple calculation shows

$$E[(S_n)^4] = nE(X_1^4) + 3n(n-1)E[(X_1^2)]^2 \leq nC + 3n^2\sigma^4$$

and by applying a Chebychev type inequality using fourth moments,

$$P[|\frac{S_n}{n}| \geq \delta\,] = P[\,|S_n| \geq n\delta\,] \leq \frac{nC + 3n^2\sigma^4}{n^4\delta^4}.$$

Because

$$\sum_{n=1}^{\infty} P[\,|\frac{S_n}{n}| \geq \delta\,] < \infty$$

we can apply the Borel-Cantelli Lemma and we are done.

Series of Independent Random variables

We wish to investigate conditions under which an infinite series with independent summands

$$S = \sum_{j=1}^{\infty} X_j$$

converges with probability 1. The basic steps are the following inequalities due to Kolomogorov and Lévy that control the behaviour of sums of independent random variables. They both deal with the problem of estimating

$$T_n(\omega) = \sup_{1 \le k \le n} |S_k(\omega)| = \sup_{1 \le k \le n} |\sum_{j=1}^{k} X_j(\omega)|$$

where $X_1, \cdots, X_n$ are n independent random variables.

Lemma 3.7. (Kolmogorov's Inequality). *Assume that* $EX_i = 0$ *and* $\mathrm{Var}(X_i) = \sigma_i^2 < \infty$ *and let* $s_n^2 = \sum_{j=1}^{n} \sigma_j^2$. *Then*

$$P\{T_n(\omega) \ge \ell\} \le \frac{s_n^2}{\ell^2}.$$

Proof. The important point here is that the estimate depends only on s_n^2 and not on the number of summands. In fact the Chebychev bound on S_n is

$$P\{|S_n| \ge \ell\} \le \frac{s_n^2}{\ell^2}$$

and the supremum does not cost anything.

Let us define the events $E_k = \{|S_1| < \ell, \cdots, |S_{k-1}| < \ell, |S_k| \ge \ell\}$ and then $\{T_n \ge \ell\} = \cup_{k=1}^{n} E_k$ is a disjoint union of E_k. If we use the independence of $S_n - S_k$ and $S_k \chi_{E_k}$ that only depends on $X_1 \cdots, X_k$

$$\begin{aligned} P\{E_k\} &\le \frac{1}{\ell^2} \int_{E_k} S_k^2 \, dP \\ &\le \frac{1}{\ell^2} \int_{E_k} [S_k^2 + (S_n - S_k)^2] \, dP \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\ell^2}\int_{E_k}[S_k^2 + 2S_k(S_n - S_k) + (S_n - S_k)^2]\,dP \\
&= \frac{1}{\ell^2}\int_{E_k} S_n^2\,dP
\end{aligned}
$$

Summing over k from 1 to n

$$P\{T_n \geq \ell\} \leq \frac{1}{\ell^2}\int_{T_n \geq \ell} S_n^2\,dP \leq \frac{s_n^2}{\ell^2}.$$

Lemma 3.8. (Lévy's Inequality). *Assume that*

$$P\{|X_i + \cdots + X_n| \geq \frac{\ell}{2}\} \leq \delta$$

for all $1 \leq i \leq n$. *Then*

$$P\{T_n \geq \ell\} \leq \frac{\delta}{1-\delta}$$

Proof. Let E_k be as in the previous lemma.

$$
\begin{aligned}
P\Big\{(T_n \geq \ell) \cap |S_n| \leq \frac{\ell}{2}\Big\} &= \sum_{k=1}^n P\Big\{E_k \cap |S_n| \leq \frac{\ell}{2}\Big\} \\
&\leq \sum_{k=1}^n P\Big\{E_k \cap |S_n - S_k| \geq \frac{\ell}{2}\Big\} \\
&= \sum_{k=1}^n P\Big\{|S_n - S_k| \geq \frac{\ell}{2}\Big\}\,P(E_k) \\
&\leq \delta\sum_{k=1}^n P(E_k) \\
&= \delta P\{T_n \geq \ell\}
\end{aligned}
$$

On the other hand,

$$P\Big\{(T_n \geq \ell) \cap |S_n| > \frac{\ell}{2}\Big\} \leq P\{|S_n| > \frac{\ell}{2}\Big\} \leq \delta$$

Adding the two,

$$P\{T_n \geq \ell\} \leq \delta P\{T_n \geq \ell\} + \delta$$

or

$$P\{T_n \geq \ell\} \leq \frac{\delta}{1-\delta}$$

We are now ready to prove

Theorem 3.9. *(Lévy's Theorem). If $X_1, X_2, \ldots, X_n, \ldots$ are independent, then the following are equivalent.*

(i) The distribution α_n of $S_n = X_1 + \cdots + X_n$ converges weakly to a probability distribution α on $\mathbf{R}$.

