

Модели систем.
Случайные величины.
Обработка результатов
моделирования.

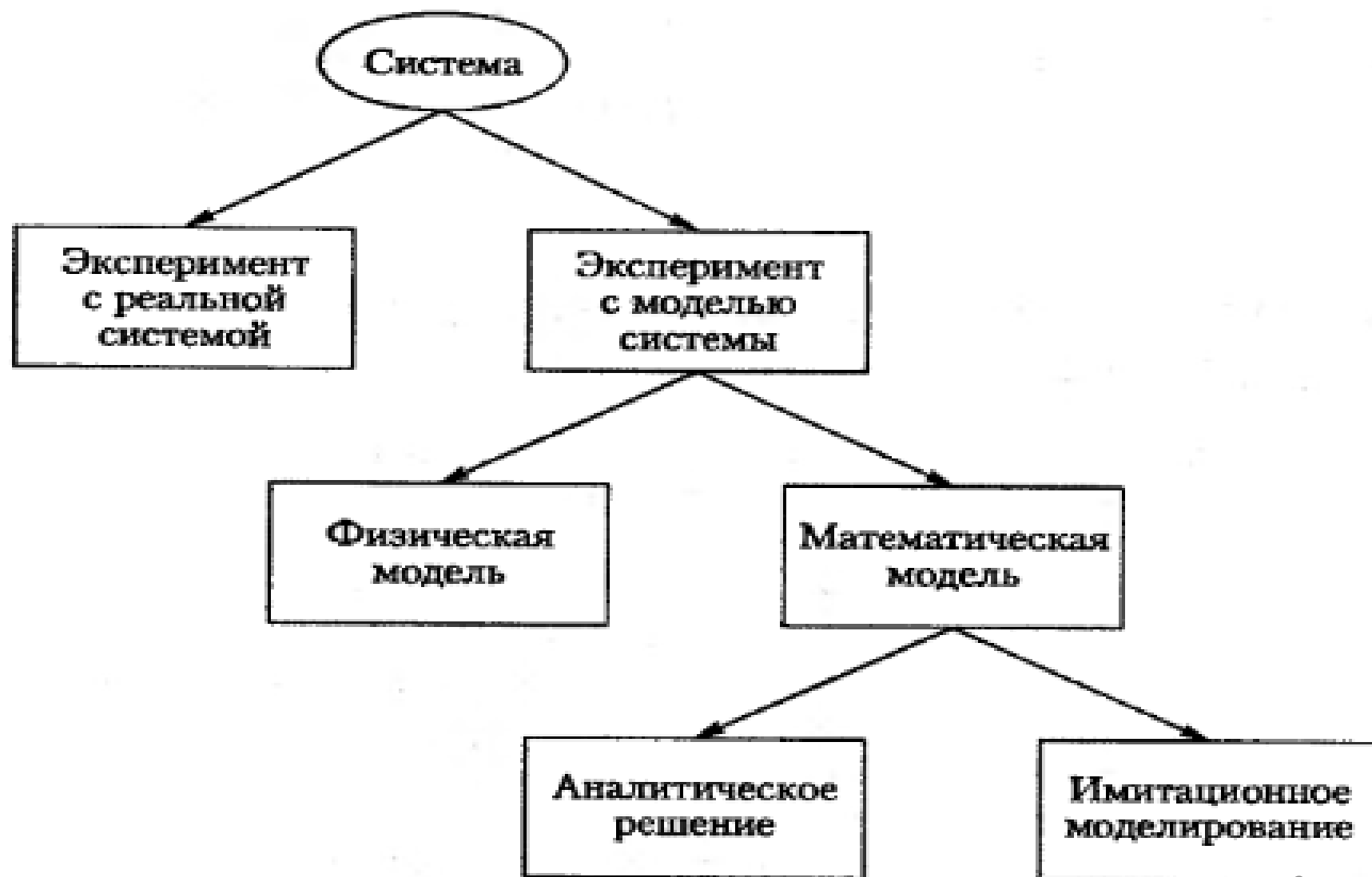
Профессор Мелентьев Олег Геннадьевич

Исследование систем

Система – совокупность объектов, функционирующих и взаимодействующих друг с другом для достижения определенной цели.

