

## Вопросы:

1. Стохастическое моделирование. Основные понятия и определения. Выборочный метод. Генеральная совокупность.
  2. Стохастическое моделирование. Основные понятия и определения. Точечные оценки параметров.
  3. Стохастическое моделирование. Основные понятия и определения. Доверительное оценивание.
  4. Стохастическое моделирование. Основные понятия и определения. Проверка статистических гипотез.
  5. Распределения, используемые для описания статистических данных эксперимента.
  6. Моделирование случайных величин и векторов с заданными законами распределения.
  7. Моделирование случайных величин и векторов с нормальным законом распределения.
  8. Методы обращения и суперпозиции, их применение к моделированию скалярных случайных величин.
  9. Общие методы моделирования случайных процессов.
  10. Понятие СМО. Классификация СМО. Процесс «гибели и размножения» и его аналитические модели. Формула Литтла.  
Рассказать про телефонную компанию, Формула Литтла
  11. Аналитические модели систем массового обслуживания: Определение Марковского случайного процесса. Описание Марковских случайных процессов.
  12. Классификация инструментальных средств моделирования. История создания и особенности инструментальных средств моделирования.
- 19 лекция на Sharelatex.com
13. Среда имитационного моделирования GPSS.  
Очередь, канал (написать про эти понятия). Привязка к основным блокам. (Queue, Depart) Достаточно сложная система оперируется небольшим числом операторным блоком. Недостаток:
  14. Объектно-ориентированное моделирование. Среда моделирования MathWorks.
  15. Введение в анализ данных. Основные понятия, методы.
  16. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта. Простая линейная регрессия и простой корреляционный анализ.  
Лекция будет выслана Домрачевой, нам нужно не уходить, она вышлет.  
Моя не понимать. Почему в билете отдельно указан анализ функциональных зависимостей, а в лекциях всё об стохастических (к примеру, Корреляционный анализ изучает на основании выборки стохастическую зависимость между случайными величинами  $X, Y$ .) аaaa
  17. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта. Метод наименьших квадратов. Пример.
  18. Показатели взаимосвязи двух ненаблюдаемыми одновременно характеристиками объекта. Подходы к оцениванию параметров исследуемой зависимости. Метод моментов. Пример.

Выборки могут быть разного объема, поэтому используются методы максимального правдоподобия, либо метод моментов.

19. Показатели взаимосвязи двух ненаблюдаемыми одновременно характеристиками объекта. Подходы к оцениванию параметров исследуемой зависимости. Метод максимального правдоподобия. Пример.
20. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта.
21. Дисперсионный анализ. Многомерный дисперсионный анализ.
22. Анализ выбросов. Пример.
23. Пошаговый дискриминантный анализ. Пример.
24. Анализ главных компонент. Пример.
25. Кластерный анализ. Иерархические методы. Пример.
26. Кластерный анализ. Эвристические методы. Пример.
27. Факторный анализ. Пример.

# 1. Стохастическое моделирование. Основные понятия и определения. Выборочный метод. Генеральная совокупность.

Моделирование:

1. Имитационное моделирование — это метод исследования, при котором изучаемая система заменяется **имитационной моделью**, которая с достаточной точностью описывает реальную систему и способна имитировать эту систему при заданных входных данных. Имитация есть численный метод, используемый для проведения вычислительного эксперимента, описывающий на компьютере поведение реальной системы без учёта конкретных физико-химических явлений, протекающих в ней.
  - a. СМО
  - b. методы статистических испытаний
  - c. **стохастические методы** -- если входные данные описаны случайными функциями ( в том числе случайными процессами )
    - i. вероятностные -- ориентированы на оценку вероятности того или иного события. В ряде случаев говорят о гипотетических моделях - модель, основанная на гипотезе, ещё не подтверждённой на практике.
    - ii. статистические -- можно отнести к классу гипотетических, но ориентированных на получение оценок точечных, доверительных, исследуемых величин.
    - iii. феномены
2. Аналитическое -- учитывает наиболее существенные физико-хим. процессы, протекающие в объекте или системе, причем воздействие легко формализуемо.
3. Эвристическое -- без строгого обоснования, но дающие приемлемое решение (основанные на догадках)

**Основные задачи мат. статистики:**

1. *Задача оценки неизвестных параметров по результатам экспериментов.*  
Как правило, нужно найти функцию от результатов эксперимента, значение

которой является достаточно хорошей оценкой неизвестного, истинного значения параметра. ( $a$  -- оценка,  $\hat{a}$  -- оценка параметра)

2. Задача интервального оценивания.

Требуется построить интервал с границами  $[a_- \leq a \leq a_+]$  таким образом, чтобы он покрывал неизвестное истинное значение параметра, с заранее заданной вероятностью  $\gamma$

3. Задачи проверки стат. гипотез.

Требуется, на основе математических экспериментов, проверить то, или иное предположение относительно вида, и параметра функции распределения с.в., и функции плотности распределения с.в.

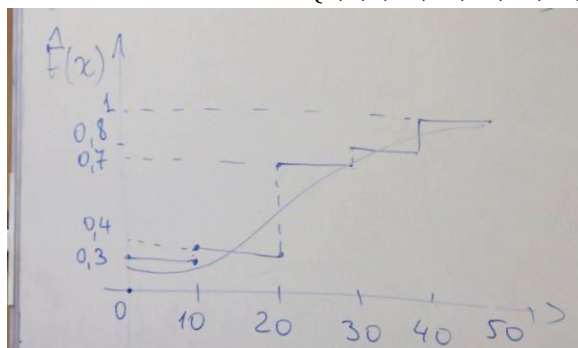
В математической статистике используется выборочная терминология, основанная на “урновой” схеме. (1)  $\{X_1, \dots, X_N\}$  - **генеральная совокупность**, объемом  $N$ . Этот набор может иметь бесконечную размерность. Из генеральной совокупности выбирается в свою очередь набор (2)  $\{x_1, \dots, x_n\}$ ,  $n \leq N$  и набор (2) называется **выборкой из генеральной совокупности** (1), объема  $n$ . Выборка может производиться с возвращением и без него. Если выборка производится с возвращением, то случайные величины в них независимы. В противном случае - зависимы. С “возвращением” тождественно равно независимой повторной выборке объема  $n$ . Терминология сохраняется и в случае с бесконечной генеральной совокупностью.

Числа выборки (2) обычно располагают в порядке убывания или возрастания (3)  $\{x^{(1)}, \dots, x^{(n)}\}$ . Набор (3) называется **вариационным рядом**. Чаще всего, в практических задачах анализируется именно вариационный ряд.

Эмпирической функцией распределения, построенной на основе выборки (3) (хотя возможно и по выборке (2)), называется функция  $\hat{F}(x) = \frac{r(x)}{n}$ ,  $n$  - общее число элементов выборки,  $r(x)$  - количество элементов  $x_i \leq x$ .

Пример:

$\{0, 0, 9, 16, 21, 24, 29, 37, 42, 48\}$



Для моделирования требуется теоретическая функция распределения случайной величины  $x$ , которая может быть оценена по эмпирической функции распределения.

$\sup_{x, n \rightarrow \infty} (|F(x) - \hat{F}_n(x)|) \rightarrow 0$ . (По теореме Гливенко-Кантелли, теоретическая и эмпирическая должны совпадать).

По эмпирической функции распределения строят эмпирический или выборочный момент.

$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  - выборочное среднее, выборочный аналог начального момента или момента ожидания.

$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  - выборочная дисперсия

$\hat{S} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$  - выборочное СКО (среднеквадратичное отклонение).

$\mu_{\xi,a} = \left( \frac{1}{n} \sum_{i=1}^n (x_i - a)^r \right)^{\frac{1}{r}}$  - выборочный момент порядка  $r$ .

В ряде случаев требуется оценить разброс результата  $R_n = |x^{(n)} - x^{(1)}|$  как размах выборки.

## 2. Стохастическое моделирование. Основные понятия и определения. Точечные оценки параметров.

Моделирование:

1. Имитационное моделирование — это метод исследования, при котором изучаемая система заменяется **имитационной моделью**, которая с достаточной точностью описывает реальную систему и способна имитировать эту систему при заданных входных данных. Имитация есть численный метод, используемый для проведения вычислительного эксперимента, описывающий на компьютере поведение реальной системы без учёта конкретных физико-химических явлений, протекающих в ней.
  - a. СМО
  - b. методы статистических испытаний
  - c. **стохастические методы** -- если входные данные описаны случайными функциями ( в том числе случайными процессами )
    - i. вероятностные -- ориентированы на оценку вероятности того или иного события. В ряде случаев говорят о гипотетических моделях - модель, основанная на гипотезе, ещё не подтверждённой на практике.
    - ii. статистические -- можно отнести к классу гипотетических, но ориентированных на получение оценок точечных, доверительных, исследуемых величин.
    - iii. феномены
2. Аналитическое -- учитывает наиболее существенные физико-хим. процессы, протекающие в объекте или системе, причем воздействие легко формализуемо.
3. Эвристическое -- без строгого обоснования, но дающие приемлемое решение (основанные на догадках)

**Основные задачи мат. статистики:**

1. *Задача оценки неизвестных параметров по результатам экспериментов.* Как правило, нужно найти функцию от результатов эксперимента, значение

которой является достаточно хорошей оценкой неизвестного, истинного значения параметра. ( $a$  -- оценка,  $\hat{a}$  -- оценка параметра)

2. *Задача интервального оценивания.*

Требуется построить интервал с границами  $[a_- \leq a \leq a_-]$  таким образом, чтобы он покрывал неизвестное истинное значение параметра, с заранее заданной вероятностью  $\gamma$

3. *Задачи проверки стат. гипотез.*

Требуется, на основе математических экспериментов, проверить то, или иное предположение относительно вида, и параметра функции распределения с.в., и функции плотности распределения с.в.

### Задача оценки неизвестных параметров по результатам эксперимента.

Как правило, нужно найти функцию от результата эксперимента, значения которой являются достаточно хорошей оценкой неизвестного истинного значения параметра.

$a$  — точный параметр

$\hat{a}$  — точечная оценка

Пусть имеется некоторая случайная величина  $\xi$  с функцией распределения  $F(x, \theta)$ ,  $f(x, \theta)$  - плотность распределения.

$$\xi \rightarrow F(x, \theta_1)$$

...

$$\xi \rightarrow F(x, \theta_n)$$

Совокупность распределения даёт параметрическое семейство распределений, в котором  $\theta$  принимает различные значения.

Вводят функцию от результатов наблюдений (4)  $\varphi = \varphi(x_1, \dots, x_n)$ , называемую статистикой. Задача построения точечной оценки параметра  $\theta$  сводится к нахождению значения статистики, такой что  $\hat{\theta} = \theta(x_1, \dots, x_n): \sup_{n \rightarrow \infty} |\hat{\theta} - \theta| \rightarrow 0$ .

Необходимо установить *эффективную оценку*, рекомендуемую в качестве результата.

Свойство оценок

$$\hat{\theta} = \theta(x_1, x_2, \dots, x_n)$$

$$\lambda = \frac{1}{\bar{x}} = \frac{1}{\frac{1}{n} \sum_{i=1}^n x_i} = \frac{n}{x_1 + \dots + x_n}$$

Оценка  $\hat{\theta}$  является не смещенной оценкой пар-ров  $\theta$ , если ее мат. ожидание  $M(\hat{\theta})$  совпадает с параметром  $\theta$ , т.е.  $M(\hat{\theta}) = \theta$ .

Оценка называется асимптотически не смещенной, если

$$\lim_{n \rightarrow \infty} M(\hat{\theta}) = \theta$$

Таким образом, на св-во оценок влияет объем выборки.

Если

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta}_n - \theta| < \epsilon\} = 1$$

$\rightarrow 1$ ,  
то говорят, что величина сходится по вероятности.

Пусть  $\hat{\theta}_n$  -- асимптотически не смещенная оценка параметра  $\theta$ , и то что ее дисперсия стремится к нулю  $\lim_{n \rightarrow \infty} S^2(\hat{\theta}_n) \rightarrow 0$ , то эта оценка является *состоятельной*.

Таким образом, асимптотическая несмещенность оценки  $\hat{\theta}_n$ , и минимизация разброса значения параметра, при  $n \rightarrow \infty$  обеспечивает состоятельность оценки (Теорема приводится без доказательства).

