What Is a Monte Carlo Simulation?

Monte Carlo simulations are used to model the probability of different outcomes in a process that cannot easily be predicted due to the intervention of random variables. The basis of a Monte Carlo simulation involves assigning multiple values to an uncertain variable to achieve multiple results and then to average the results to obtain an estimate. To determine the number of simulations, there are a large number of evaluation criteria based on checking the convergence of the distribution function to the known one for testing simple hypotheses about the membership of the analyzed sample in some fully known distribution law. But in our case, the distribution law is unknown, so I will describe several possible assumptions how this could be estimated.

Describes the process for generating ten-day returns percentiles. 1-day returns are generated by stable distribution.

We have the original expression for r_i^1.

Let's represent r_i^{10} in terms of r_i^1:

$$r_i^1 = \frac{P_{i+1}}{P_i} - 1$$

Then

$$r_i^1 + 1 = \frac{P_{i+1}}{P_i}$$

$$r_{i+1}^1 + 1 = \frac{P_{i+2}}{P_{i+1}}$$

$$\vdots$$

$$r_{i+n}^1 + 1 = \frac{P_{i+n+1}}{P_{i+n}}$$

If the obtained equations are multiplied, then all intermediate values will be reduced. Only two extreme values remain:

$$(r_i^1 + 1)(r_{i+1}^1 + 1)(r_{i+2}^1 + 1) \dots (r_{i+9}^1 + 1) = \frac{P_{i+1}}{P_i}\frac{P_{i+2}}{P_{i+1}}\frac{P_{i+3}}{P_{i+2}} \dots \frac{P_{i+10}}{P_{i+9}} = \frac{P_{i+10}}{P_i}$$

If you subtract one from the resulting fraction, you get an expression for r_i^{10}

$$\frac{P_{i+10}}{P_i} - 1 = r_i^{10}$$

Thus, the expression for $r_i^{10}$is represented as a product of known values.

$$r_i^{10} = (r_i^1 + 1)(r_{i+1}^1 + 1)(r_{i+2}^1 + 1) \dots (r_{i+9}^1 + 1) - 1$$

Estimation of the number of iterations of the Monte Carlo method

To estimate the number of Monte Carlo simulations, 3 methods can be used to increase complexity.

1. The first is a simple estimate using only theoretical data.
2. The second is an estimate using the construction of a histogram and the use of statistical numerical characteristics.
3. The third is an assessment using the Kolmogorov - Smirnov consistency test (Smirnov homogeneity test).

It is of practical interest to estimate for what sample size with a given accuracy the limiting law of statistics can be used instead of the "true" distribution.

1. **Theoretical estimate of the number of simulations of the Monte Carlo method**

Let it be required to calculate the probability P of the occurrence of some random event A. In each of the N realizations of the process, the number of occurrences of the event A is a random variable ξ taking the value 1 with probability P and the value 0 with probability 1 - P. The random variable ξ obeys the Bernoulli distribution with the probability of success P and has the mathematical expectation P and the dispersion P(P - 1).

As an estimate for the required probability P, the frequency m/N of occurrences of event A is taken for N realizations

$\frac{m}{N} = \frac{1}{N}\sum_1^N x_i$, where x_i is the number of occurrences of the event A in the implementation with the number i.

By virtue of the central limit theorem of probability theory, the frequency for sufficiently large N has a distribution close to normal

$$\sqrt{N}\frac{\frac{m}{N} - M\,\xi}{\sqrt{D\,\xi}} = \sqrt{N}\frac{\frac{m}{N} - P}{\sqrt{P(1-P)}} \to \eta \in N(0,1)$$

Then

$P\left\{\left|\sqrt{N}\frac{\frac{m}{N}-P}{\sqrt{p(1-p)}}\right| < t_\gamma\right\} = \phi(t_\gamma) - \phi(-t_\gamma) = 2\phi(t_\gamma) - 1 = \gamma$, where $t_\gamma = \phi^{-1}\left(\frac{\gamma+1}{2}\right)$ is a standard normal distribution quantile

$$P\left\{\left|\frac{m}{N} - P\right| < t_\gamma\frac{\sqrt{p(1-p)}}{\sqrt{N}}\right\} = \gamma$$

Thus, the simulation error P with a confidence level γ is

$$\varepsilon = t_\gamma\frac{\sqrt{p(1-p)}}{\sqrt{N}}$$

Hence the number of repetitions for which the error simulation does not exceed ε, is equal to

$N = \left[{t_\gamma}^2\frac{p(1-p)}{\varepsilon^2}\right] + 1$, where [×] means the whole part

The last formula can be rewritten for the relative error δ:

$$N = \left[{t_\gamma}^2\frac{1}{p(1-p)\delta^2}\right] + 1$$

For reliable estimation of probabilities, it is desirable that the relative error did not exceed 5%.