Для построения эффективных систем возникает необходимость их исследования с целью получения представления о внутренних отношениях между компонентами и производительности в заданных условиях эксплуатации.

Способы исследования систем



Задача. Оценить вероятности ошибок разной кратности в комбинации заданной длины $P(m,n)$, если вероятность ошибки в дискретном симметричном канале равна p .

Решение имитационным моделированием.

1. Используя случайный генератор с биномиальным распределением, генерируем случайный массив ошибок E с заданной вероятностью, длиной $n \cdot N_k$.
2. Задаем массив счетчиков M длиной $n+1$ и инициализируем нулями.
3. Организуем цикл в котором:
 Последовательно анализируем по n элементов массива ошибок и считаем количество ошибок в текущей комбинации m .
 Увеличиваем значение счетчика $M[m]$ на единицу
4. Делим значения всех счетчиков на количество проанализированных комбинаций

Аналитическая модель P(m,n)

Вероятность безошибочного приема блока длиной n:

$$P(0, n) = (1 - P_{\text{ош}})^n$$

Вероятность, что ошибка только в 1-м элементе:

$$P^*(1, n) = P_{\text{ош}} (1 - P_{\text{ош}})^{n-1}$$

Вероятность, что ошибка только во 2-м элементе:

$$P^*(1, n) = (1 - P_{\text{ош}}) P_{\text{ош}} (1 - P_{\text{ош}})^{n-2} = P_{\text{ош}} (1 - P_{\text{ош}})^{n-1}$$

Вероятность, что в блоке длиной n элементов одна ошибка:

$$P^*(1, n) = n \cdot P_{\text{ош}} (1 - P_{\text{ош}})^{n-1}$$

Вероятность, что ошибка в m элементах:

