КУРС «Математические и инструментальные методы принятия решений и математическое моделирование»

Семёнычев Валерий Константинович
Д.т.н., Д.э.н., профессор по кафедре «Высшая математика»

Конструкция систем поддержки принятия решений (СППР или Decision support systems, DSS)

- зависит от вида задач, для решения которых она разрабатывается, от доступных данных, информации и знаний, а также от пользователей системы;

- СППР - скоординированный набор данных, систем, инструментов и технологий, программного и аппаратного обеспечения, с помощью которого предприятие собирает и обрабатывает информацию о бизнесе и окружающей среде с целью обоснования управленческих действий.

- СППР — в большинстве случаев — это автоматизированная система, которая помогает пользователю (лицу, принимающему решения - ЛПР) использовать данные и модели для идентификации и решения задач и принятия решений

Основные характеристики СППР

- 1.СППР использует и данные, и модели;
- 2. СППР предназначены для помощи менеджерам в принятии решений для слабоструктурированных и неструктурированных задач;
- 3. СППР, в отличие от экспертных систем, поддерживают, а не заменяют человека при выработке решений;
- 4. Цель СППР — улучшение эффективности решений.

Список сфер применения СППР:

-Анализ каналов снабжения и распределения (логистика);

-Анализ производства; Анализ продаж;

-Маркетинговый анализ; Анализ клиентской базы;

-Анализ качества сервисного обслуживания; Финансовый анализ;

-Инвестиционный анализ; Анализ персонала;

-Анализ исследовательской и проектной деятельности;

-Анализ стратегии организации; Управление рисками

-Многомерный анализ; Прогнозирование; Поиск закономерностей;

- *Под методами принятия управленческих решений* понимается нахождение определенного варианта достижения поставленной цели или решения конкретной задачи, т.е. процесс разрешения проблемы путем применения совокупности приемов или операций.
- Цель применения этих моделей – *выбор наилучших действий (альтернатив), исходя из заданного критерия и ситуации, в которой принимается решение.*
- *Однако в жизни субъект управления не всегда стремится максимизировать экономический результат.* Вместо этого он принимает «удовлетворительные», «достаточно хорошие» решения.
- В этом случае при принятии решений могут использоваться такие критерии, как «приемлемая величина прибыли» и «надежное выполнение плана». Математическая теория принятия решений не дает рецептов для демонстрации того, как решения фактически должны приниматься.
- *Методы разработки управленческих решений* включают в себя способы и приемы выполнения анализа, обработки информации, выбора альтернативных вариантов действий и пр.

Методы принятия решений могут быть сведены в две большие группы:

1) Методы, основанные на расчетах, обоснованиях и доказательствах и/или на научно-практическом подходе, предполагающем выбор оптимальных решений на основе переработки больших количеств информации, помогающем обосновать принимаемые управленческие решения, требуют применения современных вычислительных средств.

Проблема моделирования и обоснования точной и/или адекватной модели предполагает необходимость всесторонней оценки конкретной ситуации ЛПР.

Ответственность за принимаемые решения предполагает самостоятельность принятия одного из нескольких вариантов возможных решений.

2) Методы, основанные на технологиях опережающего воздействия, в основе которых лежат процессы научного предвидения – прогнозирования.

Понятие горизонта прогнозирования (краткосрочное, среднесрочное, долгосрочное).

Что мы знаем о будущем?

- Точность моделирования характеризуется критериями (их абсолютными значениями или относительными).
- Наиболее известен - коэффициент детерминации (0, 1).
- Адекватность модели - ее соответствие исходным данным с точки зрения цели моделирования. Адекватность, как более общее понятие, включает в себя точность. Не всякая точная модель адекватна. Если модель неточная, то она не будет и адекватной.
- Адекватность модели определяется исходя из оценок
- точности структурной и параметрической идентификации и точности прогнозирования (обычно в % от 0 до100).
- Структурная идентификация (выбор типа модели) сложнее параметрической (оценки параметров выбранной модели).
- Модели могут быть как статическими, так и динамическими, аналитическими (решениями дифференциальных уравнений) или феноменологическими). Есть и другие классификации моделей.
- Нужен также мониторинг эволюции, учет размера выборки для применения статистики при применении идентификации, эконометрики и эконофизики.

Иллюстрация задачи поддержки решения

Предложена модель функции потребления: $\boxed{\ln C = \beta_0 + \beta_1 \ln Y + \beta_2 \ln P}$,

где C – потребление некоторого пищевого продукта на душу населения в некотором году;

Y – реальный доход на душу населения в некотором году;

P – индекс цен на этот продукт, скорректированный (дефлированный) на общий индекс стоимости жизни (относительный уровень цен);

β_i - параметры.

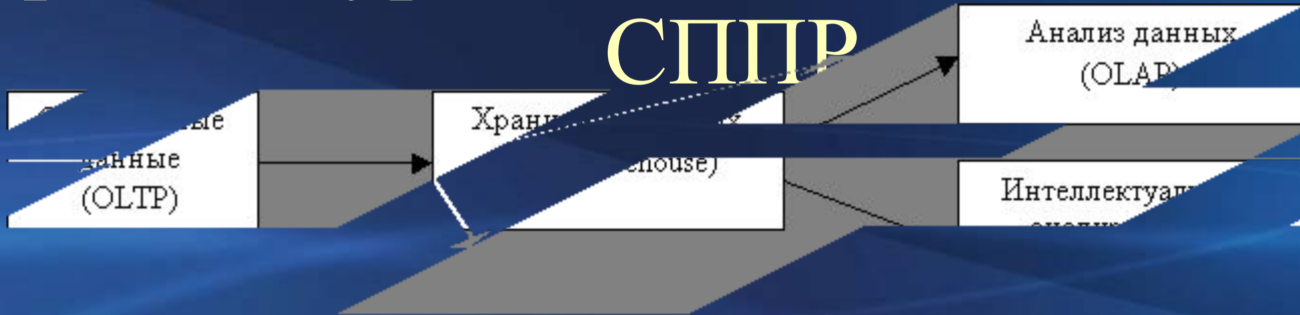
<u>**Задача «параметризации» модели - определение**</u> β_i: 1.измерение Y, P (насколько корректно измерены данные, представляют ли они то, что должны представлять по нашему мнению? 2. Погрешность измерения Y, P и C аддитивна или мультипликативна ? 3. Какой метод использовать для параметризации – МНК ?

<u>**Задача «спецификации» модели:**</u> 1. Нет ли переменных, которые следовало бы дополнительно включить в уравнение (например, цены на непродовольственные товары)?

2. Не следует ли исключить из уравнения некоторую переменную? 3. Верно ли то, что модель линейная (верны ли теоретические предпосылки модели)? 4. Является ли модель полной? Может быть необходимо учесть и уравнение предложения, кроме уравнения спроса? 5. Достаточно ли изучать макроэкономическое уравнение или необходимы такие индивидуальные (микроэкономические) данные для поставленной задачи? 7. Может быть нужно выделить факторы, которые зависят непосредственно от принятия управленческих решений данным объектом хозяйствования (предприятием, регионом), и влияние факторов, которые от менеджмента на данном хозяйствующем объекте не зависят? 8. Модель является статической. Возможно, более подходящей была бы динамическая модель. Например, можно предположить, что прошлогодний доход может влиять на текущий уровень потребления.

В централизованной экономике, эконометрика не изучалась, в отличие, скажем, от балансовых или оптимизационных методов: межотраслевого баланса и линейного программирования. Причинно-следственными связями занимается экономическая теория, а связями без выявления их причин - эконометрика. **Набор статистических методов, используемых для наблюдения за ходом развития экономики, её анализа (моделирования) и её прогнозов <u>называется эконометрикой.</u>**

Архитектурно-технологическая схема СППР



Первоначально информация хранится в оперативных базах данных (OLTP-систем). Но ее сложно использовать в процессе принятия решений.

