Chapter 1
Introduction

Discretization of stochastic processes indexed by the interval $[0,T]$ or by the half-line $[0,\infty)$ occurs very often. Historically it has been first used to deduce results on continuous-time processes from similar and often simpler results for discrete-time processes: for example Markov processes may be considered as limits of Markov chains, which are much simpler to analyze; or, stable processes as limits of random walks. This also applies to the theory of stochastic integration: the first constructions of stochastic integrals, by N. Wiener and K. Itô, were based on a Riemann-type approximation, which is a kind of discretization in time. More recently but still quite old, and a kind of archetype of what is done in this book, is the approximation of the quadratic variation process of a semimartingale by the approximate quadratic variation process: this result, due to P.A. Meyer [76] in its utmost generality, turns out to be one of the most useful results for applications.

Discretization of processes has become an increasingly popular tool in practical applications, for mainly (but not only) two reasons: one is the overwhelming extension of Monte-Carlo methods, which serve to compute numerically the expectations of a wide range of random variables which are often very complicated functions of a stochastic process: this is made available by the increasing power of computers. The second reason is related to statistics: although any stochastic process can only be observed at finitely many times, with modern techniques the frequency of observations increases steadily: in finance for example one observes and records prices every second, or even more frequently; in biology one measures electrical or chemical activity at an even higher frequency.

Let us be more specific, by describing a simple but fundamental example of some of the problems at hand. Suppose that we have a one-dimensional diffusion process X of the form

$$dX_t = a(X_t)\,dt + \sigma(X_t)\,dW_t, \qquad X_0 = x_0. \tag{1.0.1}$$

Here the initial value $x_0 \in \mathbb{R}$ is given, and W denotes a Brownian motion defined on some probability space, about which we do not care in this introduction. The drift and diffusion coefficients a and σ are nice enough, so the above equation has a unique solution.

J. Jacod, P. Protter, *Discretization of Processes*,
Stochastic Modelling and Applied Probability 67,
DOI 10.1007/978-3-642-24127-7_1,

Problem 1) We know a and σ, and we are interested in the law of the variable X_1. This law is usually not explicitly known, so to compute it, that is to compute the expected value $\mathbb{E}(f(X_1))$ for various test functions f, one may use a Monte-Carlo technique (other techniques based on PDEs are also available, especially in the one-dimensional case, but do not work so well in high dimensions). To implement this we simulate on a computer a number N of independent variables $X(j)_1$ having the law of X_1, and an approximation of $\mathbb{E}(f(X_1))$ is

$$Z_N = \frac{1}{N} \sum_{j=1}^{N} f\big(X(j)_1\big). \tag{1.0.2}$$

Indeed, by the law of large numbers the sequence Z_N converges almost surely to $\mathbb{E}(f(X_1))$ as $N \to \infty$, and moreover the central limit theorem tells us that, when f is for example bounded, the error made in replacing $\mathbb{E}(f(X_1))$ by Z_N is of order $1/\sqrt{N}$.

This presumes that one knows how to simulate X_1, which is about as scarce as the cases when $\mathbb{E}(f(X_1))$ can be explicitly computed. (More accurately some recent techniques due to A. Beskos, O. Papaspiliopoulos and G.O. Roberts, see [16] and [17] for example, allow to simulate X_1 exactly, but they require that σ does not vanish and, more important, that the dimension is 1; moreover, in contrast to what follows, they cannot be extended to equations driven by processes other than a Brownian motion.) Hence we have to rely on approximations, and the simplest way for this is to use an Euler scheme. That is, for any integer $n \geq 1$ we recursively define the approximation $X^n_{i/n}$ for $i = 1, \dots, n$, by setting

$$X^n_0 = x_0, \qquad X^n_{i/n} = X^n_{(i-1)/n} + \frac{1}{n}\, a\big(X^n_{(i-1)/n}\big) + \sigma\big(X^n_{(i-1)/n}\big)(W_{i/n} - W_{(i-1)/n}),$$

the increments of the Brownian motion being easily simulated. Other, more sophisticated, schemes can be used, but they all rely upon the same basic ideas.

Then in (1.0.2) we substitute the $X(j)_1$'s with N independent copies of the simulated variables X^n_1, giving rise to an average Z^n_N which now converges to $\mathbb{E}(f(X^n_1))$ for each given n. Therefore we need to assert how close $\mathbb{E}(f(X^n_1))$ and $\mathbb{E}(f(X_1))$ are, and this more or less amounts to estimating the difference $(X_1 - X^n_1)^2$. Some calculations show that this boils down to evaluating the difference

$$\sum_{i=1}^{n} g_n\big(\omega, (i-1)/n\big)\left((W_{i/n} - W_{(i-1)/n})^2 - \frac{1}{n}\right)$$

for suitable functions $g_n(\omega, t)$, where $\omega \mapsto g_n(\omega, t)$ is measurable with respect to the σ-field $\mathcal{F}^W_t$ of the past of W before time t. That is, we have to determine the behavior of "functionals" of the increments of W of the form above: do they converge when $n \to \infty$? And if so, what is the rate of convergence?

Problem 2) The setting is the same, that is we know a and σ, but we want to find the law of $Y = \int_0^1 h(X_s)\, ds$ for some known function h. Again, one can use a Monte-

Carlo technique, coupled with a preliminary Euler method: we set

$$Y^n = \frac{1}{n}\sum_{i=1}^{n} h\big(X^n_{i/n}\big),$$

where X^n is the Euler approximation introduced above. We can then simulate N independent versions $Y^n(1), \dots, Y^n(N)$ of the variable Y^n above, and

$$\frac{1}{N}\sum_{j=1}^{N} h\big(Y^n(j)\big)$$

is our approximation of $\mathbb{E}(h(Y))$. If X^n is a good approximation of X, then certainly Y^n is a good approximation of $\frac{1}{n}\sum_{i=1}^{n} h(X_{i/n})$, provided h satisfies some suitable smoothness assumptions. However we have an additional problem here, namely to evaluate the difference

$$\frac{1}{n}\sum_{i=1}^{n} h(X_{i/n}) - \int_0^1 h(X_s)\,ds.$$

The convergence to 0 of this difference is ensured by Riemann approximation, but the rate at which it takes place is not clear, in view of the fact that the paths of X are not smooth, albeit continuous. This is another discretization problem.

Problem 3) Suppose now that the functions a and σ are known, but depend on an additional parameter, say θ, so we have $a = a(x, \theta)$ and $\sigma = \sigma(x, \theta)$. We observe the process $X = X^\theta$, which now depends on θ, over [0, 1], and we want to infer θ. However, in any realistic situation we cannot really observe the whole path $t \mapsto X_t(\omega)$ for $t \in [0, 1]$, and we simply have "discrete" observations, say at times $0, \frac{1}{n}, \dots, \frac{n}{n}$, so we have $n + 1$ observations.