(ii) The random variable $S_n = X_1 + \cdots + X_n$ converges in probability to a limit $S(\omega)$.

(iii) The random variable $S_n = X_1 + \cdots + X_n$ converges with probability 1 to a limit $S(\omega)$.

Proof. Clearly $(iii) \Rightarrow (ii) \Rightarrow (i)$ are trivial. We will establish $(i) \Rightarrow (ii) \Rightarrow (iii)$.

$(i) \Rightarrow (ii)$**.** The characteristic functions $\phi_j(t)$ of X_j are such that

$$\phi(t) = \prod_{i=1}^{\infty} \phi_j(t)$$

is a convergent infinite product. Since the limit $\phi(t)$ is continuous at $t = 0$ and $\phi(0) = 1$ it is nonzero in some interval $|t| \leq T$ around 0. Therefore for $|t| \leq T$,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \prod_{m+1}^{n} \phi_j(t) = 1$$

By Exercise 3.9 below, this implies that for all t,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} \prod_{m+1}^{n} \phi_j(t) = 1$$

and consequently, the distribution of $S_n - S_m$ converges to the distribution degenerate at 0. This implies the convergence in probability to 0 of $S_n - S_m$ as $m, n \to \infty$. Therefore for each $\delta > 0$,

$$\lim_{\substack{n\to\infty \\ m\to\infty}} P\{|S_n - S_m| \geq \delta\} = 0$$

establishing (ii).

(ii) $\Rightarrow$ *(iii)*. To establish (iii), because of Exercise 3.10 below, we need only show that for every $\delta > 0$

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{\sup_{m<k\leq n} |S_k - S_m| \geq \delta\} = 0$$

and this follows from (ii) and Lévy's inequality.

Exercise 3.9. Prove the inequality $1 - \cos 2t \leq 4(1 - \cos t)$ for all real t. Deduce the inequality $1 - \text{Real}\,\phi(2t) \leq 4[1 - \text{Real}\,\phi(t)]$, valid for any characteristic function. Conclude that if a sequence of characteristic functions converges to 1 in an interval around 0, then it converges to 1 for all real t.

Exercise 3.10. Prove that if a sequence S_n of random variables is a Cauchy sequence in Probaility, i.e. for each $\delta > 0$,

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{|S_n - S_m| \geq \delta\} = 0$$

then there is a random variable S such that $S_n \to S$ in probability, i.e for each $\delta > 0$,

$$\lim_{n\to\infty} P\{|S_n - S| \geq \delta\} = 0.$$

Exercise 3.11. Prove that if a sequence S_n of random variables satisfies

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{\sup_{m<k\leq n} |S_k - S_m| \geq \delta\} = 0$$

for every $\delta > 0$ then there is a limiting random variable $S(\omega)$ such that

$$P\Big\{\lim_{n\to\infty} S_n(\omega) = S(\omega)\Big\} = 1$$

Exercise 3.12. Prove that whenever $X_n \to X$ in probability the distribution α_n of X_n converges weakly to the distribution α of X.

Now it is straightforward to find sufficient conditions for the convergence of an infinite series of independent random variables.

Theorem 3.10. (Kolmogorov's one series Theorem). *If $E(X_i) = 0$ and $\sum_{i=1}^{\infty} \mathrm{Var}(X_i) < \infty$, then*

$$S(\omega) = \sum_{i=1}^{\infty} X_i(\omega)$$

converges with probability 1.

Proof. By a direct application of Kolmogorov's inequality

$$\begin{aligned}\lim_{\substack{n\to\infty\\ m\to\infty}} P\{\sup_{m<k\le n} |S_n - S_m| \ge \delta\} &\le \lim_{\substack{n\to\infty\\ m\to\infty}} \sum_{j=m+1}^{n} E(X_i^2)\\ &= \lim_{\substack{n\to\infty\\ m\to\infty}} \sum_{j=m+1}^{n} \mathrm{Var}(X_i) = 0\end{aligned}$$

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{|S_n - S_m| \ge \delta\} \le 0$$

or we can first prove convergence in probability

$$\lim_{\substack{n\to\infty\\ m\to\infty}} P\{|S_n - S_m| \ge \delta\} = 0$$

by a simple application of Chebychev's inequality and then apply Lévy's Theorem.

Theorem 3.11. (Kolomogorov's two series theorem). *If $a_i = E(X_i)$ and $\sigma_i^2 = \mathrm{Var}(X_i)$ are such that $\sum_i a_i$ and $\sum_i \sigma_i^2$ converge, then the series $\sum_i X_i$ converges with probability* 1.

