МАШИННОЕ ОБУЧЕНИЕ

КУРС ЛЕКЦИЙ

(предварительный вариант)

Н. Ю. Золотых

13 октября 2010

Приложение А

Элементы теории вероятности и математической статистики

Здесь мы напоминаем некоторые определения и факты из теории вероятности и математической статистики.

А.1. Вероятность

Теория вероятностей и математическая статистика занимаются построением и анализом моделей случайных событий. Примем, что *событие* наступает (или не наступает) как исход некоторого *эксперимента*. Под экспериментом здесь понимается не только научно поставленный опыт, но и произвольное воспроизведение или наблюдение какого-либо явления в определенных условиях. В частности, таким экспериментом будет считаться, например, нагревание воды до определенной температуры или подбрасывание монеты и наблюдение, какой стороной она легла, и т. п. Эксперимент также называют испытанием.

Исход *детерминированного* эксперимента определен однозначно. Пример — нагревание воды до температуры 100 °С при нормальном атмосферном давлении. Сколько бы мы не повторяли этот эксперимент при одних и тех же условиях (в частности, атмосферное давление должно составлять 100 кПа) исход известен: вода закипит. Напротив, исход *случайного*, или *недетерминированного*, эксперимента не известен заранее. Событие, которое может произойти или не произойти в результате проведения недетерминированного эксперимента называется *случайным*. Пример случайного эксперимента — подбрасывание монеты. Здесь событие Г — монета легла вверх гербом — и событие Р — монета легла вверх решеткой — являются случайными. До проведения самого испытания мы не можем сказать с уверенностью, какой стороной ляжет монета и какое событие: Г или Р — произойдет. Другой пример — подбрасывание игральной кости. Случайными событиями являются: выпадение 6 очков, выпадение четного числа очков и т. д.

С каждым случайным экспериментом можно связать множество *элементарных* исходов. В результате проведения эксперимента наступает ровно один из них. Кроме того, с помощью множества элементарных исходов можно выразить любое интересующее нас событие, которое может наступить в результате эксперимента. В примере с подбрасыванием монеты элементарными исходами являются события Г и Р. В примере с игральной костью в качестве элементарных исходов нужно взять выпадение заданного количества очков (от 1 до 6). Событие, заключающееся в выпадении четного числа очков, очевидным образом выражается через элементарные исходы как выпадение двух, четырех или шести очков.

На множестве всех недетерминированных экспериментов можно выделить большой класс испытаний со свойством *статистической устойчивости частот*. Поясним, что это такое. Пусть A — произвольное событие, которое можно наступить или не наступить в результате эксперимента E. Проведем эксперимент в неизменных условиях n раз. Обозначим $\mu(A, n)$ количество испытаний, в которых событие A происходило. Частотой наступления события A в проведенной серии экспериментов называется величина $\mu(A, n)/n$. В статистически устойчивом эксперименте эта частота при больших n должна мало отличаться от частоты наступления того же события, если провести испытание еще n раз (или другое большое количество раз). Подбрасывание монеты — пример эксперимента со свойством статистической устойчивости частот. Действительно, если монета симметричная, то при большом числе испытаний частота наступления события Г — появление герба — мало будет отличаться от $1/2$. Аналогично для события Р — появления решетки. Статистически устойчивым является и эксперимент с подбрасыванием игральной кости. В случае, если кость симметричная, то частота выпадения 6 очков в большой серии экспериментов мало будет отличаться от $1/6$, а частота выпадения четного числа очков будет близка к $1/2$.

Итак, в статистически устойчивом эксперименте частота наступления события A должна мало отличаться от некоторого значения, которое называется вероятностью этого события. Важно подчеркнуть, чтобы при этом была возможность (по крайней мере, потенциальная) повторения эксперимента в неизменных условиях[1]

Рассмотрим теперь математическую модель этой ситуации.

Пусть Ω — некоторое непустое множество (конечное или бесконечное) и пусть $\mathbf{A} \subseteq 2^{\Omega}$. Множество $\mathbf{A}$ называется *сигма-алгеброй* над Ω, если $\Omega \in \mathbf{A}$ и $\mathbf{A}$ замкнуто относительно операций счетного объединения и дополнения, т. е. для любого A и любого конечного или счетного множества $\{A_1, A_2, \ldots\}$, где $A_i \in \mathbf{A}$ ($i = 1, 2, \ldots$), справедливо

$$\bigcup_i A_i \in \mathbf{A},$$

$$\overline{A} = \Omega \setminus A \in \mathbf{A}.$$

Тройка $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$ называется *вероятностным пространством*, если Ω — множество всех элементарных исходов, $\mathbf{A}$ — сигма-алгебра над Ω, а $\mathsf{Pr} : \mathbf{A} \to \{0, 1\}$ — *вероятностная мера* (*вероятность*). Элементы множества $\mathbf{A}$ называются *событиями*, или *исходами*, при этом $\emptyset$ называется *невозможным событием*, а Ω — *достоверным событием*. Функция Pr ставит в соответствие каждому событию A его *вероятность* $\mathsf{Pr}\ A$, так, что выполнены следующие *аксиомы Колмогорова*.

[1] Заметим, что предпринимаются попытки определять вероятности для экспериментов, повторение которых невозможно. В этой связи интересно упомянуть книгу Unwin S. D. The Probability of God: A Simple Calculation That Proves the Ultimate Truth. Crown Forum, 2003. Русс. перев. Анвин Ст. Простое вычисление, доказывающее конечную истину или вероятность Бога. АСТ, Астрель. 2008.

1. Для любого события A из $\mathbf{A}$ справедливо $\mathsf{Pr}\ A \geq 0$.
2. $\mathsf{Pr}\ \Omega = 1$.
3. Если $\{A_1, A_2, \ldots\}$ — конечное или счетное множество *несовместных* событий, т. е. $A_i \cap A_j = \emptyset$ при $i \neq j$, то
$$\mathsf{Pr}\left(\bigcup_i A_i\right) = \sum_i \mathsf{Pr}\ A_i.$$

Приведенная система аксиом непротиворечива и неполна. Непротиворечивость подтверждается существованием конкретных моделей (интерпретаций), построенных по этим аксиомам (см., в частности, примеры ниже). Неполнота данной системы означает, что конкретная модель не единственна. Модель строится согласно этой системе аксиом, так, чтобы быть адекватной той реальной ситуации в природе, технике и т. п., которую она описывает. В частности, вероятностная мера Pr должна выбираться так, чтобы $\mathsf{Pr}\ A$ хорошо приближало частоту наступления события A для любого $A \in \mathbf{A}$.