$$P(m, n) = C_n^m \cdot P_{\text{ош}}^m (1 - P_{\text{ош}})^{n-m}$$



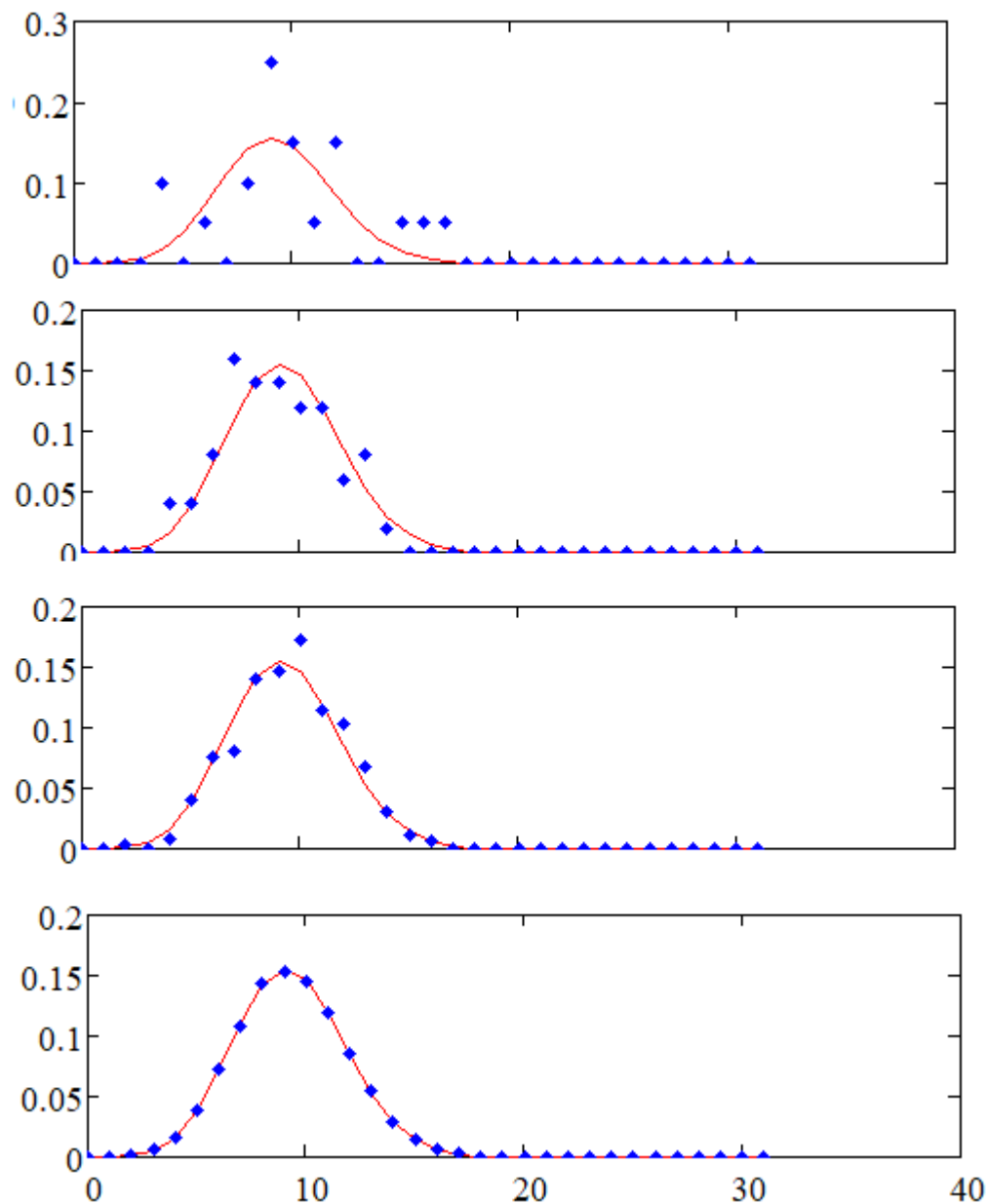
$$C_n^m = \frac{n!}{m!(n-m)!}$$

$n = 31$, $p = 0,3$.

Результаты, полученные имитационным моделированием при 20, 50, 500 и 50000 комбинаций (синие точки).

Аналитическая кривая красным цветом.

Наглядно видна зависимость точности оценок от объема выборки при имитационном моделировании.



Классификация имитационных моделей

- Статическая модель – не меняется во времени
- Динамическая модель изменяется во времени
- Детерминированная. Не содержит случайных компонентов. (Система уравнений.)
- Стохастическая – содержит случайные компоненты и выдает случайные результаты. Эти случайные результаты являются оценками истинных характеристик модели.
- Дискретные и непрерывные модели.

Дискретно-событийное моделирование.

Состояние системы – совокупность переменных необходимых для ее описания в определенный момент времени.

В дискретных системах переменные состояния меняются мгновенно в некоторые моменты времени.

Событие, возникающее в некоторый момент времени, мгновенно изменяет состояние системы.

