Тема 4. Статистические методы измерения риска. Метод Монте-Карло.

Напомним некоторые определения из теории вероятностей

Случайное событие – событие, которое может произойти, а может и не произойти.

Элементарное событие – событие, которое нельзя представить в виде множества других событий.

Случайная величина – числовая величина, которая может принимать в зависимости от будущих обстоятельств, различные значения.

Дискретная случайная величина – случайная величина, которая имеет конечное число различных значений.

Непрерывная случайная величина – случайная величина, которая может принимать бесконечное число различных значений

Пусть X - дискретная случайная величина, она может принимать различные значения: $\{x_1, x_2, x_3, \ldots, x_i, \ldots x_n\}$.

Вероятностью p наблюдения числа x_i называется отношение случаев наблюдения числа x_i (обозначим его, например, k_i) к общему количеству всех наблюдений K.

$$\mathbf{p_i = k_i/K}$$

Свойства вероятности:

1) для каждого x_i : $0 <= p_i <= 1$, $i=1,\ldots,n$.
2) $\sum p_i = 1$, $i=1,\ldots,n$.

Иногда вероятности измеряют не в долях единицы, а в процентном выражении:

$$p_i = 100*k_i/K$$

Тогда:

3) для каждого x_i : $0 <= p_i <= 100\%$, $i=1,\ldots,n$.
4) $\sum p_i = 100\%$, $i=1,\ldots,n$.

Вероятность наблюдения события может быть определена объективным или субъективным методом.

Субъективный метод основан на использовании субъективных критериев (суждение оценивающего, его личный опыт, оценка эксперта) и вероятность события в этом случае может быть разной, будучи оцененной разными экспертами.

Объективный метод определения вероятности основан на определении относительной частоты, с которой происходит данное событие.

Размах вариации *(R)* – разница между максимальным и минимальным значением фактора

$$R = x_{max} - x_{min}$$

Этот показатель дает очень грубую оценку риску, т.к. он является абсолютным показателем и зависит только от крайних значений ряда наблюдений.

Расчет величины риска по какому-либо показателю должен начинаться с определения его обычного, среднего значения. На следующем этапе переходят к расчету возможных отклонений от найденного среднего значения, которые и характеризуют риск.
В теории вероятностей усредненное значение какой-либо случайной величины характеризуется показателем математического ожидания случайной величины. Усредненное отклонение от математического ожидания называется стандартным отклонением случайной величины. Во многих областях производства, экономики, финансов именно показатель стандартного отклонения считается мерой риска.

Математическое ожидание случайной величины *X* определяется по формуле:

$$M(X) = \sum x_i p_i, \; i = 1, \ldots, n$$

где:

M – символ математического ожидания,
n – количество числовых значений, которые может принимать случайная величина;
x_i – конкретное числовое значение, которое может принимать случайная величина;
p_i – вероятность появления (наблюдения) значения x_i.

На практике результаты анализа более наглядны, если показатель разброса случайной величины выражен в тех же единицах измерения, что и сама случайная величина. Для этих целей используют стандартное (среднее квадратическое) отклонение $\sigma(X)$.

Стандартное отклонение σ случайной величины находится по правилу:

$$\sigma(X) = \sqrt{\sum_{i=1}^{n} p_i (x_i - M(X))^2}$$

Сумму под корнем называют **дисперсией случайной величины:**

$$D(X) = \sum_{i=1}^{n} p_i (x_i - M(X))^2$$

Использование дисперсии как меры риска не всегда удобно, т.к. размерность ее равна квадрату единицы измерения случайной величины.