Если $A \in \mathbf{A}$, $\omega \in A$, то говорят, что элементарный исход ω благоприятствует событию A. Для краткости вероятность $\mathsf{Pr}\ \{\omega\}$ события $\{\omega\}$, которому благоприятсвует лишь один элементарный исход ω, называют просто вероятностью этого элементарного исхода. Если множество Ω конечно или счетно, то, по аксиоме 3, вероятность события можно получить как сумму вероятностей благоприятстующих элементарных исходов.

Вероятность $\mathsf{Pr}\ \{\omega \in \Omega:\ \text{условие}\}$ часто обозначается просто как $\mathsf{Pr}\ \{\text{условие}\}$.

Из аксиом вытекают следующие основные свойства вероятности.
$$\mathsf{Pr}\ \emptyset = 0.$$
Для любого события A из $\mathbf{A}$
$$0 \leq \mathsf{Pr}\ A \leq 1.$$
Для любого события A из $\mathbf{A}$
$$\mathsf{Pr}\ \overline{A} = 1 - \mathsf{Pr}(A).$$
Правило сложения:
$$\mathsf{Pr}(A + B) = \mathsf{Pr}\ A + \mathsf{Pr}\ B - \mathsf{Pr}(A \cap B).$$
Если события *несовместные*, т. е. $A \cap B = \emptyset$, то $\mathsf{Pr}(A + B) = \mathsf{Pr}\ A + \mathsf{Pr}\ B$.

Рассмотрим некоторые примеры. Предположим, что некто непреднамеренным образом подбрасывает симметричную монету, которая может упасть либо гербом, либо решеткой. Положим $\Omega = \{\Gamma, \text{Р}\}$, где Γ — соответствует гербу, Р — решетке; $\mathbf{A} = 2^\Omega = \{\emptyset,\ \{\Gamma\}, \{\text{Р}\}, \Omega\}$. Вероятностная мера полностью определяется значениями $\mathsf{Pr}\ \{\Gamma\} = 1/2$, $\mathsf{Pr}\ \{\text{Р}\} = 1/2$.

Рассмотрим другой пример. Подбрасывается игральная кость. В качестве Ω можно взять $\Omega = \{1, 2, 3, 4, 5, 6\}$, где каждое число соответсвует выпавшему количеству очков. Пусть $\mathbf{A} = 2^\Omega$. Например, $A = \{2, 4, 6\}$ обозначает событие, заключающееся в том, что выпало четное число очков. Если кость симметричная, то следует положить $\mathsf{Pr}\,\{1\} = \mathsf{Pr}\,\{2\} = \mathsf{Pr}\,\{3\} = \mathsf{Pr}\,\{4\} = \mathsf{Pr}\,\{5\} = \mathsf{Pr}\,\{6\} = 1/6$. Вероятность других событий можно получить суммированием вероятностей благоприятствующих исходов. Например, для вероятности события A получим $\mathsf{Pr}\,A = \mathsf{Pr}\,\{2\} + \mathsf{Pr}\,\{4\} + \mathsf{Pr}\,\{6\} = 1/2$.

Пусть теперь подбрасывается две кости. Положим

$$\Omega = \{(i, j) : \ 1 \le i \le 6, \ 1 \le j \le 6\}$$

— множество всех упорядоченных пар чисел от 1 до 6. Далее, $\mathbf{A} = 2^{\Omega}$. Например,

$$A = \{(1, 5), \ (2, 4), \ (3, 3), \ (4, 2), \ (5, 1)\}$$

обозначает событие, заключающееся в том, что сумма выпавших очков на двух костях равна 6, а

$$B = \{(1, 6), \ (2, 5), \ (3, 4), \ (4, 3), \ (5, 2), \ (6, 1)\}$$

— событие, заключающееся в том, что сумма выпавших очков на двух костях равна 7. Для симметричных костей $\mathsf{Pr}\,\{(i, j)\} = 1/36$ и поэтому, например, $\mathsf{Pr}\,A = 5/36$, $\mathsf{Pr}\,B = 1/6$.

Рассмотрим эксперимент, заключающийся в том, что на отрезке $[0, 1]$ случайно выбирается точка. В качестве Ω возьмем сам отрезок $[0, 1]$. Элементарный исход ω есть координата выбранной точки. В качестве $\mathbf{A}$ возьмем множество всех измеримых (т. е. имеющих длину) подмножеств множества Ω. Например, событие $A_0 = \{\omega : \ 1/4 \le \omega \le 3/4\}$ заключается в том, что выбранная случайная точка располагается на расстоянии, не превышающем $1/4$, от середины отрезка. Теперь перейдем к вопросу задания вероятностной меры Pr. Заметим, что множество Ω несчетно, поэтому, чтобы определить Pr на всем множестве $\mathbf{A}$ не достаточно задания значений $\mathsf{Pr}\,\{\omega\}$ для всех $\omega \in \Omega$. Действительно, если событию благоприятствует несчетное число элементарных исходов (как, например, событию A_0), то невозможно определить его вероятность путем суммирования вероятностей благоприятствующих элементарных исходов. Зададим меру Pr следующим образом: пусть $\mathsf{Pr}\,A$ равно длине множества A (например, для события A_0 получим $\mathsf{Pr}\,A_0 = 1/2$). Легко видеть, что все аксиомы вероятности при этом будут выполнены. Заметим, что $\mathsf{Pr}\,\{\omega\} = 0$ для любого элементарного исхода ω, т. е. вероятность выбора заданной точки на отрезке $[0, 1]$ равна 0.

А.2. Независимые события и условная вероятность

События A и B называются *независимыми*, если

$$\mathsf{Pr}(A \cap B) = \mathsf{Pr}\,A \cdot \mathsf{Pr}\,B.$$

События $A_1, A_2, \ldots, A_s$ называются *независимыми в совокупности*, если

$$\mathsf{Pr}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_s}) = \mathsf{Pr}(A_{i_1}) \cdot \mathsf{Pr}(A_{i_2}) \cdot \ldots \cdot \mathsf{Pr}(A_{i_k}) \qquad (79)$$

для любого подмножества $\{i_1, i_2, \ldots, i_k\}$ произвольной мощности k множества $\{1, 2, \ldots, s\}$.