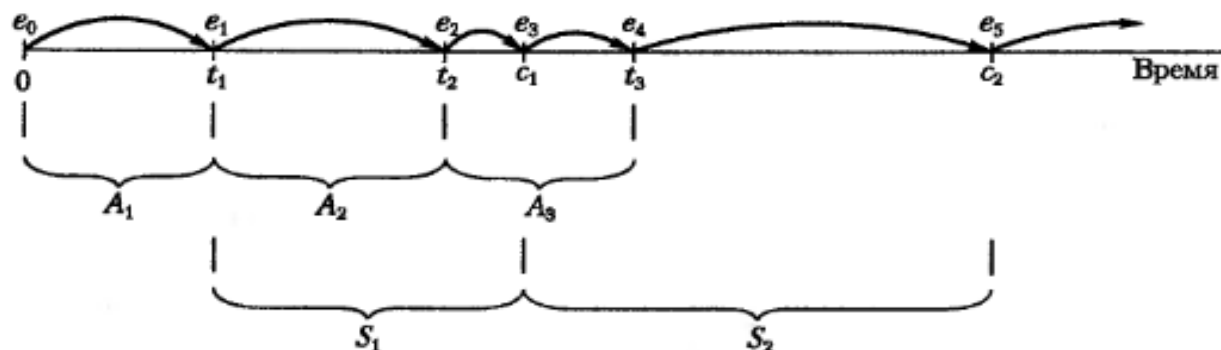
Некоторые события не меняют состояние, например, событие останова или вычисления результата моделирования.

Динамическая природа дискретно-событийного моделирования требует слежения за текущим значением имитационного времени и его модификации на протяжении всего моделирования.

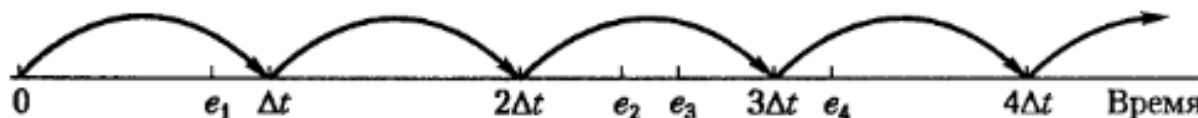
Переменная «часы модельного времени».

Существует два основных подхода к продвижению часов модельного времени:

– продвижение времени от события к событию;



– продвижение времени с постоянным шагом.



Классическим примером дискретно – событийного моделирования является имитационная модель системы массового обслуживания.

Событиями являются поступление требования в систему и покидание системы.

Генерируются:

Интервалы времени между поступающими требованиями, как независимые и одинаково распределенные случайные величины A .

Время обслуживания каждого требования, так же одинаково распределенные случайные величины B , которые не зависят от A .

Если при поступлении требования устройство свободно, требование начинает обслуживаться немедленно, минуя очередь, т.е. с нулевой задержкой.

Если устройство занято, требование ставится в очередь. При завершении обслуживания текущего требования выбирается следующее из очереди по дисциплине FIFO.

Моделирование продолжается до останова.

Оцениваются:

средняя задержка в очереди,

среднее число требований в очереди

и коэффициент занятости прибора обслуживания;

Преимущества имитационного моделирования

Большинство реальных сложных систем со стохастическими элементами не может быть точно описана аналитическими моделями, поэтому имитационное моделирование становится единственно возможным методом их исследования.

При моделировании мы можем обеспечить лучший контроль условий эксперимента, чем при экспериментировании с самой реальной системой. Моделирование позволяет оценить показатели существующей системы при фиксированных условиях эксплуатации

Позволяет сравнивать предлагаемые альтернативные варианты проектов системы для определения лучшего.

Моделирование позволяет изучить длительный интервал функционирования систем в сжатые сроки или, наоборот, изучить более подробно работу системы в развернутый интервал времени.

Недостатки

Каждый прогон имитационной модели позволяет получить только оценки характеристик модели для одного набора входных параметров. Всегда требуется много прогонов, что очень затратно при оптимизации. (Если возможно получение аналитической модели для оптимизации лучше использовать её.)

Разработка сложной имитационной модели требует много времени.

Если полученная модель не адекватна изучаемой системе результаты моделирования будут иметь мало полезной информации (какими бы убедительными они не казались).

Случайные величины и их свойства

В результате эксперимента или прогона имитационной модели мы получаем случайное число, являющееся оценкой некоторого параметра системы.