Если  $(\sigma^2(\hat{\theta}_n^1) = M(\hat{\theta}_n^1 - \theta)^2 \leq M(\hat{\theta}_n^2 - \theta)^2 = \sigma^2(\hat{\theta}_n^2))$ , то оценка  $\hat{\theta}_n^1$  является эффективной по сравнению с  $\hat{\theta}_n^2$ .

### Методы построения точечных оценок

Различаются 2 случая:

- моделирование производят на выборках  $\xi_1$  и  $\xi_2$ , наблюдаемых одновременно, в этом случае рассматривают пары точек, характеризующие  $n$  объектов в предположении, что характеристика  $\xi_1$ , а  $\xi_2$ .

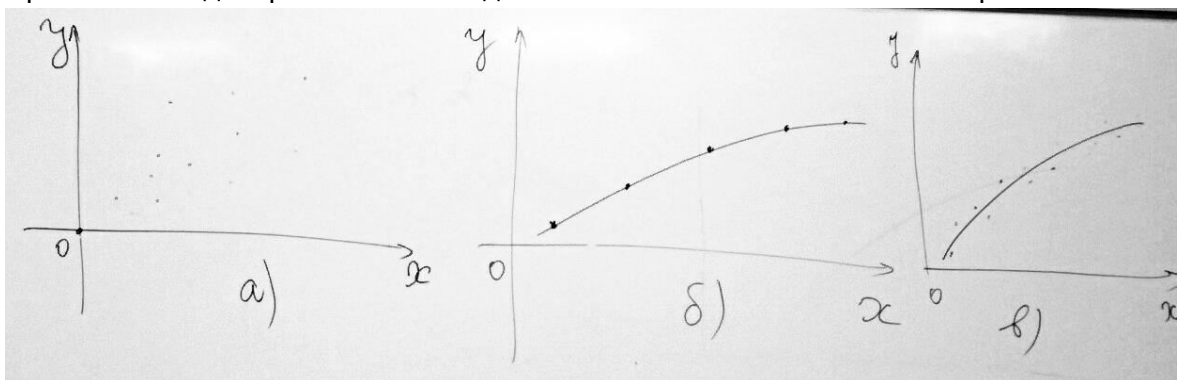
$$(x_i, y_j), i = \overline{1, n}$$

$$\xi_1: \{x_1, \dots, x_n\}$$

$$\xi_2: \{y_1, \dots, y_n\}$$

Объёмы выборок одинаковы. Как правило для определения точечных оценок параметров функции связи между случайными величинами  $x, y$ :  $y=y(x)$  используют метод наименьших квадратов.

При анализе одновременно наблюдаемых показателей возможны 3 варианта:



В случае **а)** говорят о наличии стохастической связи между двумя случайными переменными. Термин стохастическая связь был введён русским учёным Чупровым в 1926 году. *Стохастическая связь* - это такая связь, при которой значению случайной величины  $x \in X$  соответствует( $\sim$ ) одно или более значений  $y_1, \dots, y_k \in Y$ .

Таким образом, получение случайной величины  $y_i, i = \overline{1, n}$  изменяет вероятность появления других значений, но не обеспечивает их появление. Именно поэтому, говорят, что стохастическая связь не является причиной.

Рассмотрим вариант б). Функциональная связь описывается зависимостью, в которой случайной величине  $x$  из генеральной совокупности  $X$  ставится единственное значение  $y$ .

$x \in X \sim y \in Y$ , т.е. функциональная связь — причина. Предполагается отсутствие ошибок измерения.

Как правило на практике исследователи имеют дело с вариантом в):

1. присутствуют ошибки измерения
2. имеет место разброс реализаций относительно некоторой функциональной зависимости в той или иной степени достоверно описывающей входные данные. Такую функцию называют функцией тренда.

В случае достаточно тесной стохастической связи задача сводится к задаче выделения тренда и его анализа.

Как правило задачу в) решают с помощью метода наименьших квадратов.

- Исследуются не наблюдаемые одновременно параметры

$$\xi_1: \{x_1, \dots, x_n\}$$

$$\xi_2: \{y_1, \dots, y_m\}$$

### 3. Стохастическое моделирование. Основные понятия и определения. Доверительное оценивание.

Моделирование:

1. Имитационное моделирование — это метод исследования, при котором изучаемая система заменяется **имитационной моделью**, которая с достаточной точностью описывает реальную систему и способна имитировать эту систему при заданных входных данных. Имитация есть численный метод, используемый для проведения вычислительного эксперимента, описывающий на компьютере поведение реальной системы без учёта конкретных физико-химических явлений, протекающих в ней.
  - a. СМО
  - b. методы статистических испытаний
  - c. **стохастические методы** -- если входные данные описаны случайными функциями ( в том числе случайными процессами )
    - i. вероятностные -- ориентированы на оценку вероятности того или иного события. В ряде случаев говорят о гипотетических моделях - модель, основанная на гипотезе, ещё не подтверждённой на практике.
    - ii. статистические -- можно отнести к классу гипотетических, но ориентированных на получение оценок точечных, доверительных, исследуемых величин.
    - iii. феномены
2. Аналитическое -- учитывает наиболее существенные физико-хим. процессы, протекающие в объекте или системе, причем воздействие легко формализуемо.

3. Эвристическое -- без строгого обоснования, но дающие приемлемое решение (основанные на догадках)

### Основные задачи мат. статистики:

1. *Задача оценки неизвестных параметров по результатам экспериментов.*  
Как правило, нужно найти функцию от результатов эксперимента, значение которой является достаточно хорошей оценкой неизвестного, истинного значения параметра. ( $a$  -- оценка,  $\hat{a}$  -- оценка параметра)
2. *Задача интервального оценивания.*  
Требуется построить интервал с границами  $[a_- \leq a \leq a^+]$  таким образом, чтобы он покрывал неизвестное истинное значение параметра, с заранее заданной вероятностью  $\gamma$
3. *Задачи проверки стат. гипотез.*  
Требуется, на основе математических экспериментов, проверить то, или иное предположение относительно вида, и параметра функции распределения с.в., и функции плотности распределения с.в.

### Задача интервального оценивания.

$$P\{a_- \leq a \leq a^+\} = \gamma$$

Интервал строится таким образом, чтобы он покрывал неизвестное истинное значение параметра с заранее заданной вероятностью  $\gamma$  - коэффициент доверия.

### Доверительные интервалы

$$x = \{x_1, \dots, x_n\} \in X_N$$

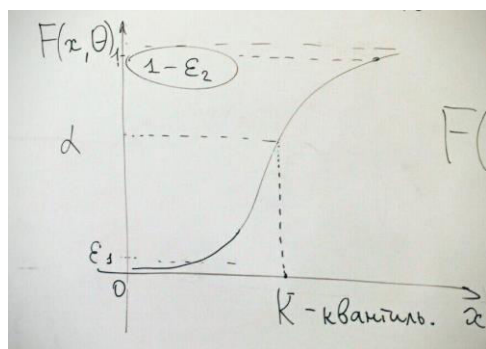
$F(x, \theta), f(x, \theta)$

Предположим, что для оцениваемого параметра  $\theta$  построен доверительный интервал  $\theta \in [\theta_-, \theta^+]$ ,  $\theta_- = \theta_-(x_1, \dots, x_n)$ ,  $\theta^+ = \theta^+(x_1, \dots, x_n)$ . Вероятность попадания в доверительный интервал  $P\{\theta_- \leq \theta \leq \theta^+\} = \gamma$ . ( $\gamma = 0.9; 0.95; 0.99; 0.995; 0.999$ )

### Методы построения доверительных интервалов

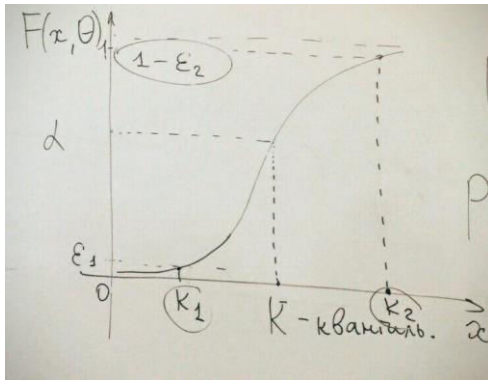
Статистикой метода называется любая функция  $T = T(x_1, \dots, x_n)$ . Если зависит ещё и от  $\theta$ , то говорят о центральной статистике, при этом выборка  $\{x_1, \dots, x_n\}$  независимая, случайная повторная из генеральной совокупности  $X_N$ . Параметр  $\theta$  - скалярная величина, но в общем случае может рассматриваться как вектор. Функция  $T$  является монотонной относительно параметра  $\theta$ .

Квантиль  $K$  уровня  $\alpha$  — функция распределения  $F(x, \theta)$ :  
 $F(K) = \alpha, F(K) = P\{x \leq K\}$ .



Зададимся двумя малыми числами  $\epsilon_1, \epsilon_2$  и определим величины  $k_1$  и  $k_2$ .





$$(*) P\{k_1 < T(x_1, \dots, x_n, \theta) \leq k_2\} = \gamma$$

$$P\{x \leq k_1\} = F(k_1) = \varepsilon_1$$

$$P\{x > k_2\} = 1 - F(k_2) = \varepsilon_2$$

$$F(k_2) - F(k_1) = 1 - \varepsilon_1 - \varepsilon_2$$

Откуда:

$$P\{x > k_1\} = 1 - \varepsilon_1$$

$$P\{x \leq k_2\} = 1 - \varepsilon_2$$

В (\*) левая граница интервала должна быть замкнутой, то есть

$$(**) P\{k_1 \leq T(x_1, \dots, x_n, \theta) \leq k_2\} = \gamma$$

Таким образом  $\gamma = 1 - \varepsilon_1 - \varepsilon_2$

$$\gamma = 1 - 2\varepsilon \Rightarrow \varepsilon = \frac{1 - \gamma}{2}$$

Далее нижняя и верхняя границы  $\theta_-, \theta^+$  доверительного интервала могут быть определены как соответственно минимальное и максимальное значения среди всех параметров  $\theta$  (среди всех возможных значений), удовлетворяющих (\*\*). Если центральная статистика монотонно возрастает по  $\theta$ , то границы доверительного интервала находят из системы:

$$(3^*) \begin{cases} T(x_1, \dots, x_n, \theta_-) = k_1 \\ T(x_1, \dots, x_n, \theta^+) = k_2 \end{cases}$$

Чаще статистику выбирают таким образом, чтобы она монотонно убывала. В этом случае  $k_1$  и  $k_2$  меняются местами. Тогда получится система:

$$(4^*) \begin{cases} T(x_1, \dots, x_n, \theta_-) = k_2 \\ T(x_1, \dots, x_n, \theta^+) = k_1 \end{cases}$$

Пример доверительного оценивания для доверительных интервалов (для нормального распределения):

$N(\mu, \sigma)$  — исследуемая функция распределения вероятности

Стандартный закон распределения:  $N_{0,1} \sim N(0,1)$  — функция распределения вероятн. значений статистики.

Для доверительного оценивания математического ожидания в качестве исходной центральной статистики используется:

$$(5^*) T = \left( \frac{\bar{x} - \mu}{\sigma} \right) \sqrt{n}.$$

Задача многопараметрического оценивания на несколько порядков (сложность повышается на порядок) при каждом новом неизвестном параметре вычислительно

сложна. В силу чего, исследователи сводят задачу многопараметрического исследования к задаче однопараметрического оценивания. Вид статистики (5\*) выбирается исследователем.