Вероятностью наступления события A при условии, что наступило событие B, называется

$$\mathsf{Pr}(A\,|\,B) = \frac{\mathsf{Pr}(A \cap B)}{\mathsf{Pr}\ B}. \qquad (80)$$

В отличие от *условной вероятности* $\mathsf{Pr}(A\,|\,B)$ вероятность $\mathsf{Pr}(A)$ называется *безусловной*, или *маргинальной*. Можно показать, что $\mathbf{A} \cap B$ является сигма-алгеброй и $\mathsf{Pr}(A\,|\,B)$ удовлетворяет всем аксиомам Колмогорова для вероятности на сигма-алгебре $\mathbf{A} \cap B$. Аналогично имеем

$$\mathsf{Pr}(B\,|\,A) = \frac{\mathsf{Pr}(A \cap B)}{\mathsf{Pr}\ A}. \qquad (81)$$

Из (80) и (81) получаем:

$$\mathsf{Pr}(B\,|\,A) = \frac{\mathsf{Pr}\ B\ \mathsf{Pr}(A\,|\,B)}{\mathsf{Pr}\ A}. \qquad (82)$$

Если $A_1, A_2, \ldots, A_s$ — полная группа несовместных событий, т. е.

$$\bigcup_{j=1}^{s} A_j = \Omega, \qquad A_i \cap A_j = \emptyset \quad (i \neq j)$$

то из (80) легко получить *формулу полной вероятности*:

$$\mathsf{Pr}(A) = \sum_{i=1}^{s} \mathsf{Pr}(A_i)\,\mathsf{Pr}(A\,|\,A_i). \qquad (83)$$

Теперь имеем

$$\mathsf{Pr}(A_j\,|\,A) = \frac{\mathsf{Pr}(A_j) \cdot \mathsf{Pr}(A\,|\,A_j)}{\sum\limits_{i=1}^{m} \mathsf{Pr}(A_i) \cdot \mathsf{Pr}(A\,|\,A_i)}. \qquad (84)$$

Равенства (82) и (84) называются *формулами Байеса*.

А.3. Случайные величины

А.3.1. Одномерные случайные величины

Борелевской сигма-алгеброй над $\mathbf{R}$ называется минимальная по включению сигма-алгебра над $\mathbf{R}$, содержащая все полуинтервалы вида $[a, b)$, где a, b — произвольные числа из R, такие, что $a < b$. Можно доказать, что борелевская сигма-алгебра над $\mathbf{R}$ существует и единственна.

Пусть F — некоторая сигма-алгебра над Ω, а $\mathscr{B}$ — борелевская сигма-алгебра над $\mathbf{R}$. Отображение $X : \Omega \to \mathbf{R}$ называется *измеримым*, если для произвольного B из $\mathscr{B}$ справедливо

$$\{\omega : \ X(\omega) \in B\} \in \mathbf{A}.$$

Для того, чтобы отображение $X : \Omega \to \mathbf{R}$ было измеримым достаточно потребовать, чтобы $\{\omega : \ X(\omega) < x\} \in \mathbf{A}$ для любого $x \in \mathbf{R}$.

Всякое измеримое отображение $\Omega \to \mathbf{R}$ называется *случайной величиной*, или *случайной переменной*.

Как правило, случайные величины мы будем обозначать большими латинскими буквами, а значения, которые они принимают — соответствующими малыми латинскими.

Интегральной (или *кумулятивной*) *функцией распределения* (или просто *функцией распределения*) случайной величины X называется

$$P_X(x) = \mathsf{Pr}\, \{\omega : \ X(\omega) < x\}\, .$$

Индекс X может опускаться, если ясно, о какой случайной величине идет речь. Интегральная функция распределения обладает следующими свойствами.

$$0 \le P(x) \le 1, \qquad \lim_{x \to +\infty} P(x) = 1, \qquad \lim_{x \to -\infty} P(x) = 0.$$

$P(x)$ — неубывающая функция.

$$\mathsf{Pr}\, \{x_1 \le X(\omega) < x_2\} = P(x_2) - P(x_1).$$

Интегральная функция распределения непрерывна слева, т. е.

$$\lim_{x \to x_0 - 0} P(x) = P(x_0),$$

но может быть разрывная справа и

$$\lim_{x \to x_0 + 0} P(x) = \mathsf{Pr}\, \{X \le x_0\}\, ,$$

откуда

$$\mathsf{Pr}\, \{X = x_0\} = \lim_{x \to x_0 + 0} P(x) - \mathsf{Pr}\, \{X \le x_0\}\, .$$

Случайная переменная X называется *дискретной*, если она принимает лишь конечное или счетное число различных значений. Если число этих значений конечно, то закон распределения можно задать с помощью таблицы вида

$$\begin{array}{|c|cccc|} \hline X & x_1 & x_2 & \ldots & x_s \\ \hline \mathsf{Pr} & p_1 & p_2 & \ldots & p_s \\ \hline \end{array} \tag{85}$$

называемой *рядом распределения*. Здесь $x_1 \le x_2 \le \ldots \le x_s$, $p_1 + p_2 + \ldots + p_s = 1$ и $\mathsf{Pr}(X = x_i) = p_i$ $(i = 1, 2, \ldots, s)$. Если на каждом конечном отрезке число значений, которые принимает X конечно, то $P_X(x)$ представляет собой ступенчатую функцию, с точками разрыва в этих значениях.

Случайная переменная X называется *непрерывной*, если найдется такая функция $p_X(x)$, что

$$P_X(x) = \int_{-\infty}^{x} p_X(x)\, dx.$$

Функция $p_X(x)$ называется *плотностью вероятности*. Если ясно, о какой случайной величине X идет речь, то индекс X опускается. Плотность вероятности обладает следующими свойствами.

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

Для любых a и b

$$\mathsf{Pr}\,\{a \le X < b\} = \int_a^b p(x)\,dx.$$

На всей числовой прямой, кроме множества точек меры Лебега 0 верно

$$p(x) = \frac{dP(x)}{dx}.$$

Функция $p(x)$ на числовой прямой имеет не более чем счетное число точек разрыва, а на любом конечном интервале — не более чем конечное число точек разрыва.

А.3.2. Многомерные случайные величины

Если случайные переменные $X_1, X_2, \ldots, X_d$ заданы на одном вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$, то упорядоченный набор $X = (X_1, X_2, \ldots, X_d)$ называется *многомерной случайной величиной*, или *многомерным случайным вектором*. Интегральной функцией распределения называется

$$P_X(x_1, x_2, \ldots, x_d) = \mathsf{Pr}\,\{X_1 < x_1,\ X_2 < x_2,\ \ldots,\ X_d < x_d\}$$

Функцию $P_X(x_1, x_2, \ldots, x_d)$ называют также (*интегральной*) *функций совместного распределения* случайных величин $X_1, X_2, \ldots, X_d$. Пусть $x = (x_1, x_2, \ldots, x_d)$, тогда функцию $P_X(x_1, x_2, \ldots, x_d)$ можно обозначить $P_X(x)$.