Proof. Define $Y_i = X_i - a_i$ and apply the previous theorem to Y_i. Of course in general random variables need not have finite expectations or variances. If $\{X_i\}$ is any sequence of random variables we can take a cut off value C and define $Y_i = X_i$ if $|X_i| \le C$, and $Y_i = 0$ otherwise. The Y_i are then independent and bounded in absolute value by C. The theorem can be applied to Y_i and if we impose the additional condition that

$$\sum_i P\{X_i \ne Y_i\} = \sum_i P\{|X_i| > C\} < \infty$$

by an application of Borel-Cantelli Lemma, with Probabilty 1, $X_i = Y_i$ for all sufficiently large i. The convergence of $\sum_i X_i$ and $\sum_i Y_i$ are therefore equivalent. We get then

Theorem 3.12. (Kolmogorov's three series theorem). *For the convergence of an infinite series of independent random variables $\sum_i X_i$ it is sufficient that the the following three infinite series converge.*

(i) For some cut off value $C > 0$, $\sum_i P\{|X_i| > C\}$ converges.

(ii) If Y_i is defined to equal X_i if $|X_i| \leq C$, and 0 otherwise, $\sum_i E(Y_i)$ converges.

(iii) With Y_i as in (ii), $\sum_i \mathrm{Var}(Y_i)$ converges.

The converse is also true. If $\sum_i X_i$ converges for a sequence of independent random variables, we must necessarily have $|X_n| \leq C$ eventually with probability 1. By Borel-Cantelli Lemma the first series must converge. This means that in order to prove the necessity we can assume with out loss of generality that $|X_i|$ are all bounded say by 1. we may also assume that $E(X_i) = 0$ for each i. Otherwise let us take independent random variables X_i' that have the same distribution as X_i. Then $\sum_i X_i$ as well as $\sum_i X_i'$ converge with probability 1 and therefore so does $\sum_i (X_i - X_i')$. The random variables $Z_i = X_i - X_i'$ are independent and bounded by 2. They have mean 0. If we can show $\sum \mathrm{Var}(Z_i)$ is convergent, since $\mathrm{Var}(Z_i) = 2\mathrm{Var}(X_i)$ we would have proved the convergence of the the third series. Now it is elementary to conclude that since both $\sum_i X_i$ as well as $\sum_i (X_i - E(X_i))$ converge, the series $\sum_i E(X_i)$ must be convergent as well. So all we need is the following

Lemma 3.13. *If $\sum_i X_i$ is convergent for a series of independent random variables with mean 0 that are individually bounded by C, then $\sum_i \mathrm{Var}(X_i)$ is convergent.*

Proof. Let $F_n = \{\omega : |S_1| \leq \ell, |S_2| \leq \ell, \cdots, |S_n| \leq \ell\}$ where $S_k = X_1 + \cdots + X_k$. If the series converges with probablity 1, we must have, for some ℓ and $\delta > 0$, $P(F_n) \geq \delta$ for all n. We have

$$\begin{aligned}
\int_{F_{n-1}} S_n^2 \, dP &= \int_{F_{n-1}} [S_{n-1} + X_n]^2 \, dP \\
&= \int_{F_{n-1}} [S_{n-1}^2 + 2S_{n-1}X_n + X_n^2] \, dP
\end{aligned}$$

$$\begin{aligned} &= \int_{F_{n-1}} S_{n-1}^2\, dP + \sigma_n^2 P(F_{n-1}) \\ &\geq \int_{F_{n-1}} S_{n-1}^2\, dP + \delta\sigma_n^2 \end{aligned}$$

and on the other hand,

$$\begin{aligned} \int_{F_{n-1}} S_n^2\, dP &= \int_{F_n} S_n^2\, dP + \int_{F_{n-1}\cap F_n^c} S_n^2\, dP \\ &\leq \int_{F_n} S_n^2\, dP + P(F_{n-1}\cap F_n^c)\,(\ell + C)^2 \end{aligned}$$

providing us with the estimate

$$\delta\sigma_n^2 \leq \int_{F_n} S_n^2\, dP - \int_{F_{n-1}} S_{n-1}^2\, dP + P(F_{n-1}\cap F_n^c)\,(\ell + C)^2$$

Since $F_{n-1}\cap F_n^c$ are disjoint and $|S_n| \leq \ell$ on F_n ,

$$\sum_{j=1}^{\infty} \sigma_j^2 \leq \frac{1}{\delta}[\ell^2 + (\ell + C)^2]$$

This concludes the proof.

