Lecture 7: Sequences of independent random variables

1. Kolmogorov 0-1 law

Let $X_1, X_2, \ldots$ be a sequence of *independent* random variables on $(\Omega, \mathcal{F}, P)$ with associated tail σ-algebra $\mathcal{T}$ as defined in the last lecture. As mentioned earlier, many important events belong to $\mathcal{T}$. For example, one such class of events of interest in the context of the law of large numbers concerns the limiting behaviour of

$$\frac{X_1 + X_2 + \cdots + X_n}{n}.$$

Theorem 1.1. *(Kolmogorov 0-1 law.) Let $X_1, X_2, \ldots$ be a sequence of* independent *random variables with σ-algebra $\mathcal{T}$. Then for any event $E \in \mathcal{T}$, either $P(E) = 1$ or $P(E) = 0$.*

Proof. (Sketch) Let $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$ and $\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \ldots)$. Then by the independence of the X_k, $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent and since $\mathcal{T} \subset \mathcal{T}_n$, $\mathcal{F}_n$ and $\mathcal{T}$ are independent. Moreover $\bigcup_n \mathcal{F}_n$ and $\mathcal{T}$ are independent. Next let $\mathcal{F}_\infty = \sigma(X_1, X_2, \ldots)$. Because $\mathcal{F}_\infty = \sigma(\bigcup_n \mathcal{F}_n)$, $\mathcal{F}_\infty$ and $\mathcal{T}$ are independent. (This last point is not trivial and requires more rigorous treatment to arrive at a proper proof, as $\bigcup_n \mathcal{F}_n$ is *not* a σ-algebra in general.) But since $\mathcal{T} \subset \mathcal{F}_\infty$, $\mathcal{T}$ is independent of itself! Therefore, for $E \in \mathcal{T}$, $P(E) = P(E \bigcap E) = P(E)^2$ and the result follows. □

Exercise: Show that any $\mathcal{T}$-measureable random variable Z must be (almost surely) trivial (i.e. deterministic), in the sense that for some constant c, $P(Z = c) = 1$.

For this reason σ-algebras with the 0-1 property are called trivial.

Kolmogorov's 0-1 law shows that for independent sequences (X_n), either

$$\begin{aligned} P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} \text{ exists}\right) &= 0 \\ \text{or } P\left(\lim_{n\to\infty} \frac{X_1 + X_2 + \cdots + X_n}{n} \text{ exists}\right) &= 1 \end{aligned}$$

and moreover, if the limit does exist, it must be constant. There are many other similar types of 0-1 laws but although they say that for a certain event E, $P(E) = 0$ or 1, they don't say which and deciding whether $P(E) = 0$ or 1 is often a very difficult problem and many such problems remain open. In the case of

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

the question is settled by the strong law of large numbers.

2. Law of large numbers

First, a very useful preliminary result which involves a nice application of the dominated convergence theorem.

Lemma 2.1. *Let X be a random variable with $E(X) = 0$ and $E(|X|^n) < \infty$ for some $n \geq 1$. Then the following asymptotic expansion holds for the characteristic function ϕ of X:*

$$\phi(\theta) = \sum_{k=1}^{n} \frac{(i\theta)^k E(X^k)}{k!} + o(\theta^n) \quad \text{as } \theta \to 0.$$

Proof. For $n \geq 0$

$$R_n(x) = e^{ix} - \sum_{k=1}^{n} \frac{x^k}{k!}.$$

Then $R_0(x) = e^{ix} - 1$ so $|R_0(x)| \leq 2$. But also $R_0(x) = \int_0^x ie^{iy}\,dy$, so $|R_0(x)| \leq |x|$. Putting these together gives

$$|R_0(x)| \leq \min(2, |x|).$$

Next, note that for $n \geq 1$

$$R_n(x) = \int_0^x iR_{n-1}(y)\,dy$$

so by induction

$$|R_n(x)| \leq \min\left(\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\right).$$

Since $\phi(\theta) = E(e^{i\theta X})$, we have

$$\phi(\theta) = \sum_{k=1}^{n} \frac{(i\theta)^k E(X^k)}{k!} + E(R_n(\theta X)).$$

It remains to prove that $E(R_n(\theta X)) \sim o(\theta^n)$.