Иногда для удобства вычислений сначала находят дисперсию случайной величины, а потом берут корень из получившегося числа.

$$\sigma(X) = \sqrt{D(X)}$$

Стандартное отклонение имеет ту же самую размерность, что и исходное случайное число, и его математическое ожидание. Для того, чтобы более ясно представить степень влияния риска на исходное случайное число, сравнивают стандартное отклонение и математическое ожидание случайного числа, умноженное на 100. Этот относительный показатель называется **коэффициент вариации** (CV):

$$CV(X) = 100 * \sigma(X) / M(X).$$

При анализе рыночных рисков часто возникает задача сравнения риска по нескольким активам. Конечно, можно рассчитать по каждому активу стандартное отклонение его цены. Однако такие величины трудно сравнивать между собой. Например, если стандартное отклонение цены одного актива составляет 10 руб., а второго – 100 руб., нельзя сказать, что второй актив «рискованнее» первого в десять раз. Если средняя стоимость первого актива составляет 1000 рублей, а второго актива - 50000 руб., то даже интуитивно понятно, что приобретая первый актив мы можем потерять 1% его стоимости, а второй актив – только 0,2%. Поэтому для сравнения ценовых рисков лучше использовать не абсолютные, а относительные величины. Для этого служит показатель, называемый **волатильность**.

Пусть нам дан ряд цен на активы за прошлые периоды (например, цены на конкретные товары за T дней). Определим, каково соотношение между изменением цены и самой ценой между двумя соседними днями:

$$\Delta_t = \frac{c_t - c_{t-1}}{c_{t-1}}, t = 2, ..., T,$$

где:

t – номер дня, t=1,…,T

c_t – цена товара в день t.

Всего можно рассчитать (T-1) таких величин. С математической точки зрения все они являются наблюдениями случайной величины Δ. Но в отличие от самой цены, Δ - это относительное изменение цены. В отличие от самой цены, которая измеряется в рублях, долларах, юанях, Δ измеряется в долях единицы. То есть эти величины, в отличие от абсолютных цен, можно сравнивать для разных товаров между собой. Поскольку нам неизвестны вероятности появления различных наблюдений Δ_t, принимается по умолчанию, что их появление равновероятно. Поскольку сумма всех вероятностей должна быть равна 1, вероятность наблюдения конкретного Δ_t, будет равна p_t=1/(T-1).

Найдем математическое ожидание величины Δ. Для этого воспользуемся формулами из раздела 3:

$$\mathrm{M}(\Delta) = \sum_{t=2}^{\mathrm{T}} \Delta_t * \mathrm{p}_t = \frac{1}{(T-1)} \sum_{t=2}^{\mathrm{T}} \Delta_t = \overline{\Delta}$$

Черта над переменной будет обозначать среднее арифметическое.

Волатильность представляет собой стандартное (среднеквадратическое) относительное отклонение цены:

$$\sigma(\Delta) = \sqrt{\sum_{t=2}^{\mathrm{T}} p_t (\Delta_t - M(\Delta))^2} = \sqrt{\frac{\sum_{t=2}^{\mathrm{T}} (\Delta_t - \overline{\Delta})^2}{\mathrm{T} - 1}}.$$

Естественно, для правильного расчёта волатильности, как и любой другой оценки случайной величины, требуется достаточно большая статистическая выборка цен. Например, для расчета однодневной волатильности желательно использовать наблюдения не менее чем за три месяца.

Если требуется определить волатильность за год, квартал однодневную волатильность умножают на квадратный корень из количества дней в году, квартале:

$$\sigma(\Delta)_{\text{период}} = \sigma(\Delta)_{\text{день}} * \sqrt{T}$$

Корреляция – мера взаимозависимости двух случайных величин. Обозначим одну случайную величину буквой *x*, а другую – буквой *y*. Будем фиксировать, какие значения принимают одновременно эти две случайные величины: это пары чисел (x_i ; y_i), i=1,…,T. В случае, если все наблюдения за случайными величинами равновероятны, формула для подсчета коэффициента корреляции такова:

$$r_{xy} = \frac{\sum_{t=1}^{\mathrm{T}} (x_t - \bar{x}) \times (y_t - \bar{y})}{\sqrt{\sum_{t=1}^{T} (x_t - \bar{x})^2} \times \sqrt{\sum_{t=1}^{T} (y_t - \bar{y})^2}}$$