Если каждая из переменных x_i принимает не более чем счетное число значений, то X называется *дискретной*.

Если найдется функция $p_X(x) = p_X(x_1, x_2, \ldots, x_d)$, зависящая от векторного аргумена $x \in \mathbf{R}^d$ (или, что то же, зависящая от d скалярных аргументов $x_1, x_2, \ldots, x_d$), такая, что

$$P_X(x) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_d} p_X(x)\,dx_1 dx_2 \ldots dx_d,$$

то случайная многомерная переменная X называется *непрерывной*, а $p_X(x)$ называется ее *плотностью вероятности*. Функцию $P_X(x_1, x_2, \ldots, x_d)$ называют также плотностью вероятности совместного распределения случайных величин $X_1, X_2, \ldots, X_d$. Плотность вероятности обладает следующими свойствами.

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

Если $G \subseteq \mathbf{R}^d$, то

$$\mathsf{Pr}\,\{X \in G\} = \int_G p(x)\,dx.$$

Пусть X_1, X_2 — случайные переменные, заданные над одним вероятностным пространством $\langle\Omega, \mathbf{A}, \mathsf{Pr}\rangle$. Эти переменные называются независимыми, если для любых борелевских множеств B_1 и B_2 из $\mathscr{B}$, где $\mathscr{B}$ — борелевская сигма-алгебра над $\mathbf{R}$, имеем

$$\mathsf{Pr}\,\{X_1 \in B_1, X_2 \in B_2\} = \mathsf{Pr}\,\{X_1 \in B_1\} \cdot \mathsf{Pr}\,\{X_2 \in B_2\}\,.$$

Для независимости необходимо и достаточно, чтобы для любых x_1, x_2

$$P(x_1, x_2) = P_{X_1}(x_1) \cdot P_{X_2}(x_2),$$

а в случае непрерывных X_1, X_2 —

$$p(x_1, x_2) = p_{X_1}(x_1) \cdot p_{X_2}(x_2).$$

Аналогично (79) вводится определение *независимости в совокупности s* случайных величин.

Пусть X — одномерная случайная функция над вероятностным пространством $\langle\Omega, \mathbf{A}, \mathsf{Pr}\rangle$ и $A \in \mathbf{A}$, $\mathsf{Pr}\,A \neq 0$. *Условной интегральной функцией распределения*, при условии, что событие A произошло, называется

$$P_X(x\,|\,A) = \mathsf{Pr}\Big(\{\omega :\ X(\omega) < x\}\ |\,A\Big) = \frac{\mathsf{Pr}\Big(\{\omega :\ X(\omega) < x\} \cap A\Big)}{\mathsf{Pr}(A)}.$$

Если X — непрерывная случайная величина, то можно определить *условную плотность вероятности*:

$$p_X(x\,|\,A) = \frac{dP_X(x\,|\,A)}{dx}.$$

В частности, если $A = \{\omega :\ Y(\omega) = y\}$, где Y — случайная переменная на том же вероятностном пространстве, то $p_X(x\,|\,A)$ обозначается $p_X(x\,|\,y)$. Можно доказать, что если $p_Y(y) \neq 0$, то

$$p_X(x\,|\,y) = \frac{p(x, y)}{p_Y(y)}.$$

Как обычно, индексы можно опустить, если понятно, о каких случайных переменных идет речь.

Справедливы следующие аналоги формулы полной вероятности и формулы Байеса:

$$\mathsf{Pr}\,A = \int_{-\infty}^{+\infty} \mathsf{Pr}(A\,|\,x) \cdot p(x)\,dx, \qquad p(x\,|\,A) = \frac{\mathsf{Pr}(A\,|\,x) \cdot p(x)}{\int\limits_{-\infty}^{+\infty} \mathsf{Pr}(A\,|\,x) \cdot p(x)\,dx}.$$

Пусть X — случайная величина, а $\varphi :\ \mathbf{R} \to \mathbf{R}$ — измеримая функция. Можно доказать, что тогда $Y = \varphi(X)$ — также случайная величина, причем если X — непрерывная величина, а φ — строго монотонная дифференцируемая функция, то

$$p_Y(y) = \frac{p_X(x)}{|\varphi'(x)|}. \tag{86}$$

А.4. Характеристики случайных величин

А.4.1. Характеристики одномерных случайных величин

Пусть закон распределения дискретной случайной величины X задан рядом распределения (85), тогда *математическим ожиданием* (или *средним значением*) случайной величины X называется

$$\mathsf{E}\,X = \sum_{i=1}^{s} x_i p_i. \tag{87}$$

Если дискретная случайная величина принимает бесконечно много различных значений, то вместо конечной суммы получаем ряд. Если ряд не сходится абсолютно, то говорят, то X не имеет математического ожидания.

Математическим ожиданием непрерывной случайной величины X называется

$$\mathsf{E}\,X = \int_{-\infty}^{+\infty} x p(x)\,dx. \tag{88}$$

Если $\int\limits_{-\infty}^{+\infty} |x|p(x)\,dx$ расходится, то говорят, что X не имеет математического ожидания.

Для случайной величины произвольного типа математическое ожидание можно определить с помощью интеграла Стильтьеса:

$$\mathsf{E}\,X = \int_{-\infty}^{+\infty} x\,dP(x).$$

Для дискретной (соответственно непрерывной) случайной величины эта формула превращается в (87) (соответственно в (88)).

Математическое ожидание обладает следующими свойствами. Если c — константа, а X — случайная величина, то

$$\mathsf{E}\,c = c, \qquad \mathsf{E}\,(c \cdot X) = c \cdot \mathsf{E}\,X,$$

Если X_1, X_2 — случайные величины, то

$$\mathsf{E}\,(X_1 + X_2) = \mathsf{E}\,X_1 + \mathsf{E}\,X_2.$$

Если X_1 и X_2 независимы, то

$$\mathsf{E}\,(X_1 \cdot X_2) = \mathsf{E}\,X_1 \cdot \mathsf{E}\,X_2. \tag{89}$$

Дисперсией случайной величины X называется

$$\mathsf{D}\,X = \mathsf{E}\,(X - \mathsf{E}\,X)^2 = \mathsf{E}\,X^2 - (\mathsf{E}\,X)^2.$$

Для дискретной случайной величины, закон распределения которой задан таблицей (85),

$$\mathsf{D}\,X = \sum_{i=1}^{s} (x_i - \mathsf{E}\,X)^2 p_i.$$

Для непрерывной случайной величины

$$\mathsf{D}\,X = \int_{-\infty}^{+\infty} (x - \mathsf{E}\,X)^2 f(x)\,dx$$

Величина $\sigma(X) = \sqrt{\mathsf{D}\,X}$ называется *средним квадратическим отклонением*.

