DISI – University of Trento

Master in Computer Science AA 2017/2018

Simulation and Performance Evaluation

Assignment 1

Understanding BitCoin Value Trends

Renato Lo Cigno, Michele Segata

March 27, 2018

You are given a dataset containing 10 traces as depicted in Fig. 1. The traces can be mathematically described as $\{y_k^j\ ;\ j = 0, \dots, 9, k = 0, \dots, 8999\}$, i.e., there are 10 different traces, and each of them has 9000 points. The traces are samples taken from a stochastic population, sample points are ordered, but there is not a "real" timestamp associated to them, so the process is discrete time. One trace is a real measure of the price (in Euro) of BitCoin. The other nine traces are synthetic traces obtained from the stochastic process described in Eq. (1).

$$\begin{aligned} x_{k+1} &= \alpha x_k + I_k \\ y_k &= x_k + V_0 \end{aligned} \tag{1}$$

where I_k are random variables taken from one of the distributions (always the same) described in Tab. 1, so that I_k is a stationary i.i.d. stochastic process that describes the increments (or decrements) in value of BitCoin, and X is a Markov-random process that describes the value trend of BitCoin, while. Finally Y describes the value of BitCoin itself, obtained adding the value V_0 of BitCoin at the beginning of the measurement period.

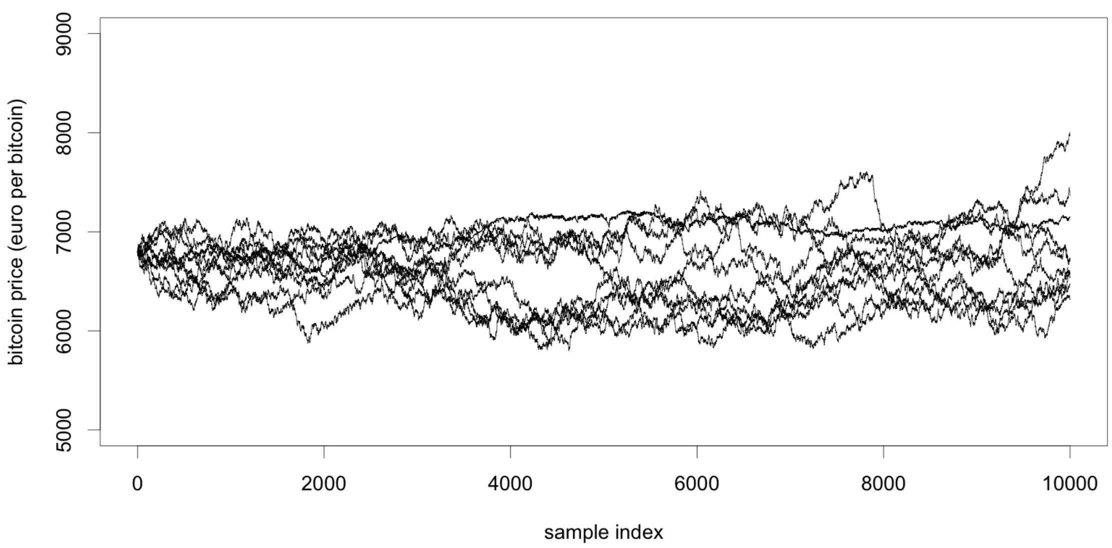

Figure 1: Plot of the 10 different traces of one dataset; no student is assigned this dataset.

The goals of the assignment are four:

1. Find which of the distributions in Tab. 1 has been used to fit the BitCoin trace;
2. Find the values of α and the parameters of the above distribution;
3. Find what is the original trace;
4. Comment on what you have obtained, if and how it can be useful to understand BitCoin, and (optional) find the time of the year (it is in 2018) when the trace has been measured.

Some hints on how to solve the assignment:

- As far as we know, there exists no `findTheBitCoinTrace` function in statistical tools, neither in R nor in other tools, so don't waste time searching for it, even if distinguishing the trace may be fairly easy;
- The goal of the assignment is mainly showing that you have learned how to manipulate data and you know how to use R to extract meaningful parameters out of a dataset;
- Analyze the traces and find the parameters that best fit your data;
- Document the theory behind the R functions you use.

To write your report use the LaTeX template[1] we suggested and do not write more than 3 pages. Deliver the PDF file of the report and the R, Matlab, or python script you used for processing and plotting as a single .zip or .tar file through Classroom; if the script does not run on a standard Linux box, we simply notify you that it does not work, and we will not attempt correction. Keep your code **CLEAN, ORGANIZED, and COMMENTED**. Do not send us your source code with unused portions commented out, blocks of code with no comments, or with monolithic pieces of code. Split your code in functions. Do not include the data set in the zip file, we already have it! The script MUST work assuming that the dataset is in the same folder of the script. DO NOT use absolute folders like

```
ds <- read.csv('/home/john.doe/Documents/spe/doe.csv')
```

but rather

```
ds <- read.csv('./doe.csv')
```

The deadline to have a correct-and-redo chance for this assignment is April 30 . . . 2018!!. If you deliver the assignment within this date, we will correct it and give you the chance to refine it before the end of the course, otherwise we will consider the work "as is" before the oral discussion is agreed upon, and give you the correction (with the evaluation) at the exam. However, if the quality of the delivery is unacceptable (e.g., no methodology is described, plots are meaningless and not explained, etc.) we will not correct it, but simply reject it, so you lose the privilege of a pre-correction.

If you have some doubts, just write us an email or ask in class.

Have Fun!

[1] `http://disi.unitn.it/locigno/teaching-duties/spe/assignment-template.zip`

Distribution name	Distribution pdf
Symmetric negative exponential	$$\frac{1}{2}\lambda e^{-\lambda\lvert x\rvert}\ ;\ \ -\infty < x < \infty$$
Gamma family	$$f_X(x) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{2\Gamma(\alpha)};\ \alpha, \lambda, -\infty < x < \infty$$ $$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy$$
Gaussian	$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$
Logistic	$$f_X(x) = \frac{e^{-\frac{x-\mu}{s}}}{s(1+e^{-\frac{x-\mu}{s}})^2}$$ $$\sigma^2 = \frac{\pi^2 s^2}{3}$$
Symmetric Pareto	$$f_X(x) = \frac{\alpha}{2x_m}\left(\frac{x_m}{\lvert x\rvert}\right)^{\alpha+1}\ ;\ -\infty < x \leq -x_m \cup x_m \leq x < \infty, x_m > 0$$

Table 1: List of the possible pdfs that may fit the increment process I_i.