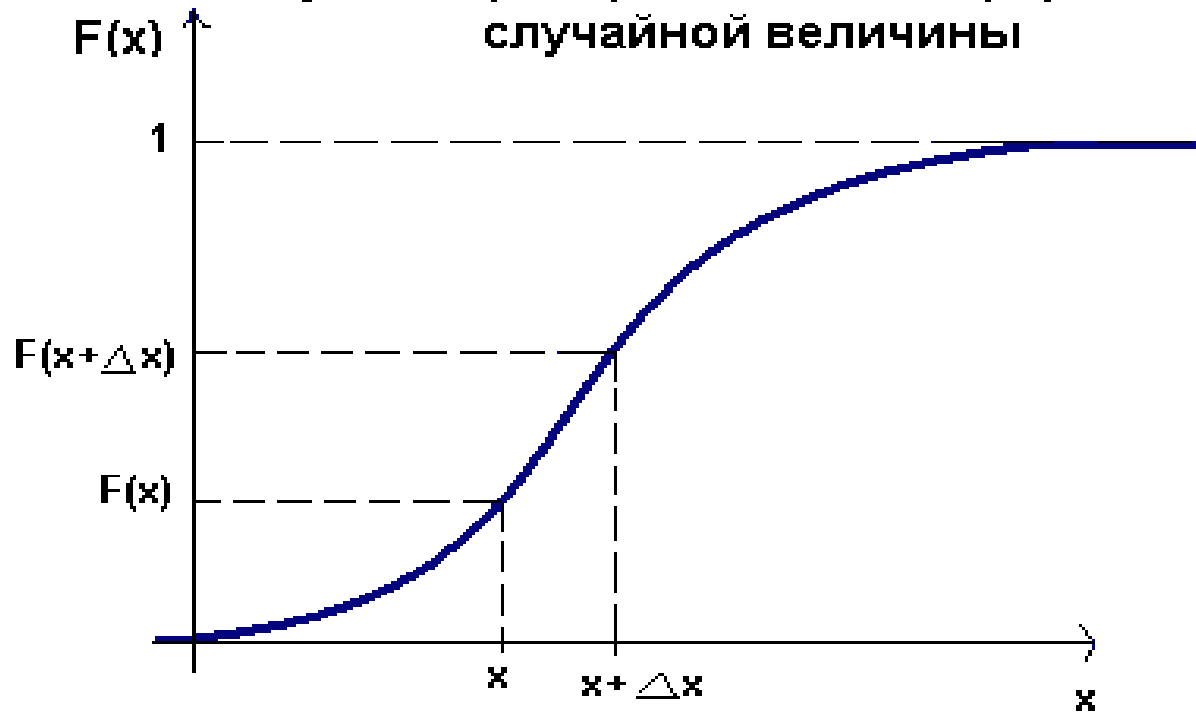
Обозначим случайную величину X , а значение которое она принимает x

Распределение случайной величины определяется интегральной функцией распределения вероятностей.