Thus, to simulate with a probability of p = 0.9 with an error of no more than 5%, **12 000 iterations are required** (t_γ = 1, 643 for γ = 0,9 [1]) [3]

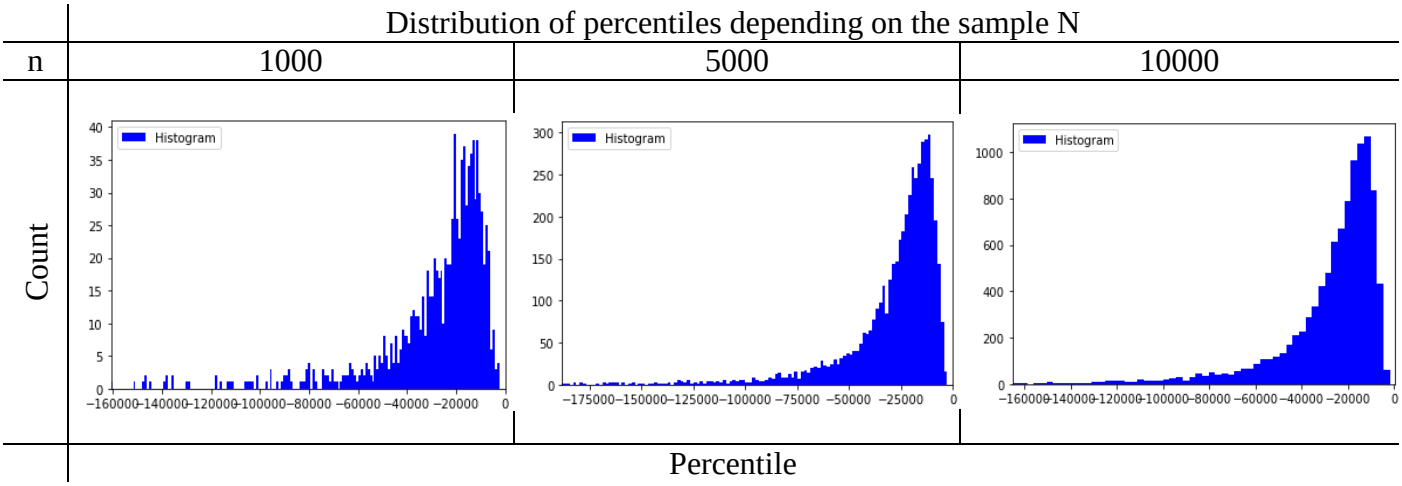
2. Estimation using the construction of a statistical series, a histogram.

Fluctuations are random deviations of the observed physical quantities from their average values. Fluctuations occur in any quantities that depend on random factors and are described by statistical methods. The simplest measure of the fluctuation of the quantity x is its dispersion σ^2_x, i.e., the mean square of the deviation of x from its mean value $\bar{x}$.

The equivalent measure of the fluctuation is the squared deviation, equal to the square root of the dispersion, or its relative value $\delta_x = \sigma_x/x$. It is known from the statistics of physics that dispersion $\sigma_x \sim \sqrt{N}$, and the relative value is proportional to $\delta_x \sim \frac{1}{\sqrt{N}}$.

We will restrict ourselves to a bin fluctuation error of 10%, that is, there are about 100 points in each bin. We will generate points until this condition is met. It is possible that some of the points will be collected with a

high probability in only a few bins, and the other points will be “smeared” with low probabilities along the whole x - axis. Then it is advisable to combine bins that have few points. Alternatively, in order not to wait for a set of 100 points in all bins, we will restrict ourselves to the number of simulations at which the histogram will not change its appearance. We will use the same bins into which the statistical material was classified for constructing the histogram. The approximate value of the random variable in each bin will be considered constant and equal to the average value, which acts as a "representative" of the bin. With a sufficient number of simulations, we will get the tendency of the arithmetic means to their math expectations. [1]



Thus, we can say that for about **10 000 is acceptable**, since for the main part of the typed values, the condition of a set of 100 points in each bin is satisfied.

3. Kolmogorov-Smirnov method

The homogeneity hypothesis is the assumption that two (or more) samples are drawn from the same general population. For this hypothesis, it does not matter what kind of distribution the general population has, therefore the “ideal” criterion of homogeneity should not depend on the form of the distribution law of a random variable. In most cases, homogeneity tests are a two-sample modification of the goodness-of-fit test. So Smirnov's homogeneity criterion corresponds to Kolmogorov's goodness-of-fit criterion.