где:

x_i - значения, принимаемые случайной величиной *x*,

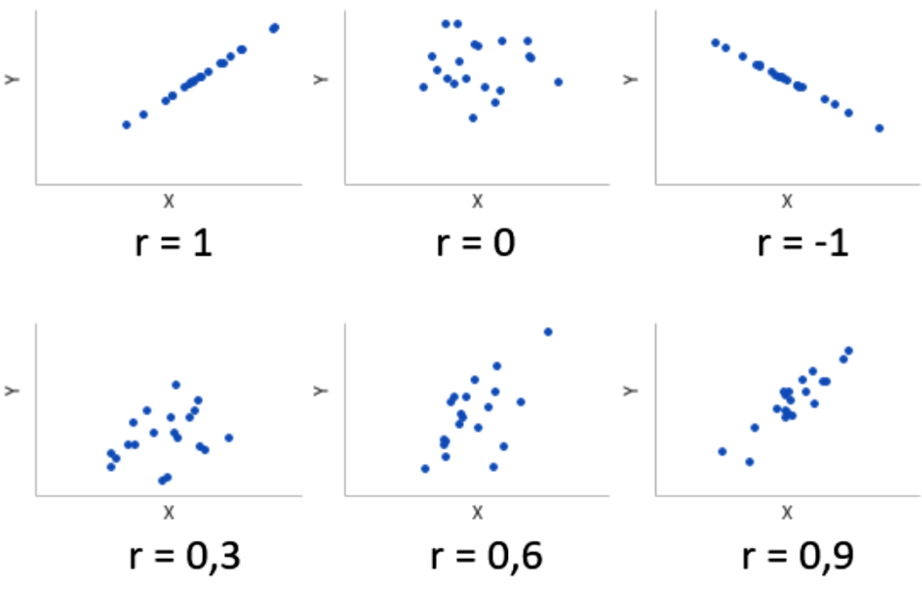
y_i - значения, принимаемые случайной величиной *y*,

$\bar{x}$ - средняя арифметическая по X,

$\bar{y}$ - средняя арифметическая по Y.

Свойства коэффициента корреляции:

1) $-1 <= r_{xy} <= 1$
2) если r_{xy} =1, то x и y зависят прямо пропорционально друг от друга
3) если случайные величины независимы, то r_{xy} =0
4) если r_{xy} = -1, то x и y зависят обратно пропорционально друг от друга



Регрессионный анализ — статистический метод поиска формулы зависимости между зависимой случайной величиной y и одной или несколькими независимыми случайными величинами $x_1, x_2, \ldots, x_m$.

$$y = f(x_1, x_2, x_3, \ldots, x_m)$$

Уравнение регрессии – это формула статистической связи между переменными. Формула статистической связи двух переменных называется *парной регрессией*, зависимость от нескольких переменных – *множественной регрессией.*

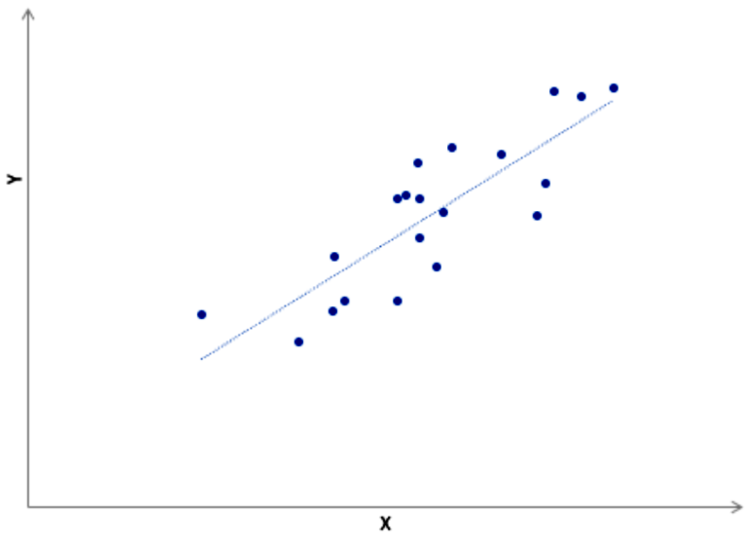
Методами регрессионного анализа можно осуществлять поиск как линейных, так и нелинейных регрессионных формул. Начальным пунктом эконометрического анализа зависимостей обычно является оценка линейной зависимости переменных. Это объясняется простотой исследования линейной зависимости. Поэтому проверка наличия такой зависимости, оценивание ее индикаторов и параметров является одним из важнейших направлений приложения математической статистики. Уравнение линейной регрессии имеет вид:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_mx_m$$

