Convergence in distribution, characteristic functions

1 Convergence in distribution

Definition Let $X, X_1, \ldots, X_n, \ldots$ be random variables with values in $\mathbb{R}^k$ (or more generally in the same metric space) which are not necessarily defined on the same probability space. We say that (X_n) converges in distribution to X if for any *continuous bounded* function $f : \mathbb{R}^k \to \mathbb{R}$ we have

$$\mathbb{E}[f(X_n)] \underset{n\to\infty}{\longrightarrow} \mathbb{E}[f(X)].$$

WARNING. The notion of convergence in distribution only involves the **law** of random variables, which is not the case for a.s. convergence, in probability and in L^p. In particular, it is an abuse of language to say that the sequence of a.v. (X_n) converges in distribution to X, because the limiting a.v. X is not uniquely defined: only its law $\mathbb{P}_X$ is.

WARNING. For convergence in distribution, the random variables involved are not necessarily defined on the same probability space, which makes convergence in distribution very different from the other convergences seen so far.

2 Characteristic functions

Characteristic functions are very useful for studying the laws of random variables with values in $\mathbb{R}^k$ (with $k \geq 1$).

Definition If X is a random variable with values in $\mathbb{R}^k$, its characteristic function is the function ϕ_X defined by

$$\begin{array}{rccc} \phi_X : & \mathbb{R}^k & \longrightarrow & \mathcal{C} \\ & u & \longmapsto & \mathbb{E}\left[e^{i\langle u,X\rangle}\right] \end{array}$$

where $\langle u, v\rangle$ denotes the scalar product of two vectors of $\mathbb{R}^k$. The characteristic function is a continuous function (even uniformly continuous).

Characterizations of laws. Characteristic functions are useful for identifying laws and showing independence:

– if X and Y are two random variables with values in $\mathbb{R}^k$,

$$X \text{ and } Y \text{ have the same law} \quad \Longleftrightarrow \quad \phi_X(u) = \phi_Y(u) \text{ for every } u \in \mathbb{R}^k.$$

In practice, to identify the law of X, we can compute the characteristic function of X and try to recognize the characteristic function of a classical law.

More generally :

– if X has values in $\mathbb{R}^k$ and Y has values in $\mathbb{R}^m$,

$$X \text{ and } Y \text{ are independent} \quad \Longleftrightarrow \quad \phi_{(X,Y)}((u,v)) = \phi_X(u) \cdot \phi_Y(v) \quad \forall u \in \mathbb{R}^k, \forall v \in \mathbb{R}^m.$$

Here $\phi_{(X,Y)}((u,v))$ stands for $\mathbb{E}\left[e^{i\langle u,X\rangle + i\langle v,Y\rangle}\right]$.

Computation of moments. Under the assumptions of existence of moments, differentiating characteristic functions at 0 gives access to moments (see the proof of the Taylor expansion of the characteristic function just before the proof of the Central Limit Theorem).

3 In practice, how to show convergence in distribution?

Let $X, X_1, \ldots, X_n, \ldots$ be random variables with values in $\mathbb{R}^k$.

Functional approach. To show that (X_n) converges in distribution to X, show that $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ for any function f that belongs to one of the following classes:

- bounded continuous functions from $\mathbb{R}^k$ into $\mathbb{R}$ (this is the definition),
- continuous functions with compact support from $\mathbb{R}^k$ into $\mathbb{R}$ (restriction of test functions),
- bounded Lipschitz functions from $\mathbb{R}^k$ into $\mathbb{R}$ (Portemanteau theorem).

Approach with cumulative distribution functions. If k =1 (i.e. we are working with real-valued random variables), denoting by F_Y the cdf of a real random variable Y, then X_n converges in distribution to X **if and only if** $F_{X_n}(u) \to F_X(u)$ at any point $u \in \mathbb{R}$ **where** F_X **is continuous**.

WARNING. If the limiting random variable X is unknown, one can compute the limit $F(u)$ of $F_{X_n}(u)$ when $n \to \infty$ and check whether there exists a random variable X such that F_X is equal to F at any point where F_X is continuous (it is sometimes useful to use the fact that the points of discontinuity of a cdf are at most countable).

This approach with cdf's is often useful for random variables involving min or max in their definition.

Approach with characteristic functions. To show that (X_n) converges in distribution to X, one can show that for all $u \in \mathbb{R}^k$, we have $\mathbb{E}\left[e^{i\langle u,X_n\rangle}\right] \to \mathbb{E}\left[e^{i\langle u,X_n\rangle}\right]$.

This approach with charactersitic functions is often useful for random variables involving sums of independent random variables.

If the limiting random variable is unknown, we try to compute the limit $\psi(u)$ of $\mathbb{E}\left[e^{i\langle u,X_n\rangle}\right]$ when $n \to \infty$. It is often possible to recognize ψ as the characteristic function of a random variable X (and then X_n converges in distribution to X).

Approach by composition (continous mapping). If (X_n) converges in distribution to X and if f is a function almost surely continuous at X, then $f(X_n)$ converges in distribution to $f(X)$.

Approach by joint convergence. If X_n converges in distribution to X and Y_n converges in distribution to Y:

- if Y is a constant random variable (i.e. there exists c such that $\mathbb{P}(Y = c) = 1$), then $(X_n, Y_n) \to (X, Y)$ in distribution (Slutsky's theorem)
- if X_n and Y_n are independent for every $n \geq 1$, then (X_n, Y_n) converges in distribution to (X, Y) with X, Y independent.

WARNING. It is not true in general that if $X_n \to X$ in distribution and if $Y_n \to Y$ in distribution, then $(X_n, Y_n) \to (X, Y)$ in distribution, in contrast with almost sure convergence and convergence in probability!

Proving stronger convergence. If X_n converges to X almost surely, in probability or in L^p, then X_n converges in distribution to X.

Central Limit Theorem approach. Assume that the random variables $(X_n)_{n\geq 1}$ are independent real random variables with the same distribution and integrable squares. Set $m = \mathbb{E}[X_1]$ and $\sigma^2 = \mathrm{Var}(X_1)$, and assume that $\sigma^2 > 0$ (otherwise the random variables are constant). Then, setting $S_n = X_1 + \cdots + X_n$,

$$\frac{S_n - nm}{\sigma\sqrt{n}}$$

converges in distribution to a standard $\mathcal{N}(0, 1)$ Gaussian random variable.

It can be shown that this convergence does not occur in probability (see Exercise sheet).