Chapter 3
Laws of Large Numbers: The Basic Results

In this chapter we prove the “Law of Large Numbers”, LLN in short, for the two types of functionals introduced in (1.1.1). By this, we mean their convergence in probability. One should perhaps call these results “weak” laws of large numbers, but in our setting there is never a result like the “strong” law of large numbers, featuring almost sure convergence. Two important points should be mentioned: unlike in the usual LLN setting, the limit is usually not deterministic, but random; and, whenever possible, we consider functional convergence (as processes).

The first type of LLNs concerns raw sums, without normalization, and the results essentially do not depend on the discretization schemes, as soon as the discretization mesh goes to 0. The second type is about normalized sums of functions of normalized increments, and it requires the underlying process to be an Itô semimartingale and also the discretization scheme to be regular (irregular schemes in this context will be studied in Chap. 14, and are much more difficult to analyze).

We start with two preliminary sections: the first one is about “general” discretization schemes, to set up notation and a few simple properties. The second one studies semimartingales which have $p$-summable jumps, meaning that $\sum_{s \leq t} \| \Delta X_s \|^p$ is almost surely finite for all $t$: this is always true for $p \geq 2$, but it may fail when $0 \leq p < 2$.

3.1 Discretization Schemes

1) A discretization grid is a strictly increasing sequence of times, starting at 0 and with limit $+\infty$, and which in practice represents the times at which an underlying process is sampled. In most cases these times are non-random, and quite often regularly spaced. In some instances it is natural to assume that they are random, perhaps independent of the underlying process, or perhaps not.

For a given discretization grid, very little can be said. Things become interesting when we consider a discretization scheme, that is a sequence of discretization grids indexed by $n$, and such that the meshes of the grids go to 0 as $n \to \infty$. This notion has already appeared at some places in the previous chapter, for example in (2.1.8)
and (2.2.26), and we formalize it as follows. Below, when we speak of a “random” discretization scheme, we assume that the filtered space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\) is given.

**Definition 3.1.1**

a) A random discretization scheme is a sequence \(T = (T_n)_{n \geq 1}\) defined as follows: each \(T_n\) consists of a strictly increasing sequence \((T(n,i) : i \geq 0)\) of finite stopping times, with \(T(n,0) = 0\) and \(T(n,i) \to \infty\) as \(i \to \infty\), and also

\[
\forall t > 0, \quad \sup_{i \geq 1} \left( (T(n,i) \wedge t - T(n,i-1) \wedge t) \right) \to 0. \quad (3.1.1)
\]

b) A discretization scheme is as above, with all \(T(n,i)\) deterministic, and to emphasize this fact we usually write \(T(n,i) = t(n,i)\) with lower case letters.

c) The scheme is called regular if \(t(n,i) = i \Delta_n\) for a sequence \(\Delta_n\) of positive numbers going to 0 as \(n \to \infty\).

The condition (3.1.1) expresses the fact that the mesh goes to 0. With any (random) discretization scheme we associate the quantities (where \(t \geq 0\) and \(i \geq 1\)):

\[
\Delta(n,i) = T(n,i) - T(n,i-1), \quad N_n(t) = \sum_{i \geq 1} 1_{\{T(n,i) \leq t\}}, \quad T_n(t) = T(n, N_n(t)), \quad I(n,i) = (T(n,i-1), T(n,i)).
\]

In the regular case \(\Delta(n,i) = \Delta_n\) and \(N_n(t) = \lfloor t / \Delta_n \rfloor\). Random schemes \(T(n,i)\) and the associated notation \(N_n(t)\) have been already encountered in Sect. 2.2.4 of Chap. 2.

Regular schemes are the most common in practice, but it is also important to consider non-regular ones to deal with “missing data” and, more important, with cases where a process is observed at irregularly spaced times, as is often the case in finance.

Note that in many applications the time horizon \(T\) is fixed and observations occur before or at time \(T\). In this context, for each \(n\) we have finitely many \(T(n,i)\) only, all smaller than \(T\). However one can always add fictitious extra observation times after \(T\) so that we are in the framework described here.

2) We consider a \(d\)-dimensional process \(X = (X_t)_{t \geq 0}\) defined on the space \((\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})\), with components \(X^i\) for \(i = 1, \ldots, d\). Suppose that a random discretization scheme \(T = (T_n)\) is given, with \(T_n = (T(n,i) : i \geq 0)\). We will use the following notation:

\[
\Delta^n_i X = X_{T(n,i)} - X_{T(n,i-1)}.
\]

Note that this notation is relative to the discretization scheme \(T\), although this does not show explicitly. We will study various sums of functions of the above increments, with or without normalization. The most basic one, called non-normalized functional, is as follows. We have an arbitrary function \(f\) on \(\mathbb{R}^d\) (it may be real-valued, or \(\mathbb{R}^q\)-valued, in which case the following should be read component by