Stochastic Machine Learning
Chapter 11 - Financial Mathematics and Deep Learning

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2024

A very short introduction to Financial mathematics

In this lecture we will shortly visit Financial mathematics.

- For all who read and understand German, I uploaded my scriptum together with the first introductory lecture.
- There is an excellent book available: Föllmer and Schied (2011).
- We are only interested in Chapters 5.1, 5.2, 5.3, 5.5 and later in chapter 8.1 and 8.2

The multi-period market model

- We consider 1 bank account and d stocks, making up $d+1$ assets.
- They are modelled through their price processes $S^0, \dots, S^d$ given by
$$S^i = (S^i_0, S^i_1, \dots, S^i_T)$$
- Our time horizon is therefore $\mathbb{T} = \{0, \dots, T\}$. $S^0 > 0$ is the bank account. We can not loose all money here.
- Trading is done as follows: I buy today, say at time $t-1$ a number of shares, say $\bar{H}$. Tomorrow, at t I sell these shares. My gain (or loss) is obviously
$$\sum_{i=0}^{d} \bar{H}^i (S^i_t - S^i_{t-1}).$$
- We now make a convention: we denote $\bar{H}$ as $\bar{H}_t$. This makes notation easy and simple. However, $\bar{H}_t$ is already known at $t-1$! Such processes are called predictable.
- This is made precise via a filtration $\mathbb{F}$. $\mathbb{F}$ is an increasing family of σ-algebras $(\mathscr{F}_t)_{t=0,\dots,T}$. A process X is called adapted if $X_t \in \mathscr{F}_t$ and predictable if $X_t \in \mathscr{F}_{t-1}$.

Definition

A trading strategy H is a predictable, $d+1$-dimensional process. It is called **self-financing** if

$$\bar{H}_t \cdot \bar{S}_t = \bar{H}_{t+1} \cdot \bar{S}_t, \quad t = 1, \dots, T-1.$$

Self-financing simply means that by selling and buying at each time t (and switching from $\bar{H}_t$ to $\bar{H}_{t+1}$) we do not gain or loose money. We can only buy shares for as much money as we have.

But of course it is possible to borrow money (through the bank account S^0). Often we will have

$$S_t^0 = \prod_{s=1}^{t} (1 + r_s)$$

with some interest rate r.

Lemma

For a self-financing trading strategy $\bar{H}$ and $t \geq 1$

$$\bar{H}_t \cdot \bar{S}_t = \bar{H}_1 \cdot \bar{S}_0 + \sum_{k=1}^{t} \bar{H}_k \cdot \Delta \bar{S}_k.$$

This is easy to show. Intuitively, the wealth at time t consists of initial wealth plus gains from trade.

Discounted prices

- A very nice trick is to consider discounted prices (note that 1 EUR tomorrow is of course different in value compared to 1 EUR today). We define

$$X_t^i := \frac{S_t^i}{S_t^0}, \quad t = 0, \dots, T, \quad i = 0, \dots, d.$$

  Then $X^0 \equiv 1$ and the representations simplify. In particular, if $\bar{H}$ is self-financing it is already defined by specifying $H = (H^1, \dots, H^d)$!

- We introduce the discounted wealth process $V = V^{\bar{H}}$

$$V_t := \bar{H}_t \cdot \bar{X}_t, \quad t = 1, \dots, T,$$

  $V_0 := \bar{H}_1 \cdot \bar{X}_0$. and the discounted gains process $G = G^{\bar{H}}$ by

$$G_t := \sum_{k=1}^{t} H_k \cdot \Delta X_k, \quad t = 1, \dots, T$$

  with $G_0 = 0$.

Proposition

For a trading strategy $\bar{H}$ t.f.a.e.

1. $\bar{H}$ is self-financing
2. $\bar{H}_t \cdot \bar{X}_t = \bar{H}_{t+1} \cdot \bar{X}_t, \quad t = 1, \dots, T-1,$
3. $V_t = V_0 + G_t \quad$ for $0 \leqslant t \leqslant T$.

The central concept is **arbitrage**. It is a risk-less gain through trading.

Definition

An arbitrage is a self-financing trading strategy H, s.t.

1. $V_0 \leqslant 0,$
2. $V_T \geqslant 0$ and
3. $P(V_T > 0) > 0.$

A market without arbitrages is called arbitrage-free.

