Laws of Large Numbers

Chebyshev's Inequality: Let X be a random variable and $a \in \mathbb{R}^+$. We assume X has density function f_X. Then

$$\begin{aligned} E(X^2) &= \int_{\mathbb{R}} x^2 f_X(x)\, dx \\ &\geq \int_{|x|\geq a} x^2 f_X(x)\, dx \\ &\geq a^2 \int_{|x|\geq a} f_X(x)\, dx = a^2 \mathrm{P}\left(|X| \geq a\right). \end{aligned}$$

That is, we have proved

$$\mathrm{P}\left(|X| \geq a\right) \leq \frac{1}{a^2} E(X^2). \tag{1}$$

We can generalize this to any moment $p > 0$:

$$\begin{aligned} E(|X|^p) &= \int_{\mathbb{R}} |x|^p f_X(x)\, dx \\ &\geq \int_{|x|\geq a} |x|^p f_X(x)\, dx \\ &\geq a^p \int_{|x|\geq a} f_X(x)\, dx = a^p \mathrm{P}\left(|X| \geq a\right). \end{aligned}$$

That is, we have proved

$$\mathrm{P}\left(|X| \geq a\right) \leq \frac{1}{a^p} E(|X|^p) \tag{2}$$

for any $p = 1, 2, \ldots$. (Of course, this assumes that $E(|X|^p) < \infty$ for otherwise the inequality would not be saying much!)

Remark: We have proved (1) and (2) assuming X has a density function f_X. However, (almost) identical proofs show the same inequalities for X having a discrete distribution.

Weak Law of Large Numbers: Let $X_1,\ X_2, X_3,\ \ldots$ be a sequence of independent random variables with common distribution function. Set $\mu = E(X_j)$ and $\sigma^2 = \mathrm{Var}(X_j)$. As usual we define

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

and let

$$S_n^* = \frac{S_n}{n} - \mu.$$

We apply Chebyshev's inequality to the random variable S_n^*. A by now routine calculation gives

$$E(S_n^*) = 0 \text{ and } \mathrm{Var}(S_n^*) = \frac{\sigma^2}{n}.$$

Then Chebyshev (1) says that for every $\varepsilon > 0$

$$\mathrm{P}\left(|S_n^*| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\mathrm{Var}(S_n^*).$$

Writing this out explicitly:

$$\mathrm{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2}\frac{\sigma^2}{n}.$$

Thus for every $\varepsilon > 0$, as $n \to \infty$

$$\mathrm{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0.$$

Borel-Cantelli Lemma: Let $A_1, A_2, \ldots$ be an infinite sequence of events in Ω. Consider the sequence of events

$$\bigcup_{n=1}^{\infty} A_n, \ \bigcup_{n=2}^{\infty} A_n, \ \bigcup_{n=3}^{\infty} A_n, \ldots.$$

Observe that this is a decreasing sequence in the sense that

$$\bigcup_{n=m+1}^{\infty} A_n \subseteq \bigcup_{n=m}^{\infty} A_n$$

for all $m = 1, 2, \ldots$. We are interested in those events ω that lie in infinitely many A_n. Such ω would lie in $\bigcup_{m=n}^{\infty}$ for every m. Thus we define

$$\mathrm{limsup} A_n = \lim_{m\to\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \text{ that are in infinitely many } A_n\}.$$

We write this event as

$$\mathrm{limsup} A_n = \{\omega \in A_n \text{ i.o.}\}$$

where "i.o." is read as "infinitely often." We can now state the Borel-Cantelli Lemma:

If $\sum_{n=1}^{\infty} \mathrm{P}(A_n) < \infty$, then $\mathrm{P}\left(\omega \in A_n \text{ i.o.}\right) = 0$.

Proof: First observe that

$$0 \leq \mathrm{limsup} A_n \subseteq \bigcup_{m=n}^{\infty} A_n$$

for every m since the sequence is a decreasing sequence of events. Thus

$$0 \leq \mathrm{P}\left(\mathrm{limsup} A_n\right) \leq \sum_{n=m}^{\infty} \mathrm{P}(A_n) \tag{3}$$

for every m. But we are assuming that the series $\sum_{n=1}^{\infty} \mathrm{P}(A_n)$ converges. This means that

$$\sum_{n=m}^{\infty} \mathrm{P}(A_n) \to 0 \text{ as } m \to \infty.$$

Taking $m \to \infty$ in (3) then gives (since the right hand side tends to zero)

$$\mathrm{P}(\mathrm{limsup} A_n) = 0.$$

Strong Law of Large Numbers: As above, let $X_1,\ X_2,\ X_3 \ldots$ denote an infinite sequence of independent random variables with common distribution. Set