где a, b_1, b_2, $b_3, \ldots, b_m$ - числа

Можно предположить и другие формы зависимости - квадратичную, гиперболическую, экспоненциальную и т.д. Однако если это реальные статистические данные, то мы никогда не получим на графике кривую, которая бы совершенно точно описывала взаимозависимость случайных переменных, какую бы форму зависимости мы ни выбрали. Связь переменных, на которую накладываются воздействия случайных факторов, называется *статистической связью*.

Выбор формулы связи переменных называется спецификацией уравнения регрессии. Наиболее простым для изучения является случай взаимосвязи двух переменных. В данном случае выбрана линейная формула: $y = a + bx$ Нам требуется оценить значения параметров a и b.



Для оценки линейных параметров регрессий используют метод наименьших квадратов, который позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений y от теоретических $\hat{y}_i$ минимальна, т.е.

$$\Delta = \sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \longrightarrow \min$$

Можно воспользоваться готовыми формулами**,** которые вытекают из решения этой задачи:

$$a = \bar{y} - b\,\bar{x}, \qquad b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sigma_x^2}$$

Черточки над переменными x и y обозначают среднеарифметические значения этих переменных.

Функция распределения случайной величины. Её свойства

Каждая случайная величина полностью определяется своей *функцией распределения*.

Если ξ- случайная величина, то функция $F(x) = F_\xi(x) = P(\xi < x)$ называется *функцией распределения* случайной величины x . Здесь $P(\xi < x)$ - вероятность того, что случайная величина ξ принимает значение, меньшее x.

Важно понимать, что функция распределения является “паспортом” случайной величины: она содержит всю информация о случайной величине и поэтому *изучение случайной величины заключается в исследовании ее функции распределения,* которую часто называют просто *распределением*.

Функция распределения любой случайной величины обладает следующими свойствами:

- $F(x)$ определена на всей числовой прямой R;
- $F(x)$ не убывает, т.е. если $x_1 \le x_2$, то $F(x_1) \le F(x_2)$;
- $F(-\infty)=0$, $F(+\infty)=1$, т.е. $\lim\limits_{x \to -\infty} F(x) = 0$ и $\lim\limits_{x \to +\infty} F(x) = 1$

Функция распределения дискретной случайной величины

Если x - дискретная случайная величина, принимающая значения $x_1 < x_2 < \ldots < x_i < \ldots$ с вероятностями $p_1 < p_2 < \ldots < p_i < \ldots$, то таблица вида

x_1	x_2	…	x_i	…
p_1	p_2	…	p_i	…

называется *распределением дискретной случайной величины*.

