**Report on the implementation of test task No. 1**

Initially, I understood the condition of the problem completely wrong. After clarifying questions, I understood it is necessary to create the desired distribution using some method of generating samples 10 daily returns from the original data. On my point of view, the Bootstrap method is well convenient for such task. Since the method is not explicitly specified in the task, I chose the most popular variation of method with the sample size coinciding with the volume of the original sample. The solution of the problem begins in the file task.py. Firstly, let's create a set of 1-day returns of 750 members. Using the function scipy.stats.levy\_stable.rvs() for this purpose. Next, I calculate the 10-day returns using the formulas given as hints in the description of the task. To calculate the prices, I set the initial value of the price. It can be arbitrary, because it will reduсed when calculating the n-day returns. This is obvious, but just in case it was checked numerically. Calculation procedure inside a function calculate\_n\_days\_returns(). Next, run a function generate\_percentile\_distribution(), that generates a given number of bootstrap pseudo-samples, calculates for each 1 percentile, and returns a list of percentile values. After that, I save the results and build graphs. Saves two files with sets of 10-day returns and with 1-percentile distribution for use in another files. In fig. 1, 2, 3 presented histograms which illustrate results of calculations for 1000001 trials. It is maximum amount of trials which I used.

The next step is to justify the sufficiency of the Monte Carlo trials. And this question is more complicated, because it is not clear enough sufficient for what. For example, to build a discrete distribution, a single test is enough. However, I suspect that a different answer is expected in the task. I suppose it means that starting from a certain number of tests a certain characteristic of distribution with a certain probability will have a certain value. For experiments with 1-percentile distribution characteristics were created a file proof.py. It generates many distributions of 1 percentile from one sample of 10-day returns with different numbers of Monte Carlo trials. After that, statistical parameters are calculated for each distribution obtained, which allows us to build the dependence of the parameter from the number of tests. Results for two initial distribution of 10-day returns are presented in figures 4 – 6. The information content of the experiments was not too high. First conclusion of performed experiments is that the median and especially mode is not convenient metrics for analysis. Indeed, the median either behaves as a constant (fig. 5.a ) or randomly oscillates in a certain interval (fig. 5.b ). Other quantiles behave in a similar way. In principle, it is not always possible to determine mode in the resulting distributions. Often, they look like multimode. Another conclusion is that with an increase of the number of trials, the dependence of mean and SE on their number decreases. It is clear that the number of tests will affect the scatter of the resulting metrics. The lower the number of tests, the higher the scatter. Let’s analyze this dependence. Using the file proof\_2.py. Obtained 100 values for the mean and standard deviation of the percentile distributions obtained from one sample of 10-day returns using the same number of trials. I used number of trials 100, 1000, 10000. Results presented on the fig. 7-12 and table 1.

|  |
| --- |
| Table 1 SE values of metric distributions |
| |  |  |  | | --- | --- | --- | | Number of trial | Mean | SE | | 100 | 596.6 | 794.18 | | 1000 | 174.97 | 238.72 | | 10000 | 64.88 | 77.34 | |

It is easy to notice that with an increase of the number of trials by 10 times, the standard deviation of the resulting sample decreases by approximately 3 times, which approximately corresponds to the root of 10. By extrapolating this dependence, we can assume for the same samples of the percentile distribution metrics consisting of one million trials, the standard deviation will be 10 times lower than for 10000 trials. Approximately 6.5 for means and 7.7 for standard deviation. Thus, we can choose the number of sufficient trials based on a predetermined standard error of the selected metric.

Remark 1: The chosen method of generating the initial sample (10-day returns) leads to poor reproducibility of the samples.

Remark 2: The selected method of forming the initial sample (10-day returns) is not physical, leading to results that are not observed in reality. If you set the initial price equal to 1. And try to calculate prices based on profitability data. Values of the order of -1021 or 1018 can be obtained, while they can often change the sign from day to day. I do not know a product whose price has ever behaved in this way.

Remark 3: Sorry for the too confused statement of thoughts.

|  |
| --- |
|  |
| Figure 1 Black line is a probability density function for stable distribution with the given parameters |
|  |
| Figure 2 |
|  |
| Figure 3 Black line is a probability density function for normal distribution with the given parameters |
|  |
| Figure 4.a Mean set\_of\_10-day\_returns\_1000000\_trials.data |
|  |
| Figure 4.b Mean set\_of\_10-day\_returns\_1000001\_trials.data |
|  |
| Figure 5.a Median set\_of\_10-day\_returns\_1000000\_trials.data |
|  |
| Figure 5.b Median set\_of\_10-day\_returns\_1000001\_trials.data |
|  |
| Figure 6.a SE set\_of\_10-day\_returns\_1000000\_trials.data |
|  |
| Figure 6.b SE set\_of\_10-day\_returns\_1000001\_trials.data |
|  |
| Figure 7 |
|  |
| Figure 8 |
|  |
| Figure 9 |
|  |
| Figure 10 |
|  |
| Figure 11 |
|  |
| Figure 12 |