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

Let $\mu = E(X_j)$ and $\sigma^2 = \mathrm{Var}(X_j)$. The weak law of large numbers says that for every sufficiently large *fixed* n the average S_n/n is likely to be near μ. The strong law of large numbers ask the question in what sense can we say

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu. \tag{4}$$

Clearly, (4) cannot be true for *all* $\omega \in \Omega$. (Take, for instance, in coining tossing the elementary event $\omega = HHHH...$ for which $S_n(\omega) = 1$ for every n and hence $\lim_{n\to\infty} S_n(\omega)/n = 1$.) Thus we want to look at the event

$$\mathcal{E} = \left\{\omega \in \Omega : \lim_{n\to\infty} \frac{S_n(\omega)}{n} = \mu\right\}.$$

The Strong Law of Large Numbers says that

$$\mathrm{P}\left(\mathcal{E}\right) = 1.$$

We will prove this under the additional restriction that $\sigma^2 = E(X_j^2) < \infty$ and $E(X_j^4) < \infty$.

It is no loss of generality to assume $\mu = 0$. (Simply look at the new random variables $Y_j = X_j - \mu$.) Now if

$$\lim_{n\to\infty} \frac{S_n(\omega)}{n} \neq 0,$$

then there exist $\varepsilon > 0$ such that for infinitely many n

$$\left| \frac{S_n(\omega)}{n} \right| > \varepsilon.$$

Thus to prove the theorem we prove that for every $\varepsilon > 0$

$$\mathrm{P}\left(|S_n| > n\varepsilon \ \text{i.o.}\right) = 0.$$

This then shows (by looking at the complement of this event) that

$$\mathrm{P}(\mathcal{E}) = \mathrm{P}\left(\frac{S_n}{n} = 0\right) = 1.$$

We use the Borel-Cantelli lemma applied to the events

$$A_n = \left\{\omega \in \Omega : |S_n| \geq n\varepsilon\right\}.$$

To estimate $\mathrm{P}(A_n)$ we use the generalized Chebyshev inequality (2) with $p = 4$. Thus we must compute $E(S_n^4)$ which equals

$$E\left(\sum_{1\leq i,j,k,\ell\leq n} X_i X_j X_k X_\ell \right).$$

When the sums are multiplied out there will be terms of the form

$$E(X_i^3 X_j),\ E(X_i^2 X_j X_k),\ E(X_i X_j X_k X_\ell)$$

with i, j, k, ℓ all distinct. These terms are all equal to zero since $E(X_i) = 0$ and the random variables are independent (and the subscripts are distinct). (Recall $E(XY) = E(X)E(Y)$ when X and Y are independent.) Thus the nonzero terms in the above sum are

$$E(X_i^4) \ \text{and} \ E(X_i^2 X_j^2) = \left(E(X_i^2)\right)^2$$

There are n terms of the form $E(X_i^4)$. The number of terms of the form $E(X_i^2 X_j^2)$ is $3n(n-1)$.[1] Thus we have shown

$$E(S_n^4) = nE(X_1^4) + 3n(n-1)\sigma^4.$$

For n sufficiently large there exists a constant C such that

$$3\sigma^4 n^2 + \left(E(X_1^4) - 3\sigma^4\right) n \leq Cn^2.$$

(For n sufficiently large, C can be chosen to be $3\sigma^4 + 1$.) That is,

$$E(S_n^4) \leq Cn^2. \tag{5}$$

Then the Chebyshev inequality (2) (with $p = 4$) together with (5) gives

$$\mathrm{P}\left(|S_n| \geq n\varepsilon\right) \leq \frac{1}{(n\varepsilon)^4} E(S_n^4) \leq \frac{C}{\varepsilon^4 n^2}.$$

Thus

$$\sum_{n \geq n_0} \mathrm{P}\left(|S_n| \geq n\varepsilon\right) \leq \sum_{n \geq n_0} \frac{C}{\varepsilon^4 n^2} < \infty.$$

(Here n_0 is the first n so that the inequality (5) holds. Since we are neglecting a finite set of terms in the sum, this cannot affect the convergence or divergence of the infinite series.) Thus by the Borel-Cantelli lemma

$$\mathrm{P}\left(|S_n| \geq n\varepsilon \text{ i.o.}\right) = 0.$$

Since this holds for every $\varepsilon > 0$ we have proved the strong law of large numbers.

Remarks: Our proof assumed that the moments $E(X_i^4)$ and $E(X_i^2)$ are finite. It can be shown that the strong law of large numbers holds only under the assumption $E(|X_i|) < \infty$. Of course, we are still taking X_i to be independent with common distribution.

[1] There are $\binom{n}{2}$ ways to choose the indices i and j and once these are chosen there are 6 terms giving $X_i^2 X_j^2$.