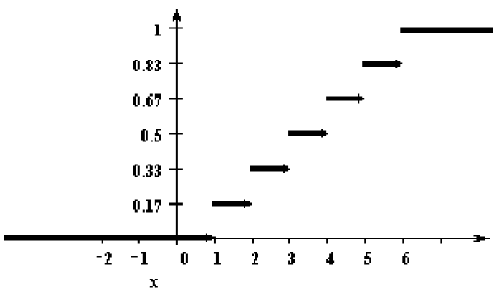
Функция распределения случайной величины, с таким распределением, имеет вид

$$F(x)=\begin{cases}0, & при\ x<x_1,\\ p_1, & при\ x_1\le x<x_2,\\ p_1+p_2, & при\ x_2\le x<x_3,\\ \ldots, & \\ p_1+p_2+\ldots+p_{n-1}, & при\ x_{n-1}\le x<x_n,\\ 1, & при\ x\ge x_n.\end{cases}$$

У дискретной случайной величины функция распределения ступенчатая. Например, для случайного числа очков, выпавших при одном бросании игральной кости, распределение, функция распределения и график функции распределения имеют вид:

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

$$F(x)=\begin{cases}0,\ при -\infty<x<1,\\ \frac{1}{6},\ при\ 1\le x<2,\\ \frac{1}{6}+\frac{1}{6},\ при\ 2\le x<3,\\ \frac{1}{6}+\frac{1}{6}+\frac{1}{6},\ при\ 3\le x<4,\\ \frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6},\ при\ 4\le x<5,\\ \frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6}+\frac{1}{6},\ при\ 5\le x<6,\\ 1,\ при\ x\ge 6\end{cases}$$



Функция распределения и плотность вероятности непрерывной случайной величины

Если функция распределения $F_\xi(x)$ непрерывна, то случайная величина ξ называется *непрерывной случайной величиной.*

Если функция распределения непрерывной случайной величины *дифференцируема*, то более наглядное представление о случайной величине дает *плотность вероятности случайной величины* $p_x(x)$, которая связана с функцией распределения $F_\xi(x)$ формулами

$$F_\xi(x)=\int_{-\infty}^{x}p_\xi(t)\,dt$$

$$p_\xi(x)=\frac{dF_\xi(x)}{dx}.$$

Отсюда, в частности, следует, что для любой случайной величины

$$\int_{-\infty}^{\infty}p(x)dx=1.$$

Квантили

При решении практических задач часто требуется найти значение x, при котором функция распределения $F_\xi(x)$ случайной величины x принимает заданное значение p, т.е. требуется решить уравнение $F_\xi(x) = p$. Решения такого уравнения (соответствующие значения x) в теории вероятностей называются *квантилями.*

Квантилью x_p (p-квантилью, квантилью уровня p) случайной величины ξ, имеющей функцию распределения $F_\xi(x)$, называют решение x_p уравнения $F_\xi(x) = p$, $p \subset (0, 1)$..
Квантили, наиболее часто встречающиеся в практических задачах, имеют свои названия:
медиана - квантиль уровня 0.5;
нижняя квартиль - квантиль уровня 0.25;
верхняя квартиль - квантиль уровня 0.75;
децили - квантили уровней 0.1, 0.2, …, 0.9;
процентили - квантили уровней 0.01, 0.02, …, 0.99.

Вероятность попадания наблюдений случайной величины в интервал
Вероятность того, что значение случайной величины $F_\xi(x)$ попадает в интервал (a, b), равная $P(a < x < b) = F_\xi(b) - F_\xi(a)$, вычисляется по формулам:

$$P(a < \xi < b) = \int_a^b p_\xi(x)\,dx = F(b) - F(a)$$

- для непрерывной случайной величины и

$$P(a < \xi < b) = \sum_{i: x_i \in (a, b)} p_i = F(b) - F(a)$$

- для дискретной случайной величины.

Если $a = -\infty$, то $P(a < \xi < b) = P(\xi < b) = F(b)$,
если $b = \infty$, то $P(a < \xi < b) = P(a < \xi) = 1 - P(\xi < a) = 1 - F(a)$.

Нормальное распределение случайной величины— одно из важнейших распределений в теории вероятностей; встречаются также названия: *закон Гаусса, гауссовское распределение, распределение Гаусса-Лапласа*; Распределение вероятностей случайной величины ξ называется *нормальным*, если оно имеет плотность вероятности вида

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2}.$$

Семейство *нормальных распределений* зависит от двух параметров m и $\sigma > 0$, где

$m = \mathbf{E}\xi$— математическое ожидание,
$\sigma^2 = \mathbf{D}\xi$— дисперсия нормальной случайной величины.