We are here in the classical setting of a parametric statistical problem. For any given n there is no way exactly to infer θ, unless a and σ have a very special form. But we may hope for good asymptotic estimators as $n \to \infty$. All estimation methods, and there are many, are based on the behavior of functionals of the form

$$\sum_{i=1}^{n} f_n\big(\theta, \omega, (i-1)/n, X_{i/n} - X_{(i-1)/n}\big) \tag{1.0.3}$$

for suitable functions $f_n(\theta, \omega, t, x)$, where again $\omega \mapsto f_n(\theta, \omega, t, x)$ is $\mathcal{F}^W_t$ measurable. The consistency of the estimators is deduced from the convergence of functionals as above, and rates of convergence are deduced from associated central limit theorems for those functionals.

Problem 4) Here the functions a and σ are unknown, and they may additionally depend on (ω, t), as for example $\sigma = \sigma(\omega, t, x)$. We observe X at the same discrete times $0, \frac{1}{n}, \dots, \frac{n}{n}$ as above. We want to infer some knowledge about the coefficients a and σ. As is well known, we usually can say nothing about a in this setting, but the convergence of the approximate quadratic variation mentioned before says that:

$$\sum_{i=1}^{[nt]} (X_{i/n} - X_{(i-1)/n})^2 \;\to\; \int_0^t \sigma(X_s)^2\,ds$$

(convergence in probability, for each t; here, $[nt]$ denotes the integer part of the real nt). This allows us in principle to determine asymptotically the function $t \mapsto \sigma(\omega, t, X_t(\omega))$ on $[0, 1]$, and under suitable assumptions we even have rates of convergence. Here again, everything hinges upon functionals as in the left side above. Note that here we have a statistical problem similar to Problem 3, except that we do not want to infer a parameter θ but a quantity which is fundamentally random: this occurs for example in finance, for the estimation of the so-called stochastic volatility.

Problem 5) A more basic problem is perhaps the following one, which deals directly with discretized processes. Namely, let us call an n-discretized process of X the process defined by $X_t^{(n)} = X_{[nt]/n}$. Then of course $X^{(n)} \to X$ pointwise in ω, locally uniformly in time when X is continuous and for the Skorokhod topology when X is right-continuous and with left limits. But, what is the rate of convergence?

The common feature of all the problems described above, as different as they may appear, is the need to consider the asymptotic behavior of functionals like (1.0.3). And, when the process X is discontinuous, many other problems about the jumps can also be solved by using functionals of the same type.

1.1 Content and Organization of the Book

In the whole book we consider a basic underlying d-dimensional process X, always a *semimartingale*. This process is sampled at discrete times, most of the time regularly spaced: that is, we have a mesh $\Delta_n > 0$ and we consider the increments

$$\Delta_i^n X \;=\; X_{i\Delta_n} - X_{(i-1)\Delta_n}$$

and two types of *functionals*, where f is a function on $\mathbb{R}^d$:

$$\begin{aligned} V^n(f, X)_t &= \sum_{i=1}^{[t/\Delta_n]} f\big(\Delta_i^n X\big) && \text{“\textit{non-normalized functional}”} \\ V'^n(f, X)_t &= \Delta_n \sum_{i=1}^{[t/\Delta_n]} f\big(\Delta_i^n X/\sqrt{\Delta_n}\big) && \text{“\textit{normalized functional}”.} \end{aligned} \tag{1.1.1}$$

The aim of this book is to provide a comprehensive treatment of the mathematical results about functionals of this form, when the mesh Δ_n goes to 0. We will not restrict ourselves to the simple case of (1.1.1), and will also consider more general (but similar) types of functionals:

- f may depend on k successive increments of X for $k \geq 2$.
- $f = f_n$ may depend on n, and also on k_n successive increments, with $k_n \to \infty$.
- $f = f(\omega, t, x)$ may be a function on $\Omega \times \mathbb{R}_+ \times \mathbb{R}^d$, so that $f(\Delta_i^n X)$ is replaced by $f(\omega, (i-1)\Delta_n, \Delta_i^n X)$ in the first formula (1.1.1), for example.
- The sampling times are not necessarily equally spaced.

Basically, there are two different levels of results:

Level 1: We have (under appropriate assumptions, of course, and sometimes after normalization) convergence of the functionals to a limiting process, say for example $V^n(f, X) \to V(f, X)$. This convergence typically takes place in probability, either for a fixed time t, or "functionally" for the local uniform (in time) topology, or for the Skorokhod topology. We call this type of convergence a *Law of Large Numbers*, or LLN.

Level 2: There is a "second order" type of results, which we qualify as *Central Limit Theorems*, or CLT. Namely, for a proper normalizing factor $u_n \to \infty$ the sequence $u_n(V^n(f, X) - V(f, X))$ for example converges to a limiting process. In this case, the convergence (for a given time t, or functionally as above) is typically in law, or more accurately "stably in law" (the definition of stable convergence in law is recalled in detail in Chap. 2).

In connection with the previous examples, it should be emphasized that, even though the mathematical results given below have some interest from a purely theoretical viewpoint, the main motivation is *practical*. This motivation is stressed by the fact that the last section of most chapters contains a brief account of possible applications. These applications have indeed been the reason for which all this theory has been developed.

As it is written, one can hardly consider this book as "applied". Nevertheless, we hope that the reader will get some feeling about the applications, through the last sections mentioned above. In particular, the problem of estimating the *volatility* is recurrent through the whole book, and appears in Chaps. 3, 5, 8, 9, 11, 13, 14 and 16.

Two last general comments are in order:

1. A special feature of this book is that it concentrates on the case where the underlying process X has a non-trivial continuous martingale part X^c, which is $X_t^c = \int_0^t \sigma(X_s)\, dW_s$ in the case of (1.0.1). All results are of course still true in the degenerate situation where the continuous martingale part vanishes identically, but most of them become "trivial", in the sense that the limiting processes are also vanishing. That is, in this degenerate situation one should employ other normalization, and use different techniques for the proofs.

2. We are concerned with the behavior of functionals like (1.1.1) as $\Delta_n \to 0$, but *not* as the time t goes to infinity. That is, we only consider the "finite horizon" case. When $t \to \infty$ the results for these functionals requires some ergodicity assumptions on the process X: the results, as well as the techniques needed for the proofs, are then fundamentally different.

Synopsis of the Book: Chapter 2 is devoted to recalling the basic necessary results about semimartingales and the various notions of convergence used later (Skorokhod topology, stable convergence in law, and a few useful convergence criteria). The rest of the book is divided into four main parts:

Part II: This part is about the "simple" functionals, as introduced in (1.1.1):

- Chapter 3 is devoted to the Laws of Large Numbers (first level).
- Chapter 4 contains the technical results needed for Central Limit Theorems. To avoid fastidious repetitions, these technical results are general enough to provide for the proofs of the CLTs for more general functionals than those of (1.1.1).
- Chapter 5 is about Central Limit Theorems (second level). For $V'^n(f, X)$ it requires few assumptions on the function f but quite a lot about the jumps of X, if any; for $V^n(f, X)$ it requires little of X, but (in, say, the one-dimensional case) it basically needs either $f(x) \sim x^2$ or $f(x)/|x|^3 \to 0$ as $x \to 0$.
- Chapter 6 gives another kind of Central Limit Theorems (in the extended sense used in this book) for $V^n(f, X)$, when $f(x) = x$: this is a case left out in the previous Chap. 5, but it is also important because $V^n(f, X)_t$ is then $X_t^{(\Delta_n)} - X_0$, where $X^{(\Delta_n)}$ is the "discretized process" $X_t^{(\Delta_n)} = X_{\Delta_n[t/\Delta_n]}$.