Дисперсия обладает следующими свойствами. Если c — константа, а X — случайная величина, то

$$\mathsf{D}\,c = 0, \qquad \mathsf{D}\,(c \cdot X) = c^2 \cdot \mathsf{D}\,X,$$

Если X_1, X_2 — независимые случайные величины, то

$$\mathsf{D}\,(X_1 + X_2) = \mathsf{D}\,(X_1) + \mathsf{D}\,(X_2).$$

Для любого положительного ε справедливо *неравенство Чебышева*:

$$\mathsf{Pr}\,\{|X - \mathsf{E}\,X| > \varepsilon\} \le \frac{\mathsf{D}\,X}{\varepsilon^2}.$$

Пусть k — натуральное число. *Начальным моментом* k-го порядка случайной величины X называется $\mathsf{E}\,X^k$. *Центральным моментом* k-го порядка случайной величины X называется $\mathsf{E}\,(X - \mathsf{E}\,)^k$.

Пусть p — действительной число из отрезка $[0, 1]$. *Квантилем* уровня p случайной величины X называется число $x \in \mathbf{R}$, такое, что

$$P(x) = \mathsf{Pr}\,\{X < x\} = p.$$

Квантиль уровня $\frac{1}{2}$ называется *медианой*, квантиль уровня $\frac{1}{4}$ — *нижним квартилем*, а квантиль уровня $\frac{3}{4}$ — *верхним квартилем*. *Процентилем*, или *процентной точкой*, уровня p называется число $x \in \mathbf{R}$, такое, что

$$\mathsf{Pr}\,\{|X - \mathsf{E}\,X| < x\} = p.$$

А.4.2. Ковариация и корреляция случайных величин

Ковариацией двух случайных величин X, Y, заданных на одном вероятностном пространстве $\langle \Omega,\ \mathbf{A},\ \mathsf{Pr} \rangle$, называется

$$\mathsf{Cov}(X, Y) = \mathsf{E}\,\big((X - \mathsf{E}\,X) \cdot (Y - \mathsf{E}\,Y)\big) = \mathsf{E}\,(X \cdot Y) - \mathsf{E}\,X \cdot \mathsf{E}\,Y.$$

Очевидно, что если случайные величины независимы, то ковариация равна нулю. Обратное в общем случае не верно. Величины называются *некоррелированными*, если их ковариация равна нулю. Очевидно, что для таких величин

$$\mathsf{E}\,(XY) = \mathsf{E}\,X \cdot \mathsf{E}\,Y.$$

Корреляцией называется

$$\mathsf{Corr}(X, Y) = \frac{\mathsf{Cov}(X, Y)}{\sigma X \cdot \sigma Y}.$$

Можно доказать, что

$$-1 \leq \mathsf{Corr}(X, Y) \leq 1, \tag{90}$$

причем $|\mathsf{Corr}(X, Y)| = 1$ тогда и только тогда, когда с вероятностью 1 между X и Y имеется линейная связь, т. е. существуют постоянные α и β, такие, что

$$Y = \alpha + \beta X \quad \text{или} \quad X = \alpha + \beta Y. \tag{91}$$

Значения корреляции, близкие к 1 или -1, указывают, что между величинами X и Y имеется связь, близкая к линейной. Значения, близкие к 0, указывают, что связь между величинами слаба или связь носит нелинейный характер[2].

А.4.3. Характеристики многомерных случайных величин

Пусть $X = (X_1, X_2, \dots, X_d)$ — случайный вектор. Матрица $\mathsf{Cov}(X_1, X_2, \dots, X_d) = (c_{ij})$, в которой c_{ij} есть ковариация случайных переменных X_i и X_j, называется *матрицей ковариации*. Предполагая, что X — вектор-столбец, матрицу ковариации можно записать, используя матричные обозначения:

$$\mathsf{Cov}\, X = \mathsf{E}\left((X - \mathsf{E}\, X) \cdot (X - \mathsf{E}\, X)^\top\right).$$

Математическим ожиданием случайного вектора $X = (X_1, X_2, \dots, X_d)$ называется вектор $\mathsf{E}\, X = (\mathsf{E}\, X_1, \mathsf{E}\, X_2, \dots, \mathsf{E}\, X_d)$.

Рассмотрим случайный вектор $X - \mathsf{E}\, X$. Пусть α — вектор-столбец в $\mathbf{R}^d$, $\|\alpha\| = 1$. Проекцию вектора $X - \mathsf{E}\, X$ на направление α обозначим Y_α. Величина Y есть одномерная случайная величина. Оказывается,

$$\mathsf{D}\, Y_\alpha = \alpha^\top \cdot \mathsf{Cov}(X_1, X_2, \dots, X_d) \cdot \alpha.$$

Пусть $\varphi = (\varphi_1, \dots, \varphi_d) : \ \mathbf{R}^d \to \mathbf{R}^d$, а X и Y — непрерывные многомерные случайные величины, причем $X = \varphi(Y)$. Обозначим

$$J = \begin{pmatrix} \dfrac{\partial \varphi_1}{\partial x_1} & \dots & \dfrac{\partial \varphi_d}{\partial x_1} \\ \dots & \dots & \dots \\ \dfrac{\partial \varphi_1}{\partial x_d} & \dots & \dfrac{\partial \varphi_d}{\partial x_d} \end{pmatrix}$$

[2] Рассмотрим множество всех случайных величин с нулевым математическим ожиданием и конечной дисперсией, определенных на одном и том же вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$. Будем отождествлять случайные величины X и Y, для которых $\mathsf{Pr}\,\{X = Y\} = 1$. Нетрудно проверить, что указанное множество образует линейное пространство. Следуя общепринятой терминологии, будем называть элементы этого пространства (т. е. случайные величины) векторами. Функция $(X, Y) = \mathsf{Cov}(X, Y)$ удовлетворяет аксиомам скалярного произведения и поэтому данное линейное пространство с введенным подобным образом скалярным произведением является евклидовым. Заметим тогда, что норма $\|X\|$ вектора X есть σX, а $\mathsf{Corr}(X, Y)$ равна косинусу угла между векторами X и Y. Согласно неравенству Коши–Буняковского, имеем $|(X, Y)| \leq \|X\| \cdot \|Y\|$. Записывая это неравенство для случайных величин $X - \mathsf{E}\, X$ и $Y - \mathsf{E}\, Y$ (мы уже не предполагаем, что $\mathsf{E}\, X = \mathsf{E}\, Y = 0$), получаем неравенство (90). Известно, что равенство в неравенстве Коши–Буняковского достигается тогда и только тогда, когда векторы X и Y коллинеарны, откуда имеем (91).