Затем она используется в процедурах многомерного анализа (OLAP) и для интеллектуального анализа данных (ИАД), который реализует:
1.выявление закономерностей (моделирование;
2.использование выявленных закономерностей для прогноза;
3.анализ исключений (аномалий или выбросов), в найденных закономерностях.

ИАД реализуют экспертные и интеллектуальные системы, методы искусственного интеллекта, базы знаний, базы данных, компьютерное моделирование, нейронные сети, нечеткие системы. социально-экономического развития города является метод имитационного моделирования.

Рекомендуемая литература

1. Аникин П.В., Королев В.А., Тороповцев Е.А Математические и инструментальные методы. Изд-во «Кнорус». 2014. Можно скачать.

2. Красс М.С., Чупринов Б.П. Математика в экономике. Математические модели и методы. Изд-во «Кнорус». Можно скачать.

3. Семенычев В.К., Семенычев Е.В. Параметрическая идентификация рядов динамики: структуры, модели, эволюция. - Самара. Изд-во «СамНЦ РАН», 2011. – 346 с.

4. Семенычев В.К., Коробецкая А.А., Кожухова В.Н. Предложения эконометрического инструментария моделирования и прогнозирования эволюционных процессов. - Самара. САГМУ. – 384 с.

5. Конюховский П.В. Микроэкономическое моделирование в банковской деятельности. - Спб. Питер.-2001. - 224 с.

7. Эконометрика/Под ред. И. И. Елисеевой. - М.: Финансы и статистика, 2005. - 575 с. (и все более поздние издания).

8.Бородич С.А. Эконометрика. - Минск: Новое знание. 2001. - 408 с.

9. Кондратьевские волны. Под редакцией Л.Е. Гринина, А.В. Коротаева. – Волгоград: Учитель. 2014. – 360 с.

10. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. Айвазян А.А., Мхитарян В.С.-- М.; ЮНИТИ-ДАНА. 2001.- 656 с.

11. Дайитбегов Д.М. Компьютерные технологии анализа данных в Эконометрике. – М. ИНФРА-М. 2008.- 578 с.

12. Кожухова В.Н., Коробецкая А.А., Семенычев В.К. Свободная программная среда R. Практикум. Самара. Изд-во «САГМУ». 2016.-48 с.

Скачать

Статистический анализ структуры социально-экономических процессов и явлений (Сивелькин В.А., Кузнецова В.Е.).

Статистические методы и модели (Костин В.Н., Тишина Н.А.)

Общая постановка задачи разработки СППР

- СЭС: Макроуровень, Мезоуровень, Микроуровень, Наноуровень

Многомерные факторные модели показателей:

$$\vec{Y} = \vec{Y}(\vec{x})$$

$\vec{Y}$ -вектор определяемых (эндогенных) показателей, формируемых в процессе и внутри функционирования СЭС, $\vec{x}$ - вектор факторных или экзогенных переменных, которые или часть из которых поддаются регистрации и планированию.

Обычно принимают структуру статической модели в виде

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \ldots + \alpha_{10} X_{10},$$

как множественную линейную регрессию (линейную форму факторных переменных) при выборе одного определяемого показателя. Редукция связи факторов (например, с 10 до 2 или даже 1 , которым может быть и время t).

Имеем «мнимую точность» из-за возможной связи факторов при аддитивной (т.е. независимой) форме структуры модели, а динамика показателей зачастую требует нелинейных моделей (модель эволюции всегда нелинейная).

«Проклятие размерности» проявляется и при расчете моделей на разных метриках при редукции факторов с учетом стохастических погрешностей

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \varepsilon \neq \beta X_1 + \gamma X_{2,3} + \varepsilon(\mu).$$

Пространственные ряды данных – одномоментные наблюдения определяемого показателя (значения прожиточного минимума для различных областей, уровень цен на определенный продукт в магазинах города, зависимость дохода предприятия от квалификации сотрудников, от количества конкурентов, от стоимости блага, от уровня инфляции в стране и т.д.).

Временные ряды (динамические ряды, ряды динамики) – упорядоченная во времени последовательность уровня показателя (дохода, ВВП, ВРП региона, уровня инфляции, динамики курсов валют и ценных бумаг на валютных рынках и т.д.).

Временные ряды могут быть эквидистантными и неэквидистантными (от этого может зависеть выбор применяемого метода идентификации модели.

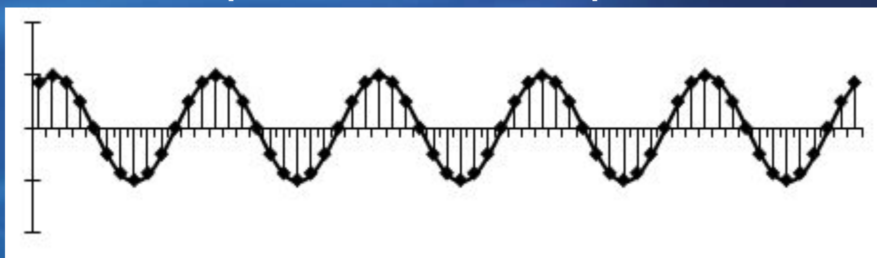
Порядок наблюдений во временных рядах, в отличие от пространственных рядов, нельзя менять местами.

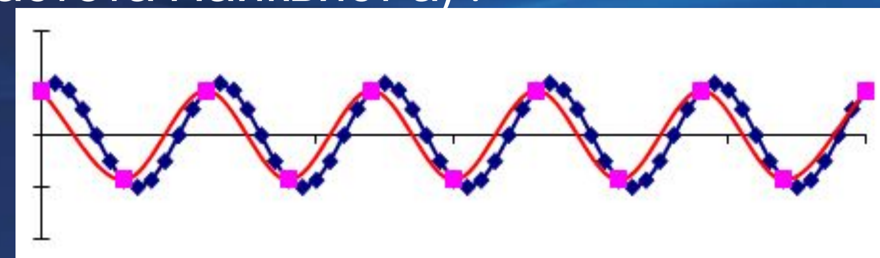
Формирование временных рядов

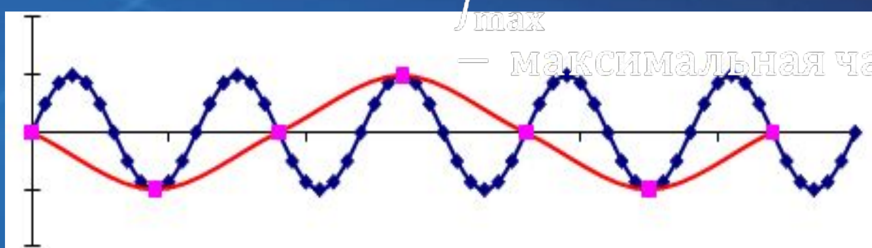
f_{max} – максимальная частота спектра траектории , частота Δ дискретизации .

$$5 < \frac{T}{\Delta} < 10$$

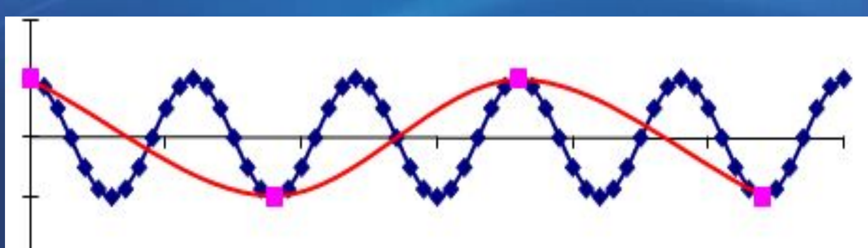
Выбор частоты дискретизации (частота Найквист а) :

