

Laws of Large Numbers ¹

Let $X_1, X_2 \dots$ be a sequence of IID random variables on $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}X_n = \mu < \infty$, and let A_n denote the average:

$$A_n = (X_1 + \dots + X_n)/n.$$

Observe that $\mathbb{E}A_n = n\mu/n = \mu$. That is, the *expected* frequency of successes is μ for *any* n . The content of the laws of large numbers is that for large n the *realized* frequency of successes is close to μ with high probability.

1 Markov and Chebyshev inequalities

Proposition (The basic inequality). *For $X \geq 0$ and $c > 0$ we have $P(X \geq c) \leq E(X)/c$.*

Proof. We have

$$\begin{aligned} E(X) &= E(X \cdot I(X \geq c)) + E(X \cdot I(X < c)) \geq E(X \cdot I(X \geq c)) \\ &\geq E(c \cdot I(X \geq c)) = c \cdot E(I(X \geq c)) = c \cdot P(X \geq c). \end{aligned} \quad \square$$

Applying this to $|X|^k$ (and observing that $|X| \geq c \iff |X|^k \geq c^k$) we get the *Markov inequality*

$$P(|X| \geq c) \leq E(|X|^k)/c^k.$$

And applying this to $|X - E(X)|$ we get the *Chebyshev inequality*

$$P(|X - E(X)| \geq k) \leq \text{Var}(X)/k^2.$$

2 The Weak Law of Large Numbers

Theorem (WLLN). *Assume $\text{Var}(X_n) = \sigma^2 < \infty$ for all n . Then for any fixed $\epsilon > 0$*

$$P(|A_n - \mu| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Proof. We know that $\mathbb{E}A_n = \mu$ and $\text{Var}(A_n) = \sigma^2/n$, so the Chebyshev inequality directly implies $P(|A_n - \mu| > \epsilon) \leq \sigma^2/(n\epsilon^2) \rightarrow 0$. $\square$

¹S Modica 2025. Main references: Williamson *Probability with Martingales* and Billingsley *Probability and Measure*.

3 The Strong Law of Large Numbers

The limits will always be taken as $n \rightarrow \infty$. The Strong Law asserts that the event $A_n \rightarrow \mu$ has probability 1. But one has to show that the event is in $\mathcal{F}$ in the first place.

Lemma. $\{A_n \rightarrow \mu\} \equiv \{\omega : \lim_{n \rightarrow \infty} A_n(\omega) = \mu\} \in \mathcal{F}$.

Proof. Since $A_n \rightarrow \mu \iff A_n - \mu \rightarrow 0$ it suffices to take $\mu = 0$. By definition $\{\omega : \lim_{n \rightarrow \infty} A_n(\omega) = 0\}$ is the set of ω 's such that

$$\forall h > 0 \quad \exists n \text{ such that } |A_m(\omega)| < 1/h \quad \forall m \geq n.$$

In other words

$$\{A_n \rightarrow 0\} = \bigcap_{h \geq 1} \bigcup_n \bigcap_{m \geq n} \{|A_m(\omega)| < 1/h\} \quad (1)$$

and this is in $\mathcal{F}$ since $\{|A_m(\omega)| < 1/h\} \in \mathcal{F}$ for all m . $\square$

Now we can state the result. A proof is presented in Section 5.

Theorem (SLLN). $P(A_n \rightarrow \mu) = 1$.

4 The Strong Law is stronger than the Weak Law

We abbreviate $\lim_{n \rightarrow \infty} a_n = b$ with $a_n \rightarrow b$. We show first that the strong law implies the weak law. Convergence $X_n \rightarrow 0$ in probability is equivalent to

$$P(|X_n| > 1/m) \rightarrow 0 \quad \forall m \in \mathbb{N}.$$

Assume the strong law, that $P(X_n \rightarrow 0) = 1$. This is equivalent to $P(X_n \not\rightarrow 0) = 0$. The event $\{X_n \not\rightarrow 0\}$ is that for some $m > 0$ it is $\{|X_n| \geq 1/m\}$ infinitely often, that is

$$\{X_n \not\rightarrow 0\} = \bigcup_{m > 0} I_m \quad \text{with} \quad I_m = \bigcap_{n > 0} \bigcup_{k \geq n} \{|X_k| \geq 1/m\}.$$

Now $P(X_n \not\rightarrow 0) = P(\bigcup_{m > 0} I_m) = 0$ clearly implies $P(I_m) = 0 \quad \forall m > 0$. Fix m ; letting $E_n^m = \bigcup_{k \geq n} \{|X_k| \geq 1/m\}$ so that $I_m = \bigcap_n E_n^m$; the sequence E_n^m is decreasing in n , therefore by continuity $P(E_n^m) \rightarrow P(\bigcap_n E_n^m) = P(I_m) = 0$. On the other hand obviously $P(|X_n| \geq 1/m) \leq P(\bigcup_{k \geq n} \{|X_k| \geq 1/m\}) = P(E_n^m)$. The weak law follows.