Part III: This part concerns various extensions of the Law of Large Numbers:

- In Chap. 7 the test function f is random, that is, it depends on (ω, t, x).
- In Chap. 8 the test function $f = f_n$ may depend on n and on k (fixed) or k_n (going to infinity) successive increments.
- In Chap. 9 the test function f is truncated at a level u_n, with u_n going to 0 as Δ_n does; that is, instead of $f(\Delta_i^n X)$ we consider $f(\Delta_i^n X) 1_{\{|\Delta_i^n X| \le u_n\}}$ or $f(\Delta_i^n X) 1_{\{|\Delta_i^n X| > u_n\}}$, for example. The function f can also depend on several successive increments.

Part IV: In this part we study the Central Limit Theorems associated with the extended LLNs of the previous part:

- Chapter 10 gives the CLTs associated with Chap. 7 (random test functions).
- Chapter 11 gives the CLTs associated with Chap. 8 when the test function depends on k successive increments.
- Chapter 12 gives the CLTs associated with Chap. 8 when the test function depends on k_n successive increments, with $k_n \to \infty$.
- Chapter 13 gives the CLTs associated with Chap. 9 (truncated test functions).

Part V: The last part is devoted to three problems which do not fall within the scope of the previous chapters, but are of interest for applications:

- In Chap. 14 we consider the situation where the discretization scheme is not regular. This is of fundamental importance for applications, but only very partial results are provided here, and only when the process X is continuous.
- In Chap. 15 we study some degenerate situations where the rate of convergence is not the standard $1/\sqrt{\Delta_n}$ one.
- In Chap. 16 we consider a situation motivated again by practical applications: we replace the process X by a "noisy" version, that is by $Z_t = X_t + \varepsilon_t$ where ε_t is a noise, not necessarily white but subject to some specifications. Then we examine how the functionals (based on the observations $Z_{i\Delta_n}$ instead of $X_{i\Delta_n}$) should be modified, in order to obtain limits which are basically the same as in the non-noisy case, and in particular do not depend on the noise.

1.2 When X is a Brownian Motion

Before proceeding to the main stream of the book, we give in some detail and with heuristic explanations, but without formal proofs, the simplest form of the results: we suppose that the one-dimensional process X is either a Brownian motion, or a Brownian motion with a drift, or a Brownian motion plus a drift plus a compound Poisson process.

Although elementary, these examples essentially show most qualitative features found later on, although of course the simple structure accounts for much simpler statements. So the remainder of this chapter may be skipped without harm, and its aim is to exhibit the class of results given in this book, and their variety, in an especially simple situation.

We start with the Brownian case, that is

$$X = \sigma W, \qquad \text{where } W \text{ is a Brownian motion and } \sigma > 0; \text{ we set } c = \sigma^2. \qquad (1.2.1)$$

We will also use, for any process Y, its "discretized" version at stage n:

$$Y_t^{(\Delta_n)} = Y_{\Delta_n [t/\Delta_n]}.$$

1.2.1 The Normalized Functionals $V'^n(f, X)$

Recalling (1.1.1), the functionals $V'^n(f, X)$ are easier than $V^n(f, X)$ to analyze. Indeed, the summands $f(\Delta_i^n X/\sqrt{\Delta_n})$ are not only i.i.d. as i varies, but they also have the same law as n varies. We let ρ_c be the centered Gaussian law $\mathcal{N}(0, c)$ and write $\rho_c(f) = \int f(x)\rho_c(dx)$ when the integral exists. Then, as soon as f is Borel and integrable, resp. square integrable, with respect to ρ_c, then $f(\Delta_i^n X/\sqrt{\Delta_n})$ has expectation $\rho_c(f)$ and variance $\rho_c(f^2) - \rho_c(f)^2$.

The ordinary Law of Large Numbers (LLN) and Central Limit Theorem (CLT) readily give us the following two convergence results:

$$\begin{aligned} &V'^n(f,X)_t \xrightarrow{\mathbb{P}} t\rho_c(f) \\ &\frac{1}{\sqrt{\Delta_n}}\big(V'^n(f,X)_t - t\rho_c(f)\big) \xrightarrow{\mathcal{L}} \mathcal{N}\big(0, t\big(\rho_c\big(f^2\big) - \rho_c(f)^2\big)\big), \end{aligned} \tag{1.2.2}$$

where $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{\mathcal{L}}$ stand for the convergence in probability and the convergence in law, respectively. This example shows why we have put the normalizing factor $1/\sqrt{\Delta_n}$ inside the function f.

The first subtle point we encounter, even in this basic case, is that, contrary to the usual LLN, we get convergence in probability but *not* almost surely in the first part of (1.2.2). The reason is as follows: let ζ_i be a sequence of i.i.d. variables with the same law as $f(X_1)$. The LLN implies that $Z_n = \frac{t}{[t/\Delta_n]}\sum_{i=1}^{[t/\Delta_n]}\zeta_i$ converges a.s. to $t\rho_c(f)$. Since $V'^n(f,X)_t$ has the same law as Z_n we deduce the convergence in probability in (1.2.2) because, for a deterministic limit, convergence in probability and convergence in law are equivalent. However the variables $V'^n(f,X)_t$ are connected one with the others in a way we do not really control when n varies, so we cannot conclude that $V'^n(f,X)_t \to t\rho_c(f)$ a.s.

(1.2.2) gives us the convergence for any time t, but we also have a "functional" convergence:

1) First, recall that a sequence g_n of nonnegative increasing functions on $\mathbb{R}_+$ converging pointwise to a *continuous* function g also converges locally uniformly; then, from the first part of (1.2.2) applied separately for the positive and negative parts f^+ and f^- of f and using a "subsequence principle" for the convergence in probability, we obtain

$$V'^n(f,X)_t \overset{\text{u.c.p.}}{\Longrightarrow} t\rho_c(f) \tag{1.2.3}$$

where $Z_t^n \overset{\text{u.c.p.}}{\Longrightarrow} Z_t$ means "convergence in probability, locally uniformly in time": that is, $\sup_{s\le t}|Z_s^n - Z_s| \xrightarrow{\mathbb{P}} 0$ for all t finite.

2) Next, if instead of the one-dimensional CLT we use the "functional CLT", or Donsker's Theorem, we obtain

$$\left(\frac{1}{\sqrt{\Delta_n}}\big(V'^n(f,X)_t - t\rho_c(f)\big)\right)_{t\ge 0} \overset{\mathcal{L}}{\Longrightarrow} \sqrt{\rho_c\big(f^2\big) - \rho_c(f)^2}\; B \tag{1.2.4}$$

where B is another standard Brownian motion, and $\overset{\mathcal{L}}{\Longrightarrow}$ stands for the convergence in law of processes (for the Skorokhod topology, see later for details on this topology, even though in this special case we could also use the "local uniform topology", since the limit is continuous).