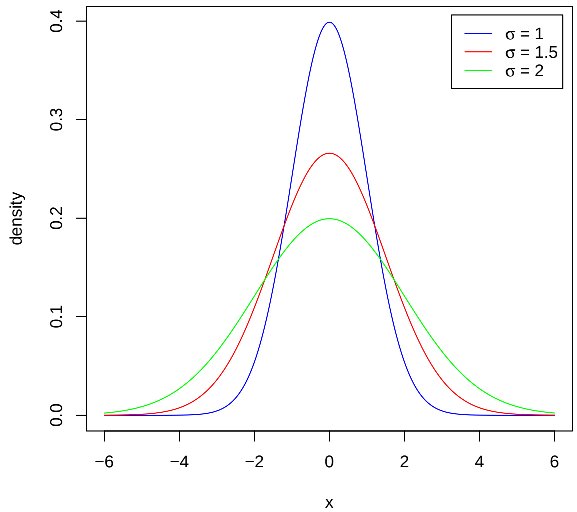

Рис. A.1. Графики плотностей нормального распределения для разных значений σ.

Тогда если якобиан $\det J$ не равен нулю для всех x, тогда

$$p_Y(y) = \frac{p_X(x)}{|\det J|}$$

(ср. с (86)). В частности, в случае линейного преобразования $Y = Ax + \mu$ получаем

$$p_Y(y) = \frac{p_X\left(A^{-1}(y-\mu)\right)}{|\det A|}.$$

A.5. Некоторые типы распределений

A.5.1. Нормальное распределение

Говорят, что непрерывная случайная величина имеет *нормальное распределение*, если ее плотность вероятности равна

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\cdot\frac{(x-\mu)^2}{\sigma^2}}.$$

Нетрудно проверить, что $\mathsf{E}\,X = \mu$, $\mathsf{D}\,X = \sigma^2$. Для нормального распределения с математическим ожиданием μ и среднеквадратическим отклонением σ будем использовать обозначение $N(\mu, \sigma)$.

Графики плотностей нормального распределения для разных значений σ приведены на рис. A.1.

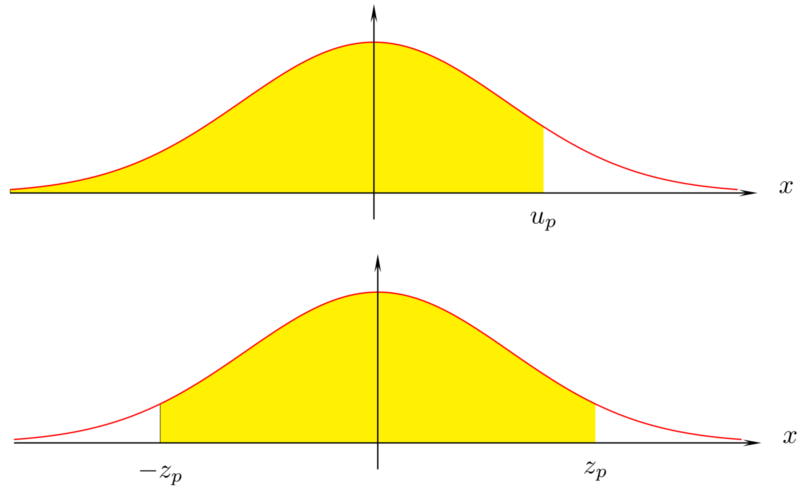

Рис. А.2. Графики плотности вероятности нормально распределенной случайной величины с математическим ожиданием $\mu = 0$ и среднеквадратичным отклонением $\sigma = 1$. С вероятностью p значение этой случайной величины располагается на интервале $(-\infty, u_p)$. С вероятностью p значение этой случайной величины располагается на интервале $(-z_p, z_p)$.

Квантили u_p и процентили z_p нормального распределения $N(0, 1)$:

p	0.90	0.95	0.975	0.99	0.995	0.999
u_p	1.282	1.645	1.960	2.326	2.576	3.090
z_p	1.645	1.960	2.326	2.576	3.090	3.291

Пусть $X_1, X_2, \ldots, X_d$ — независимые случайные переменные, заданные на вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$, распределенные по одному и тому же закону $N(0, 1)$. Тогда плотность вероятности совместного распределения величин $X_1, X_2, \ldots, X_d$ определяется равенством

$$p(x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} p_{X_i}(x_i) = \frac{1}{(2\pi)^{d/2}} \cdot e^{-\frac{1}{2}(x_1^2 + x_2^2 + \ldots + x_d^2)}.$$

Преобразование $Y = AX + \mu$ приводит к

$$p(y) = \frac{1}{(2\pi)^{d/2} \det A} \cdot e^{-\frac{1}{2}(y-\mu)^\top (A^{-1})^\top A^{-1}(y-\mu)}.$$

Матрица ковариации равна

$$\Sigma = \mathsf{E}\left((y-\mu)^\top (y-\mu)\right) = AA^\top,$$

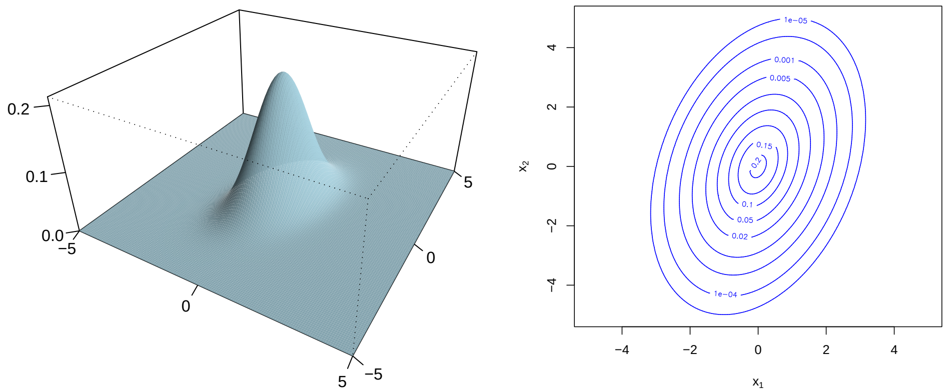

Рис. А.3. График плотности вероятности и линии уровня для двумерного нормального распределения.