- We can now show that every market is free of arbitrage, if and only if every single-period market (S_t, S_{t+1}) is free of arbtirage (a key result).
- Note that Hans Föllmer and Alexander Schied require positivity in their book for this step, although it is not really necessary.
- We are able to classify arbitrage-free marktes through martingales.

Definition

A stochastic process M is a Q-**martingale**, if

1. M is adapted
2. $E_Q[|M_t|] < \infty$ for $t = 0, \dots, T$,
3. $M_s = E_Q[M_t|\mathscr{F}_s]$ for $0 \leqslant s \leqslant t \leqslant T$.

We call two measures P and Q equivalent ($P \sim Q$) if for all $F \in \mathscr{F}$

$$P(F) = 0 \Leftrightarrow Q(F) = 0.$$

We call a measure Q a martingale measure if X is a martingale under Q.

The first fundamental theorem

Theorem

A financial market is free of arbitrage, if and only if there exist an equivalent martingale measure

European claims

- We are most interested in derivatives like calls and puts. A call on a stock S^i offers the payoff
$$(S_T - K)^+$$
at maturity T when the so-called strike is K.
- More generally, we call any $\mathscr{F}_T$-measurable random variable C an European contingent claim.
- How can we define arbitrage-free prices for C? We choose an equivalent martingale measure Q and price the claim via
$$C_t = E_Q[C_T|\mathscr{F}_t].$$
- Then, the price process (S, C) is again a Q-martingale and the market is still free of arbtirage !
- Please check the Black-Scholes formel in the scriptum or in the book.

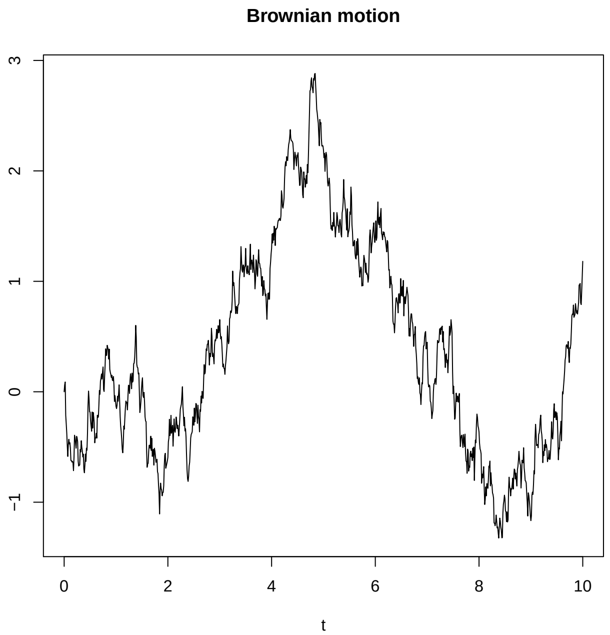
The Black-Scholes model

- We will first define a Brownian motion: this is a continuous process B with independent and stationary (and, by the central limit theorem) normally distributed increments, starting in 0.
- Hence, $B_t - B_s$ is independent from $B_s = B_s - 0$.
- We call B **standard** when $B_t \sim \mathcal{N}(0, t)$. We say B has volatility σ if $B_t \sim \mathcal{N}(0, \sigma^2 t)$.
- Intuitively, we can discretize B and see it as a limit of the discretization

$$B_{t_n} = \sum_{i=1}^{n} \sqrt{t_i - t_{i-1}}\, \xi_i,$$

with some time-points $t_0 < t_1 < \cdots$ and (ξ_i) i.i.d. $\mathcal{N}(0, 1)$.

Simulation



- The Black-Scholes model is a geometric Brownian motion.
- It is often described through a stochastic differential equation

$$dS_t = S_t \mu dt + S_t \sigma dB_t,$$

  with initial value $S_0 = 0$.
- Note that there is no differential - B is not differentiable. It is merely an abbreviation for the integral equatiion

$$S_t = S_0 + \int_0^t S_s \mu ds + \int_0^t S_s \sigma dB_s.$$

- The first integral is a classical integral. The second integral is a **stochastic integral**. It is obtained as an appropriate limit of the elementary sums

$$\sum_{i=1}^{n} S_{t_{i-1}} \sigma (B_{t_i} - B_{t_{i-1}}).$$

- An important consequence is that a function $f \in C^2$ of Brownian motion can be represented via the stochastic integral. This is the important Ito-formula.