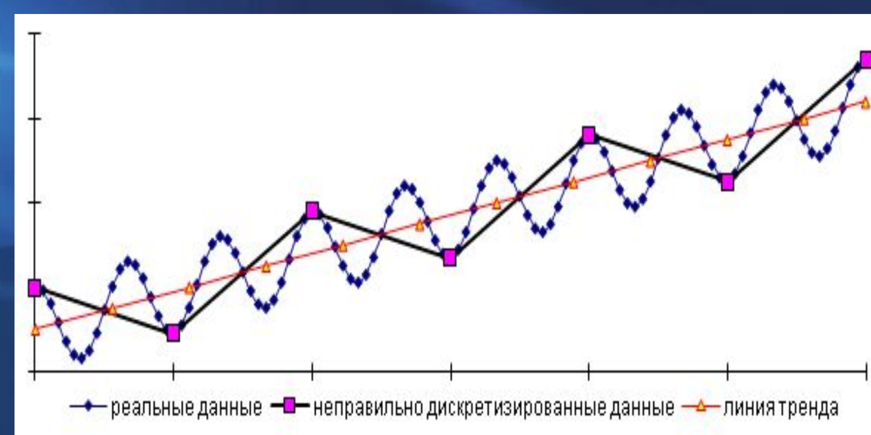






f_{max} – максимальная частота спектра траектор





Трендовые модели для отражения динамики определяемого показателя и эволюции

Известные подходы к построению моделей:

- <u>аналитический</u> или параметрический (достоинства - малые выборки, возможность прогноза);

- <u>непараметрический</u> или алгоритмический (простота, но длинные выборки, прогноз лишь при придании ему адаптивного характера).

<u>В аналитическом подходе считают</u>, что модель детерминированной части траектории D_t (или отсчета временного ряда D_k) зависит от набора параметров μ :

$$D_t = f(\mu, t).$$

<u>Эмпирическая (стохастическая) трендовая модель</u> имеет случайную погрешность e_t , зависящую от постулируемой модели , принятой декомпозиции структуры модели и реализованного метода идентификации

(где D_t^* - модельные (случайные) уровни предложенной модели):

$$D_t^* = f(\mu^*, t).$$

Оценки точности прогнозирования (экстраполяции)

Второй коэффициент Тейла $k_{T2} = \sqrt{\left(\sum_{k=n}^{n+l-1} (Y_k - Y_k^o)^2\right) \Big/ \left(\sum_{k=n}^{n+l-1} Y_k^2 + \sum_{k=n}^{n+l-1} (Y_k^o)^2\right)} \cdot 100\%.$

удобен при малых значениях Y

$MAPE$ – оценка:

$$MAPE = \frac{1}{n+l} \sum_{k=1}^{n+l} \left| \frac{Y_k - Y_k^*}{Y_k} \right| \cdot 100\%$$

не чувствительна к выбросам

погрешность
$\leq$ 10 – отличная,
$\leq$ 20 – хорошая

<u>общий объем анализируемой выборки $(n + l)$</u>

Модели:

Детерминированные Вероятностные Нечеткой логики

Теория вероятностей – наука о закономерностях массовых случайных явлений

(Лаплас, Пуассон, Гаусс, Бернулли, П.Л. Чебышев, А.М. Ляпунов, А.А. Марков, А.Н. Колмогоров и др.)

- социально - экономическая статистика;
- многомерные статистические методы;
- *<u>эконометрика; эконометрическое моделирование</u>;*
- методы социально-экономического прогнозирования;
- страхование и актуарные расчеты;
- теория риска и моделирования рисковых ситуаций;
- маркетинг; теория массового обслуживания;
- технический и фундаментальный анализ,
- теория планирования эксперимента;
- теория надежности;
- теория информации (статистическая радиотехника),
- выборочный контроль качества и др.

Классическое определение вероятности

Испытание. Бросание монеты: A – герб, B – цифра;

Испытание. Выстрел: A - попадание, $\bar{A}$ – промах (**противоположное событие**). A и $\bar{A}$ - образуют полную группу событий при одном выстреле.

Испытание. Покупают 23 лотерейных билета: событие А – выигрыш выпал на первый билет, отсутствие выигрыша на втором; событие В – выигрыш не на первом, а на втором билете; событие С – выигрыш обоих билетов; событие D – выигрыш не выпал ни на один билет.

Данные события образуют полную группу событий, если проверяются два билета. **А при проверке трех билетов, что образует полную группу событий?**

Испытание - игральная кость: событие А – появление числа 1, событие В $\Rightarrow 2$, событие С $\Rightarrow 3$, D $\Rightarrow 4$, E $\Rightarrow 5$, F $\Rightarrow 6$. События равновероятны? (**если**…).

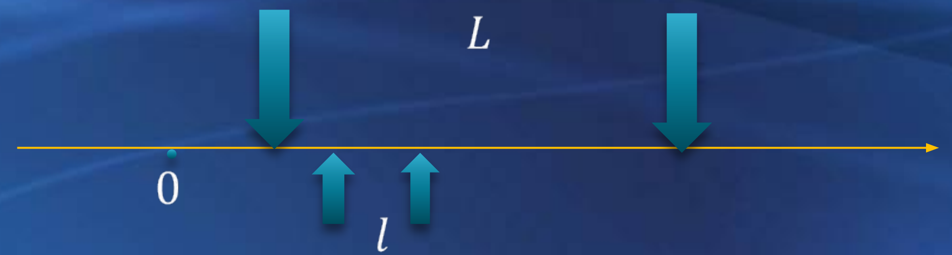
Вероятность – мера объективной возможности появления события:

$$P(A) = \frac{m}{n}$$

где m – число исходов испытания, которое благоприятствует появлению события А,
n – общее число исходов испытания.

Недостатки классического определения вероятности

- Неприменимость при бесконечном числе исходов $(n \to \infty)$. Выход находят путем введения геометрической вероятности (отношения площадей, длин отрезков).



- Примеры: вероятность попадания отрезка длиной l в отрезок длиной L, рулетка.

$$P(A) = \frac{l}{L}$$

Априори бывает трудно обосновать то, что элементарные события равновероятны. Тогда вводят статистическую (апостериорную) вероятность:

$$P^*(A) = \frac{m}{n},$$

m − число исходов, в которых событие А появилось, n − общее число исходов

$$P^*(A) \to P(A) \qquad \lim_{n\to\infty} |P^*(A) - P(A)| \prec \varepsilon, \forall\varepsilon$$

Алгебра событий

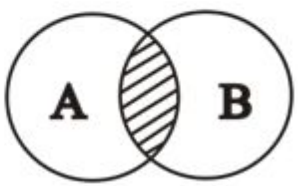
1. Суммой двух событий A и B называют событие $(A+B)$, состоящее в появлении или A, или B, или обоих (если они **совместны**).

Пример. Для двух выстрелов сумма событий: A – попадание при первом выстреле, B попадание при втором выстреле, AB – попадание при обоих выстрелах, т.е. $A+B = A+B+A\cdot B$

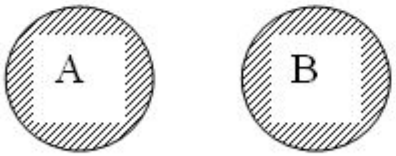
Суммой нескольких событий называют событие, заключающееся в появлении хотя бы одного из них. $\boxed{A+B+C = A+B+C+AB+AC+BC+ABC}$

При несовместных событиях: $A+B = A+B$, $\boxed{A+B+C = A+B+C}$.