Выбранная (5\*) обладает следующими свойствами:

- Монотонно убывающая по  $\mu$  функция, имеющая стандартное нормальное распределение. Вводится  $\varepsilon$  равная  $\frac{1-\gamma}{2}$  и квантиль  $k_\varepsilon$ .

$$\begin{cases} \left( \frac{\bar{x} - \mu_-}{\sigma} \right) \sqrt{n} = u_{1-\varepsilon} = k_2 \\ \left( \frac{\bar{x} - \mu^-}{\sigma} \right) \sqrt{n} = u_\varepsilon = k_1 \\ \left\{ \begin{array}{l} \mu_- = \bar{x} - \sigma \frac{u_{1-\varepsilon}}{\sqrt{n}} \\ \mu^- = \bar{x} - \sigma \frac{u_\varepsilon}{\sqrt{n}} \end{array} \right. \end{cases}$$

#### 4. Стохастическое моделирование. Основные понятия и определения. Проверка статистических гипотез.

Моделирование:

1. Имитационное моделирование — это метод исследования, при котором изучаемая система заменяется **имитационной моделью**, которая с достаточной точностью описывает реальную систему и способна имитировать эту систему при заданных входных данных. Имитация есть численный метод, используемый для проведения вычислительного эксперимента, описывающий на компьютере поведение реальной системы без учёта конкретных физико-химических явлений, протекающих в ней.
  - a. СМО
  - b. методы статистических испытаний
  - c. **стохастические методы** -- если входные данные описаны случайными функциями ( в том числе случайными процессами )
    - i. вероятностные -- ориентированы на оценку вероятности того или иного события. В ряде случаев говорят о гипотетических моделях - модель, основанная на гипотезе, ещё не подтверждённой на практике.
    - ii. статистические -- можно отнести к классу гипотетических, но ориентированных на получение оценок точечных, доверительных, исследуемых величин.
    - iii. феномены
2. Аналитическое -- учитывает наиболее существенные физико-хим. процессы, протекающие в объекте или системе, причем воздействие легко формализуемо.
3. Эвристическое -- без строгого обоснования, но дающие приемлемое решение (основанные на догадках)

#### Основные задачи мат. статистики:

1. *Задача оценки неизвестных параметров по результатам экспериментов.* Как правило, нужно найти функцию от результатов эксперимента, значение которой является достаточно хорошей оценкой неизвестного, истинного значения параметра. ( $a$  -- оценка,  $\hat{a}$  -- оценка параметра)

2. *Задача интервального оценивания.*  
Требуется построить интервал с границами  $[a_- \leq a \leq a^+]$  таким образом, чтобы он покрывал неизвестное истинное значение параметра, с заранее заданной вероятностью  $\gamma$
3. *Задачи проверки стат. гипотез.*  
Требуется, на основе математических экспериментов, проверить то, или иное предположение относительно вида, и параметра функции распределения с.в., и функции плотности распределения с.в.

**Статистической гипотезой  $H$**  называют любое утверждение относительно функции распределения  $F(x)$  случайной величины  $x$ , касающееся видов функции распределения и значения её параметра.

Гипотезы  $H$  проверяются путем сопоставления выдвинутых предположений с результатами эксперимента, которые в статистике представляют собой  $n$  независимых повторных случайных наблюдений над случайной величиной  $x$ .

Различают две постановки задач:

1. Теоретическая функция распределения считается известной, в этом случае проверяемая гипотеза называется простой;
2. Теоретическая функция распределения неизвестна, проверку статистической гипотезы осуществляют для семейства функций распределения, отличающихся значениями параметров (сложная гипотеза).

Статистические гипотезы проверяют с помощью статистических критериев.

$H_0$  — **основная гипотеза** ( $H_0$  — нуль-гипотеза)

$H_A$  — **альтернативная гипотеза** ( $H_1$  — первая гипотеза)

Гипотеза, справедливость которой проверяется в результате эксперимента, называется основной, в зависимости от того, какие альтернативы основной гипотезе возможны в предметной области, формулируют альтернативную (конкурирующую) гипотезу.

Статистический критерий — совокупность правил, позволяющих по полученной выборке принять основную гипотезу и отвергнуть альтернативную или наоборот.

Имеется общий принцип построения статистических критериев:  $\underline{x} = \frac{1}{n} \sum_{i=1}^n x_i$ .

1. задаётся некоторая функция  $S = S(x_1, \dots, x_n)$  называемая статистикой критерия
2.  $\Omega; T_0, T_{кр}, T_{и} \in \Omega$  (множество всех значений статистики  $\Omega$  разбивается на  $T_0, T_{кр}, T_{и}$ ).

$T_0$  — **множество принятия решений** (соответствует основной гипотезе)

$T_{кр}$  — **критическое множество** (соответствует альтернативной гипотезе). Если критерий таков, что критическими являются как малые, так и большие значения статистики, то критерий называется **двухсторонним**, в противном случае — **односторонним**.

Множество, отделяющее основную гипотезу и альтернативную называют **зоной индифферентности** ( $T_{и}$ ).

3. Если  $S \in T_0$ , то  $H_0$ ;  
Если  $S \in T_{кр}$ , то  $H_A$ .

В силу того, что  $S = S(x_1, \dots, x_n)$  является случайной (случайной величиной), событие  $S \in T_{кр}$  может произойти как при справедливости  $H_0$ , так и при справедливости  $H_A$ .

**Ошибкой I рода** называется возможность принятия альтернативной гипотезы  $H_A$ , когда верна  $H_0$ .

**Ошибкой II рода** называется вероятность принятия  $H_0$ , когда верна  $H_A$ .

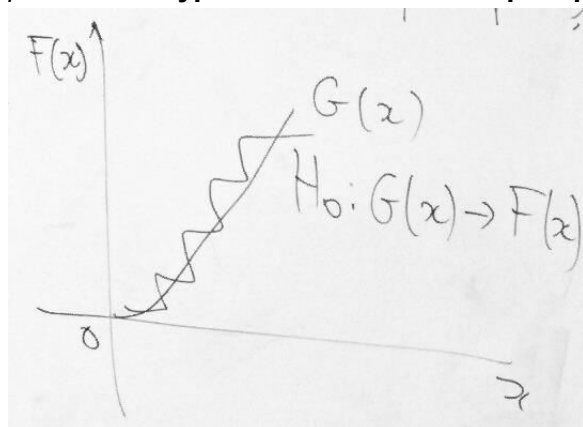
$\alpha = P(S(x_1, \dots, x_n) \in T_{кр} | H_0)$  — вероятность ошибки первого рода;

$\beta = P(S(x_1, x_2, \dots, x_n) \in T_0 | H_A)$  — вероятность ошибки второго рода.

$\alpha$  и  $\beta$  незначительные величины (в идеале вообще равны нулю) цель исследователя минимизировать эти значения, но одновременно  $\alpha$  и  $\beta$  не минимизируются (они связаны функционально). Вводят (псевдо-) величину  $\gamma = 1 - \beta$ , называемую мощностью критерия.  $\alpha$  задаётся исследователем (как правило, незначительная величина  $\sim 0.01$ ), при этом величина  $\gamma$  максимизируется.

$\alpha$  называют **размером критерия**. Чем ниже диапазон построения, тем ниже уровень доверия.

$\beta$  называют **уровнем значимости критерия**.



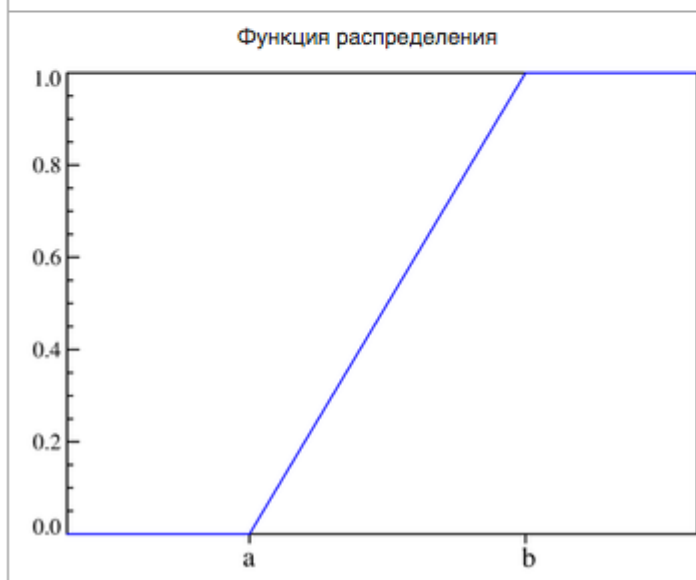
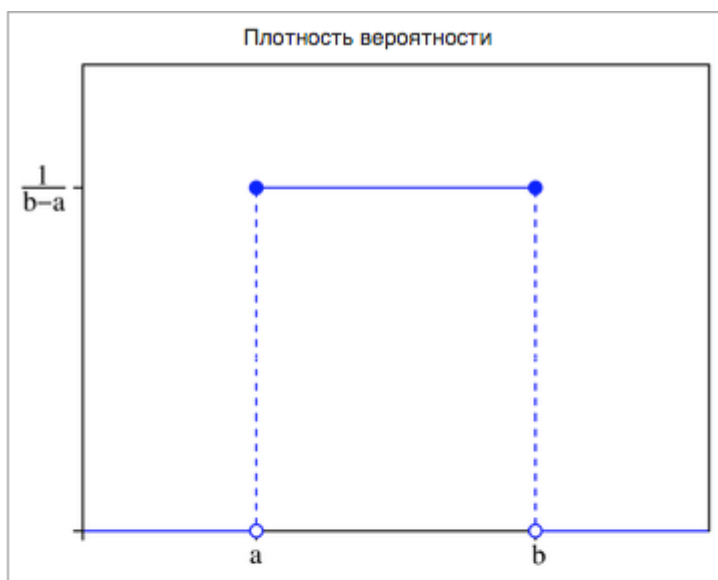
Если известен закон распределения случайной величины, статистические выводы оказываются достаточно точными, в то же время необходимо проверить соответствуют ли экспериментальные данные этому распределению. Для проверки используют критерий согласия, в частности, критерий Колмогорова-Смирнова и критерий  $\omega^2$ .

## 5. Распределения, используемые для описания статистических данных эксперимента.

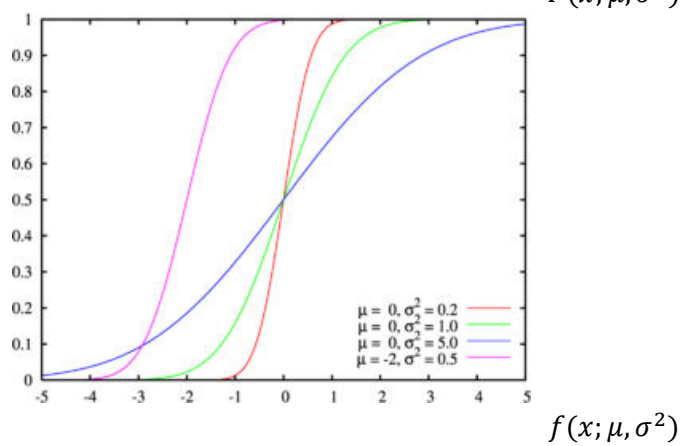
Для описания статистических данных эксперимента чаще всего используются следующие распределения ( $F$  -- функция распределения,  $f$  -- плотность):

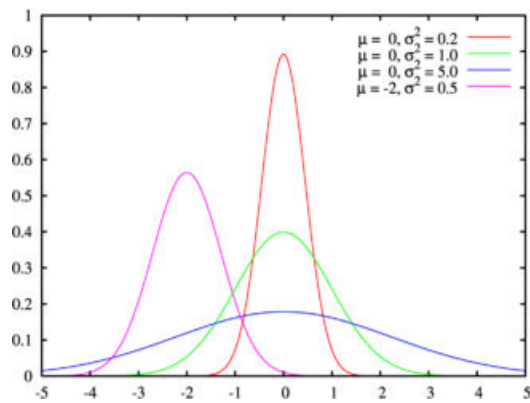
- Равномерное распределение

$$f_X(x) = \frac{1}{b-a}, x \in [a, b] \quad F_X(x) = \frac{x-a}{b-a}, x \in [a, b]$$



- Нормальное распределение (распределение Гаусса)  $N(\mu, \sigma^2)$

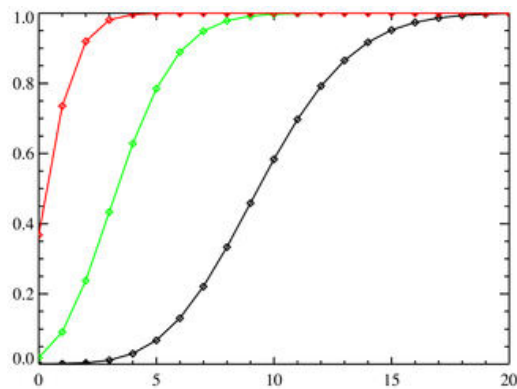




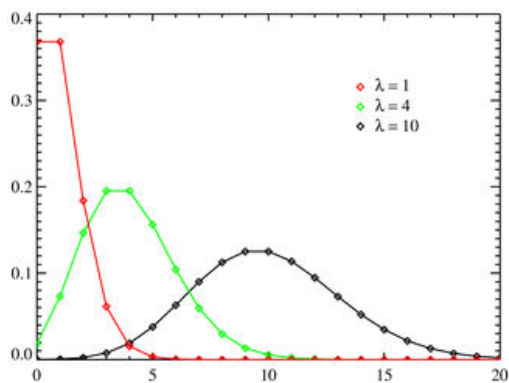
**Стандартное нормальное распределение** -- нормальное распределение с математическим ожиданием  $\mu = 0$  и стандартным отклонением  $\sigma = 1$ .