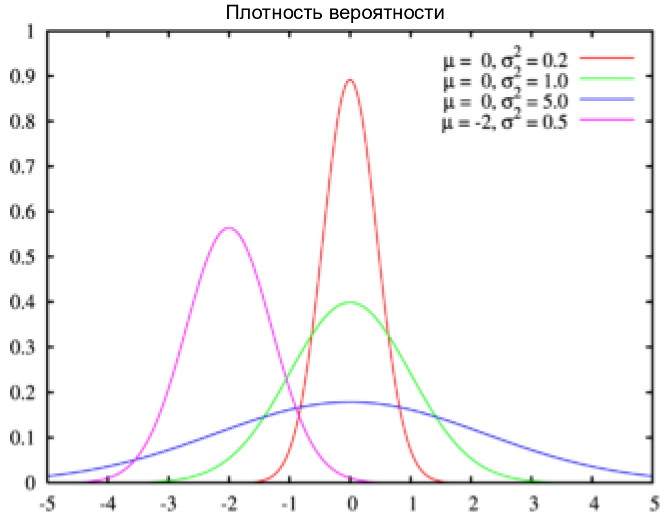
Кривая нормальной плотности $y = f(x)$ симметрична относительно ординаты, проходящей через точку $x = m$, и имеет в этой точке единственный максимум, равный $1/\sqrt{2\pi}\sigma$. Изменение σ меняет форму кривой: с уменьшением σ кривая нормальной плотности становится более островершинной; изменение m при постоянном σ вызывает смещение кривой вдоль оси абсцисс, не меняя формы кривой. Площадь, заключённая под кривой нормальной плотности всегда равна единице.

При $m = 0$, $\sigma = 1$ распределение называется *стандартным нормальным распределением* и соответствующая функция распределения принимает вид

$$\Phi(x) = \int_{-\infty}^{x} e^{-u^2/2} du.$$

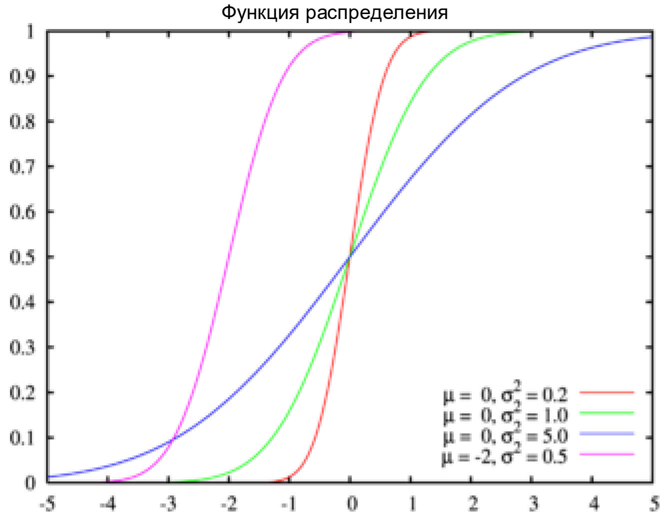
Внешний вид функции плотности вероятности и функции распределения случайной нормально распределенной величины показан на рисунке:

Плотность вероятности



Зелёная линия соответствует стандартному нормальному распределению

Функция распределения



Цвета на этом графике соответствуют графику наверху

Гипотеза о нормальном распределении применяется в большом числе практических задач. Теоретическое объяснение этого факта дают теоремы теории вероятностей, из которых следует, что нормальное распределение служит хорошим приближением каждый раз, когда исследуемая случайная величина представляет собой сумму большого числа независимых случайных величин, максимальная из которых мала по сравнению со всей суммой. В рыночной экономике многие показатели (например, суммарные объемы продаж за период, себестоимость продукции, количество посещений магазинов, цены товаров на бирже) описываются с помощью нормального распределения.