$$|E(R_n(\theta X))| \leq E(|R_n(\theta X)|) \leq \theta^n E\left[\min\left(\frac{2|X|^n}{n!}, \frac{|\theta X|^{n+1}}{(n+1)!}\right)\right].$$

The integrand inside $E(\cdot)$ on the right-hand side above is bounded above by $2|X|^n/n!$ which has finite expectation by assumption and tends to 0 as $\theta \to 0$. Hence by dominated convergence

$$\frac{|E(R_n(\theta X))|}{\theta^n} \leq E\left[\min\left(\frac{2|X|^n}{n!}, \frac{|\theta X|^{n+1}}{(n+1)!}\right)\right] \to 0$$

as $\theta \to 0$, or in other words, $E(R_n(\theta X)) \sim o(\theta^n)$. □

Recall from Proposition 2.4 that $X_n \to X$ in probability implies $X_n \to X$ in distribution. The following is a converse of this result in the case where the limit X is a constant.

Lemma 2.2. *Suppose that $X_n \to c$ in distribution, where c is a constant. Then $X_n \to c$ in probability.*

Proof. This is left as an exercise.

Theorem 2.3. *(Weak law of large numbers I.) Let $X_1, X_2, \ldots$ be independent and identically distributed with $\mu = E(X_1)$ exists and is finite. Let $S_n = X_1 + X_2 + \cdots + X_n$. Then*

$$\frac{S_n}{n} \to \mu \quad \textit{in distribution} \tag{2.1a}$$

$$\frac{S_n}{n} \to \mu \quad \textit{in probability.} \tag{2.1b}$$

Proof. We need only prove (2.1a) because (2.1b) follows from (2.1a) by Lemma 2.2. Let ϕ_X be the common characteristic function of the X_k and let ϕ_n be the characteristic function of S_n/n. Then

$$\phi_n(\theta) = \prod_{k=1}^{n} E(e^{i\theta X_k/n}) = \phi_X(\theta/n)^n.$$

As $n \to \infty$, Lemma 2.1 gives

$$\phi_X(\theta/n) = 1 + i\theta\mu/n + o(1/n)$$

so that

$$\phi_n(\theta) = (1 + i\theta\mu/n + o(1/n))^n \to e^{i\theta\mu}$$

as $n \to \infty$. But $e^{i\theta\mu}$ is the characteristic function of the constant μ. □

Theorem 2.4. *(Weak law of large numbers II.) Let $X_1, X_2, \dots$ and S_n be as before. Suppose that in addition $E(X_1^2) < \infty$. Then $S_n/n \to \mu$ in L^2 (and hence also in probability).*

Proof. This is left as an exercise.

Theorem 2.5. *(Strong law of large numbers I.) Let $X_1, X_2, \dots$ be independent with $\mu = E(X_k)$ for all k. Suppose in addition that for some constant C, $E(X_k^4) \leq C$ for all k. Then $S_n/n \to \mu$ almost surely.*

Proof. Without loss of generality, we can assume that $\mu = 0$ (otherwise replace X_k by $X_k - \mu$). By independence and the fact that $E(X_i) = 0$,

$$E(X_i X_j^3) = E(X_i X_j X_k^2) = E(X_i X_j X_k X_l) = 0$$

for distinct i, j, k, l. Hence

$$\begin{aligned} E(S_n^4) &= E((X_1 + X_2 + \cdots + X_n)^4) \\ &= E\left(\sum_{k=1}^n X_k^4 + 3\sum_{i=1}^n \sum_{j \neq i} X_i^2 X_j^2\right). \end{aligned}$$

By Jensen's inequality, $E(X_i^2)^2 \leq E((X_i^2)^2) = E(X_i^4) \leq C$. By independence $E(X_i^2 X_j^2) = E(X_i^2)E(X_j)^2 \leq C$ for $i \neq j$. Putting these together yields

$$E(S_n^4) \leq nC + 3n(n-1)C \leq SCn^2$$

hence

$$E\left(\sum_n (S_n/n)^4\right) \leq 3C \sum_n n^{-2} < \infty$$

which implies that $\sum_n (S_n/n)^4 < \infty$ almost surely and hence $S_n/n \to 0$ almost surely. □

Note that there is no assumption in Theorem 2.5 that the X_k are i.i.d. The assumption that $E(X_k^4) < \infty$ is made purely because it allows a much simpler proof - the strong law is true without this assumption. Indeed the best version of the strong law in the i.i.d case is the following:

Theorem 2.6. *(Strong law of large numbers II.) Let $X_1, X_2, \dots$ be i.i.d. with $\mu = E(X_1)$. Then $S_n/n \to \mu$ almost surely and in L^1*

The proof requires certain (backward) martingale techniques as well as making use of uniform integrability properties and Kolmogorov's 0-1 law

.