Thus almost sure convergence implies convergence in probability. *The converse is false*, as the following well known example shows.

Example. Let $\{X_n\}$ be an independent sequence (on some $(\Omega, \mathcal{F}, P)$) with $P(X_n = 1) = 1/n$ and $P(X_n = 0) = 1 - 1/n$. For any $\epsilon > 0$ it is $P(|X_n| \geq \epsilon) \leq 1/n \rightarrow 0$ so $X_n \rightarrow 0$ in probability. On the other hand $\sum_{n \geq 1} P(X_n = 1) = \sum_{n \geq 1} 1/n = \infty$ and the events $\{X_n = 1\}$ are independent, so by the second Borel-Cantelli lemma $P(\limsup \{X_n = 1\}) = 1$, which says that with probability 1 it is $X_n = 1$ infinitely often, that is $P(X_n \rightarrow 0) = 0$ (not 1!).

5 Proof of the Strong Law

Sequences of sets. Let $\{E_n\}_{n \in \mathbb{N}}$ be a sequence of sets in $\mathcal{F}$. The event $\limsup E_n$ is defined as the event that E_n occurs *infinitely often*, abbreviated E_n *i.o.*; the event $\liminf E_n$ is the event that E_n occurs *eventually*, abbreviated E_n *ev.* Formally (with k also integer)

$$\limsup E_n = \bigcap_n \bigcup_{k \geq n} E_k \quad \text{and} \quad \liminf E_n = \bigcup_n \bigcap_{k \geq n} E_k.$$

Both $\limsup E_n$ and $\liminf E_n$ are in $\mathcal{F}$ because $\mathcal{F}$ is a σ -algebra. Note that if E_n occurs eventually it occurs infinitely often, that is $\liminf E_n \subseteq \limsup E_n$. Also, E_n does not occur eventually iff E_n^c occurs infinitely often: indeed $(\liminf E_n)^c = \limsup E_n^c$.

Series of numbers. If $a_n \geq 0$, by definition $\sum_{n=1}^{\infty} a_n = \lim_{N \rightarrow \infty} \sum_{n=1}^N a_n \leq \infty$. Write $\sum_{n=1}^{\infty} a_n$ as $\sum_n a_n$ to ease notation. If $\sum_n a_n < \infty$ we say that the series converges; if $\sum_n a_n = \infty$ the series diverges. Suppose the series converges, say $\sum_n a_n = s$. Observe that

$$\sum_{n \geq m} a_n = s - \sum_{n=1}^{m-1} a_n$$

and $\lim_{m \rightarrow \infty} \sum_{n=1}^{m-1} a_n = s$, so $\lim_{m \rightarrow \infty} \sum_{n \geq m} a_n = 0$ if the series converges. We will use the fact that $\sum_{n=1}^{\infty} 1/n^2$ converges,² and that on the contrary $\sum_{n \geq 1} 1/n = \infty$.³

²Proof of this: observe that for all $n > 1$ we have $1/n^2 < 1/[(n-1)n]$; but $1/[(n-1)n] = 1/(n-1) - 1/n$, therefore

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{(n-1)n} = 1 - \frac{1}{n} \rightarrow 1.$$

³One way to prove this is the following. Let $S_n = \sum_{1 \leq h \leq n} 1/h$ be the partial sum. We know from calculus (if not, just draw) that $e^x > 1 + x$. Now

$$\begin{aligned} e^{S_n} &= \exp(1 + 1/2 + \cdots + 1/n) = e^1 \cdot e^{1/2} \cdot \cdots \cdot e^{1/n} \\ &> (1+1) \cdot (1+1/2) \cdot \cdots \cdot (1+1/n) \\ &= \frac{2}{1} \cdot \frac{3}{2} \cdot \cdots \cdot \frac{n+1}{n} = n+1 \end{aligned}$$

so $S_n = \ln e^{S_n} > \ln(n+1) \rightarrow \infty$.