$$F(x) = P(X \leq x) \quad \text{для} \quad -\infty < x < \infty$$

Это вероятность того, что случайная величина получит значение не превышающее x

Функция распределения непрерывной случайной величины



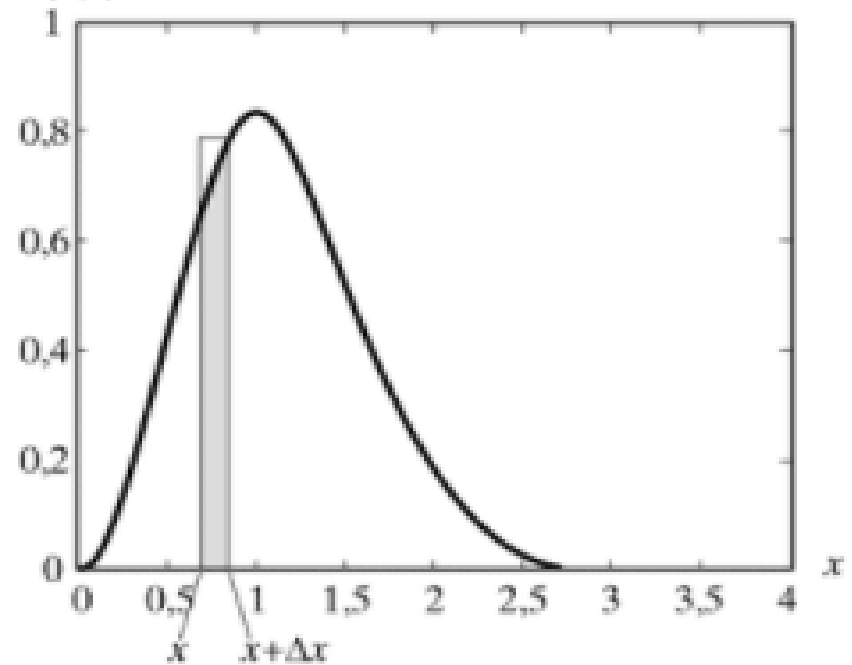
Функция распределения вероятностей может принимать значения от 0 до 1 и является неубывающей.

Для случайной величины существует функция плотности распределения непрерывной случайной величины $f(x)$

$$P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} f(y) dy$$

Площадь под функцией плотности равна единице

$$\int_{-\infty}^{\infty} f(y) dy = 1$$



Можно сказать, что плотность является производной от интегральной функции распределения вероятности.

Вероятность попадания случайной величины в интервал $I = [a, b]$

$$P(X \in I) = \int_a^b f(y) dy = F(b) - F(a)$$

Случайная величина имеет математическое ожидание

μ или $E(X)$

$$\mu = \begin{cases} \sum_{j=1}^{\infty} x_j p(x_j) & \text{для дискретной} \\ \int_{-\infty}^{\infty} x \cdot f(x) dx & \text{для непрерывной} \end{cases}$$

Для оценки мат.ожидания по результатам n прогонов модели мы будем использовать выборочное среднее

$$\overline{X}(n) = \frac{\sum_{i=1}^n X_i}{n}$$

Дисперсия случайной величины X обозначается σ^2 или $Var(X_i)$

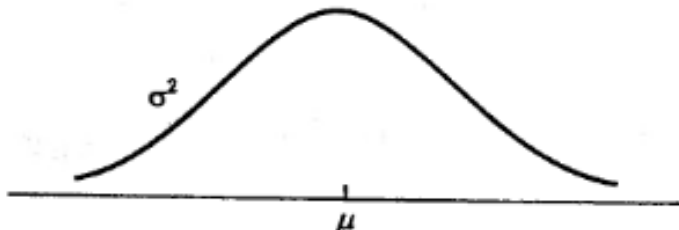
$$\sigma^2 = E[(X_i - \mu_i)^2] = E(X_i)^2 - \mu_i^2$$

Дисперсия характеризует рассеяние случайной величины по отношению к среднему значению.

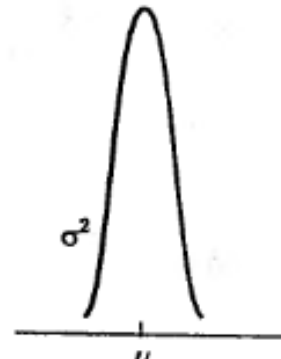
Для оценки дисперсии по результатам n прогонов модели мы будем использовать выражение

$$S^2(n) = \frac{\sum_{i=1}^n (X_i - \bar{X}(n))^2}{n-1}$$

большая



маленькая



Показателем зависимости двух случайных величин является ковариация

$$\text{Cov}(X, Y) = E[(X_i - \mu_x)(Y_j - \mu_y)] = E(XY) - \mu_x \mu_y$$

Если случайные величины независимы, то ковариация равна нулю.

(обратное утверждение не всегда верно!!!)

Если $\text{Cov}(X, Y) > 0$

величины считаются положительно коррелированы. В этом случае если одна величина превышает свое математическое ожидание, то и другая с высокой степенью вероятности тоже превысит свое.