Strong Law of Large Numbers

We will now prove the theorem that if $X_1 \cdots, X_n, \cdots$ is a sequence of i.i.d. random variables with $E|X_i| < \infty$ and $E(X_i) = 0$, then

Theorem 3.14. *With probability* 1,

$$\lim_{n\to\infty} \frac{S_n}{n} = 0$$

where $S_n = X_1 + \cdots + X_n$.

Proof. We define

$$\begin{cases} Y_n = X_n & \text{if } \ |X_n| \leq n \\ Y_n = 0 & \text{otherwise} \end{cases}$$

$$a_n = P[X_n \neq Y_n], \quad b_n = E[Y_n] \ \text{ and } \ c_n = \mathrm{Var}(Y_n).$$

First we note that

$$\sum_n a_n = \sum_n P[|X_1| > n\,] \le E|X_1| < \infty$$

$$\lim_{n\to\infty} b_n = 0$$

and

$$\begin{aligned}\sum_n \frac{c_n}{n^2} \le \sum_n \frac{E[Y_n^2]}{n^2} &= \sum_n \int_{|x|\le n} \frac{x^2}{n^2}\, d\alpha \\ &= \int x^2 \Big(\sum_{n\ge x} \frac{1}{n^2}\Big)\, d\alpha \le C\int |x|\, d\alpha < \infty\end{aligned}$$

where α is the common distribution of X_i. From the three series theorem and the Borel-Cantelli Lemma, we conclude that $\sum_n \frac{Y_n - b_n}{n}$ as well as $\sum_n \frac{X_n - b_n}{n}$ converge almost surely. It is elementary to verify that for any series $\sum_n \frac{x_n}{n}$ that converges, $\frac{x_1+\cdots+x_n}{n} \to 0$ as $n \to \infty$. We therefore conclude that

$$P\Big\{\lim_{n\to\infty} [\frac{X_1+\cdots+X_n}{n} - \frac{b_1+\cdots+b_n}{n}] = 0\Big\} = 1$$

Since $b_n \to 0$ as $n \to \infty$, the theorem is proved.

Exercise 3.13. Show that for a sequence of i.i.d. random variables $X_1, \cdots, X_n, \cdots$, if the strong law of large numbers holds with some limit, i.e.

$$P[\lim_{n\to\infty} \frac{S_n}{n} = \xi\,] = 1$$

for some random variable ξ, which may or may not be a constant with probability 1, then necessarily $E|X_i| < \infty$. Consequently $\xi = E(X_i)$ with probabilty 1.

One may ask why the limit cannot be a proper random variable. There is a general theorem that forbids it called Kolmogorov's Zero-One law. Let us look at the space Ω of real sequences $\{x_n : n \ge 1\}$. We have the σ-field $\mathcal{B}$, the product σ-field on Ω. In addition we have the sub σ-fields $\mathcal{B}^n$ generated by $\{x_j : j \ge n\}$. $\mathcal{B}^n$ are $\downarrow$ with n and $\mathcal{B}^\infty = \cap_n \mathcal{B}^n$ which is also a σ-field is called the tail σ-field. The typical set in $\mathcal{B}^\infty$ is set depending on the tail behavior of the sequence. For example the sets $\{\omega : x_n \text{ is bounded }\}$, $\{\omega : \limsup_n x_n = 1\}$ are in $\mathcal{B}^\infty$ where as $\{\omega : \sup_n |x_n| = 1\}$ is not.

Theorem 3.15. (Kolmogorov's Zero-One Law). *If $A \in \mathcal{B}^\infty$ and P is any product measure (not necessarily with identical components) $P(A) = 0$ or 1.*

Proof. The proof depends on showing that A is independent of itself so that $P(A) = P(A \cap A) = P(A)P(A) = [P(A)]^2$ and therefore equals 0 or 1. The proof is elementary. Since $A \in \mathcal{B}^\infty \subset \mathcal{B}^{n+1}$ and P is a product measure, A is independent of $\mathcal{B}_n$ = σ-field generated by $\{x_j : 1 \leq j \leq n\}$. It is therefore independent of sets in the field $\mathcal{F} = \cup_n \mathcal{B}_n$. The class of sets $\mathcal{A}$ that are independent of A is a monotone class. Since it contains the field $\mathcal{F}$ it contains the σ-field $\mathcal{B}$ generated by $\mathcal{F}$. In particular since $A \in \mathcal{B}$, A is independent of itself.

Corollary 3.16. *Any random variable measurable with respect to the tail σ-field $\mathcal{B}^\infty$ is equal with probaility 1 to a constant relative to any given product measure.*

Warning. For different product measures the constants can be different.

Exercise 3.14. How can that happen?