3. Central limit theorem

Theorem 3.1. *Let $X_1, X_2, \dots$ be i.i.d. with $\mu = E(X_1)$ and $\sigma^2 = Var(X_1) < \infty$ and let $S_n = X_1 + X_2 + \cdots + X_n$ and*

$$Z_n = \frac{S_n - \mu n}{\sqrt{\sigma^2 n}}.$$

Then as $n \to \infty$,

$$P(Z_n \le x) \to \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-y^2/2}\, dy.$$

(i.e. $Z_n \to N(0,1)$ in distribution.)

Proof. Assume without loss of generality that $\mu = 0$ (otherwise replace X_k by $X_k - \mu$). Let ϕ_{Z_n}, ϕ_{S_n} and ϕ_X denote the characteristic functions of the respective random variables. By Lemma 2.1, for small θ

$$\phi_X(\theta) = 1 - \frac{1}{2}\sigma^2\theta^2 + o(\theta^2).$$

Therefore

$$\begin{aligned} \phi_{Z_n}(\theta) &= \phi_{S_n}(\theta/\sqrt{\sigma^2 n}) = \phi_X(\theta/\sqrt{\sigma^2 n})^n \\ &= \left(1 - \frac{\theta^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-\theta^2/2} \end{aligned}$$

which is the characteristic function of $N(0,1)$. □

The central limit theorem is behind nearly all normal approximation results: for example, the normal approximations to the binomial and Poisson distributions are both applications of it.

One striking feature of the proof is that it involves only a very simple analysis of the asymptotic expansion of characteristic functions up to the 2nd order term. For this reason, the result lends itself to many possible variations and extensions.

Example 3.1. Here, we look at a situation where a large number of independent but non-i.i.d random variables are summed.

Let $E_1, E_2, \dots$ be a sequence of independent events with $P(E_n) = 1/n$. Let

$$S_n = I_{E_1} + I_{E_2} + \cdots + I_{E_n}.$$

Then

$$\begin{aligned} E(S_n) &= \sum_{k=1}^{n} \frac{1}{k} \sim \log n \\ \mathrm{Var}(S_n) &= \sum_{k=1}^{n} \frac{1}{k}\left(1 - \frac{1}{k}\right) \sim \log n. \end{aligned}$$

It is therefore natural to expect

$$Z_n = \frac{S_n - \log n}{\sqrt{\log n}}$$

to converge in distribution to $N(0,1)$. We have

$$\phi_{Z_n}(\theta) = e^{-i\theta\sqrt{\log n}}\phi_{S_n}(\theta/\sqrt{\log n})$$

where

$$\phi_{S_n}(t) = \prod_{k=1}^{n} \phi_{I_k}(t) = \prod_{k=1}^{n} \left(1 - \frac{1}{k} + \frac{1}{k}e^{it}\right).$$

Putting $t = \theta/\sqrt{n}$ and letting $n \to \infty$ (i.e. $t \to 0$),

$$\begin{aligned}
\log \phi_{Z_n}(\theta) &= -it\log n + \sum_{k=1}^{n} \log\left(1 + \frac{1}{k}(e^{it} - 1)\right) \\
&= -it\log n + \sum_{k=1}^{n} \log\left(1 + \frac{1}{k}\left(it - \frac{1}{2}t^2 + o(t^2)\right)\right) \\
&= -it\log n + \sum_{k=1}^{n} \left\{\frac{1}{k}\left(it - \frac{1}{2}t^2 + o(t^2)\right) + O(t^2/k^2)\right\} \\
&\sim -it\log n + \left(it - \frac{1}{2}t^2 + o(t^2)\right)\log n + O(t^2) \\
&\sim -\frac{1}{2}\theta^2 + o(1) \to -\frac{1}{2}\theta^2.
\end{aligned}$$