$$f(B_t) = f(B_0) + \int_0^t f'(B_s) dB_s + \frac{1}{2} \int_0^t f''(B_s) ds.$$

- With this formula, we can show that the solution of the above SDE is

$$S_t = S_0 \exp\Big(\sigma B_t + \frac{(\mu - \sigma^2)t}{2}\Big).$$

- You have already looked at the Black-Scholes formula. It says that

$$E[(S_t - K)^+] = S_0 \Phi(d_1) - K \Phi(d_2),$$

where Φ is the cdf of a standard normal distribution and

$$d_{1/2} = \frac{\log(\frac{S}{K}) \pm \frac{\sigma^2 t}{2}}{\sqrt{\sigma^2 t}}$$

(we have assumed zero interest rate here).

- Our central question will be: how to apply deep learning in this setting ?

Deep Hedging

- Deep hedging is one of the fascinating success story of AI in Finance
- To achieve big data, one simulates from a model and therefore is exposed to **model risk**
- Our aim is to raise this problem and provide a solution in the realm of (generalized) affine processes
- This is a parametrized Markov process and clearly the parameters are not known but need to be estimated
- How can this model risk be incorporated in a concise way ?
- We will propose a solution in the realm of **Knightian uncertainy** and provide a **Monte-Carlo scheme** for simulating processes under uncertainty with an application to deep hedging.

Robust Deep Hedging

- Now to the deep hedging: Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a non-constant function, called **activation function**.
- A (feed-forward) neural network with input dimension $d_{\text{in}} \in \mathbb{N}$, output dimension $d_{\text{out}} \in \mathbb{N}$, $l \in \mathbb{N}$ layers, and activation function φ is a function of the form

  $$\begin{aligned} \mathbb{R}^{d_{\text{in}}} &\to \mathbb{R}^{d_{\text{out}}} \\ x &\mapsto A_l \circ \varphi \circ A_{l-1} \circ \cdots \circ \varphi \circ A_0(x), \end{aligned}$$

  where $(A_i)_{i=0,\dots,l}$ are affine functions $A_i : \mathbb{R}^{h_i} \to \mathbb{R}^{h_{i+1}}$, and where the activation function is applied component-wise.

- A **derivative** is given by its payoff Φ_T (which also could be path-dependent)
- Our aim is to determine hedging strategies $(h_t)_{0\le t\le T}$ and cash positions $d \in \mathbb{R}$ such that the quadratic error is minimized

$$\min_{(h_t)_{0\le t\le T}, d\in\mathbb{R}} \mathbb{E}^P\left[\left(d + \int_0^T h_t \; dX_t - \Phi_T\right)^2\right] \tag{1}$$

  for all $P \in \mathcal{A}(0, x_0, \Theta)$.
- This formulation is a consequence of the considered model ambiguity, under which every measure from $\mathcal{A}(0, x_0, \Theta)$ is taken into account.
- We aim at a suitable modification of the deep hedging in **buehler2019deep**.

Numerical Experiments

- To gather some experience we provide a number of numerical experiments in the paper.
- Consider a nonlinear generalized affine process with parameters specified through

$$\begin{aligned} x_0 &= 10 \\ a_0 &\in [0.3, 0.7], \quad a_1 \in [0.4, 0.6], \\ b_0 &\in [-0.2, 0.2], \quad b_1 \in [-0.1, 0.1], \\ \gamma &\in [0.5, 1.5]. \end{aligned} \tag{2}$$

Algorithm 1: Computation of Optimal Hedging Strategies

for iter $= 1, \ldots, N_{\text{iter}}$ **do**

  for $b = 1, \ldots, B$ **do**

    Generate paths of the generalized affine process:

    $X_0^b := x_0$, $\Delta t_i := t_{i+1} - t_i$

    for $i = 0, \ldots, n-1$ **do**

      Generate $\Delta W_i \sim N(0, \Delta t_i)$;