The Borel-Cantelli Lemmas. They are two ubiquitous results:

Proposition (Borel-Cantelli Lemmas). *Let $\{E_n\}$ be a sequence of events in $\mathcal{F}$.*

1st. *If $\sum_n P(E_n)$ converges then $P(\limsup E_n) = 0$.*

2nd. *For independent sequences,⁴ if $\sum_n P(E_n)$ diverges then $P(\limsup E_n) = 1$.*

Proof. First Lemma. For each m we have $\limsup E_n \subseteq \cup_{k \geq m} E_k$, so

$$P(\limsup E_n) \leq P(\cup_{k \geq m} E_k) \leq \sum_{k \geq m} P(E_k)$$

and this tends to zero as $m \rightarrow \infty$ if $\sum_n P(E_n)$ converges.

Second Lemma. The complement of $\limsup E_n$ is $\liminf E_n^c$ so it suffices to prove that $P(\liminf E_n^c) = 0$. Since $\liminf E_n^c = \cup_{n \geq 1} \cap_{k \geq n} E_k^c$ it is enough to prove that $P(\cap_{k \geq n} E_k^c) = 0$ for all n . Fix n . Since $1 - x \leq e^{-x}$ for any $x \in \mathbb{R}$,⁵ we have (equality below by independence)

$$P(\cap_{k=n}^{n+j} E_k^c) = \prod_{k=n}^{n+j} (1 - P(E_k)) \leq e^{-\sum_{k=n}^{n+j} P(E_k)}.$$

Since $\sum_n P(E_n)$ diverges $\lim_{j \rightarrow \infty} \sum_{k=n}^{n+j} P(E_k) = \infty$, so

$$0 = \lim_{j \rightarrow \infty} e^{-\sum_{k=n}^{n+j} P(E_k)} = \lim_{j \rightarrow \infty} P(\cap_{k=n}^{n+j} E_k^c) = P(\cap_{k \geq n} E_k^c). \quad \square$$

Proof of the Theorem. Following Billingsley we shall assume $\mathbb{E}X_n^4 < \infty$ all n (the general proof is considerably more involved). Since $A_n \rightarrow \mu \iff A_n - \mu \rightarrow 0$ we may focus on $\mu = 0$ without loss of generality. First observe that by equation (1) $P(A_n \rightarrow 0) = 1$ iff for each $h \geq 1$ it is $P(\{|A_n(\omega)| < 1/h \text{ ev.}\}) = 1$, equivalently iff for each $h \geq 1$ the complement has probability zero, that is $P(\{|A_n(\omega)| \geq 1/h \text{ i.o.}\}) = 0$. Fix $h \geq 1$. Let $S_n = nA_n = \sum_{i=1}^n X_i$; observe that $E(S_n^4) = \sum E(X_a X_b X_c X_d)$ with a, b, c, d ranging independently from 1 to n . Since $E(X_i) = 0$, by independence the terms where one index is different from the other three is zero. This leaves the n terms $E(X_i^4) \equiv \xi^4$ and the terms $E(X_i^2 X_j^2) = E(X_i^2)E(X_j^2) = \sigma^4$ (we know that $E(X_i^2) = \sigma^2$ since $E(X_i) = 0$), of which there are $\binom{n}{2} \binom{4}{2} = 3n(n-1)$ (pick two indexes i, j and place them anywhere in the n factors, and for each i, j choose the two places for i). Therefore

$$E(S_n^4) = n\xi^4 + 3n(n-1)\sigma^4 \leq Kn^2$$

for some constant K . Using Markov inequality for $k = 4$ we get

$$P(|A_n(\omega)| \geq 1/h) = P(|S_n(\omega)| \geq n/h) \leq E(S_n^4) h^4 / n^4 \leq Kh^4 / n^2.$$

⁴A sequence of sets is independent if any finite subsequence is independent.

⁵If you don't know it just draw the functions, observing that $De^{-x} = -1$ for $x = 0$ meaning that e^{-x} is tangent to the line at zero.

Therefore $\sum_n P(|A_n(\omega)| \geq 1/h)$ converges, and then by the first Borel-Cantelli Lemma we conclude that $P(\{|A_n(\omega)| \geq 1/h \text{ i.o.}\}) = 0$.