author Evgeny Kozyrev
19.06.2024

Monte-Carlo task

Return of greater degree can be expressed through the values of return of the first degree, as shown in formula (1).

$$r^1_i = p_{i+1}/p_i - 1 \qquad r^1_{i+1} = p_{i+2}/p_{i+1} - 1$$
$$r^2_i = p_{i+2}/p_i - 1 = (r^1_{i+1} + 1)\cdot p_{i+1}/p_i - 1 = (r^1_{i+1} + 1)\cdot r^1_i$$
$$r^3_i = p_{i+3}/p_i - 1 = (r^1_{i+2} + 1)\cdot r^2_i = (r^1_{i+2} + 1)\cdot(r^1_{i+1} + 1)\cdot r^1_i$$

$$r^n_i = p_{i+n}/p_i - 1 = (r^1_{i+n-1}+1)\cdot r^{n-1}_i = (r^1_{i+n-1}+1)\cdot(r^1_{i+n-2}+1)\cdot \ldots \cdot (r^1_{i+1} + 1)\cdot r^1_i \qquad (1)$$

Below, in Fig. 1(a), a stable distribution of $r1$ with parameters (α=1.7;β=0.0;γ=1.0;δ=1.0) with 750 events is presented. Fig. 1(b) shows the distribution of overlapping $r10$, calculated according to formula (1). The distribution of $r10$ is significantly wider, as it is determined by the product of ten random variables.

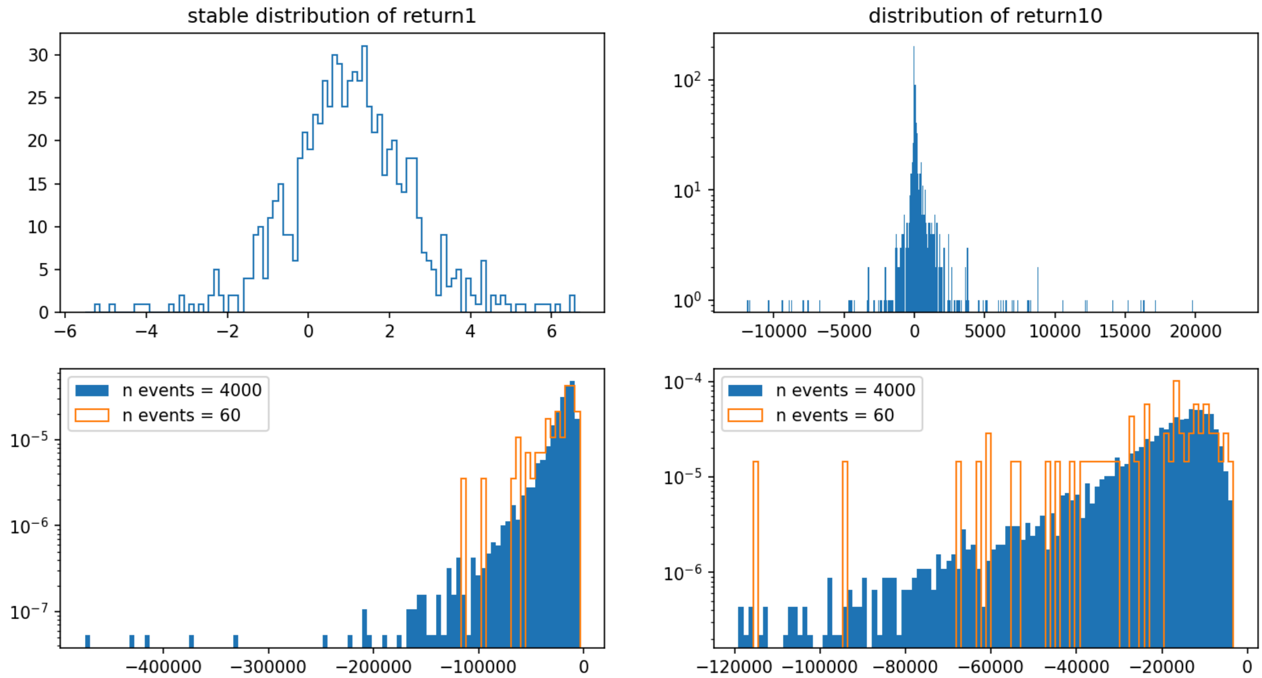

Fig 1. (a) The spectrum of return1; (b) the spectrum of return10; (c,d) the spectrum of 0.01 quantiles.

The np.quantile(..., q=0.01, method='hazen') method was used to determine the quantile coordinate. The Hazen method involves piecewise linear interpolation of the sample distribution function $p_k = (k-1/2)/n$, where k is the rank of the sorted event and n is the total number of events [1]. Other methods of determining the quantile can also be validly used.

It should be noted that the width of the quantile spectrum is determined, first of all, by the number of events in the r_{10} spectrum. With a larger number of events, the position of the quantiles can be determined much more accurately, and their spectrum will be narrower. Secondly, the sample spectrum of r_{10} could be approximated using parametric functions (e.g., a sum of two beta functions), allowing for a more precise determination of the quantile position, although this would introduce model uncertainty.

The obtained spectrum of 1% quantiles is shown in Fig. 1(c,d). The generated sample P1 with 4000 events is shown in blue. The distribution function of this spectrum is close to the function of the general population, although the left tail of the spectrum still requires more events.

The task includes a question regarding what sample size is sufficient. The answer to this question depends on the requirements for the values that are calculated using the desired sample distribution function. At the same time, the number of simulations should be minimized to reduce execution time and costs. Suppose that in this case, it is important to calculate the average value of the 1% quantiles of the r10 distribution accurately and quickly. Then, a criterion for stopping the generation process can be the requirement that the average value of the spectrum stops fluctuating. If we require that the standard deviation of the last 10 mean values is less than 1% of the mean, the generation will stop after 60 events. Such a spectrum is shown in orange in Fig. 1(c,d).

The two spectra shown satisfy the Kolmogorov-Smirnov criterion and are characterized by values of statistic=0.127 and pvalue=0.26. It's obvious for two samples with common origin and high and less numbers of events.

Conclusion:
The desired distribution of 1% quantiles is shown in orange in Fig. 1 (c,d) and exhibits asymmetric behavior with a long left tail.

Note to the Conclusion:

The original problem of finding the quantile could be solved by reconstructing the distribution function of $r10$ (instead of simulating it). To achieve this, we can change variables and the calculate a multidimensional integral over the coordinates of $r1$, on which $r10$ depends. However, in practice, the probability density of $r1$ cannot be known accurately and likely varies over time. Therefore, the accuracy of determining the quantile is either limited by the statistics—the number of days with measured $r1$, or by the knowledge of the initial probability density of $r1$.

Additionally, if we assume that $r^1_i = p_{i+1}/p_i - 1$ and the price cannot be negative, then $r1$ belongs to the interval $[-1,+\infty]$. In this case, simulating events according to a stable distribution from $-\infty$ to $+\infty$ is not appropriate.

[1] R. J. Hyndman and Y. Fan, “Sample quantiles in statistical packages,” The American Statistician, 50(4), pp. 361-365, 1996.