In (1.2.4) we see a new Brownian motion B appear. What is its connection with the basic underlying Brownian motion W? To study that, one can try to prove the "joint convergence" of the processes on the left side of (1.2.4) together with W (or equivalently X) itself.

This is an easy task: consider the 2-dimensional process Z^n whose first component is W and the second component is the left side of (1.2.4). The discretized version of Z^n is $(Z^n)^{(\Delta_n)}_t = \sqrt{\Delta_n}\,\sum_{i=1}^{[t/\Delta_n]} \zeta^n_i$, where the ζ^n_i are 2-dimensional i.i.d. variables as i varies, with the same distribution as $(W_1, f(\sigma W_1) - \rho_c(f))$. Then the 2-dimensional version of Donsker's Theorem gives us that the pair of processes with components $W^{(\Delta_n)}$ and $\frac{1}{\sqrt{\Delta_n}}(\Delta_n V'^n(f,X)_t - t\rho_c(f))$ converges in law to a 2-dimensional Brownian motion with variance-covariance matrix at time 1 given by

$$\begin{pmatrix} 1 & \rho_c(g) \\ \rho_c(g) & \rho_c(f^2) - \rho_c(f)^2 \end{pmatrix}, \quad \text{where } g(x) = xf(x)/\sigma.$$

We write this as

$$\left(W^{(\Delta_n)}_t, \frac{1}{\sqrt{\Delta_n}}\big(V'^n(f,X)_t - t\rho_c(f)\big)\right)_{t\geq 0} \stackrel{\mathcal{L}}{\Longrightarrow} \big(W, aW + a'W'\big),$$
$$\text{where } a = \rho_c(g),\ a' = \big(\rho_c\big(f^2\big) - \rho_c(f)^2 - \rho_c(g)^2\big)^{1/2}, \tag{1.2.5}$$

where W' is a standard Brownian motion independent of W.

In (1.2.5) we could have used another symbol in place of W since what really matters is the joint law of the pair (W, W'). However for the first component, not only do we have convergence in law but pathwise convergence $W^{(\Delta_n)} \to W$. This explains why we use the notation W here, and in fact this results in a stronger form of convergence for the second component as well. This mode of convergence, called *stable convergence in law*, will be explained in detail in the next chapter.

Remark 1.2.1 We can even make $f = f_n$ depend on n, in such a way that f_n converges to some limit f fast enough. This is straightforward, and useful in some applications.

Remark 1.2.2 (1.2.5) is stated in a unified way, but there are really two—quite different—types of results here, according to the parity properties of f:

a) If f is an even function then $\rho_c(f) \neq 0$ in general, and $a = 0$. The limit in the CLT is $(W, a'W')$, with two independent components.

b) If f is an odd function then $\rho_c(f) = 0$ and $a \neq 0$ in general. The limit in the CLT has two dependent components. A special case is $f(x) = x$: then $a = \sigma$ and $a' = 0$, so the limit is $(W, X) = (W, \sigma W)$. This was to be anticipated, since in this case $V'^n(f, X) = \sqrt{\Delta_n}\,X^{(\Delta_n)}$, and the convergence in (1.2.5) takes place not only in law, but even in probability.

In general, the structure of the limit is thus much simpler in case (a), and most applications use this convergence for test functions f which are even.

1.2.2 The Non-normalized Functionals $V^n(f, X)$

We now turn to the processes $V^n(f,X)$. Their behavior results from the behavior of the processes $V'^n(f,X)$, but already in this simple case they show some distinctive features that will be encountered in more general situations. Basically, all increments $\Delta_i^n X$ become small as n increases, so the behavior of f near 0 is of the utmost importance, and in fact it conditions the normalization we have to use for the convergence.

To begin with, we consider power functions:

$$f_r(x) = |x|^r, \qquad \overline{f}_r(x) = |x|^r \operatorname{sign}(x),$$

where $r > 0$ and where $\operatorname{sign}(x)$ takes the value $+1$, 0 or -1, according to whether $x > 0$, $x = 0$ or $x < 0$. Note that

$$V^n(f_r, X) = \Delta_n^{r/2-1}\, V'^n(f_r, X)$$

and the same for $\overline{f}_r$. Moreover, if m_p denotes the p absolute moment of $\mathcal{N}(0,1)$, that is $m_p = \rho_1(f_p)$, and if $h_r(x) = x f_r(x)/\sigma$ and $\overline{h}_r(x) = x f_r'(x)/\sigma$ (recall $\sigma > 0$), we have

$$\begin{array}{lll} \rho_c(f_r) = m_r\sigma^r, & \rho_c\big(f_r^2\big) = m_{2r}\sigma^{2r}, & \rho_c(h_r) = 0, \\ \rho_c(\overline{f}_r) = 0, & \rho_c\big(\overline{f}_r^2\big) = m_{2r}\sigma^{2r}, & \rho_c(\overline{h}_r) = m_{r+1}\sigma^r. \end{array}$$

Hence we can rewrite (1.2.3) and (1.2.5) as follows, where W' denotes a standard Brownian motion independent of W (we single out the two cases f_r and $\overline{f}_r$, which correspond to cases (a) and (b) in Remark 1.2.2):

$$\begin{aligned} &\Delta_n^{1-r/2} V^n(f_r, X)_t \stackrel{\text{u.c.p.}}{\Longrightarrow} t m_r \sigma^r, \\ &\Big(W_t^{(\Delta_n)}, \frac{1}{\sqrt{\Delta_n}}\big(\Delta_n^{1-r/2} V^n(f_r, X)_t - t m_r\sigma^r\big)\Big)_{t\geq 0} \stackrel{\mathcal{L}}{\Longrightarrow} \big(W, \sigma^r\sqrt{m_{2r} - m_r^2}\, W'\big), \end{aligned} \tag{1.2.6}$$

$$\begin{aligned} &\Delta_n^{1-r/2} V^n(\overline{f}_r, X) \stackrel{\text{u.c.p.}}{\Longrightarrow} 0, \\ &\big(W^{(\Delta_n)}, \Delta_n^{1/2-r/2} V^n(\overline{f}_r, X)\big) \stackrel{\mathcal{L}}{\Longrightarrow} \big(W, \sigma^r\big(m_{r+1} W + \sqrt{m_{2r} - m_{r+1}^2}\, W'\big)\big). \end{aligned} \tag{1.2.7}$$

Note that the second statement implies the first one in these two properties.