Ковариация имеет размерность квадрата случайной величины, что не удобно.

Поэтому в качестве основного показателя линейной зависимости используют корреляцию

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{\sigma_i^2 \sigma_j^2}}$$

Корреляция безразмерная величина, принимает значение от -1 до 1 и имеет тот же знак, что и ковариация.

Доверительный интервал

В результате моделирования мы получаем статистическую оценку какого-либо параметра системы. Хотелось бы понимать насколько полученное случайное число близко к истинному значению оцениваемого параметра.

Для ответа на этот вопрос используют понятия доверительный интервал и доверительная вероятность.

Доверительным называют интервал, который покрывает неизвестный параметр с заданной надёжностью.

Положим для оценки математического ожидания нормально распределенной случайной величины мы провели 10 наблюдений в которых получили следующие результаты:

1.2 1.5 1.68 1.89 0.95 1.49 1.58 1.55 0.5 1.09.

$$\bar{X}(10) = \frac{\sum_{i=1}^{10} X_i}{10} = 1,343$$

$$S^2(10) = \frac{\sum_{i=1}^{10} (X_i - \bar{X}(10))^2}{9} = 0.168$$

$$\bar{X}(n) \pm t_{n-1, 1-\alpha/2} \sqrt{\frac{S^2(n)}{n}}$$

$t_{n-1, 1-\alpha/2}$ – верхняя критическая точка t распределения с $(n-1)$ степенями свободы.

Зададимся доверительной вероятностью 0,95 ($\alpha=0,05$). Тогда при степени свободы $n-1=9$, получим $t_{n-1, 1-\alpha/2} = 1,833$

Тогда доверительный интервал $[1.106 - 1.58]$ покрывает истинное значение мат ожидания с вероятностью 0,95.

Можно сказать, что истинное значение мат ожидания попадет в указанный интервал с вероятностью 0,95.

В Mathcad есть функция $qt(p, d)$ - Возвращает обратное кумулятивное распределение Стюдента со степенями свободы d и доверительной вероятностью p .

Уточнение! Ранее мы рассматривали вероятность попадания случайной величины в заданный неслучайный интервал. Здесь дело обстоит иначе: величина не случайна, зато случаен интервал. Случайно его положение на оси абсцисс, определяемое его центром; случайна вообще и длина интервала, так как величина вычисляется, как правило, по экспериментальным данным. Поэтому в данном случае лучше будет толковать величину не как вероятность «попадания» точки в интервал, а как вероятность того, что случайный интервал накроет точку.

Построение Гистограмм.

Для определения закона распределения наблюдаемой случайной величины целесообразно сгруппировать данные и построить гистограмму, которая позволяет наглядно увидеть очертание плотности распределения вероятностей.

Для построения гистограмм в маткаде есть функция `hist(intvls,A)`.

Данная функция возвращает вектор, элементами которого являются частоты, с которыми значения данных попадают в интервалы, заданные параметром `intvls`.

Параметр `intvls` может быть вектором конечных точек интервалов или целым числом подинтервалов равной длины.

Пусть полученные случайные данные находятся в массиве A .
Определяем минимальное и максимальное значение случайной величины.
Формируем вектор границ интервалов Int .

Функция $hist(Int, A)$ вернет вектор элементами которого будет количество попаданий в интервалы.

Границы интервалов могут быть выбраны так, что не все точки попадут в заданные интервалы.

Точки вектора A , лежащие за пределами границ интервалов не считаются, но при получении вероятностей попадания в интервалы, делить нужно на все количество сгенерированных точек (включая не попавшие).

Для получения распределения плотности вероятности необходимо дополнительно разделить на ширину интервала.

При построении гистограммы необходимо принять меры, чтобы полученные значения плотностей вероятностей соответствовали серединам интервалов.

Пример гистограммы нормально распределенной случайной величины