- **Распределение Пуассона**

$$F(x; \lambda, k)$$

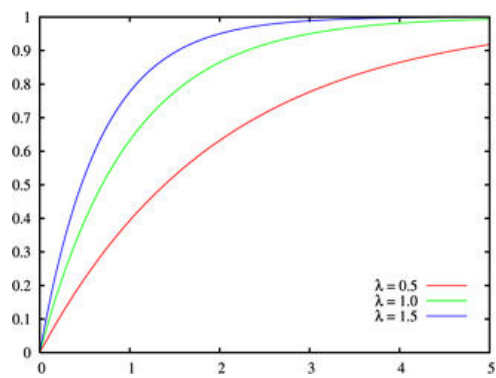


$$f(x; \lambda, k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

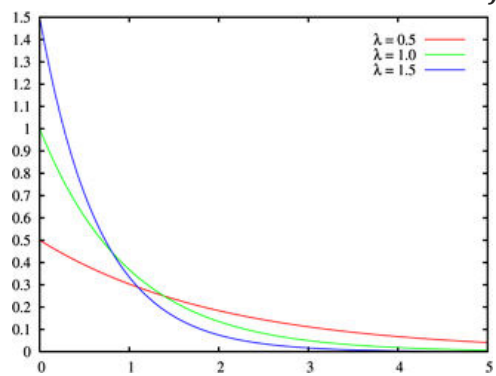


- Экспоненциальное распределение (частный случай Р. Вейбулла)

$$F(x; \lambda) = 1 - e^{-\lambda x}$$

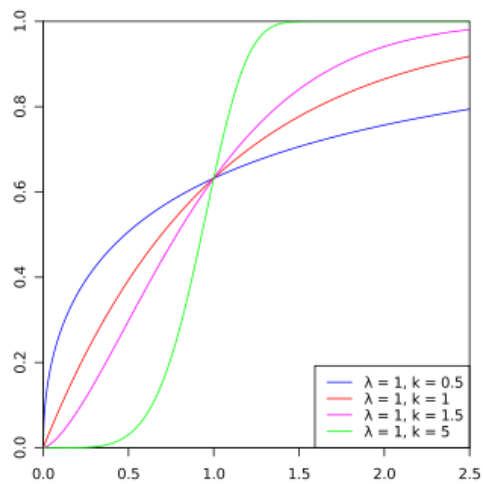


$$f(x; \lambda) = \lambda e^{-\lambda x}$$

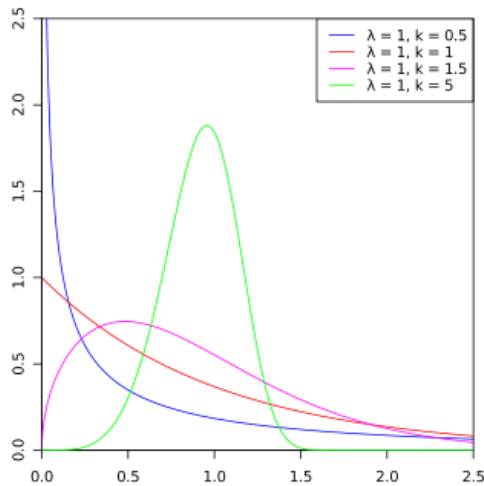


- Распределение Вейбулла

$$F(x; k, \lambda) = 1 - e^{-(x/\lambda)^k} \quad (F(x; k, \lambda) = 1 - e^{-\lambda x^k})$$



$$f(x; k, \lambda) = (k/\lambda)(x/\lambda)^{k-1} e^{-(x/\lambda)^k} \quad (f(x; k, \lambda) = \lambda k x^{k-1} e^{-\lambda x^k})$$



## 6. Моделирование случайных величин и векторов с заданными законами распределения.

**Метод Неймана.** Подход является универсальным для произвольной функции распределения.  $r_1, r_2$  -- берутся с датчиков.  $f$  -- плотность распределения.

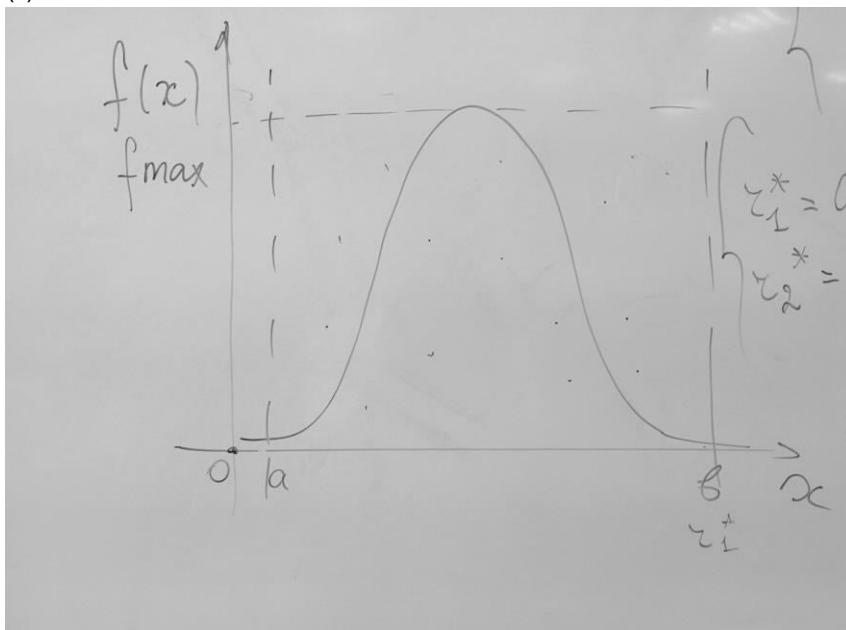
$$r_1: R[0,1] \rightarrow r_1^*: R[a, b]$$

$$r_2: R[0,1] \rightarrow r_2^*: R[0, f_{\max}]$$

$$r_1^* = a + (b - a) \cdot r_1$$

$$r_2^* = f_{\max} \cdot r_2$$

$$(*) \begin{cases} r_2^* \leq f(r_1^*) \\ r_1^* = a + (b - a) \cdot r_1 \\ r_2^* = f_{\max} \cdot r_2 \end{cases}$$



Если пары точек удовлетворяют условию ограничения системы (\*), то соответствующие им значения  $r_1^*$  составляют выборку из генеральной совокупности, распределенной по заданному закону, в противном случае — отбрасываются.



## 7. Моделирование случайных величин и векторов с нормальным законом распределения.

**Метод на основе центральной предельной теореме (ЦПТ).** Известно, что сумма нескольких независимых случайных величин, равномерно распределенных в интервале  $[0,1]$  асимптотически стремится к нормальному распределению, то есть имеет асимптотически нормальное распределение.

$$r_i: R[0,1]$$

$$x = \sum_{i=1}^n r_i, x \in N(\mu, \sigma)$$

$$[2,1,5,3,4,6]$$

$$\frac{1}{6}(1.5^2 + 2.5^2 + 1.5^2 + 0.5^2 + 0.5^2 + 2.5^2) = \frac{1}{6}(4.5 + 0.5 + 12.5) \\ = 17.5/6 \approx 2.91$$

$$\mu = 3.5, \sigma \approx 1.7$$

1. генерация шести случайных величин по равномерному распределению на  $[0,1]$  20 раз,  $x_j = \sum_{i=1}^6 r_i, j = 1, \dots, 20$ , получили  $\{x_j\}$
2. Рассчитываем выборочное среднее  $\bar{x}$  и выборочную дисперсию  $\sigma_x^2$  для выборки  $\{x_j\}$
3. Переходим от выборки  $\{x_j\}: N(\bar{x}, \sigma_x^2) \rightarrow \{x_j'\}: N(0,1)$

$$x_j' = \frac{x_j - \bar{x}}{\sigma_x}$$

4. Переходим от выборки  $\{x_j'\}: N(0,1) \rightarrow \{x_j''\}: N(\mu, \sigma^2)$  с заданными параметрами  $\mu$  и  $\sigma$ .

$$x_j'' = (x_j' + \mu)\sigma$$

## 8. Методы обращения и суперпозиции, их применение к моделированию скалярных случайных величин.

### Метод обращений

Предположим, что:  $x = \omega(y)$ .

$\omega(y) = \lambda e^{-\lambda y}$  -- плотность,  $x$  -- функция распределения

Тогда

$$e^{-\lambda y} = \frac{x}{\lambda}$$

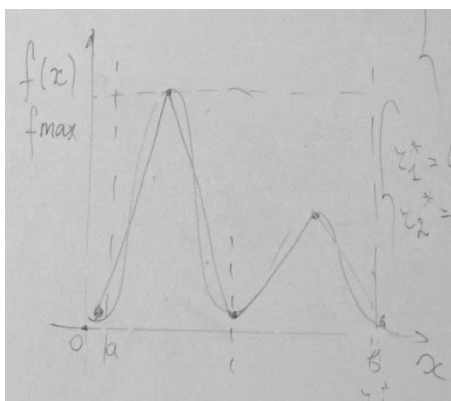
$$\ln\left(\frac{x}{\lambda}\right) = -\lambda y$$

$$y = -\frac{1}{\lambda} \ln \frac{x}{\lambda}$$

Если  $x$  -- равномерно распределена, то  $y$  -- экспоненциально.

Метод обращений реализуем для небольшого круга функций и, как правило, используется для экспоненциального закона.

### Метод суперпозиции



Если распределение имеет сложный вид, его декомпозируют на простые по форме составляющие, интерполируют простыми распределением, чаще равномерным. Такой метод называется методом суперпозиции.

## 9. Общие методы моделирования случайных процессов.

В предположении, что случайный вектор представляет собой набор зависимых случайных величин с заданным законом распределения, моделирование случайного вектора не отличается от подхода к моделированию случайных процессов, т.к. случайные процессы (функции) на ЭВМ дискретизируются.

Рассмотрим моделирование случайного процесса с показательным распределением, в этом случае получают последовательность  $n$  независимых случайных величин (с нормальным стандартным распределением), после чего строится следующая рекуррентная зависимость:

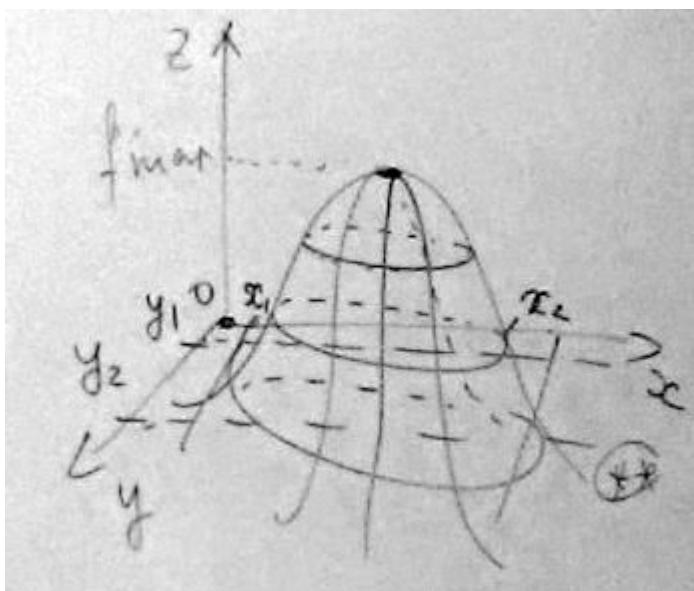
$$\xi(t_n) = \rho_n \xi(t_{n-1}) + \sqrt{1 - \rho_n^2} x[n],$$

$$\rho_n = e^{-\lambda(t_n - t_{n-1})},$$

$$\lambda = \frac{\pi}{2\Delta t}.$$

Показательное распределение (экспоненциальный закон) является примером моделирования стационарного случайного процесса с заданным законом распределения. Другие законы распределения предполагают более сложную систему описания, в силу чего в ряде случаев в задаче моделируются случайные процессы в иной классификации.