Next, we consider functions f which vanish on a neighborhood of 0, say over some interval $[-\varepsilon, \varepsilon]$. Since X is continuous, we have $\sup_{i\leq[t/\Delta_n]} |\Delta_i^n X| \to 0$ pointwise for all t, and thus for each t there is a (random) integer A_t such that

$$n \geq A_t \;\Rightarrow\; V^n(f, X)_s = 0 \quad \forall s \leq t. \tag{1.2.8}$$

Finally, we consider "general" functions f, say Borel and with polynomial growth. If we combine (1.2.6) or (1.2.7) with (1.2.8), we see that the behavior of f

far from 0 does not matter at all, whereas the behavior near 0 is crucial for $V^n(f,X)$ to converge (with or without normalization). So it is no wonder that we get the following result:

$$\begin{array}{lll} f(x) \sim f_r(x) \text{ as } x \to 0 & \Rightarrow & \Delta_n^{1-r/2} V^n(f,X)_t \stackrel{\text{u.c.p.}}{\Longrightarrow} t m_r \sigma^r, \\ f(x) \sim \overline{f}_r(x) \text{ as } x \to 0 & \Rightarrow & \Delta_n^{1-r/2} V^n(f,X)_t \stackrel{\text{u.c.p.}}{\Longrightarrow} 0. \end{array} \tag{1.2.9}$$

These results are trivial consequences of the previous ones when f coincides with f_r or $\overline{f}_r$ on a neighborhood of 0, whereas if they are only equivalent one needs an (easy) additional argument. As for the CLT, we need f to coincide with f_r or $\overline{f}_r$ on a neighborhood of 0 ("close enough" would be sufficient, but how "close" is difficult to express, and "equivalent" is not enough). So we have, for any $\varepsilon > 0$ (recall that f is of polynomial growth):

$$\begin{aligned} & f(x) = f_r(x) \text{ if } |x| \le \varepsilon \quad \Rightarrow \\ & \left(W_t^{(\Delta_n)}, \frac{1}{\sqrt{\Delta_n}} \left(\Delta_n^{1-r/2} V^n(f,X)_t - t m_r \sigma^r \right) \right)_{t \ge 0} \stackrel{\mathcal{L}}{\Longrightarrow} \left(W, \sigma^r \sqrt{m_{2r} - m_r^2}\, W' \right), \end{aligned} \tag{1.2.10}$$

$$\begin{aligned} & f(x) = \overline{f}_r(x) \text{ if } |x| \le \varepsilon \quad \Rightarrow \\ & \left(W^{(\Delta_n)}, \Delta_n^{1/2-r/2} V^n(f,X) \right) \stackrel{\mathcal{L}}{\Longrightarrow} \left(W, \sigma^r \left(m_{r+1} W + \sqrt{m_{2r} - m_{r+1}^2}\, W' \right) \right) \end{aligned} \tag{1.2.11}$$

where again W' is a standard Brownian motion independent of W.

These results do not exhaust all possibilities for the convergence of $V^n(f,X)$. For example one can prove the following:

$$f(x) = |x|^r \log |x| \quad \Rightarrow \quad \frac{\Delta_n^{1-r/2}}{\log(1/\Delta_n)}\, V^n(f,X) \stackrel{\text{u.c.p.}}{\Longrightarrow} -\frac{1}{2}\, t m_r \sigma^r,$$

and a CLT is also available in this situation. Or, we could consider functions f which behave like x^r as $x \downarrow\downarrow 0$ and like $(-x)^{r'}$ as $x \uparrow\uparrow 0$. However, we essentially restrict our attention to functions behaving like f_r or $\overline{f}_r$ near the origin: for simplicity, and because more general functions do not really occur in applications, and also because the extension to processes X more general than the Brownian motion is not easy, or not available at all, for other functions.

Example 1.2.3 Convergence of the approximate quadratic variation. The functional

$$V^n(f_2,X)_t = \sum_{i=1}^{[t/\Delta_n]} \left(\Delta_i^n X \right)^2$$

is called the "approximate quadratic variation", and "realized quadratic variation" or "realized volatility" in the econometrics literature. It is of course well known, and a consequence of (1.2.6), that it converges in probability, locally uniformly in time, to the "true" quadratic variation which here is $\sigma^2 t$. Then (1.2.6) also gives the rate of convergence, namely that $\frac{1}{\sqrt{\Delta_n}}(V^n(f_2,X)_t - t\sigma^2)$ converges in law to $2\sigma^4 W'$; and we even have the joint convergence with X itself, and in the limit W' and X (or W) are independent.

1.3 When X is a Brownian Motion Plus Drift

Here we replace (1.2.1) by

$$X_t \;=\; bt + \sigma W_t, \qquad \text{where } \sigma \geq 0 \text{ and } b \neq 0.$$

1.3.1 The Normalized Functionals $V'^n(f,X)$

We first assume that $\sigma > 0$. The normalized increments $\Delta_i^n X/\sqrt{\Delta_n}$ are still i.i.d. when i varies, but now their laws depend on n. However, $\Delta_i^n X/\sqrt{\Delta_n} = Y_i^n + b\sqrt{\Delta_n}$ with Y_i^n being $\mathcal{N}(0,\sigma^2)$ distributed. Then, clearly enough, $f(\Delta_i^n X/\sqrt{\Delta_n})$ and $f(Y_i^n)$ are almost the same, at least when f is continuous, and it is no wonder that (1.2.3) remains valid (with the same limit) here, that is

$$V'^n(f,X)_t \stackrel{\text{u.c.p.}}{\Longrightarrow} t\rho_c(f).$$

Moreover, it turns out that the continuity of f is not even necessary for this, being Borel with some growth condition is again enough.

For the CLT, things are more complicated. When $X = \sigma W$ the CLT (1.2.4) boils down to the ordinary (functional) CLT, or Donsker's theorem, for the i.i.d. centered variables $\zeta_i^n = f(\Delta_i^n X/\sqrt{\Delta_n}) - \rho_c(f)$, but now while these variables are still i.i.d. when i varies, they are no longer centered, and their laws depend on n.

In fact ζ_i^n is distributed as $f(\sigma U + b\sqrt{\Delta_n}) - \rho_c(f)$, where U denotes an $\mathcal{N}(0,1)$ variable. Now, assume that f is C^1, with a derivative f' having at most polynomial growth. Then $f(\sigma U + b\sqrt{\Delta_n}) - f(\sigma U)$ is approximately equal to $f'(\sigma U)b\sqrt{\Delta_n}$. It follows that the variables ζ_i^n satisfy

$$\begin{aligned}
&\mathbb{E}(\zeta_i^n) = \sqrt{\Delta_n}\big(b\rho_c(f') + \mathrm{o}(1)\big)\\
&\mathbb{E}\big((\zeta_i^n)^2\big) = \rho_c(f^2) - \rho_c(f)^2 + \mathrm{o}(1)\\
&\mathbb{E}\big((\zeta_i^n)^4\big) = \mathrm{O}(1).
\end{aligned}$$

A CLT for triangular arrays of i.i.d. variables (see the next chapter) gives us

$$\left(\frac{1}{\sqrt{\Delta_n}}\big(V'^n(f,X)_t - t\rho_c(f)\big)\right)_{t\geq 0} \stackrel{\mathcal{L}}{\Longrightarrow} \Big(b\,\rho_c(f')\,t + \sqrt{\rho_c(f^2) - \rho_c(f)^2}\,B_t\Big)_{t\geq 0}. \tag{1.3.1}$$

Comparing with (1.2.4), we see an additional bias coming in here. Exactly as in (1.2.5), we also have a joint convergence (and stable convergence in law as well). With the notation a, a' and W' of (1.2.5), the expression is

$$\left(W_t^{(\Delta_n)}, \frac{1}{\sqrt{\Delta_n}}\big(bV'^n(f,X)_t - t\rho_c(f)\big)\right)_{t\geq 0} \stackrel{\mathcal{L}}{\Longrightarrow} \big(W_t, b\,\rho_c\big(f'\big)\,t + aW_t + a'W'_t\big)_{t\geq 0}. \tag{1.3.2}$$

Remark 1.3.1 We have the same dichotomy as in Remark 1.2.2. When f is an even function, the limit in (1.3.2) is simply $(W, a'W')$, with $a' = \sqrt{\rho_c(f^2) - \rho_c(f)^2}$, and in particular there is no bias (observe that f' is then odd, so $\rho_c(f') = 0$). When f is an odd function, we do have $\rho_c(f') \neq 0$ in general, and the bias does appear. A special case again is when $f(x) = x$, so $a = \sigma$ and $a' = 0$ and $\rho_c(f') = 1$, so the limit is (W, X) again, as it should be from the property $V'^n(f,X) = \sqrt{\Delta_n}\, X^{(\Delta_n)}$.