поэтому можно записать

$$p(y) = \frac{1}{(2\pi)^{d/2}\sqrt{\det\Sigma}} \cdot e^{-\frac{1}{2}(y-\mu)^\top \Sigma^{-1}(y-\mu)}.$$

Это *многомерное нормальное распределение*.

На рис. А.3 представлен график плотности вероятности и линии уровня для двумерного нормального распределения, где

$$A = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}, \quad \Sigma = AA^\top = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{5}{4} \end{pmatrix}, \quad \mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Напомним, что если переменные независимы, то они некоррелированы. Обратное в общем случае не верно. Оказывается, для нормально распределенных случайных величин обратное утверждение верно: если две переменные имеют совместное нормальное распределение и они некоррелированы, то они независимы.

Рассмотрим *центральную предельную теорему* (в упрощенной форме Ляпунова), объясняющую важную роль, которую играет нормальное распределение. Пусть на вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$ задана последовательность независимых в совокупности случайных величин X_i $(i \geq 1)$ с одинаковым законом распределения. Обозначим $\mathsf{E}\, X_i = \mu$, $\mathsf{D}\, X_i = \sigma^2$. Тогда последовательность случайных величин

$$Y_N = \frac{\sum\limits_{i=1}^{N} X_i - N\mu}{\sqrt{N}\sigma}$$

сходится по распределению к нормальному распределению $N(0,1)$. *Сходимость по распределению* означает, что $P_{Y_N}(y)$ сходится поточечно к интегральной функции нормального распределения.

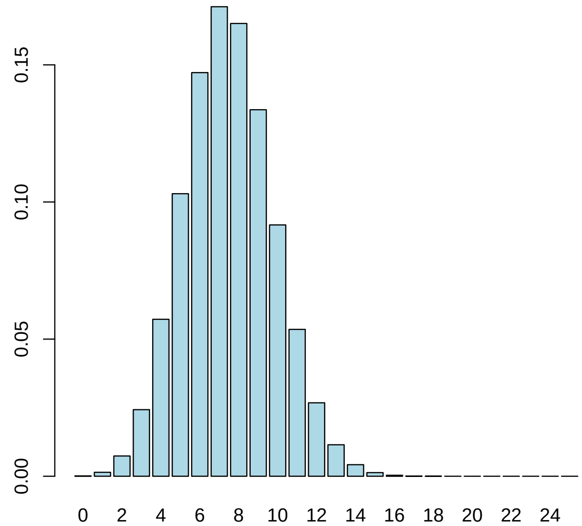

Рис. А.4. Биноминальное распределение для $n = 25$, $p = 0.3$.

Таким образом, если X_i распределены независимо по одному закону, то при достаточно больших значениях N случайная величина $\frac{1}{N}\sum_{i=1}^{N} X_i$ имеет распределение, близкое нормальному $N\left(\mu, \frac{\sigma}{\sqrt{N}}\right)$. Часто используемое эвристическое правило: $N \geq 30$.

А.5.2. Биномиальное распределение

Рассмотрим схему испытаний Бернулли. Пусть некоторый эксперимент проводился при одинаковых условиях независимым образом n раз. Предположим, что вероятность наступления события A в каждом таком эксперименте равна p. Пусть X — (дискретная) случайная величина, равная количеству наступления события A в n экспериментах. Нетрудно видеть, что

$$\mathsf{Pr}\,\{X = m\} = \binom{n}{m} p^m (1-p)^{n-m}.$$

Это — *биномиальное распределение*. Для случайной величины, распределенной по такому закону, $\mathsf{E}\,X = np$, $\mathsf{D}\,X = np(1-p)$.

Из центральной предельной теоремы следует при $npq \to \infty$

$$\mathsf{Pr}\,\{m_1 \leq X < m_2\} = \Phi\left(\frac{m_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{m_1 - np}{\sqrt{npq}}\right) + O\left(\frac{1}{\sqrt{npq}}\right)$$

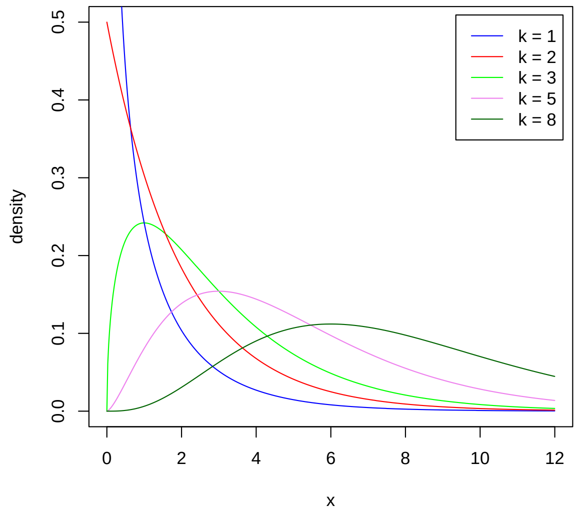

Рис. A.5. Графики плотностей χ^2-распределение для разных значений степени свободы k.

(*интегральная теорема Муавра–Лапласа*),

$$\mathsf{Pr}\,\{X = m\} = \frac{1}{\sqrt{2\pi npq}} \cdot e^{-\dfrac{(m-np)^2}{2npq}} + O\left(\frac{1}{\sqrt{npq}}\right)$$

(*локальная теорема Муавра–Лапласа*).

A.5.3. χ^2-распределение

Распределением χ^2 *с* k *степенями свободы* называется распределение случайной величины $\chi^2(k)$, равной сумме квадратов k независимых распределенных по закону $N(0,1)$ случайных величин U_i ($i = 1, 2, \ldots, k$):

$$\chi^2(k) = U_1^2 + U_2^2 + \ldots + U_k^2.$$

χ^2-распределение с k степенями свободы там, где это не вызывает недоразумений, также обозначается $\chi^2(k)$. Плотность вероятности случайной величины $\chi^2(k)$ равна

$$p_{\chi^2(k)}(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{1}{2^{k/2} \cdot \Gamma(k/2)} \cdot x^{(k-2)/2} e^{-x/2}, & x > 0. \end{cases}$$

Графики этой функции для разных значений k приведены на рис. A.5.