Метод Монте-Карло

Метод Монте-Карло (методы Монте-Карло, ММК) — общее название группы численных методов, основанных на получении большого числа реализаций стохастического (случайного) процесса, который формируется таким образом, чтобы его вероятностные характеристики совпадали с аналогичными величинами решаемой задачи. Используется для решения задач в различных областях физики, химии, математики, экономики, оптимизации, теории управления и др. суть метода состоит в том, что с помощью датчика случайных чисел многократно генерируются случайные колебания входных показателей на входе модели некоторого процесса, и изучаются статистические характеристики итоговых показателей на выходе этой модели.

Датой рождения метода Монте-Карло принято считать 1949 г., когда американские ученые Н. Метрополис и С. Улам опубликовали статью «Метод Монте-Карло», в которой систематически его изложили. Название метода связано с названием города Монте-Карло, где в игорных домах (казино) играют в рулетку - одно из простейших устройств для получения случайных чисел, на использовании которых основан этот метод.

ЭВМ позволяют легко получать так называемые псевдослучайные числа (при решении задач их применяют вместо случайных чисел); это привело к широкому внедрению метода во многие области науки и техники (статистическая физика, теория массового обслуживания, теория игр и др.). Метод Монте-Карло используют в том числе для исследования различного рода сложных экономических систем.

Сущность метода Монте-Карло состоит в следующем: требуется найти значение a некоторой изучаемой величины. Для этого выбирают такую случайную величину $\boldsymbol{X}$, математическое ожидание которой равно a:

$$M(X) = a.$$

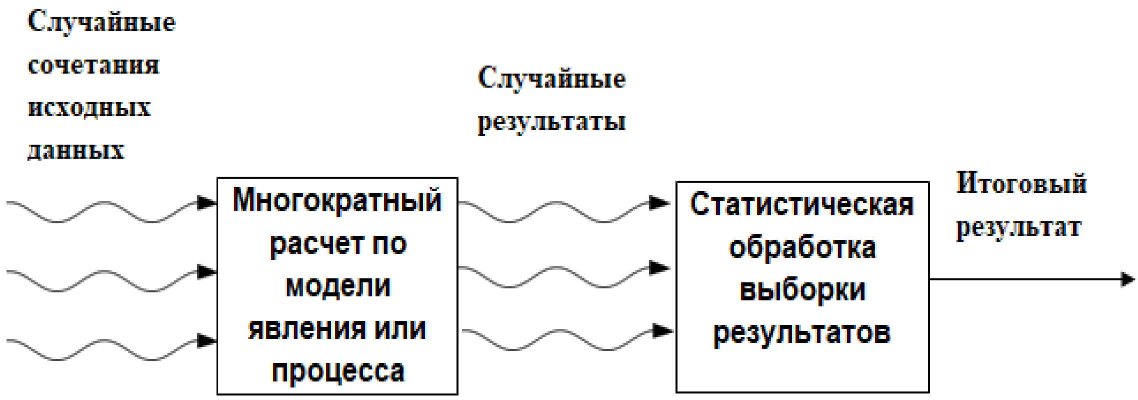
Практически же поступают так: производят n испытаний, в результате которых получают n возможных значений X, вычисляют их среднее арифметическое и принимают его в качестве оценки (приближенного значения) a^* искомого числа a.

Поскольку метод Монте-Карло требует проведения большого числа испытаний, его часто называют также *методом статистических испытаний*. Отыскание возможных значений случайной величины X (моделирование) называют «разыгрыванием случайной величины». Общая схема проведения расчетов с использованием модельных расчетов показана на схеме ниже:

Схема проведения модельного эксперимента

А) Детерминированный случай (без участия случайных величин):

Б) Метод Монте-Карло (с участием случайных величин):



Рассмотрим числовой пример, иллюстрирующий применение метод статистических испытаний.