Диаграммы Эйлера-Венна



$$P(A+B) = P(A) + P(B) - P(AB)$$



$$P(A+B) = P(A) + P(B)$$

Пример. В урне 30 шаров: 10 красных, 5 синих, 15 белых. Найти вероятность появления цветного шара при вынимании из урны одного шара. С появлением цветного шара свяжем или событие A - вынимание красного шара, или событие B вынимание синего шара. $P(A+B) = P(A) + P(B) \Rightarrow P(A) = \frac{10}{30}$, $P(B) = \frac{5}{30} \Rightarrow P(A+B) = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$.

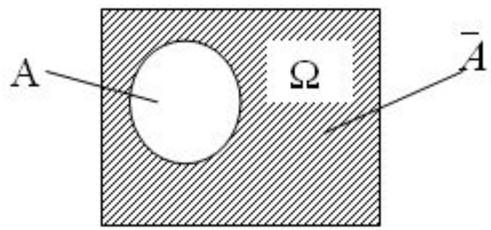
Сумма вероятностей событий $A_1, A_2, \ldots, A_n$, образующих полную группу, равна 1 (**условие полноты группы событий**).

$$\boxed{P(A_1) + P(A_2) + \ldots + P(A_n) = 1}.$$

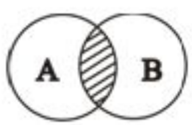
2.<u>Противоположными называют</u> два единственно возможных события, образующих полную группу.

$$P(A) + P(\bar{A}) = 1,$$

$$\boxed{P(\bar{A}) = 1 - P(A).}$$



3.<u>Произведением двух событий</u> A и B называют событие $A \cdot B$, состоящее в совместном появлении этих событий.



<u>Пример:</u> A – деталь годна, B – деталь окрашена $\Rightarrow A \cdot B$ – деталь годная и окрашенная.

<u>Произведением нескольких событий</u> называют событие, состоящее в совместном появлении всех этих событий.

<u>Пример</u>: событие A – появление герба при первом броске монеты, B - появление герба при втором броске, C - появление герба при третьем броске, $A \cdot B \cdot C$ – появление герба при всех трех бросках.

4. Условной вероятностью $P(A|B)$ (или $P_B(A)$) называют вероятность события A, вычисленную в предположении, что событие B наступило.

Пример. В урне 3 белых шаров и 3 черных шара. Из урны дважды вынимают по одному шару, не возвращая обратно. Найти вероятность появления белого шара, при втором испытании (событие B), если при первом испытании (событие A) был извлечен черный шар. $P(B \mid A) = 3/5,\ P(B) = 1/2 \Rightarrow P(B \mid A) \neq P(B)$.

События A и B в этом примере зависимые.

$P(B)$ - безусловная вероятность, а $P(B \mid A)$ - условная вероятность.

5. Формула вероятности произведения двух зависимых событий.

$$\boxed{P(AB) = P(A) \cdot P(B \mid A) = P(B) \cdot P(A \mid B)} \Rightarrow \boxed{P(B \mid A) = \frac{P(AB)}{P(A)}} \text{ и } \boxed{P(A \mid B) = \frac{P(AB)}{P(B)}}.$$

6. Формула вероятности произведения двух независимых событий:

$P(AB) = P(A) \cdot P(B) \Rightarrow$ Условие независимости: $P(AB) = P(A) \cdot P(B)$.

7. Формула вероятности произведения нескольких зависимых событий:

$$\boxed{P(A_1, A_2, A_3, \ldots, A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 A_2) \cdot \ldots \cdot P(A_n \mid A_1 A_2 \ldots A_{n-1})}.$$

Пример. В урне 5 белых шаров, 4 черных и 3 синих шара. Наудачу извлекают по одному шару, не возвращая обратно. Найти вероятность того, что при первом испытании появится белый шар (событие А), при втором – черный (событие В), при третьем–синий (событие С).

$P(A) = 5/12,\ P(B \mid A) = 4/11,\ P(C \mid AB) = 3/10 \Rightarrow P(A \cdot B \cdot C) = (5/12)(4/11)(3/10) = 1/22.$

Замечания: 1. Порядок, в котором расположены события, может быть выбран любым, т.е. безразлично какие события можно считать первыми, вторыми и т.д.

2. Независимость событий взаимна, если A не зависит от B, то и B не зависит от A.

Формула полной вероятности

Пусть событие A может наступить при условии появления одного из несовместных событий $H_1, H_2, ..., H_n$, называемых **гипотезами** и образующих полную группу событий. При этом вероятности $p(H_i)$ и $p(A|H_i)$ будем считать известными. Тогда справедлива **формула полной вероятности** $\boxed{P(A) = \sum_{i=1}^{n} P(H_i) \cdot P(A|H_i).}$

Формула полной вероятности – «удобная схема» (форма) расчета вероятности событий.

Пример 1. Имеются 2 набора деталей. Вероятность того, что деталь первого набора стандартна, равна 0,8,

Пример 2. В первом наборе 20 деталей и из них 18 стандартны. Во втором наборе 10 деталей и из них 9 - стандартны. Из второго набора наудачу взята деталь и переложена в первый. Найти вероятность того, что деталь, наудачу извлеченная из первого набора, будет стандартна. Введем обозначения для событий: A – из первого набора извлечена стандартная деталь, H_1 – из второго набора извлечена стандартная деталь, H_2 – из второго набора извлечена нестандартная деталь. **Можно H_1 и H_2 - считать гипотезами?**

$$P(H_1) = \frac{9}{10},\ P(H_2) = \frac{1}{10},\ P(A|H_1) = \frac{19}{21},\ P(A|H_2) = \frac{18}{21} \Rightarrow$$

$$P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) = \frac{9}{10} \cdot \frac{19}{21} + \frac{1}{10} \cdot \frac{18}{21} = 0,9.$$

полной вероятности $P(A) = P(H_1) \cdot P(A|H_1) + P(H_2) \cdot P(A|H_2) = \frac{1}{2} \cdot 0,8 + \frac{1}{2} \cdot 0,9 = 0,85.$

Формула вероятности гипотез

До сих пор рассматривалась - **априорная, безусловная** вероятность гипотез $P(H_i)$. Опытные данные (реальные события) могут уточнить характеристику объекта анализа, определить $P(H_i \mid A)$ - **условную (апостериорную) вероятность гипотезы**.

По формуле произведения вероятности событий можем записать:

$$P(AH_i) = P(A)P(H_i \mid A) = P(H_i) \cdot P(A \mid H_i).$$

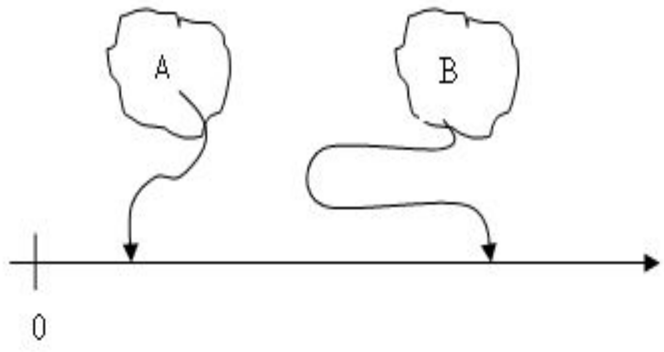
Откуда выразим **условную апостериорную** (условие - событие A произошло) вероятность

$$P(H_i / A) = \frac{P(H_i) \cdot P(A / H_i)}{P(A)}.$$

Знаменатель раскроем по формуле полной вероятности и получим **формулу вероятности гипотез (формулу Байеса**, лежащую в основе «байесовского подхода» в проверке гипотез):

$$\boxed{P(H_i \mid A) = \frac{P(H_i) \cdot P(A \mid H_i)}{\sum_i P(H_i) \cdot P(A \mid H_i)}}.$$ Например, $$P(H_1 / A) = \frac{P(H_1) \cdot P(A / H_1)}{\sum_i P(H_i) \cdot P(A / H_i)}.$$