Универсальным подходом к моделированию стационарного случайного процесса с заданным законом распределения является обобщением метода фон Неймана на  $n$ -мерный вектор. Формируется набор из  $n + 1$  случайной величины, распределённый равномерно на интервале  $[0,1]$ .



Проводится преобразование:  $r_1 \rightarrow r_1^*, \dots, r_n \rightarrow r_n^*$ , где значения со звездой изменяются в любом заданном диапазоне.

$$(**) r_{n+1}^* \leq f(r_1^*, \dots, r_n^*)$$

Если выполняется условие (\*\*), то набор сохраняется и координата  $r_{n+1}$  считается распределённой по определённому закону в диапазоне  $[0,1]$ . В противном случае набор отбрасывается. Недостатками метода являются усечение функции плотности распределения и получение набора независимых случайных величин.

Здесь можно упомянуть про Марковские процессы: моделируем без учета закона распределения набора случайных величин.

**Марковский случайный процесс** — это случайный процесс, реализация которого в заданный момент времени известна. Переход процесса в новое состояние при известном значении текущей реализации не зависит от его прошлых состояний. Различают понятие марковской цепи (частный случай марковского процесса), когда пространство состояний дискретно.

Кроме того, марковский процесс рассматривают как авторегрессию первого порядка.

Определение марковского процесса по Венцель: “будущее” процесса зависит от “прошлого” только через его “настоящее”.

$$x_{n+1} = ax_n + b$$

#### **10. Понятие СМО. Классификация СМО. Процесс «гибели и размножения» и его аналитические модели. Формула Литла.**

Пусть имеется телефонный узел (*устройство, прибор* в терминологии ТМО), на котором телефонистки соединяют пары телефонных абонентов.

При небольшом количестве звонков соединение не требует ожидания.

При интенсивном увеличении говорят об *СМО с ожиданием*.

Ожидающие удовлетворения заявки (*транзакты*) помещаются в *очередь*.

Очередь может быть ограничена ( $N$  заявок).

В этом случае говорят о возможности потери заявок (всего  $L$  заявок).

Если считать заявки (транзакты) равноправными (актуальными являются только моменты поступления заявок), то поток заявок считается однородным.

Если поток однороден и после их обработки дисциплина функционирования системы не меняется, то говорят о потоке *без последдействия*.

**Формула Литтла (среднее количество заявок в системе):**

$$N = \lambda T$$

- $T$  -- время обработки;
- $\lambda$  -- интенсивность потока.

Основными элементами СМО (помимо входного потока заявок, очереди заявок) являются каналы (несколько однотипных приборов обслуживания), выходной поток обслуженных заявок, фаза обслуживания.

**Классификация:**

1. одноканальные / многоканальные
2. с очередями ожиданием / с очередями отказов
3. с приоритетом / без приоритета
4. однофазные / многофазные
5. по взаимосвязи с потоками заявок: открытые (разомкнутые) / замкнутые

Если интенсивность входного потока заявок не зависит ни от количества заявок в СМО, ни от количества уже обслуженных заявок, то говорят об *открытой СМО*. Система, сочетающая в себе свойства многоканальности, многофазности, разомкнутости классифицируется как **сеть** массового обслуживания}.

Моделируемые системы должны быть эффективными, в связи с чем вводятся показатели эффективности СМО.

- **Абсолютная пропускная способность СМО:** среднее количество заявок, обслуживаемых системой в единицу времени.
- **Относительная пропускная способность СМО:** отношение среднего количества заявок, обслуживаемых системой в единицу времени к среднему числу всех заявок, поступивших за это время в СМО.

Среднее число занятых каналов и коэффициента их занятости (отношение числа занятых каналов к общему числу каналов) формализуются как показатель эффективности СМО.

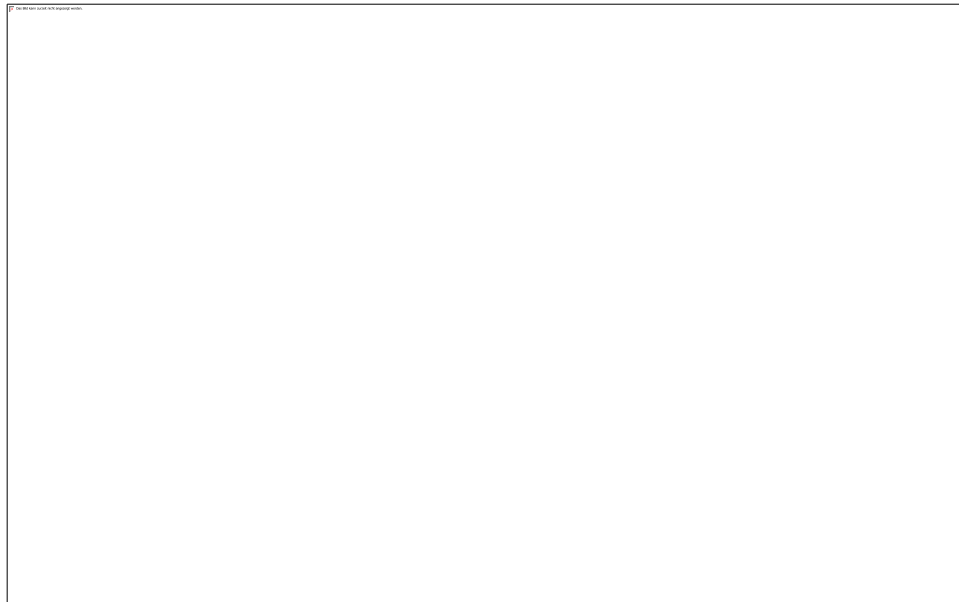
Среднее число свободных каналов и коэффициент простоя (формализуются по показателям занятости (использования) каналов).

Среднее время нахождения заявки в очереди, среднее время нахождения заявки в СМО, а также дисперсия числа заявок в очереди и в целом, в СМО, также являются показателями эффективности СМО.

Указанные показатели легко формализуются и являются аналитическими моделями СМО.

**Процессы гибели и размножения.**

Марковский процесс с дискретными состояниями называют *процессом гибели-размножения*, если он имеет размеченный граф состояний вида:



$\sum_j p_{ij} = 1$  (для состояния  $i$ )

$$S_0 \rightarrow S_1$$

Предполагают, что вероятности  $p_{ij}, j = 1, \dots, k, i = 1, \dots, k$

$$S_0 \rightarrow S_1$$

$$\lambda_{01}p_{01} = \lambda_{10}p_{10}$$

$$\{\lambda_{01}p_{01}\} + \lambda_{12}p_{12} = \{\lambda_{10}p_{10}\} + \lambda_{21}p_{21}$$

$$\{\lambda_{01}p_{01}\} + \{\lambda_{12}p_{12}\} + \lambda_{23}p_{23} = \{\lambda_{10}p_{10}\} + \{\lambda_{21}p_{21}\} + \lambda_{32}p_{32}$$

...

$$\lambda_{ij}p_{ij} = \lambda_{ji}p_{ji}$$

...

Мы можем последовательно сначала сократить элементы 01 и 10, затем 12 и 21 и т.д. В итоге получим (\*):

$$\lambda_{01}p_{01} = \lambda_{10}p_{10}$$

$$\lambda_{12}p_{12} = \lambda_{21}p_{21}$$

$$\lambda_{23}p_{23} = \lambda_{32}p_{32}$$

...

$$\lambda_{k-1,k}p_{k-1,k} = \lambda_{k,k-1}p_{k,k-1}$$

В системе (\*)  $p_{ij} \neq p_{ji}$  для большинства практических приложений. В силу чего система имеет  $k$  уравнений и  $2k$  неизвестных. Неизвестных в два раза больше чем уравнений. Для сокращения числа неизвестных в 2 раза предполагают равновероятными переходы из одного и того же состояния  $s_1$  в предыдущее и следующее. То есть равновероятным является переход в новое состояние на 1 шаг и возврат в предыдущее.

(\*\*)

$$\lambda_{01}p_1 = \lambda_{10}p_2$$

$$\lambda_{12}p_2 = \lambda_{21}p_3$$

$$\lambda_{23}p_3 = \lambda_{32}p_4$$

...

$$\lambda_{k-1,k}p_k = \lambda_{k,k-1}p_1$$

Система (\*\*) называется системой уравнений гибели-размножения. Все вероятности выражаются через вероятность  $p_1$ . Из последнего уравнения (\*\*) выразим  $p_1$ .

$$p_1(\cdot) = \frac{1}{\frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{10}\lambda_{21}} + \dots + \frac{\lambda_{k-2,k-1}\lambda_{k-3,k-2} \dots \lambda_{01}}{\lambda_{k-2,k-3} \dots \lambda_{10}}}$$

(\*\*\*)

$$p_1(\cdot)$$

$$p_2 = \lambda_{01}\lambda_{10}p_1$$

...

$$p_k = \frac{\lambda_{k-2,k-1} \dots \lambda_{01}}{\lambda_{k-2,k-3} \dots \lambda_{10}} p_1$$

Предположения, наложенные на модель относительно вероятности перехода, оказываются достаточно жёсткими. И в целом, уравнения системы (\*\*) применительно к практической задаче могут модифицироваться. В общем случае, уравнения для переходной функции марковского случайного процесса описываются как дифференциальные уравнения Колмогорова, в предположении, что функция перехода из состояния  $i$  в состояние  $j$  зависит от моментов перехода  $s$  и  $t$ .

$$1) \frac{\partial p_{ij}(s,t)}{\partial s} = \sum_k \alpha_{jk}(s) \cdot p_{kj}(s,t)$$

$$2) \frac{\partial p_{ij}(s,t)}{\partial t} = \sum_k p_{ik}(s,t) \alpha_{kj}(t)$$

$$3) \alpha_{ij}(s) = \lim_{t \rightarrow s} \frac{[p_{ij}(s,t)_{(1-\sigma_{ij})}]}{t-s}, t > s$$

$$\sigma = 1, i = j$$

$$\sigma = 0, i \neq j$$

Можно сформулировать правила составления уравнений Колмогорова по размеченному графу состояний непрерывной марковской цепи.

- Число уравнений в системе **равно** числу вершин графа;
- Вероятность состояния  $p_i$  соответствует как переходу в последующее, так и возврату в предыдущее состояние.
- Система уравнений имеет форму Коши;
- Число слагаемых в правой части равно числу дуг в графе, интерпретирующих переход из  $i$ -го состояния в любое другое, кроме самого себя;

- Переход в новое состояние соответствует слагаемое со знаком “+”;
- Возврату в предыдущее состояние соответствует слагаемое со знаком “-”;
- Каждое слагаемое представляет собой произведение вероятности  $i$ -го состояния и плотности вероятности перехода по данной дуге;
- Начальные условия для постановки задачи Коши определяются непосредственно начальным состоянием системы;

Например, в цепочке  $s_0, s_1, \dots, s_k$  старт начинается из состояния  $s_2$ :

$$p_0(0) = 0; p_1(0) = 0; p_2(0) = 1; \dots; p_k(0) = 0$$

#### 11. Аналитические модели систем массового обслуживания: Определение Марковского случайного процесса. Описание Марковских случайных процессов.

Имитация предполагает Марковость процесса. Что такое Марковский процесс, описать теорию по этому процессу.

*Марковский случайный процесс* — это случайный процесс, реализация которого в заданный момент времени известна. Переход процесса в новое состояние при известном значении текущей реализации не зависит от его прошлых состояний. Различают понятие марковской цепи (частный случай марковского процесса), когда пространство состояний дискретно.

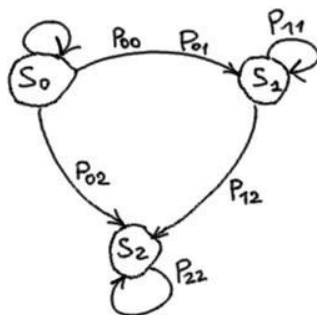
Кроме того, марковский процесс рассматривают как авторегрессию первого порядка. Определение марковского процесса по Венцель: “будущее” процесса зависит от “прошлого” только через его “настоящее”.

$$x_{n+1} = ax_n + b$$

#### Пример:

Известны вероятности перехода из одного состояния в другое ( $s_0$  --- новая мишень,  $s_1$  --- повреждённая мишень,  $s_2$  --- поражённая мишень).