Система контроля качества продукции состоит из трех приборов. Вероятность безотказной работы каждого из них в течение времени *T* равна 5/6. Приборы выходят из строя независимо друг от друга. При отказе хотя бы одного прибора вся система перестает работать. Найти вероятность *Pотк* того, что система откажет за время *T*.

Решим задачу аналитически и методом статистических испытаний.

Аналитическое решение. Событие *A* (выход из строя хотя бы одного из трех приборов за время Т) и событие $\bar{A}$ (ни один из трех приборов не выйдет из строя за время Т) - противоположные. Вероятность $P(\bar{A}) = (5/6)^3$. Искомая вероятность:

$$p_{отк} = P(A) = 1 - P(\overline{A}) = 1 - \left(\frac{5}{6}\right)^3 = 0{,}42.$$

Теперь решим задачу методом статистических испытаний. Напомним, что при использовании данного метода возможны два подхода: либо непосредственно проводят эксперименты, либо имитируют их другими экспериментами, имеющими с исходными одинаковую вероятностную структуру. В условиях данной задачи «натуральный» эксперимент- наблюдение за работой системы в течение времени *T*. Многократное повторение этого эксперимента может оказаться трудноосуществимым или просто невозможным. Заменим этот эксперимент другим.

Для определения того, выйдет или не выйдет из строя за время *T* один из приборов, будем подбрасывать три игральных кости. Если на одной из костей выпадет одно очко, то будем считать, что прибор вышел из строя; если два, три, ..., шесть очков, то будем считать, что прибор работал безотказно. Вероятность того, что выпадет одно очко, так же как и вероятность выхода прибора из строя, равна 1/6, а вероятность того, что выпадет любое другое число очков, как и вероятность безотказной работы прибора, равна 5/6.

Чтобы определить, откажет или нет вся система за время *T*, будем подбрасывать три игральные кости (или одну кость три раза). Если хотя бы на одной из трех костей выпадет одно очко, то это будет означать, что система отказала.

Повторим испытание, состоящее в подбрасывании трех игральных костей, много раз подряд и найдем отношение числа *m* «отказов» системы к общему числу *n* проведенных испытаний. Вероятность отказа

$$p_{отк} = \frac{m}{n}.$$

Попробуем решить более сложную задачу: определить с помощью метода Монте-Карло ожидаемое значение величины себестоимости продукции. Как известно, себестоимость продукции складывается из многих видов затрат. Будущие удельные расходы отдельных видов расходов, являются, вообще говоря, случайными величинами. В таблице ниже приведены статьи затрат и возможные границы, в которых меняются отдельные виды расходов:

Статьи затрат	Минимальная оценка затрат	Максимальная оценка затрат
1. Сырье и материалы	Xmin	Xmax
2. Заработная плата	Ymin	Ymax
3. Амортизация ОС	Zmin	Zmax
4. Прочие расходы	Qmin	Qmax
Итого себестоимость, S:	?	

Пусть у нас имеется датчик случайных чисел, с помощью которого можно получать случайные числа, равномерно распределенные в интервале между 0 и 1. Обозначим конкретные случайные числа, полученные в интервале [0; 1] через α_i, β_i, γ_i, δ_i. С помощью датчика случайных чисел нужно получить длинный ряд случайных чисел: i=1,…,N.

Будем моделировать случайные значения статей затрат следующим образом:

$Xi = Xmin + \alpha_i (Xmax – Xmin)$,

$Yi = Ymin + \beta_i (Ymax – Ymin)$,

$Zi = Ymin + \gamma_i (Ymax – Ymin)$,

$Qi = Qmin + \delta_i (Qmax – Qmin)$, i=1,…,N.

В итоге каждого статистического испытания получается уникальное значение себестоимости:

$Si=Xi+Yi+Zi+Qi$, i=1,…,N.

Накопив достаточно большое количество реализаций случайной величины Si, можно получить в итоге среднее, приближенное к математическому ожиданию, ожидаемое значение себестоимости S:

$$S=\sum_i Si / N$$