Случайные одномерные величины



$\left.\begin{matrix}A\\B\\C\end{matrix}\right\}$ случайные события, $\Rightarrow$ $\left.\begin{matrix}X\\Y\\Z\end{matrix}\right\}$ случайные величины (с.в.) $\Rightarrow$

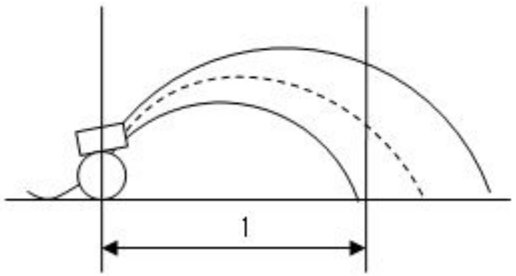
$\left.\begin{matrix}x\\y\\z\end{matrix}\right\}$ значения (в том числе и отрицательные) с.в. на числовой оси.

Случайной называют величину, которая в результате испытания примет одно и только одно возможное значение, априори неизвестное и независящее от случайных причин, которые заранее не могут быть учтены.

С.в. может быть непрерывной (аналоговой) или дискретной (прерывистой).

Случайные одномерные величины

<u>Пример непрерывной с.в.</u> - расстояние, которое пролетает снаряд из орудия, зависит от большого количества факторов (ветра, t^0, угла прицела, изменений количества и состояния пороха и т.д.), точности измерений (шаги, метры, микроны и т.д.). Имеет место и «эллипс» рассеяния (влево-вправо).



Непрерывная с.в. имеет <u>бесконечное число значений</u> и несчетно. Вероятность <u>отдельного конкретного значения случайной величины равна 0</u> (но это событие возможно). Можно говорить о вероятности диапазона значений непрерывной с.в.

<u>Дискретная с.в. может быть конечной или бесконечной</u>, но она <u>счетна</u> (ей можно поставить в соответствие натуральный ряд чисел).

<u>Примеры дискретной случайной величины</u>:

1.Число родившихся мальчиков из 100 новорожденных случайно = 0, 1, 2, 3, ... 100. Однако оно всегда дискретно и устойчиво больше, чем число родившихся девочек. С возрастом соотношение выравнивается, затем становится меньше.

2. Вероятность числа «успехов» в схеме Бернулли.

<u>Характеристики с.в.:</u>

1.Графики, таблицы.

2.Аналитические функции (интегральная и дифференциальная функция распределения), функционалы от них (числовые характеристики), квантили.

Случайные одномерные величины

Распределения вероятностей дискретной случайной величины

$X \Rightarrow x_i \Rightarrow \boxed{P(x_i)=p_i} \quad i=1,n; \sum_{i=1}^{n} p_i = 1 \Rightarrow$ Таблицы

x	x_1	x_2	...	x_n
p	p_1	p_2	...	p_n

ряд распределения,

если $x_{i+1} \geq x_i$ => вариационный ряд

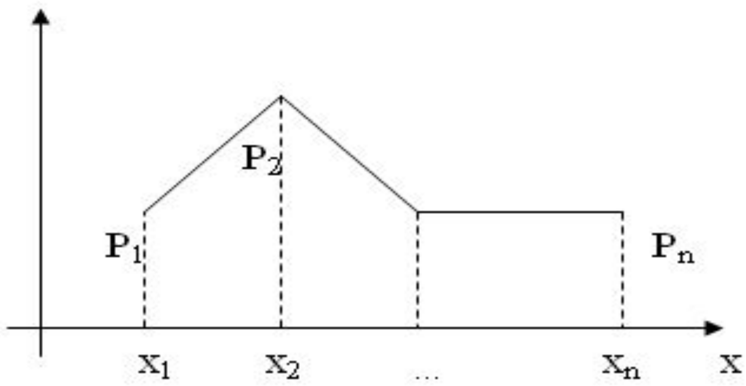


График: многоугольник или полигон распределения

Пример. В денежной лотерее выпущены 100 билетов. При этом 1 выигрыш – 50 руб., 10 выигрышей по 1 руб. Найти закон распределения. Пусть x – величина возможного выигрыша при покупке 1 билета. x_1=50 руб.$\Rightarrow p_1$=1/100;

x	0	1	50
p	0,89	0,1	0,01

x_1=10 руб.$\Rightarrow p_1$=10/100;

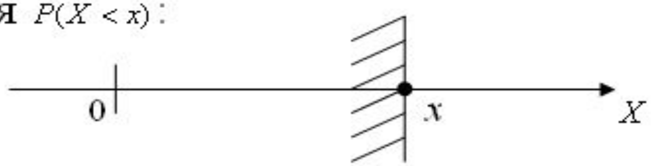
x_1=0 руб.$\Rightarrow p_1$=89/100.

Случайные одномерные величины

Функция (интегральный закон) распределения с.в.

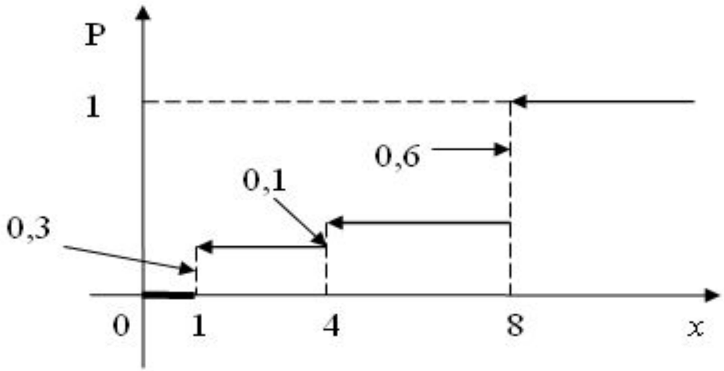
$F(x)$ – является универсальной характеристикой для непрерывных и дискретных одномерных с.в. и описывает вероятность события $P(X < x)$:

$$\boxed{F(x) = P(X < x)}.$$



<u>Пример.</u> Пусть случайная величина X задана таблицей распределения (т.е. X – дискретная случайная величина).

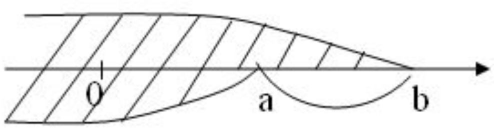
x	1	4	8
p	0,3	0,1	0,6



$$P(X < 1) = 0;\ \ P(X < 4) = 0,3;\ P(X < 8) = 0,4;\ P(X < 9) = 1;\ P(X < 10) = 1 \Rightarrow \sum_i P_i = 1.$$

СВОЙСТВА

1. $0 \le F(x) \le 1$;
2. $F(+\infty) = P(x < +\infty) = 1$;
3. $F(-\infty) = P(x < -\infty) = 0$;
4. $P(X \ge x) = 1 - F(x)$ - <u>вероятность противоположного события</u>;
5. $P(a \le x \le b) = F(b) - F(a)$ - <u>вероятность попадания в интервал значений</u>.
6. $x_2 < x_1 \Rightarrow F(x_2) \ge F(x_1)$

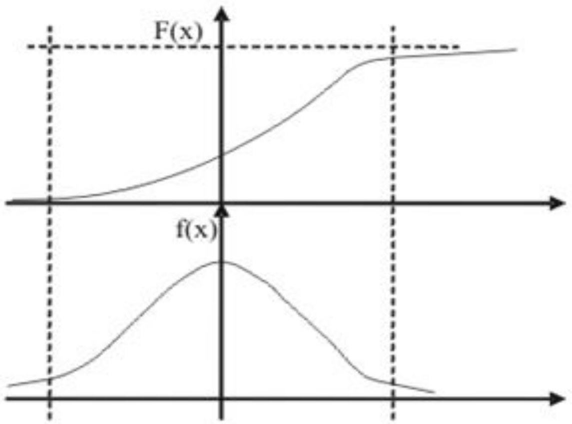


Случайные одномерные величины

Плотность вероятности

(дифференциальный закон) распределения непрерывной с.в.