Необходимо определить  $k$  --- среднее количество снарядов, необходимых для поражения цели.



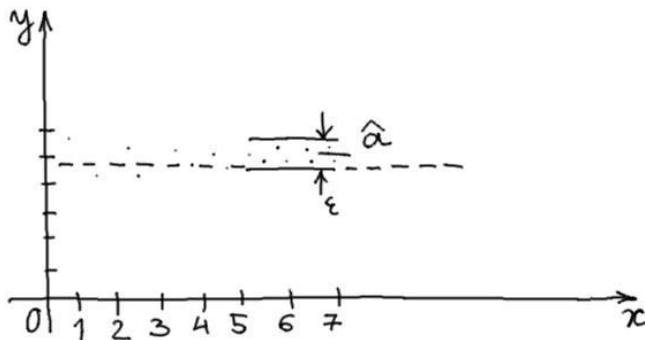
$$1) s_0 \rightarrow s_0 \rightarrow s_1 \rightarrow s_1 \rightarrow s_1 \rightarrow s_2 : 5$$

$$2) s_0 \rightarrow s_0 \rightarrow s_0 \rightarrow s_2 : 3$$

...

$$8) S_0 \rightarrow S_2: 1$$

При увеличении числа реализаций случайного процесса оцениваемая величина стремится к величине  $a$  и при бесконечном числе реализации достигает его. Для остановки счёта на ЭВМ пользуются следующим правилом:



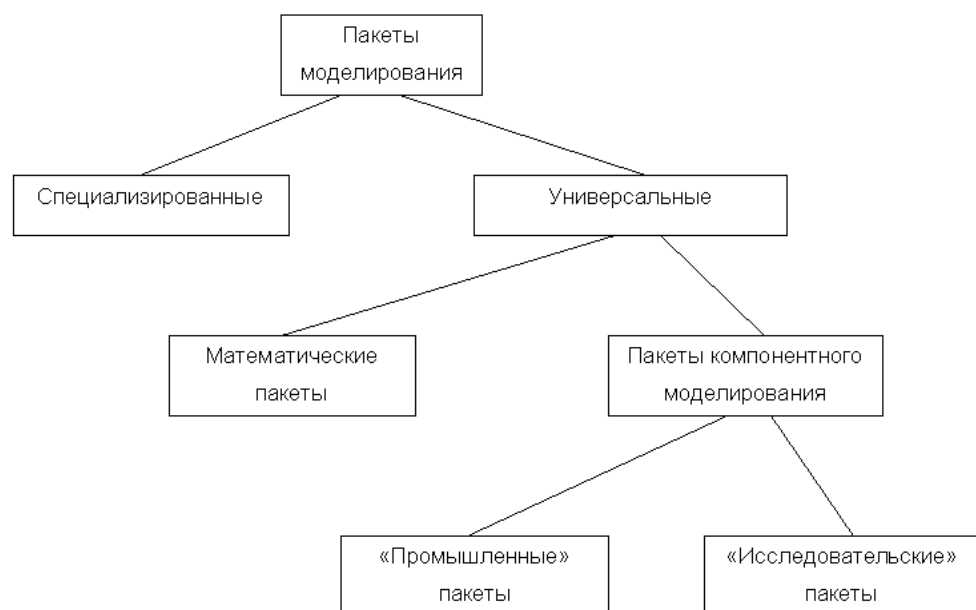
Если 3 последние точки попадают в заданный диапазон с разницей границ, не превышающей  $\epsilon$ , то счёт завершается, а оценкой  $a$  является середина заданного диапазона.

В данном случае величина равна 4.25 (диапазон  $[4, 4.5]$ ). Т. к. речь идёт о ресурсах, то округление в большую сторону.

**Ответ:** в среднем потребуется 5 снарядов. (значение на оси Y на графике выше).

**12. Классификация инструментальных средств моделирования. История создания и особенности инструментальных средств моделирования.** (latex - 63, gdrive - 43)

а. **Классификация** - в лекциях ни слова, нашёл следующую классификацию:



Инструментальные средства делятся на две группы: специализированные, ориентированные на специфические понятия конкретной прикладной области,

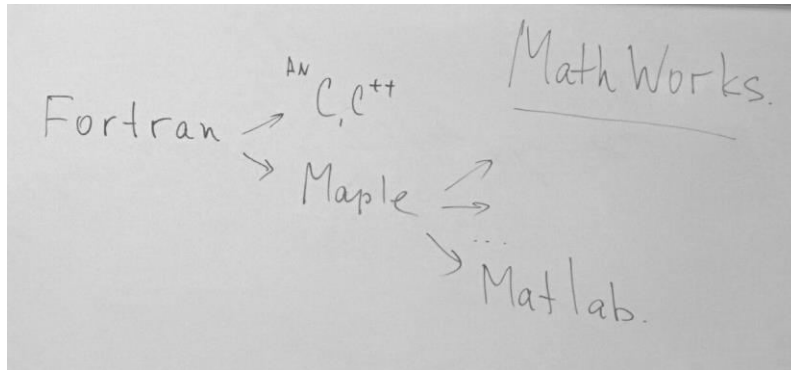


и, «универсальные» пакеты, ориентированные на определенный класс математических моделей (чаще используются на практике).

«Универсальные» пакеты обычно разделяют на «математические» пакеты (Mathematica, MathCAD, MatLab, Maple), предполагается, что математическая модель уже построена и ее требуется исследовать, и пакеты компонентного моделирования, предполагается, что описание моделируемой системы строится из компонентов, а совокупная математическая модель формируется пакетом автоматически. На практике используют компонентное, так как практически невозможно строить и поддерживать мат. модели реальных систем.

В свою очередь пакеты компонентного моделирования делятся на промышленные, предназначенные для решения сложных промышленных и научно-исследовательских задач большими производственными или научными коллективами (основной акцент на организации работ), и исследовательские, предназначенные для предварительных исследований, выполняемых отдельными учеными или проектировщиками (акцент на быстрой обратной связи).

b. **История:**



Какая-нибудь вольная интерпретация картинки, а вкратце: есть фортран, с него пиздим либы, и пользуем везде.

c. **Особенности инструментальных средств моделирования**

i. хз что она имела ввиду, ничего валидного не гуглится - смотрим лекции.

ii. я тоже хз, я просто оставлю это тут:

Решение компании MathWorks представляют современный инструментарий проектирования и имитации сложных систем на основе заданной или построенной математической модели, позволяет создавать сложные *многодоменные* объекты, включающие приложения для обработки изображений сигналов, математических расчетов, расчетов технической оснастки проекта, обмена данными, и т.д.

*Многодоменные* --- подразумевается имитация асцилография, компьютера, и для каждого оборудования производится имитация.

Mathworks предполагает описание всех аппаратных, программных и алгоритмических средств участвующих в моделировании --- как объектов, характеризуя их наборами данных, в т.ч. моделями, функциями и возникающими, в процессе функционирования системы, событиями.

Такой подход называется *{объектно ориентированное проектирование}*.

### 13. Среда имитационного моделирования GPSS.

для описания имитационной модели на языке GPSS как правило модель представляется в виде схемы, на которой отображаются элементы СМО (системы массового обслуживания) (устройства, источники, поглотители заявок и т.п.). Каждому блоку схемы ставится в соответствие оператор языка GPSS (а если в ходе выполнения не ставится - то значит такую нотацию странную выбрали).

Описание на языке GPSS есть совокупность операторных блоков, характеризующих процессы обработки заявок, инициализации, задержки при обработке, изменения параметров заявок и т.п.

каждый транзакт может иметь 12 параметров. Существует группа условных операторных блоков, исполнение которых, зависит от состояния параметров транзактов.

При продвижении заявок в системе пользуются специальными операторами передачи управления, условными и безусловными (а значит у каждого операторного блока есть метка)

#### Структура операторов GPSS

[метка] [название][операторный блок][поле переменных][; комментарии]□□□□□□□□

---- тут я выходила ----

FNK FUNCTION RN1, C, 4

Описание функции название-"FNK", Название операторного блока-"FUNCTION", поле переменных""

аргументом является случайная величина (на что указывает аргумент RN1 - зарезервированная переменная - нормально распределенная случайная величина), распределенная равномерно в диапазоне 0,1

функция является непрерывной числовой (на что указывает параметр C).

число 4 указывает на то, что функция задана таблично четырьмя точками

(0,0) (0.1, 0.8) (1.2, 1.7) (1.4, 1.9)

FNK FUNCTION \*2, D

\*2 означает, что аргументом функции является значение второго параметра транзакта, для которого вычисляется значение дискретной величины D

SEIZE PLOT

занятие устройства plot приходящим на вход транзактом, если устройство занято, то транзакт оказывается в очереди

RELEASE PLOT

освобождения устройства plot обслуженным транзактом

QUEUE R

опреатор организации очереди, дляяи очереди R увеличивается на 1 при прибытии нового транзакта

DEPART R

ENTER NA, \*2

{войти название\_накопителя второй\_оператор\_транзакта=k}  
занятие транзактом k единиц емкости в накопителе (не очереди)

LEAVE NA, \*2

если емкость накопителя существенно меньше, то для его организации используется задержка транзакта на время, определенная содержимым поля переменных (поле переменных соответствует полю операторского блока GENERATE, соответствует времени исполнения в реальной системе)

TREANSFER ,MET

безусловная передача управления оператору с меткой MET

TREANSFER BOTH, LAB1, UNN

передача управлени оператору с меткой lab1, если он не возможно - то к оператору с меткой unn, если и он невозможно, то транзакт задерживается до момента коррекции подельного времени с повторной проверкой возможности перехода

both - вероятность перехода по основной метке (SICK - равновероятный переход между метками)

во всех остальных случаях осуществляется переход на основную метку lab1

...тогда операторный блок осуществляет переход к идентификатору блока AAA5

LOOP 6, MET - не успела записать

ASSEMBLE 4

ASSIGN

SPLIT 3, MET, 6

копированиетранзактов (3 копии) исходный транзакт отправляется в следующий по порядку операторный блок, а созданные копии в блок с меткой LLL. При этом параметр "б" увеичивается на единицу, а транзакции копий на 2, на 3, на 4 соответственно

удаление транзакта из системы, при этом счетчик транзактов уменьшается на 3, моделирование завершается, если знаечение ...

SAVEVALUE PR, \*3

велечина, хранимая как третий параметр транзакта записывается в переменную PR поле .... может содержать различные элементы, в том числе ссылки и номера параметров

TEST E V7, k256, LAB

если v7 равно 256 - идем к метке lab, в противном случае моделирование осуществляется в естественном порядке  
уловие проверяется  
G - больше  
L - меньше  
E - равно  
NE - не равно  
LE - меньше или равно  
GE - больше или равно

для любой программы gpss должна быть задана начальная карта программы, если разработчик намерен выполнить прогон модели (получить отчет с результатами статистического анализа) то программа начинается с исполнения операторного блока simulate  
если такой карты нет, проводится поиск синтаксических ошибок в программе, но прогон не осуществляется  
одновременно simulate используется операторный блок start аргументом которого является содержимое счетчика тактов модельного времени  
при функционировании программы счетчик уменьшается с каждым тактом времени на единицу, пока не обнулится

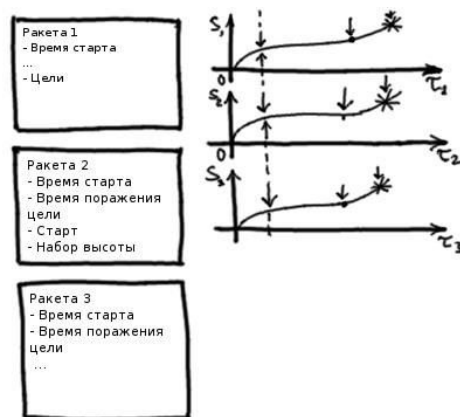
#### Основные команды интерпретатора gpsspc

@ имя\_файла - загрузка исходного текста модели  
save имя\_файла - сохранить текст модели  
edit номер\_строки - редактировать указанную строку  
clear стирает все статистическую информацию, включая таймеры модельного времени, возвращая все транзакты в пассивный буфер  
end - завершает работу интерпретатора

### **14. Объектно-ориентированное моделирование. Среда моделирования MathWorks.**

**Объектно-ориентированное моделирование** позволяет в короткий период времени создать прототип системы, нормализовать его, оптимизировать, а также, верифицировать, обеспечить валидацию требований и осуществлять тестирование модели.

