Limit theorems and inference for high frequency financial data

Mark Podolskij

We present a detailed overview on probabilistic and statistical methods for high frequency data that typically appear in finance. Nowadays price processes may be observed at ultra high frequencies, such as one second or higher, which motivated researchers to come up with probabilistic tools to analyze the fine structure of price models using *infill* asymptotics. Mathematically speaking, we consider equidistant observations $X_{i\Delta_n}$, $i = 0, \ldots, [T/\Delta_n]$, of an asset price X on a *fixed* time interval $[0, T]$, and assume that $\Delta_n \to 0$. Due to classical no arbitrage conditions, the process X is usually assumed to be a (Itô) semimartingale of the form

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t \sigma_s dW_s + \text{Jumps}_t, \qquad t \in [0, T],$$

where $(a_s)_{s\in[0,T]}$ is a drift process, $(\sigma_s)_{s\in[0,T]}$ is the volatility, W denotes a Brownian motion and the jump part can exhibit finitely or infinitely many jumps on compact intervals. For the purpose of modeling and forecasting one would like to infer as much as possible information from discrete observations $X_{i\Delta_n}$. Some of the most important estimation and testing problems are:

(i) How to estimate the quadratic variation $[X]_t = \int_0^t \sigma_s^2 ds + \sum_{s\leq t} |X_s - X_{s-}|^2$?

(ii) Does the unobserved path $(X_s(\omega))_{s\in[0,T]}$ exhibit jumps? Does $(X_s(\omega))_{s\in[0,T]}$ contain a Brownian part?

(iii) How can volatility be separated from jumps?

Quite surprisingly, those questions can be answered in a (almost) fully non-parametric way, i.e. only very mild assumptions on the random processes $(a_s)_{s\in[0,T]}$ and $(\sigma_s)_{s\in[0,T]}$ are required to provide consistent estimation and test procedures. Almost all statistical methods are based on the functionals of the type

$$V(X, f)_t^n = \Delta_n \sum_{i=1}^{[t/\Delta_n]-k+1} f\Big(\Delta_i^n X/\sqrt{\Delta_n}, \ldots, \Delta_{i+k-1}^n X/\sqrt{\Delta_n}\Big),$$

$\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$, where $f : \mathbb{R}^k \to \mathbb{R}$ is a smooth function. In this course we determine the asymptotic behaviour of the partial sums process $V(X, f)_t^n$ (including law of large numbers and mixed normal limit theorems) and discuss various applications including the afore-mentioned problems. We will face some non-classical (in particular, non-ergodic) limit theorems and learn how to use them for practical questions. If the time allows we will also consider some non-semimartingale models used in finance and discuss the statistical inference in those frameworks.