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x}$$



СВОЙСТВА

1. $f(x) \geq 0$;
2. $\int\limits_{-\infty}^{+\infty} f(x)dx = 1$ - условие нормировки (единичная площадь под кривой распределения, полнота группы событий);
3. $F(x) = \int\limits_{-\infty}^{x} f(t)dt$ - выражение функции распределения через плотность.
4. $P(a \leq x \leq b) = \int\limits_{a}^{b} f(x)dx = F(b) - F(a)$ - вероятность попадания в интервал значений.

Функции $F(x)$, $f(x)$, ряды распределения p_i исчерпывающим образом описывают одномерную с.в., однако они довольно сложны, а информация, содержащаяся в них, зачастую избыточна, поэтому широко используют числовые характеристики.

Числовые характеристики распределения с.в.

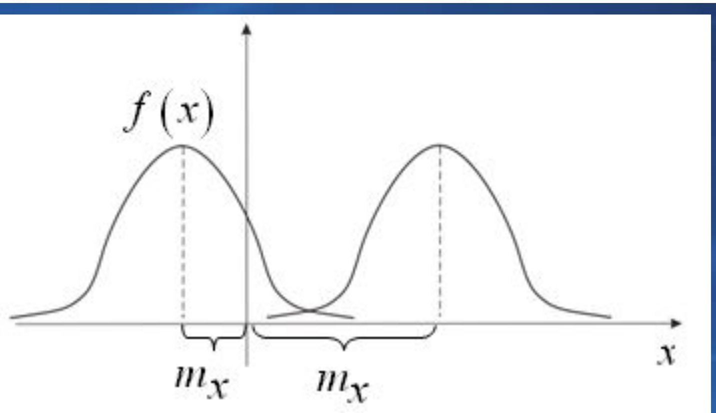
1. Математическое (безусловное) ожидание с.в.

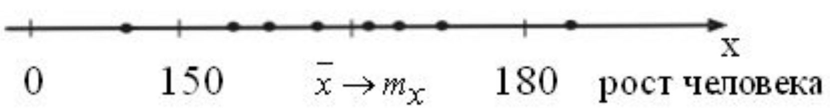
$$M[x] = m_x = \begin{cases} \sum\limits_i x_i p_i \\ \int\limits_{-\infty}^{+\infty} x f(x)dx \end{cases},$$

p_i - вероятность,

$f(x)dx = dp$ - элемент вероятности (вероятность попадания в dx),

$M[\ ]$ - оператор математического ожидания





m_x - **центр группирования, «среднее значение»,** центр тяжести плоской фигуры, ограниченной кривой $f(x)$.

Свойства матожидания

1. $M[x] = c \cdot 1 = c$;
2. $M[cx] = cm_x$;
3. $M[x \pm y] = m_x \pm m_y$;
4. $M[ax + b] = am_x + b$;
5. $M[xy] = m_x m_y$

если x и y независимые с.в.

$$\alpha_k[x] = \begin{cases} \sum\limits_i x_i \cdot P_i \\ \int\limits_{-\infty}^{\infty} x^k f(x)dx \end{cases}$$ - начальный момент

k – того порядка

$\boxed{\alpha_0 = 1}$

$\boxed{\alpha_1[x] = m_x}$

Числовые характеристики распределения с.в.

2. Дисперсия (безусловная) с.в.

$$\boxed{D_x = D[x] = \begin{cases} \sum\limits_i (x_i - m_x)^2 \cdot P_i \\ \int\limits_{-\infty}^{\infty} (x - m_x)^2 \cdot f(x)dx \end{cases}}$$

где $x_i - m_x = \overset{0}{x}$ - центрированная с.в.;

D_x - **мера разброса (рассеяния) вокруг матожидания** (в квадратных единицах),

$D[.]$ - оператор дисперсии

Свойства дисперсии

1. $D[c] = (c - c)^2 \cdot 1 = 0$
2. $D[cx] = c^2 D_x$
3. $D[x \pm y] = D_x + D_y$
4. $D[ax + b] = a^2 D_x$

$\sigma_x = \sqrt{D_x}$ - среднее квадратическое отклонение (в линейных единицах)

$k_в = \dfrac{\sigma_x}{m_x}$ - коэффициент вариации (изменчивости) (в относительных единицах)

Замечание. Коэффициент вариации применяют при анализе риска инвестирования.

Удобная формула расчета дисперсии:

$$\boxed{D[X] = M[(x - m_x)^2]} = M[x^2 - 2xm_x + m_x^2] = M[x^2] - 2m_x M[x] + m_x^2 = \boxed{M[x^2] - m_x^2}.$$

$$\boxed{\mu_k[x] = \begin{cases} \sum\limits_i (\overset{0}{x}_i)^k \; p_i \\ \int\limits_{-\infty}^{\infty} (\overset{0}{x})^k f(x)dx \end{cases}}$$ - центральный момент k – того порядка