Целью процесса является выявление ошибок на ранних этапах моделирования. Для каждой компоненты системы вводится так называемая спецификация, включающая помимо атрибутов и операций наборы входных\выходных данных в заданный или формируемый момент времени.



Верификация выполняется на всех этапах функционирования системы для каждой компоненты. Верификация должна учитывать взаимное влияние компонент и изменения условий моделирования. Процесс довольно трудоёмкий и может быть унифицирован за счёт наложения однотипных условий на однотипные компоненты (и процессы).

К проблемам тестирования модели относятся неверно сформулированные требования. В том числе и противоречивые требования. Возникает необходимость в ранней валидации требований. Как правило, модель содержит ограничения, накладываемые на переменные с учетом физической природы, изменений окружающей среды или изменений природы, а также накопление вычислительной погрешности при реализации модели.

В случае, если эти составляющие на практике учесть представляется не возможным или сложно, то исследователь вводит доверительный интервал с заданным коэффициентом доверия для исследуемой величины. В ряде случаев исследователь затрудняется в оценке границ доверительного интервала с высоким коэффициентом доверия — в этом случае, необходимо построить систему тестов для контроля за исследуемыми величинами. В первую очередь создаются тесты системного уровня, обеспечивающие тестирование модели в соответствии с системными требованиями, генерируется пространство случайных параметров, при реализации модели величины, исследуемые на таких пространствах должны попадать в заданный диапазон. В MathWorks реализован модуль System Test, который на основе продукта Simulink Verification & Validation, разработчик может связать свою схему модели и диапазоны изменений величин со стандартными процедурами и тестами этих продуктов

Решения компании MathWorks представляют собой современный инструментарий проектирования и имитации сложных систем на основе заданной или построенной математической модели. Позволяют создавать сложные многодоменные проекты, включающие приложения для обработки изображений и сигналов, математических расчётов, расчёта технической оснастки проекта, обмена данными и т.д. MathWorks предполагает описание всех аппаратных программных и алгоритмических средств, участвующих в моделировании как объектов, характеризуя их наборами данных (а также моделями данных), функциями и возникающими в процессе функционирования системы событиями. Такой подход называется объектно-ориентированным проектированием.

#### Построение модели сложных систем:

1. Наследуя принципы среды моделирования Simulink, основанные на

графических блок-схемах, MathWorks позволяет выбрать структуру описываемой системы и установить связи (разного типа, как в UML) между компонентами системы.

2. Вводится набор дескрипторов, определяющие каждую компоненту как объект. То есть каждый блок компоненты может быть описан физически, математическими объектами, дискретными событиями или графиками состояния.

3. Оптимизация схемы — удаление/объединение блоков, функционирование которых не влияет на решение задачи.

4. Выбирается способ имитации системы, реализуются алгоритмы посредством библиотеки алгоритмов, наследованной от MATLAB и Simulink.

#### Достоинства пакета MathWorks:

1. интеграция большого объема вычислительных методов;
2. удобный графический интерфейс;
3. интерфейс с другими языками программирования, в том числе импорт/экспорт программ из/в C, C++;
4. возможна автоматическая генерация на C, C++, код генерируется непосредственно из модели сложной системы, что удобно для развёртывания и прототипирования;
5. встроенный язык для распараллеливания вычислений, в том числе распараллеливание реально поступающих данных с датчиков;
6. встроенная система тестирования.

#### Недостатки пакета MathWorks:

1. низкая точность вычислений;
2. значительное количество ошибок возникает при распараллеливании вычислений и данных;
3. ограничен интерфейс с иными средствами программирования.

### **15. Введение в анализ данных. Основные понятия, методы.**

а. зzzzzzzz

*Функциональная зависимость* - это закон, ставящий в соответствие каждому действительному числу  $X$  из множества действительное число  $Y$  из множества.

*Стохастическая (случайная) зависимость* между величинами  $X$  и  $Y$  - это зависимость, при которой строго определенному значению величины  $X$  может соответствовать множество значений величины  $Y$ .

### **16. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта. Простая линейная регрессия и простой корреляционный анализ.**

Пары точек наносятся на координатную сетку. Из этого получают общее предварительное представление о рассеянии облака точек (1 - нет ошибок измерения, 2- есть ошибки измерения, 3 - присутствуют ошибки измерения и изменения, то есть отсутствует явно выраженный тренд). В случае 1 - функциональная зависимость, в случае 2,3 - стохастическая.

*Функциональная зависимость* - это закон, ставящий в соответствие каждому действительному числу  $X$  из множества действительное число  $Y$  из множества.

*Стохастическая (случайная) зависимость* между величинами  $X$  и  $Y$  - это зависимость, при которой строго определенному значению величины  $X$  может соответствовать множество значений величины  $Y$ . Зависимость носит вероятностный характер, то есть СВ  $Y$  принимает разные значения с некоторой вероятностью. Термин “стохастическая” связь впервые введен русскими статистиком А.А. Чупровым в 1926 г.

По отношению к событиям функциональная зависимость - причинна, то есть наступление одного события влечет за собой другое. Стохастическая связь при этом, наоборот, не является причинной. Одно событие изменяет вероятность наступления другого, однако не всегда вызывает его появление. Функциональная зависимость является предельным случаем стохастической - при наиболее тесной связи. Другой крайний случай - полная независимость величины  $Y$  от величины  $X$ . Таким образом стохастическая связь может быть более или менее тесной.

При оценивании стохастической зависимости различают корреляцию (существует ли взаимосвязь между переменными) и регрессию (какая зависимость).

Совокупность характеристик стохастической связи исследователи делят условно на две группы: в первую входят характеристики, позволяющие дифференцировать линейную и нелинейную зависимости, во вторую - характеристики степени тесноты связи (меры связи).

В [Мирский Г.Я. Характеристики стохастической взаимосвязи и их измерения. - М.: Энергоиздат, 1982] приводятся и подробно обсуждаются вероятностные характеристики, с помощью которых описывается стохастическая связь случайных процессов и случайных величин, а также способы их измерения, в том числе:

1. корреляционная функция и взаимная корреляционная функция;
2. нормированная и нормированная взаимная корреляционные функции;
3. коэффициент корреляции (полный, частный, сводный);
4. корреляционное отношение;
5. моментная функция и взаимная моментная функция;
6. функция регрессии;
7. функция когерентности;
8. условная функция распределения вероятностей;
9. условная плотность распределения вероятностей;
10. коэффициент коллигации и функция коллигации;
11. ранговый коэффициент корреляции Спирмэна и Кендалла и др.

Можно утверждать, что связанные случайные процессы не всегда являются взаимосвязанными. Так на изменение кардиограммы оказывают влияние колебания атмосферного давления, изменение состояния окружающей среды и т. д., хотя сами климатические условия от кардиограммы не зависят.

Наиболее часто используют такие характеристики стохастической связи как коэффициент и функция корреляции. Корреляционный анализ [Фестер Э., Ренц Б. Методы регрессионного и корреляционного анализа. - М.: Финансы и статистика, 1983], дающий хорошие результаты при линейных зависимостях, недостаточно эффективен, однако, в случае нелинейной связи. Поэтому представляется полезным исследовать коэффициент взаимосвязанности (коллигации), примененный академиком С.Н. Бернштейном для оценки степени зависимости двух случайных событий и . Коэффициент коллигации служит характеристикой жесткости связи при любом ее характере - линейном и нелинейном, позволяет обнаруживать и достоверно констатировать независимость случайных величин.

Корреляционный анализ изучает на основании выборки стохастическую зависимость между случайными величинами  $X, Y$ .

Ковариация определяется как математическое ожидание произведения отклонений случайных величин:

$$cov_{XY} = M[(X - M(X))(Y - M(Y))],$$

Смешанный центральный момент второго порядка

Для устранения недостатка ковариации был введён **линейный коэффициент корреляции** (или **коэффициент корреляции Пирсона**), который разработал Карл Пирсон в 90-х годах XIX века. Коэффициент корреляции рассчитывается по формуле:

$$r_{XY} = \frac{cov_{XY}}{\sigma_X \sigma_Y} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}$$

где  $\bar{X}$ ,  $\bar{Y}$ , — среднее значение выборок.

Коэффициент корреляции изменяется в пределах от минус единицы до плюс единицы

Функция от моментов второго порядка называется коэффициентом корреляции

В математической статистике регрессионным анализом называют совокупность приемов для установления связей между независимой переменной  $Y$  и одной или несколькими переменными.

Регрессия - условное математическое ожидание случайной переменной  $Y$  при условии, что другая условная переменная  $X$  приняла значение  $x$ .

Моделью линейной регрессии является модель, в которой теоретическое среднее значение наблюдаемой величины  $y$  является линейной комбинацией независимых переменных

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots \quad (1)$$

для случая, когда в модель включаются  $k$  переменных  $x$ .

Множители  $a_i$ ,  $i=1,2,\dots$  представляют собой параметры модели, значения которых должны быть установлены. Они называются коэффициентами регрессии, а  $a_0$  называется свободным или постоянным членом.

Модель, более чем с одной переменной  $x$  (1) называется моделью множественной регрессии.

## 17. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта. Метод наименьших квадратов.

### Пример.

- а. Метод, применяемый для решения большинства задач (например аппроксимации функции), и основанный на минимизации суммы квадратов отклонений некоторой функции от искомых переменных.  
[Википедия]

$$\sum_{i=1}^n (y_i - f(x_i))^2 \rightarrow \min_x$$

- і. Пример: (Лабораторная №1 по Басарабу) решение задачи Коши, путем представления значения функции  $y(x)$  (с заданными также граничными условиями), как линейную комбинацию базисных функций:

1.  $y(x) = \varphi_0 \sum_{i=1}^n c_i \varphi_i$ , где  $\varphi_i$ - базисный вектор.

Применение метода заключается в поиске коэффициентов  $c_i$  в ряде разложения.



$$\sum_{m=1}^K c_m \int_{\Omega} A\varphi_k(\mathbf{x}) A\varphi_m(\mathbf{x}) d\mathbf{x} = \int_{\Omega} A\varphi_k(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \\ (k = \overline{1, K})$$

2.

ii. Пример 2: Поиск линии тренда.

**18. Показатели взаимосвязи двух ненаблюдаемыми одновременно характеристиками объекта. Подходы к оцениванию параметров исследуемой зависимости. Метод моментов. Пример.**

- Применительно к выборкам  $\xi_1, \xi_2$ , ненаблюдаемых одновременно, следует сказать что размер выборки может отличаться.
  - $\xi_1 = \{x_1, \dots, x_n\}$
  - $\xi_2 = \{x_1, \dots, x_m\}$
- Под оценкой параметров зависимости понимается оценка параметров установленной связи. Если согласовывать с лабораторной №1, то рассматриваем степенную зависимость:  $\xi_i = \alpha \xi_i^\beta$ , поэтому оцениваем  $\alpha$  и  $\beta$  (также можно, например рассмотреть линейную зависимость  $\xi_i = k \xi_i$ ). В случае степенной зависимости, метод моментов здесь требуется для вычисления (на основе мат. ожиданий  $M_i$  и дисперсий  $D$ ) параметров зависимости:

$$\begin{aligned} - \beta &= \sqrt{\frac{D(\ln \xi_1)}{D(\ln \xi_2)}} \\ - \alpha &= e^{(M_1 - \beta M_2)} \end{aligned}$$

- Полный Метод моментов (sharelatex с.10 (п.3.2.1), gdrive с.6 (2.2.1))
  - Пример: рассмотреть 1-ую лабораторную работу, про поиск коэффициентов степенной зависимости ( $\alpha, \beta$ ) между двумя стохастически зависимыми (изменение распределения одной величины ведет к изменению значений другой величины) случайными выборками  $\xi_1, \xi_2$ . Привести результаты лабораторной.