Suppose now $\sigma = 0$, that is $X_t = bt$. Then of course there is no more randomness, and all results ought to be elementary, but they are different from the previous ones. For example if f is differentiable at 0, we have

$$\frac{1}{\sqrt{\Delta_n}}\big(V'^n(f,X)_t - tf(0)\big) \;\to\; b\,f'(0)\,t,$$

locally uniformly in t. This can be considered as a special case of (1.3.1), with ρ_0 being the Dirac mass at 0. Note that the normalization $1/\sqrt{\Delta_n}$ inside the test function f is not really adapted to this situation, a more natural normalization would be $1/\Delta_n$.

1.3.2 The Non-normalized Functionals $V^n(f,X)$

For the functionals $V^n(f,X)$ we deduce the results from the previous subsection, exactly as for Brownian motion, at least when $\sigma > 0$. We have (1.2.8) when f vanishes on a neighborhood of 0, because this property holds for any continuous process X. Then we have (1.2.9), and also (1.2.10) when $r \geq 1$ (use Remark 1.3.1, the condition $r \geq 1$ ensures that f_r is C^1, except at 0 when $r = 1$). Only (1.2.11) needs to be modified, as follows, and again with $r \geq 1$:

$$\begin{aligned} f(x) = \overline{f}_r(x) \text{ if } |x| \leq \varepsilon \quad &\Rightarrow \quad \big(W^{(\Delta_n)}, \Delta_n^{1/2-r/2}V^n(f,X)\big) \\ &\stackrel{\mathcal{L}}{\Longrightarrow} \Big(W_t, r\,m_{r-1}bt + \sigma^r\Big(m_{r+1}W_t + \sqrt{m_{2r} - m_{r+1}^2}\;W'_t\Big)\Big)_{t\geq 0}. \end{aligned} \tag{1.3.3}$$

The case of the approximate quadratic variation is exactly as in Example 1.2.3.

Finally when $\sigma = 0$ we have $V^n(f,X)_t = f(b\Delta_n)\,\Delta_n[t/\Delta_n]$, and thus trivially

$$f \text{ differentiable at } 0 \;\Rightarrow\; \frac{1}{\Delta_n}\big(V^n(f,X)_t - f(0)\,t\big) \;\to\; b\,f'(0)\,t.$$

1.4 When X is a Brownian Motion Plus Drift Plus a Compound Poisson Process

In this section the structure of the process X is

$$X = Y + Z, \qquad Y_t = bt + \sigma W_t, \qquad Z_t = \sum_{n\geq 1} \Psi_n 1_{\{T_n \leq t\}}, \tag{1.4.1}$$

where $b \in \mathbb{R}$, $\sigma \geq 0$ and W is a Brownian motion, and Z is a compound Poisson process: that is, the times $T_1 < T_2 < \cdots$ are the arrival times of a Poisson process on $\mathbb{R}_+$, say with parameter $\lambda > 0$, and independent of W, and the Ψ_n's are i.i.d. variables with law F, say, and independent of everything else. For convenience, we put $T_0 = 0$ and $N_t = \sum_{n\geq 1} 1_{\{T_n \leq t\}}$ (which is the Poisson process mentioned above). To avoid trivial complications, we assume $\lambda > 0$ and $F(\{0\}) = 0$.

Before proceeding, we state an important remark:

Remark 1.4.1 The Poisson process N, hence X as well, has a.s. infinitely many jumps on the whole of $\mathbb{R}_+$. However, in practice we are usually interested in the behavior of our functionals on a given fixed finite interval $[0, T]$. Then the subset Ω_T of Ω on which N and X have no jump on this interval has a positive probability. On Ω_T we have $X_t = Y_t$ for all $t \leq T$, hence for example $V^n(f, X)_t = V^n(f, Y)_t$ for $t \leq T$ as well. Then, *in restriction to the set* Ω_T, $(V^n(f, X)_t)_{t\in[0,T]}$ behaves as $(V^n(f, Y)_t)_{t\in[0,T]}$, as described in the previous section: there is no problem for (1.2.9) since the convergence in probability is well defined in restriction to the subset Ω_T. For the convergence in law in (1.2.10) and (1.2.11) saying that it holds "in restriction to Ω_T" makes *a priori* no sense; however, as mentioned before, we do have also the stronger stable convergence in law, for which it makes sense to speak of the convergence in restriction to Ω_T: this is our first example of the importance of stable convergence, from a purely theoretical viewpoint.

The functionals $V'^n(f, X)$ are particularly ill-suited when X has jumps, because the normalized increment $\Delta_i^n X/\sqrt{\Delta_n}$ "explodes" as $n \to \infty$ if we take $i = i_n$ such that the interval $((i-1)\Delta_n, i\Delta_n]$ contains a jump. More precisely, $\Delta_i^n X/\sqrt{\Delta_n}$ is equivalent to $\Psi/\sqrt{\Delta_n}$ if Ψ is the size of the jump occurring in this interval. So general results for these functionals ask for very specific properties of f near infinity. Therefore, below we restrict our attention to $V^n(f, X)$.

1.4.1 The Law of Large Numbers

The key point now is that (1.2.8) fails. In the situation at hand, for any t there are at most finitely many q's with $T_q \leq t$, or equivalently $N_t < \infty$. The difference $V^n(f, X)_t - V^n(f, Y)_t$ is constant in t on each interval $[i\Delta_n, j\Delta_n)$ such that $(i\Delta_n, (j-1)\Delta_n]$ contains no jump. Moreover, let us denote by Ω_t^n the subset of

Ω on which $T_q - T_{q-1} \geq \Delta_n$ for all q such that $T_q \leq t$, and by $i(n,q)$ the unique (random) integer i such that $(i-1)\Delta_n < T_q \leq i\Delta_n$. Note that Ω_t^n tends to Ω as $n \to \infty$, for all t. Then if we set

$$\zeta_q^n = f\big(\Psi_q + \Delta_{i(n,q)}^n Y\big) - f\big(\Delta_{i(n,q)}^n Y\big), \qquad \overline{V}^n(f)_t = \sum_{q=1}^{N_t^{(\Delta_n)}} \zeta_q^n,$$

where Ψ_q is as in (1.4.1), we have

$$V^n(f,X)_s = V^n(f,Y)_s + \overline{V}^n(f)_s, \quad \forall s \leq t, \quad \text{on the set } \Omega_t^n. \tag{1.4.2}$$