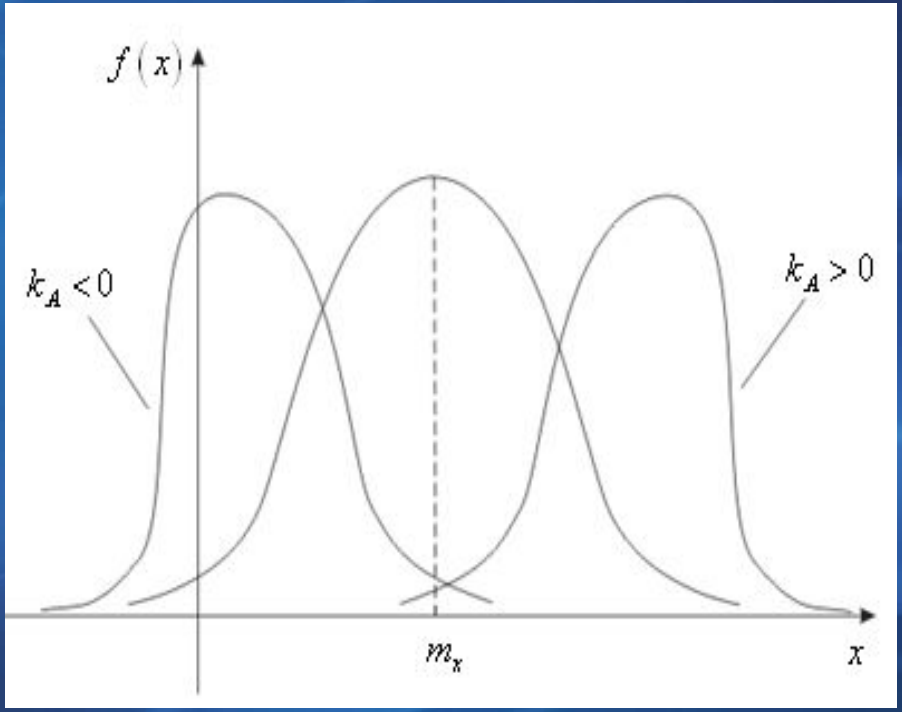
$\mu_2[x] = D_x$

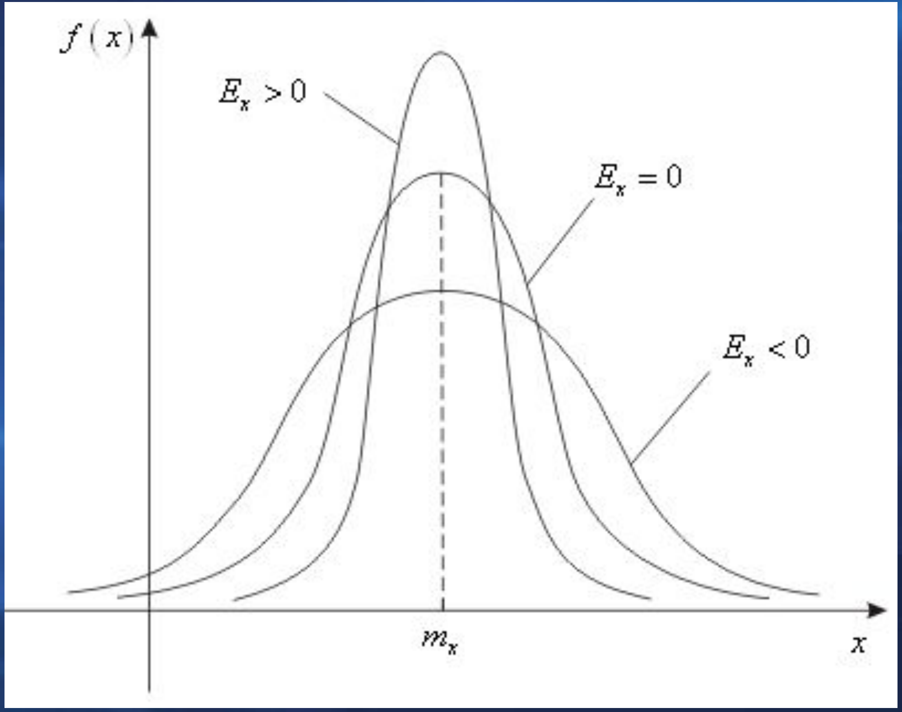
Числовые характеристики распределения с. в.

3. $k_A = \dfrac{\mu_3[x]}{\sigma_x^3}$ - **коэффициент асимметрии;**

4. λ - **эксцесс (коэффициент островершинности)**

$$\left.\begin{matrix} k_A = 0 \\ E_x = 0 \end{matrix}\right\} \Rightarrow$$ нормальное (Гауссово) распределение выступает в качестве меры (эталона):



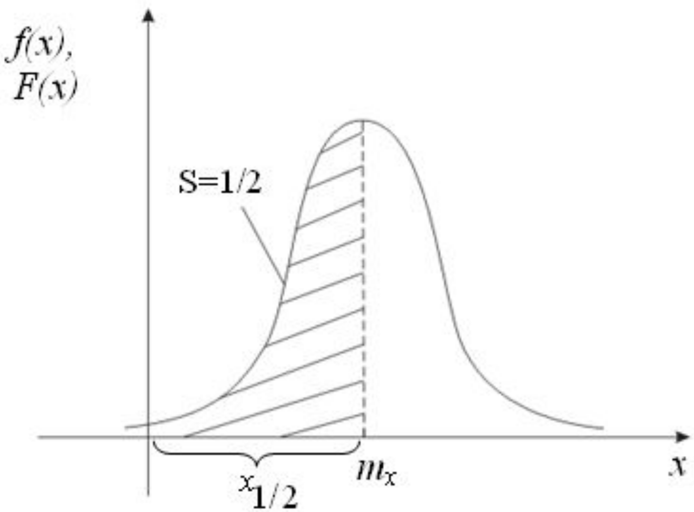


5.Квантили распределения

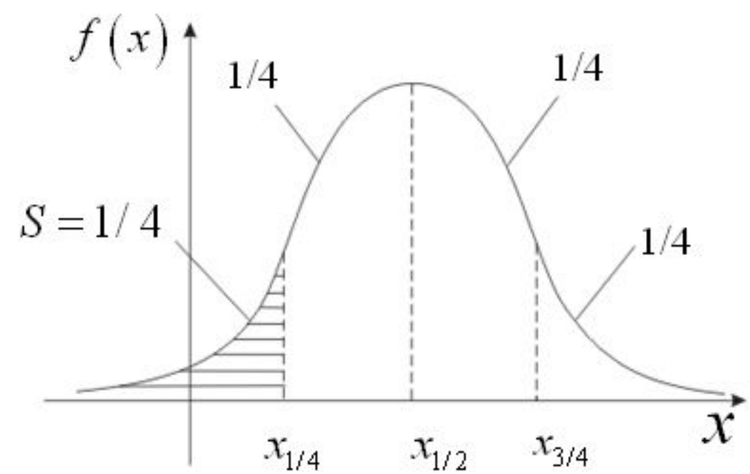
x_p - квантиль уровня (вероятности) p определяется как корень уравнения $\boxed{F(x_p)=p}$ или, соответственно, $\boxed{P(X<x_p)=p}$

Квантили

$x_{1/2}=\mu e$ (медиана) | квартили | децили | процентили



Квартили



Квартили - $x_{1/4}, x_{2/4}, x_{3/4}$; Децили - $x_{1/10}, x_{2/10}, ..., x_{9/10}$; Процентили - $x_{1/100}, x_{2/100}, ..., x_{99/100}$.

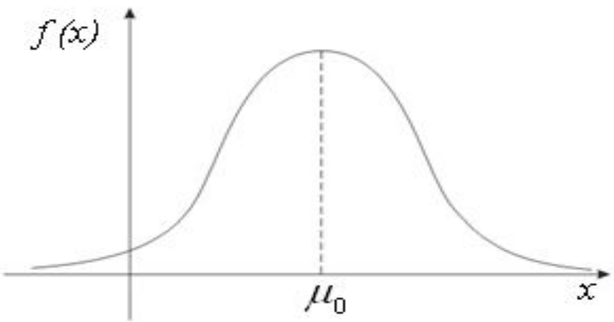
Квантили широко используется в робастной статистике (robust):

178; 187; 1655, 175; 178 $\Rightarrow m_x = 474,6$. Наблюдение 1655 – выброс (промах).

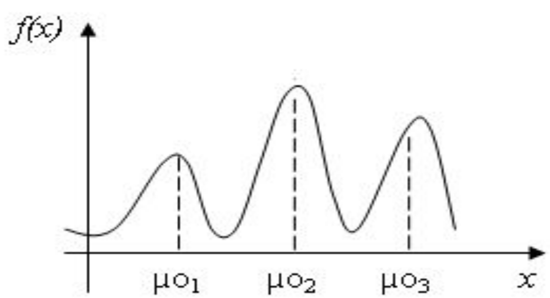
175; 178; 178; 187; 1655 $\Rightarrow \mu e = 178$

Числовые характеристики распределения с.в.