Tests for hypotheses on the distributions of two sets of data as a rule employ either the Smirnov test.

One can check the homogeneity of two ranked sets of size m ≤ n as represented by the variation series

$$x_1 < x_2 < \ldots < x_m$$

$$y_1 < y_2 < \ldots < y_n$$

and testing hypothesis H0 that the two samples derived from the same population, i.e., F(x) = G(x) for any x = y.

Smirnov Test. This test involves the assumption that the distributions F(x) and G(x) are continuous. The Smirnov test statistic characterizes the maximum difference between the empirical distributions:

$$D_{m,n} = \sup_x \left| G_m(x) - F_n(x) \right|$$

If H0 is correct, i.e.,

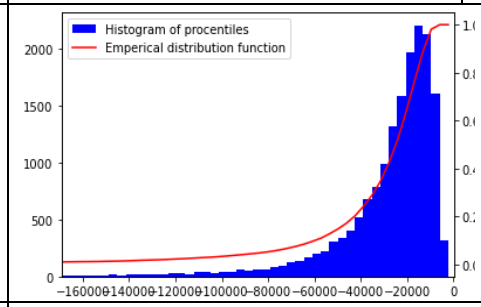
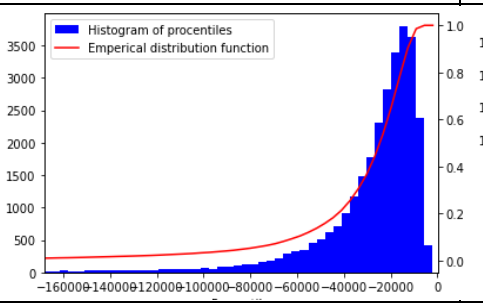
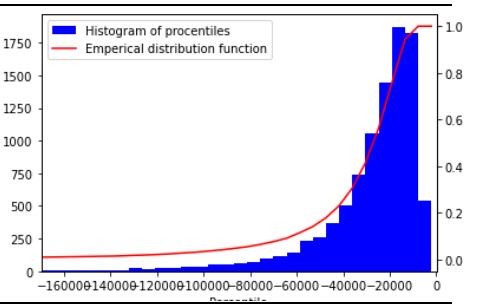
$$\lim_{m\to\infty} P\left\{ \sqrt{\frac{mn}{m+n}} D_{m,n} < s \right\} = K(s)$$

and the statistic

$$S_C = \sqrt{\frac{mn}{m+n}} D_{m,n}$$

in the limit follows a Kolmogorov distribution K(s). The set of possible values constitutes a lattice with pitch 1/k, in which k is the least common multiple of m and n. [2]

The distribution function in the general case is a graph of a non-decreasing function from 0 to 1. The distribution function changes abruptly, and the magnitude of the jump is equal to the probability of this value. As the number of possible values of random variables increases and the intervals between them decrease, the number of jumps becomes larger, and the jumps themselves become smaller. The stepped curve becomes smoother, the random variable gradually approaches a continuous value, and its distribution function becomes a continuous function.

	N = 8500 Err = 0,01 Pvalue = 0,8	N = 14770 error = 0.01 Pvalue = 0.5	N = 4860 error = 0.01 Pvalue = 0.97
Count			
	Percentile		

Thus, observing several simulations, we can restrict ourselves to an estimate of the average number of required iterations. For 10 different observations, about 10 000 iterations were obtained.

Conclusion

In the process, the Monte Carlo method generated 1% percentiles of ten-day returns. The ten-day returns were calculated based on the one-day returns given by a stable distribution.

There are also several ways to estimate the number of iterations by the Monte Carlo method. A selection of 10,000 points is sufficient.

Literature

1. Теория вероятностей. Вентцель Е.С. [Electronic resource]. URL: https://may.alleng.org/d/math/math350.htm (accessed: 06.07.2021).
2. Lemeshko B., Lemeshko S. Statistical distribution convergence and homogeneity test power for Smirnov and Lehmann–Rosenblatt tests // Meas. Tech. 2005. Vol. 48. P. 1159–1166.
3. Постовалов С. Н. ПРИМЕНЕНИЕ КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ ДЛЯ РАСШИРЕНИЯ ПРИКЛАДНЫХ ВОЗМОЖНОСТЕЙ КЛАССИЧЕСКИХ МЕТОДОВ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ//2013.