Observe that $\Delta_{i(n,q)}^n Y \to 0$ for all q, because Y is continuous. Then as soon as f is continuous and vanishes at 0, we have $\zeta_q^n \to f(\Psi_q)$, hence $\zeta_q'^n \to f(\Psi_q)$ as well. Since $N_t^{(\Delta_n)} \leq N_t < \infty$ and since $\mathbb{P}(\Delta X_t \neq 0) = 0$ for any given t (because the Poisson process N has no fixed time of discontinuity), we deduce

$$\overline{V}^n(f)_t \xrightarrow{\text{a.s.}} \sum_{q=1}^{N_t} f(\Psi_q) = \sum_{s \leq t} f(\Delta X_s),$$

where $\Delta X_s = X_s - X_{s-}$ denotes the size of the jump of X at time s. This convergence is not local uniform in time. However, it holds for the Skorokhod topology (see Chap. 2 for details), and we write

$$\overline{V}^n(f)_t \overset{\text{a.s.}}{\Longrightarrow} \sum_{s \leq t} f(\Delta X_s). \tag{1.4.3}$$

When f vanishes on a neighborhood of 0 and is continuous, and if we combine the above with (1.2.8) for Y, with (1.4.2) and with $\Omega_t^n \to \Omega$, we see that (1.2.8) ought to be replaced by

$$V^n(f,X)_t \overset{\mathbb{P}}{\Longrightarrow} \sum_{s \leq t} f(\Delta X_s) \tag{1.4.4}$$

(convergence in probability for the Skorokhod topology).

The general case is also a combination of (1.4.2) and (1.4.4) with (1.2.9) applied to the process Y: it all depends on the behavior of the normalizing factor $\Delta_n^{1-r/2}$ in front of $V^n(f,Y)$, which ensures the convergence. If $r > 2$ the normalizing factor blows up, so $V^n(f,Y)$ goes to 0; when $r < 2$ then $V^n(f,Y)$ blows up (at least in the first case of (1.2.9)) and when $r = 2$ the functionals $V^n(f,Y)$ go to a limit, without normalization. Therefore we end up with the following LLNs (we always suppose f continuous, and is of polynomial growth in the last statement below; this means that $|f(x)| \leq K(1+|x|^p)$ for some constants K and p):

$$f(x) = \mathrm{o}\big(|x|^2\big) \text{ as } x \to 0 \ \Rightarrow \ V^n(f,X)_t \overset{\mathbb{P}}{\Longrightarrow} \sum_{s \leq t} f(\Delta X_s)$$

$$
\begin{aligned}
&f(x) \sim x^2 \text{ as } x \to 0 \;\Rightarrow\; V^n(f,X)_t \stackrel{\mathbb{P}}{\Longrightarrow} ct + \sum_{s\le t} f(\Delta X_s) \\
&f(x) \sim |x|^r \text{ as } x \to 0 \;\Rightarrow\; \begin{cases} V^n(f,X)_t \stackrel{\mathbb{P}}{\longrightarrow} +\infty & \text{if } r\in(0,2) \text{ and } t>0 \\ \Delta_n^{1-r/2} V^n(f,X)_t \stackrel{\text{u.c.p.}}{\Longrightarrow} t m_r \sigma^r. \end{cases}
\end{aligned}
\tag{1.4.5}
$$

Once more, this does not cover all possible test functions f.

1.4.2 The Central Limit Theorem

We have different CLTs associated with the different LLNs in (1.4.5). The results rely again upon the decomposition (1.4.2). In view of (1.4.2), and since we already have the CLT for $V^n(f,Y)$, we basically need to establish a CLT for $\overline{V}^n(f)$, for which the LLN takes the form (1.4.3). Due to some peculiarity of the Skorokhod topology, (1.4.3) does *not* imply that the difference $\overline{V}^n(f)_t - \sum_{s\le t} f(\Delta X_s)$ goes to 0 for this topology. However we do have Skorokhod convergence to 0 of the discretized processes, that is

$$
\widehat{V}^n(f)_t := \overline{V}^n(f)_t - \sum_{s\le \Delta_n[t/\Delta_n]} f(\Delta X_s) \stackrel{\text{a.s.}}{\Longrightarrow} 0,
$$

and we are looking for a CLT for these processes $\widehat{V}^n(f)$.

The key steps of the argument are as follows:

Step 1) We rewrite $\widehat{V}^n(f)_t$ as $\widehat{V}^n(f)_t = \sum_{q=1}^{N_t^{(\Delta_n)}} \eta_q^n$, where

$$
\eta_q^n = f\big(\Psi_q + \Delta_{i(n,q)}^n Y\big) - f(\Psi_q) - f\big(\Delta_{i(n,q)}^n Y\big).
$$

Assuming that f is C^1 with $f(0)=0$, and recalling $\Delta_{i(n,q)}^n Y \to 0$, a Taylor expansion gives

$$
\eta_q^n = \big(f'(\Psi_q) - f'(0)\big)\, \Delta_{i(n,q)}^n Y \,\big(1 + \mathrm{o}\big(\Delta_{i(n,q)}^n Y\big)\big).
$$

Since $\Delta_{i(n,q)}^n Y = b\Delta_n + \sigma\sqrt{\Delta_n}\; \Delta_{i(n,q)}^n W$, we deduce (this has to be justified, of course):

$$
\eta_q^n = \big(f'(\Psi_q) - f'(0)\big)\, \sigma\, \Delta_{i(n,q)}^n W + \mathrm{o}(\sqrt{\Delta_n}). \tag{1.4.6}
$$

Step 2) The jump times T_q and sizes Ψ_q, hence the random integers $i(n,q)$, are independent of W. Moreover one can check that the sequence $(\Delta_{i(n,q)}^n W)_{q\ge 1}$ is asymptotically independent of the process X as $n\to\infty$, whereas in restriction to the set Ω_t^n the variables $\Delta_{i(n,q)}^n W$ for $q=1,\dots,N_t$ are independent and $\mathcal{N}(0,\Delta_n)$.

Therefore, if $(\Phi_q)_{q\geq 1}$ denotes a sequence of independent $\mathcal{N}(0,1)$ variables, independent of the process X, we deduce the following joint convergence in law, as $n\to\infty$:

$$\left(X,\left(\frac{1}{\sqrt{\Delta_n}}\,\eta_q^n\right)_{q\geq 1}\right)\xrightarrow{\mathcal{L}}\left(X,\left(\left(f'(\Psi_q)-f'(0)\right)\sigma\,\Phi_q\right)_{q\geq 1}\right).$$

Step 3) The previous step and (1.4.6) give

$$\left(X,\frac{1}{\sqrt{\Delta_n}}\,\widehat{V}^n(f)\right)\stackrel{\mathcal{L}}{\Longrightarrow}\left(X,\widehat{V}(f)\right),\quad \text{where } \widehat{V}(f)_t=\sum_{q=1}^{N_t}\left(f'(\Psi_q)-f'(0)\right)\sigma\,\Phi_q \tag{1.4.7}$$

(we also have the stable convergence in law). This is the desired CLT for $\widehat{V}^n$.