**19. Показатели взаимосвязи двух ненаблюдаемыми одновременно характеристиками объекта. Подходы к оцениванию параметров исследуемой зависимости. Метод максимального правдоподобия. Пример.**

- Подходы к оцениванию:
  - Метод моментов (определив характер взаимосвязи, используем этот метод для поиска параметров):
    - $k$  -- линейная связь
    - $\alpha, \beta$  -- в случае степенной зависимости.
  - Метод максимального правдоподобия:
    - Сносно описан в ее лекциях (П.3.2.2 в sharelatex). Доработал описание.

## 20. Анализ функциональных зависимостей между наблюдаемыми одновременно характеристиками объекта.

При анализе одновременно наблюдаемых показателей, возможны 3 варианта:

В случае (а), говорят о наличии стохастической связи между 2-мя случайными переменными. Термин *стохастическая связь* был введен Чупровым в 1926 году. *Стохастическая связь* --- это такая связь, при которой значению величины  $x \in X$  соответствует одно или более значений  $y_1 \dots y_k \in Y$ .

Таким образом, получение с.в.  $y_i$  изменяет вероятность появления других значений, но не обеспечивает их появление. Именно поэтому говорят, что стохастическая связь не является причинной.

(б) Ф-циональная связь описывается зависимостью, в которой с.в  $x \in X, y \in Y$ , соответствует только одно значение  $y \in Y$ . Функциональная связь причинна, т.е. для каждого  $x_i$  соответствует конкретная реализация  $y_i$ . Предполагает взаимнооднозначное преобразование.

Как правило, на практике исследователи имеют дело с вариантом В. Присутствуют ошибки измерения, имеет место разброс реализации относительно некоторой ф-циональной зависимости в той или иной степени достоверно описывающей входные данные. Такую ф-цию называют *функцией тренда*. В случае достаточно тесной стохастической связи, задача сводится к выделению тренда и его анализа.

Задачу В решают с помощью метода наименьших квадратов.

## 21. Дисперсионный анализ. Многомерный дисперсионный анализ.

Дисперсионный Анализ предполагает вклад двух факторов и эффект их взаимодействия.

Пример: Вытачиваем деталь на станке. Деталь зависела от марки стали и  $t$  нагрева. Соответственно, влияет: материал,  $t$ , и взаимодействие двух параметров (диаметр поэтому варьируется). Можно установить, насколько влияет материал,  $t$ , и их взаимодействие. Как оказалось (при анализе этой задачи), наибольшее воздействие на диаметр оказывает взаимодействие двух параметров.

Многомерный: Вклад поэлементного взаимодействия и общее.

## 22. Анализ выбросов. Пример.

**Цензурирование выборки** — замена реальных значений случайных величин выбранными аналогами.

**Анализ выбросов** — изъятие значений случайных величин из результатов эксперимента на основе статистического анализа данных.

В случае, если предполагаемые реальные значения близки к заданному значению левой границы выборки  $a$ , но не могут быть измерены, реальные значения заменяются на величину  $a$  (**цензурирование слева**). Если предполагаемые значения близки к правой заданной границе выборки  $b$ , но не могут быть измерены, то они заменяются

на значения  $b$ (цензурирование справа). Возможна ситуация, когда выборке требуется **цензурирование слева и справа**.

Исследователь может руководствоваться сведениями о виде функции, при этом, как правило, под цензурирование и слева, с права попадает не более 10% процентов исследуемых значений, в противном случае рекомендуется изменить условия проведения эксперимента.

Анализ выбросов является процедурой альтернативной цензурированию и позволяет выявить точки, искажающие решения задач.

Существует ряд подходов к анализу выбросов, которые можно условно классифицировать:

- А. базирующиеся на виде закона распределения;
- В. базирующиеся на построении доверительного интервала.

К группе А относятся:

1. Метод, основанный на предварительном знании вида функции распределения случайной величины, наблюдаемой в рамках эксперимента. В курсе лекций был рассмотрен пример о проверке равномерности выборки генеральной совокупности, распределенной нормально с выбранными параметрами  $\mu$  и  $\sigma$ .
2. Если закон распределения выборки  $x \in X: N_X(\mu, \sigma)$ , то анализ выбросов производится на основе правила трёх  $\sigma$  (практически все значения нормально распределённой случайной величины лежат в интервале  $(\bar{x} - 3\sigma; \bar{x} + 3\sigma)$ . Более строго — приблизительно с вероятностью 0,9973 значение нормально распределённой случайной величины лежит в указанном интервале). Процедур, относящихся к классу А немного, так как они специфичны для конкретного закона распределения. Кроме того закон распределения может быть неизвестен.

К группе В относятся:

1. Альтернативой, обеспечивающей анализ выбросов является подход, связанный с построением доверительных интервалов. Если значительная часть элементов выборки оказывается вне интервала, то говорят о наличии выбросов.  $[\underline{x} - k\hat{S}, \underline{x} + k\hat{S}], \gamma = 0.99, k = 1.55 + 0.8 \lg(\frac{N}{10})\sqrt{e-1}(1)$
2. Существует достаточно много вариантов оценок коэффициента  $k$ , в ряде случаев анализируют выборочную функцию распределения и фиксируют квантили заданного уровня

Если при проведении другого эксперимента значимо большее количество элементов выборки, чем в контрольной, оказывается вне диапазона  $[K(\alpha_1), K(\alpha_2)]$ , то говорят о наличии выбросов.

Недостатком методов группы В является выявление значительного количества выбросов в выборке.

В качестве примера можно рассказать как мы удаляли точку в лабораторной с чешским метром.

Можно про лампочки ещё рассказать, там цензурирование.

### **23. Пошаговый дискриминантный анализ. Пример.**

Если наблюдаемые одновременно величины, то устанавливаем функцию вида  $N(0,1)$ .

Когда функция уточняется на каждом шаге с появлением новой точки --- в этом суть пошаговости алгоритма.

Для решения задач классификации с двумя группами как правило используется дискриминантный анализ. Дискриминантный анализ предполагает построение дискриминирующей функции, аргументами которой являются измеряемые величины, далее выделяются области, при попадании в которые объект относится к первой или второй группе. Если требуется на разделение на 3 и более групп и/или нет сведений о характеристиках объектов, определяющих группу, используются методы кластерного анализа.

Короче, хуячим подвыборку. По ней делим два множества прямой. Координаты - рассматриваемые признаки, предварительно фильтруем из модели не влияющие (или слабовлияющие). Потом хуячим по разделённым множествам классификационные прямые (если два множества - то из две). По ним для каждого нового элемента считаем значение, выбираем максимальное из них и по нему выбираем к какой группе относится.

Пример -- кто понял, пусть напишет.

#### 24. Анализ главных компонент. Пример.

**Метод главных компонент** (англ. *principal component analysis, PCA*) — один из основных способов уменьшить **размерность** данных, потеряв наименьшее количество **информации**.

см билет 27, про факторный анализ, там идея метода описана.

С примером всё туго.

#### 25. Кластерный анализ. Иерархические методы. Пример.

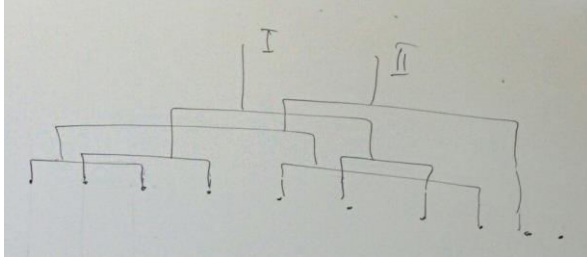
(latekc - 26, gdrive - 19)

**Кластерный анализ.** Пусть  $X$  множество объектов,  $Y$  - множество меток(имён) кластеров. Известна функция расстояния между объектами  $\rho(x', x''), x', x'' \in X$ . Имеется конечная (обучающая) выборка объектов  $X_m \subset X$ ,  $X_m = \{x_1, \dots, x_m\}$ . Требуется разбить множество объектов на непересекающиеся подмножества, называемые кластерами, так, чтобы каждый кластер состоял из объектов, близких друг к другу по метрике  $\rho$ , при этом сами кластеры удалены друг от друга по метрике  $\rho$ .  $x_i \in X: x_i \sim y_j$ ,  $y_j$  - метка кластера,  $y_j \in Y$ .  $i$ - бесконечный счётчик. При обучении:  $i = 1, \dots, m$ , при кластеризации:  $i = 1, 2, \dots, j = 1, \dots, n$   $n$ - задано или вычислено.

Алгоритм кластеризации - это функция  $a: X \rightarrow Y$ , при этом каждому из элементов множества  $X$  ставится в соответствие элемент из множества  $Y$ .

Кластер имеет следующие математические характеристики: центр, радиус, среднеквадратичное отклонение (СКО), может быть задан размер кластера. Центр кластера - среднее геометрическое место точек в пространстве переменных. Размер кластера (радиус) - максимальное расстояние от точек кластера до его центра. СКО - величина квадратично зависящая от расстояния между всеми точками кластера и его центром (на самом деле квадратично зависима дисперсия).

**Иерархические методы.** Суть иерархии состоит в последовательном объединении кластеров меньшего размера в большие. Или разделение больших по размеру кластеров на меньшие. В начале работы все объекты являются отдельными кластерами. На первом этапе наиболее похожие объекты объединяются в кластер. На последующих шагах объединение продолжается до тех пор, пока объекты не будут объединены в заданное число кластеров. Подход называется агломеративным. На рисунке представлена иерархия дендрограммы:



Сам подход называется агломеративным. Количество кластеров задаётся.

Альтернативой агломеративным подходам является дивизимные методы, то есть в начале работы все объекты принадлежат одному кластеру, на каждом шаге группы расщепляются. Процесс продолжается до достижения заданного числа кластеров.

**Иерархические методы:**

- метод Варда
- метод ближайшего соседа
- метод наиболее удалённого соседа

**Не иерархические** (чтобы ни было записано в лекции, итеративный он):

- k-means

## 26. Кластерный анализ. Эвристические методы. Пример.

В двух словах: “билет удалён”. Пример: пусть рождаются мальчики и девочки с ростовесовыми хар-ками. Мы их разделяем по росту на 2 группы, и сравниваем с равной статистикой, мальчики это или девочки. Далее, разбиваем на 3 группы (при рождении) (мальчики, девочки, аномальные).

## 27. Факторный анализ. Пример.

(latex - 50, gdrive - 34)

$$y = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_n x_n$$

Полагается, что значение каждого признака  $x_i$  могут быть выражены (\*), как взвешенные суммы других признаков (латентные признаки/переменные/факторы), количество которых может быть меньше, чем число исходных признаков.  $\varepsilon_i$  (не специфический фактор) остаточный член ряда, определяющий влияние неучтённых факторов. Можно говорить, что дисперсия  $x_i$  зависит от  $\sigma_i(\varepsilon_i)$ .

$$(*) \begin{cases} x_i = \beta_0 + \dots \beta_{i-1} x_{i-1} + \beta_{i+1} x_{i+1} \dots \beta_k x_k + \varepsilon_i \\ \sigma_i = \sigma(\varepsilon_i) \end{cases}$$

Коэффициенты при переменных называются **нагрузкой фактора**, переменные  $x_i$  — **факторами**. Величины  $\varepsilon_i, i = 1, \dots, k$  независимы друг от друга и от любого фактора  $x_i$ .

Можно наложить условие для  $n$  признаков, **оптимальное число факторов**  $k$  определяется эмпирической формулой:  $n - k < \frac{n+k}{2}$ .

Сумма квадратов нагрузок основной модели анализа называют **общностью соответствующего признака**  $x_i$ . Чем больше это значение, тем лучше описывается решение задачи выделенным фактором.

Формально, описанная система имеет  $k$  неизвестных, и  $k$  уравнений, т.е. задача распределения нагрузки решается однозначно. С другой стороны, наложение доп. условий может привести к увеличению числа неизвестных и неоднозначному решению. В этом случае, можно уменьшить количество факторов за счет вращения системы в заданной системе координат. Во избежание снижения точности, осуществляют поворот гиперболы данных в пространстве, при переходе к новой, ортогональной системе координат. В этом случае, разброс ряда переменных в новой системе координат, оказывается не значительным, в силу чего изменением этой координаты можно пренебречь. Такой подход называется **вращением факторов**, и лежит в основе **метода главных компонент**.