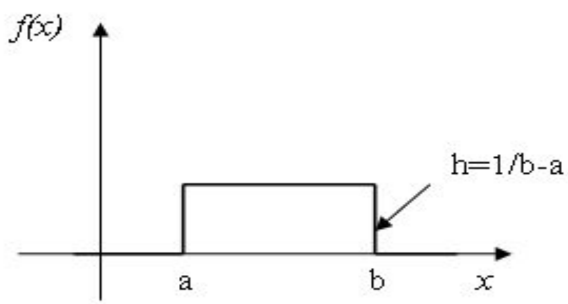
6. Мода распределения - $\boxed{\mu o = \arg\max f(x)}$



Одномодальное (унимодальное) распределение

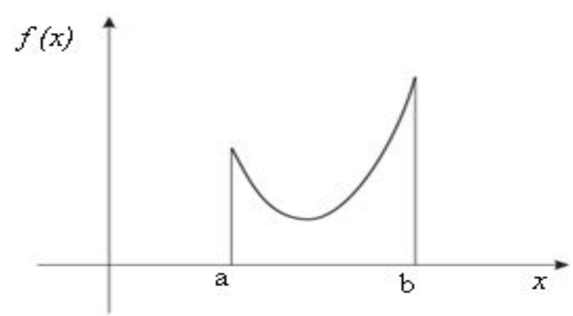


Полимодальное распределение



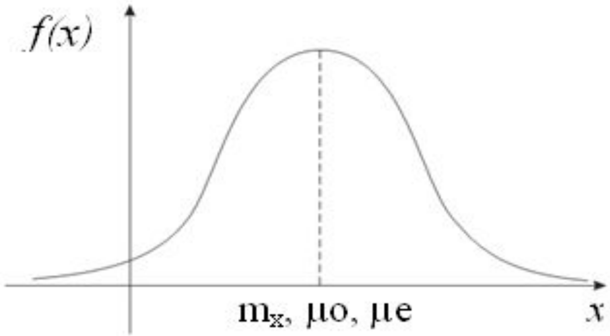
Амодальное распределение

Пример использования моды: размеры одежды, перепись населения.



Антимодальное распределение

Нормальное (Гауссово) распределение



$$f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{\frac{-(x-m_x)^2}{2\sigma_x^2}}, \quad F(x) = \frac{1}{\sigma_x\sqrt{2\pi}} \cdot \int_{-\infty}^{x} e^{\frac{-(t-m_t)^2}{2\sigma_t^2}} dt$$

$f(x) = f(x, m_x, \sigma_x)$ - двухпараметрическое распределение,

$x \sim N(m_x, \sigma_x)$ - обозначение с. в., имеющей нормальное распределение

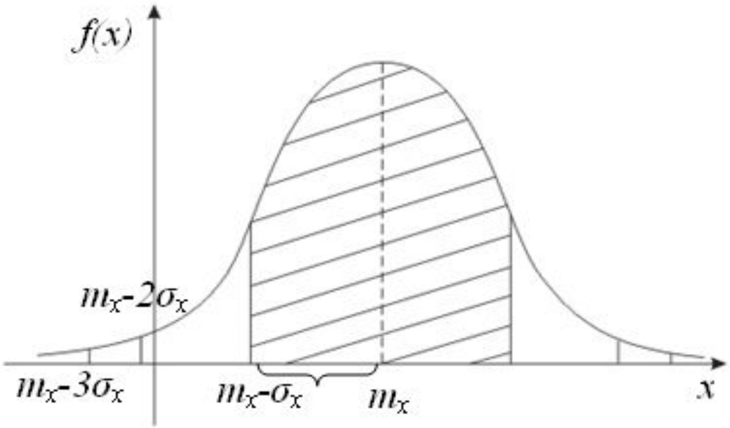
$u \sim N(0,1)$ - $\left.\begin{matrix} m_x = 0 \\ \sigma_x = 1 \end{matrix}\right\}$ - обозначение с.в., имеющей нормированное (стандартное) нормальное распределение

Правило 3-х сигм

$$P(x \in [m_x - \sigma_x, m_x + \sigma_x]) = 0,68$$

$$P(x \in [m_x - 2\sigma_x, m_x + 2\sigma_x]) = 0,95$$

$$P(x \in [m_x - 3\sigma_x, m_x + 3\sigma_x]) = 0,99$$



$m_x \pm 3\sigma_x$ - зона практического рассеивания

Операции нормирования и центрирования:

$$\frac{x-m_x}{\sigma_x} \Rightarrow M\left[\frac{x-m_x}{\sigma_x}\right] = \frac{1}{\sigma_x}\{M[x] - m_x\} = 0; \; D\left[\frac{x-m_x}{\sigma_x}\right] = \frac{1}{\sigma_x^2} \cdot \sigma_x^2 = 1,$$

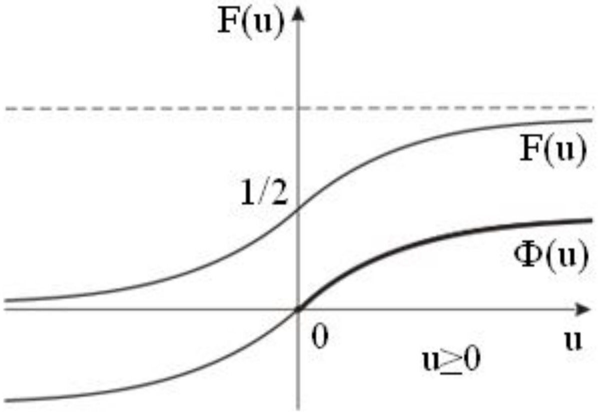
т.к. $D[x] = M\left[(x-m)^2\right] = D_x = \sigma_x^2$.

Функция Лапласа

$$u = \frac{x - m_x}{\sigma_x}, \quad f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}, \quad F(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{t^2}{2}} dt \text{ -}$$

нормальное распределение нормированной с.в. от $-\infty$ до u.

Проще рассматривать случай $u \geq 0$.



$$\boxed{\Phi(u) = F(u) - 1/2 = \frac{1}{\sqrt{2\pi}} \int_{0}^{u} e^{-\frac{t^2}{2}} dt}$$

- табулированная функция Лапласа, широко используемая для определения вероятности попадания в диапазон значений.

$$\boxed{P(a \leq x \leq b) = \Phi\left(\frac{b - m_x}{\sigma_x}\right) - \Phi\left(\frac{a - m_x}{\sigma_x}\right)}.$$

Роль нормального распределения. Теоретики и практики («мифы»)

Центральная теорема: Сумма конечного числа нормальных с.в. есть нормальная с. в.

Центральная предельная теорема: Сумма бесконечного числа с.в. с любыми законами распределения, но с примерно одинаковыми дисперсиями, имеет нормальное распределение.

Пример. В результате длительных наблюдений определено, что дивиденды X' и Y' по акциям фирм A и B являются нормальными с. в.: $X' \sim N(5;5)$; $Y' \sim N(15;15)$. Стоимость каждой акции равна 100$. Инвестор хочет приобрести акции на 1000 $.

а) Какие законы распределения имеют доходы X, Y от вложения всей суммы в акции фирмы A или B? б) Каков закон распределения имеет доход Z от покупки акций в пропорции 2/3? в) Построить графики функций случайных величин X, Y, Z. г) Какова вероятность того, что полученный доход Z от вложения будет лежать в пределах от 110$ до 150$?

а) $X = 10X' \Rightarrow X \sim N(50; \sqrt{10^2 \cdot 25} = 50)$ **или** $Y = 10Y' \Rightarrow Y \sim N(150, \sqrt{10^2 \cdot 125} = 150)$;

б) $Z = 4 \cdot X' + 6 \cdot Y' \Rightarrow Z \sim N(m_z = 4 \cdot 5 + 6 \cdot 15 = 110,$
$\sigma_z = \sqrt{16 \cdot 25 + 36 \cdot 225} = 92{,}2)$;

в) При построении графиков целесообразно пользоваться правилом 3-х сигм и соблюдать условие нормировки (площадь под кривой распределения одна и та же - равна единице).

г) $P(110 \le Z \le 150) = \Phi(0{,}43) - \Phi(0) = 0{,}16$ (используются таблицы функции Лапласа).