Computational Finance, Fall 2017
Computer Lab 4

The aim of the Lab is to learn some possibilities for identifying market model parameters from historical data.

Consider the Black-Scholes market model

$$dS(t) = S(t)(\mu(t)\,dt + \sigma(S(t),t)\,dB(t).$$

We have shown in the last lecture that then the logarithm of the stock price satisfies

$$d(\ln S(t)) = (\mu(t) - \frac{\sigma(S(t),t)^2}{2})\,dt + \sigma(S(t),t)\,dB(t). \tag{1}$$

In the case of constant parameters μ and σ it now follows (using the property $\ln a - \ln b = \ln \frac{a}{b}$), that for any time moments t_1 and t_2 we have

$$\ln \frac{S(t_2)}{S(t_1)} = (\mu - \frac{\sigma^2}{2})\,(t_2 - t_1) + \sigma(B(t_2) - B(t_1)).$$

Consequently (using the notation that $N(a, b^2)$ denotes the normal distribution with mean a and standard deviation b)

$$\ln \frac{S(t+\Delta t)}{S(t)} \sim N((\mu - \frac{\sigma^2}{2})\Delta t, \sigma^2 \Delta t).$$

So, if $S_i,\ i = 0, 1, 2, \ldots$ are closing prices of a stock at consecutive trading days, then $x_i = \ln \frac{S_{i+1}}{S_i}$ are values of normally distributed iid random variables. Usually time is measured in years, thus $\Delta t = \frac{1}{\text{trading days in a year}}$. By computing the mean and standard deviation of x_i we can find estimates for the constants μ and σ as follows:

$$\hat{\sigma} = \frac{\text{std}(x)}{\sqrt{\Delta t}},\ \hat{\mu} = \frac{\text{mean}(x)}{\Delta t} + \frac{\hat{\sigma}^2}{2}$$

Exercise 1. Download from http://finance.yahoo.com/ historical prices of the stock of Apple corporation for one year as a *.csv file. Write a script, that reads the adjusted closing prices from the file and uses them to estimate μ and σ (assuming the BS model with constant parameters holds). NB! pay attention to the order the data is in the file!

Unfortunately stock prices usually do not behave according to the model with constant coefficients. In order to see if the model suits for a particular stocks we can use various tests to verify the hypothesis that x_i come from a normal distributions. Two such tests are Anderson-Darling test and Shapiro-Wilk test (the functions `stats.anderson()` and `stats.shapiro()` in the `stats` subpackage of the `scipy` package). Actually, the tests never tell us that the model is valid but they warn us if it is very unlikely that the model is correct.

Exercise 2. Read the help information of the normality tests. Check if we can assume that the Apple stock prices follow the Black-Scholes model with constant coefficients.

If we do not want to assume that the parameters are constant, we may start with approximating the equation 1 according to the ideas of the Euler method:

$$\ln \frac{S(t_{i+1})}{S(t_i)} \approx (\mu(t_i) - \frac{\sigma(S(t_i),t_i)^2}{2})\,(t_{i+1} - t_i) + \sigma(S(t_i),t_i)(B(t_{i+1}) - B(t_i)).$$

Next, we introduce a finite number of unknown parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ and make an assumption how the functions $\mu = \mu_\theta$ and $\sigma = \sigma_\theta$ depend on those parameters. One way to find those parameters is to maximize the log-likelihood function: if Y_i are random variables with (conditional on the known values from earlier time moments) probability density functions f_i, then the log-likelihood function of the observed values y_i is

$$\sum_i \ln f_i(y_i).$$

Recall that the probability density function of the normal distribution $N(\mu, \sigma^2)$ is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}.$$

Since in our case the random variables $Y_i = \ln \frac{S_{i+1}}{S_i}$ are according to the approximate market model normally distributed with the mean $m_i = (\mu_\theta(t_i) - \frac{\sigma_\theta(S_i,t_i)^2}{2})\Delta t$ and the standard deviation $\sigma_\theta(S_i, t_i)\sqrt{\Delta t}$ we have to maximize the function

$$\text{loglike}(\theta) = -\sum_i \left(\frac{(Y_i - m_i)^2}{2\sigma_\theta(S_i, t_i)^2 \Delta t} + \ln \sigma_\theta(S_i, t_i) + \ln \sqrt{2\pi\Delta t} \right).$$

Since adding a constant to a function does not change the location of its maximum points, we can leave out the terms $\ln \sqrt{2\pi\Delta t}$.

Often there are no procedures for maximizing a function in software packages but minimization procedures are available. Fortunately this does not pose any problems, since maximization of a function $f(\theta)$ is equivalent to minimization of the function $g(\theta) = -f(\theta)$. Therefore, in Python, we are going to minimize the function

$$g(\theta) = \sum_i \left(\frac{(Y_i - m_i)^2}{2\sigma_\theta(S_i, t_i)^2 \Delta t} + \ln \sigma_\theta(S_i, t_i) \right).$$

In Python there are various commands for finding the minimum of a function; we try `optimize.fmin()` and `optimize.fmin_cg()`.

Exercise 3. Assume that

$$\mu_\theta(t_i) = \theta_0 + \theta_1 \cdot \frac{S_i - S_{i-2}}{S_i}$$

and

$$\sigma_\theta(s, t) = \theta_2.$$

Use the maximal likelihood method to determine the optimal values of θ.

Exercise 4. We can again test the validity of our assumptions. Namely, if our assumptions are correct, then $\frac{Y_i - m_i}{\sigma_\theta(S_i, t_i)}$ should be independent random variables from a normal distribution for the estimated parameters θ. So we can again use the tests for normality. Please test the validity of the model fitted in the previous exercise.

Exercise 5. Often we can not find a model for which statistical tests confirm the validity of it. Then we have to decide, which of the models we have considered is the best one. One possibility for choosing the best model is to use Akaike information criteria (AIC) - one can argue that the best model should have the lowest value of

$$AIC = 2 \cdot k - 2 \cdot \text{log\_likelihood},$$

where k is the number of parameters used in the model. Since our function f computes the negative of logarithmic likelihood, we can easily compute the AIC values for the model with constant μ and σ, and for the model with three parameters. Determine, which of the two models can be considered to be the best one according to AIC.