      Generate $\gamma^{(i)} \sim U\left([\underline{\gamma}, \overline{\gamma}]\right)$, $a_0^{(i)} \sim U\left([\underline{a_0}, \overline{a_0}]\right)$,
      $a_1^{(i)} \sim U\left([\underline{a_1}, \overline{a_1}]\right)$,
      $b_0^{(i)} \sim U\left([\underline{b_0}, \overline{b_0}]\right)$, $b_1^{(i)} \sim U\left([\underline{b_1}, \overline{b_1}]\right)$;

      set $X_{i+1}^b := X_i^b + (b_0^{(i)} + b_1^{(i)} X_i^b)\Delta t_i + (a_0^{(i)} + a_1^{(i)} X_i^+)^{\gamma^{(i)}} \Delta W_i$

    end

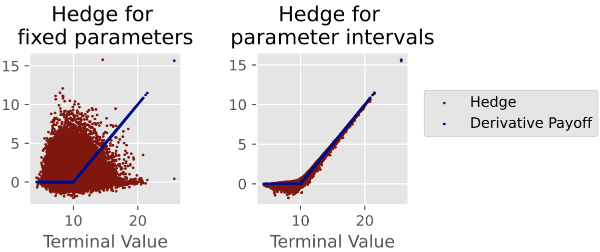
  end

  Apply stochastic gradient descent to minimize the loss

$$\sum_{b=1}^{B} \left(d + \sum_{i=0}^{n-1} h(t_i, X_i^b)(X_{i+1}^b - X_i^b) - \Phi\left((X_i^b)_{i=1,\ldots,n} \right) \right)^2$$

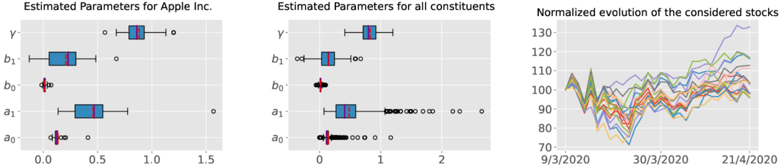
  w.r.t. the parameters of h and w.r.t. d

end



- Hedging of an at-the-money call option.
- left: The optimal hedge for fixed parameters $a_0 = 0.5$, $a_1 = 0.5$, $b_0 = 0$, $b_1 = 0$, $\gamma = 1$ (in the middle of the intervals) evaluated on $50,000$ paths.
- right: The robust deep hedge evaluated on $50,000$ paths.

Real-world data



- Left: The maximum-likelihood-estimated parameters of a generalized affine process when assuming the stock of Apple Inc. is modelled through a generalized affine process. The estimations are performed every 100 days for a lookback window of 250 days.
- Middle: The maximum-likelihood-estimated parameters of all considered 20 constituents of the $S\&P$ 500.
- Right: The normalized (to initial value 100) evolution of the considered 20 constituents of the $S\&P$ 500-index in the considered time period from 09 **March** 2020 **until** 21 **April** 2020 (In the pandemic crisis).

Crisis period

Parameters	fixed	robust	a_0 fixed	a_1 fixed	b_0 fixed	b_1 fixed	γ fixed	Black–Scholes
mean	5.8939	**0.9014**	0.9369	5.1408	0.9030	0.9433	2.1390	5.7859
std. dev.	2.7666	0.7705	0.7410	3.0596	0.7545	0.8073	1.8934	2.8477

- Relative hedging errors for an Asian at-the money put option $\left(x_0 - \frac{1}{30}\sum_{t=1}^{30} X_t\right)^+$ of trained hedging strategies of the considered 20 constituents.
- Each column represents another trained strategy which considers either fixed parameters (from the most recent maximum-likelihood estimation), robust parameter intervals (determined by the most extreme maximum-likelihood estimations), or robust intervals except for a single parameter which is still fixed.
- The rightmost column shows the hedging error when assuming an underlying Black–Scholes model. This is comparable to a NGA model with fixed parameters (first row)
- In a non-crisis period, the robust deep hedge is a bit more expensive in comparision (see paper)

- Summarizing, the change to uncertainty might seem like a small step, but it is a change in paradigm:
- Completely different techniques have to be used
- Where do we get the uncertainty intervals from ? A still open question.
- We showed that recent deep hedging techniques can be applied in this setting as well and outperform classical strategies on some data examples.

Many thanks for your attention!

More details may be found in:

- Fadina, Neufeld, Schmidt: Affine processes under Parameter Uncertainty (2019). Probability, Uncertainty and Quantitative Risk.
- Lütkebohmert, Schmidt, Sester: Robust Deep Hedging (2022). Quantitative Finance.

Föllmer, Hans and Alexander Schied (2011). **Stochastic finance: an introduction in discrete time**. Walter de Gruyter.