Step 4) It remains to combine (1.4.7) with the result of the previous section, in the light of the decomposition (1.4.2). In order to stay simple, although keeping the variety of possible results, we only consider the absolute power functions $f_r(x)=|x|^r$. The results strongly depend on r, as did the LLNs (1.4.5) already, but here we have more cases.

For getting a clear picture of what happens, it is useful to rewrite (1.4.7) in a somewhat loose form (in particular the "equality" below is in law only), as follows, at least when $r>1$ so f_r is C^1 and $f_r'(0)=0$:

$$\begin{aligned}\overline{V}^n(f_r)_t &= A_t^n+B_t^n+\mathrm{o}(\sqrt{\Delta_n})\quad \text{"in law", where}\\ A_t^n &= \sum_{q=1}^{N_t^{(\Delta_n)}} f_r(\Psi_q),\qquad B_t^n=\sqrt{\Delta_n}\sum_{q=1}^{N_t} f_r'(\Psi_q)\,\sigma\,\Phi_q.\end{aligned}\tag{1.4.8}$$

Analogously, we can rewrite (1.2.10) for Y as follows:

$$\begin{aligned}V^n(f_r,Y)_t &= A_t'^n+B_t'^n+\mathrm{o}\left(\Delta_n^{r/2-1/2}\right)\quad \text{"in law", where}\\ A_t'^n &= \Delta_n^{r/2-1}m_r\sigma^r\,t,\quad B_t'^n=\Delta_n^{r/2-1/2}\sigma^r\sqrt{m_{2r}-m_r^2}\;W_t'.\end{aligned}$$

Note that $A_t^n\gg B_t^n$ (meaning $B_t^n/A_t^n\stackrel{\mathbb{P}}{\longrightarrow}0$ as $n\to\infty$), and $A_t'^n\gg B_t'^n$. Then we can single out seven (!) different cases. For simplicity we do not write the joint convergence with the process X itself, but this joint convergence nevertheless always holds.

1) If $r>3$**:** We have $B_t^n\gg A_t'^n$, hence

$$\frac{1}{\sqrt{\Delta_n}}\left(V^n(f_r,X)_t^n-\sum_{q=1}^{N_t^{(\Delta_n)}} f_r(\Psi_q)\right)\stackrel{\mathcal{L}}{\Longrightarrow}\sum_{q=1}^{N_t} f_r'(\Psi_q)\,\sigma\,\Phi_q.$$

2) If $r = 3$**:** Both terms B_t^n and $A_t'^n$ are of the same order of magnitude, hence

$$\frac{1}{\sqrt{\Delta_n}}\left(V^n(f_3, X)_t^n - \sum_{q=1}^{N_t^{(\Delta n)}} f_3(\Psi_q)\right) \stackrel{\mathcal{L}}{\Longrightarrow} m_3\sigma^3 t + \sum_{q=1}^{N_t} f_3'(\Psi_q)\,\sigma\,\Phi_q.$$

3) If $2 < r < 3$**:** We have $A_t^n \gg A_t'^n \gg B_t^n$. Then we do *not* have a proper CLT here, but the following two properties:

$$\frac{1}{\Delta_n^{r/2-1}}\left(V^n(f_r, X)_t^n - \sum_{q=1}^{N_t^{(\Delta n)}} f_r(\Psi_q)\right) \stackrel{\text{u.c.p.}}{\Longrightarrow} m_r\sigma^r\, t,$$

$$\frac{1}{\sqrt{\Delta_n}}\left(V^n(f_r, X)_t^n - \sum_{q=1}^{N_t^{(\Delta n)}} f_r(\Psi_q) - \Delta_n^{r/2-1} m_r\sigma^r\, t\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{q=1}^{N_t} f_r'(\Psi_q)\,\sigma\,\Phi_q.$$

4) If $r = 2$**:** Both terms A_t^n and $A_t'^n$, resp. B_t^n and $B_t'^n$, are of the same order of magnitude, and one can show that

$$\frac{1}{\sqrt{\Delta_n}}\left(V^n(f_2, X)_t^n - \sigma^2 t + \sum_{q=1}^{N_t^{(\Delta n)}} (\Psi_q)^2\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sqrt{2}\,\sigma^2\, W_t' + 2\sum_{q=1}^{N_t} \Psi_q\,\sigma\,\Phi_q$$

(recall $m_2 = 1$ and $m_4 = 3$ and $f_2'(x) = 2x$). Here W' is a Brownian motion independent of X, and also of the sequence (Φ_q). Note that, if we replace t by $\Delta_n[t/\Delta_n]$ in the left side above, which does not affect the convergence, this left side is the difference between the approximate quadratic variation and the discretized true quadratic variation.

5) If $1 < r < 2$**:** We have $A_t'^n \gg A_t^n \gg B_t'^n \gg B_t^n$. Then as in Case 3 we have two results:

$$\frac{1}{\Delta_n^{1-r/2}}\left(\Delta_n^{1-r/2} V^n(f_r, X)_t^n - m_r\sigma^r\, t\right) \stackrel{\mathbb{P}}{\Longrightarrow} \sum_{q=1}^{N_t} f_r(\Psi_q),$$

$$\frac{1}{\Delta_n^{r/2-1/2}}\left(V^n(f_r, X)_t^n - \Delta_n^{r/2-1} m_r\sigma^r\, t - \sum_{q=1}^{N_t^{(\Delta n)}} f_r(\Psi_q)\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sigma^r\sqrt{m_{2r} - m_r^2}\; W_t'.$$

6) If $r = 1$**:** The function f_1 is not differentiable at 0, but one can show that $\overline{V}^n(f_1)$ has a decomposition (1.4.8) with the same A_t^n and with a B_t^n satisfying $A_t^n \gg B_t^n$. Now, A_t^n and $B_t'^n$ have the same order of magnitude, so we get

$$\frac{1}{\sqrt{\Delta_n}}\left(\sqrt{\Delta_n}\; V^n(f_1, X)_t^n - m_1\sigma\, t\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sum_{q=1}^{N_t} |\Psi_q| + |\sigma|\sqrt{1 - m_1^2}\; W_t'.$$

7) If $0 < r < 1$**:** Again the function f_r is not differentiable at 0, but obviously $\overline{V}^n(f_r)_t$ stays bounded in probability. Then we have:

$$\frac{1}{\sqrt{\Delta_n}}\left(\Delta_n^{1-r/2}\, V^n(f_r, X)_t^n - m_r\sigma^r\, t\right) \stackrel{\mathcal{L}}{\Longrightarrow} \sigma^r\sqrt{m_{2r} - m_r^2}\, W_t'.$$

The jumps have disappeared from the picture in this case, which is as in (1.2.10).

From this brief description, we are able to conclude a moral that pervades the theory: including processes with jumps complicate matters more than one might naively suspect.

http://www.springer.com/978-3-642-24126-0

Discretization of Processes
Jacod, J.; Protter, P.E.
2012, XVI, 596 p., Hardcover
ISBN: 978-3-642-24126-0