Для случайной величины $\chi^2(k)$ справедливо $\mathsf{E}\,\chi^2(k) = k$, $\mathsf{D}\,\chi^2(k) = 2k$. При больших значениях k ($k > 30$) распределение $\chi^2(k)$ с хорошей точностью аппроксимируется нормальным рапределением $N(k, \sqrt{2k})$.

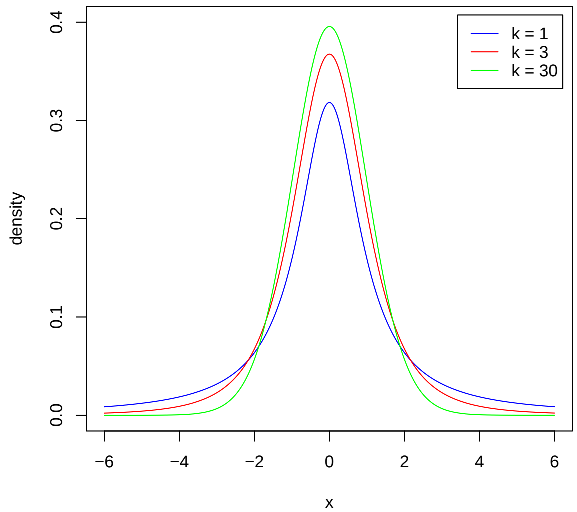

Рис. A.6. Графики плотностей t-распределения Стьюдента для разных значений степени свободы k.

A.5.4. t-распределение Стьюдента

Распределением Стьюдента, или *t-распределением*, *с k степенями свободы* называется распределение случайной величины $T(k)$, равной отношению независимых случайных величин U и V, где

$$U \sim N(0,1), \qquad V \sim \chi^2(k)/k.$$

Распределение Стьюдента с k степенями свободы обозначается $T(k)$. Ее плотность вероятности равна

$$p_{T(k)}(x) = \frac{\Gamma\big((k+1)/2\big)}{\Gamma(k/2)\sqrt{\pi k}} \left(1 + \frac{x^2}{k}\right)^{-(k+1)/2}.$$

Графики этой функции для разных значений k приведены на рис. A.6.

Для случайной величины $T(k)$ справедливо $\mathsf{E}\,T(k) = 0$, $\mathsf{D}\,T(k) = \dfrac{k}{k-2}$ $(k > 2)$. При больших значениях k $(k > 30)$ распределение $T(k)$ с хорошей точностью аппроксимируется нормальным рапределением $N(0,1)$.

A.5.5. F-распределение Фишера

Распределением Фишера, или *F-распределением*, *со степенями свободы k_1, k_2* называется распределение случайной величины $F(k_1, k_2)$, равной отношению независимых случайных величин U и V, где

$$U \sim \chi^2(k_1)/k_1, \qquad V \sim \chi^2(k_2)/k_2.$$

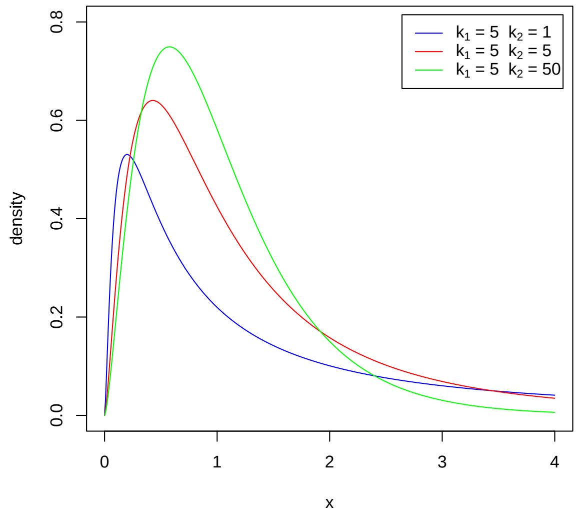

Рис. A.7. Графики плотностей F-распределения Фишера для разных значений k_1, k_2.

Распределение Фишера со степенями свободы k_1, k_2 также обозначается $F(k_1, k_2)$. Плотность вероятности случайной величины $F(k_1, k_2)$ равна

$$p_{F(k_1,k_2)}(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{\Gamma\big((k_1+k_2)/2\big)}{\Gamma(k_1/2)\Gamma(k_2/2)}(k_1/k_2)^{k_1/2}\dfrac{x^{k_1/2-1}}{(1+k_1/k_2)^{(k_1+k_2)/2}}, & x > 0. \end{cases}$$

Графики этой функции для разных значений k_1, k_2 приведены на рис. A.7.

Для случайной величины $F(k_1, k_2)$ справедливо $\mathsf{E}\, F(k_1, k_2) = \dfrac{k_2}{k_2 - 2}$ $(k > 2)$.

A.6. Основные понятия математической статистики

В математической статистике *генеральной совокупностью* называют множество всех наблюдаемых объектов или множество наблюдений за одним объектом, проводимых в неизменных условиях. Предполагается, что каждый элемент генеральной совокупности есть реализация в независимых экспериментах некоторой (неизвестной) случайной величины X (генеральной случайной величины), заданной на вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$. Величина X может быть многомерной.

Выборочной совокупностью, или просто *выборкой*, называют конечную последовательность $x_1, x_2, \ldots, x_N$ из N реализаций случайных величин $X_1, X_2, \ldots, X_N$, заданных на вероятностном пространстве $\langle \Omega, \mathbf{A}, \mathsf{Pr} \rangle$ и совпадающих с X. Случайные величины $X_1, X_2, \ldots, X_N$ также называются выборочной совокупностью, выборкой или наблюдениями. Величины $x_1, x_2, \ldots, x_N$ часто называют *наблюдениями*, а X — *статистически наблюдаемой* случайной величиной. Любая неслучайная функция от $X_1, X_2, \ldots, X_N$ называется *статистикой*.

Задача математической статистики заключается в изучении свойств функции распределения $P_X(x)$ случайной величины X по имеющейся выборочной совокупности, если, возможно, известно, что она принадлежит некоторому заданному классу. Изучение свойств функции $P_X(x)$ может включать определение характеристик, восстановление параметров, полное восстановлении P_X и т. п.

Статистическим рядом называется таблица вида

1	2	$\ldots$	N
x_1	x_2	$\ldots$	x_N

,

в которой в первой строке записывается номер элемента выборочной совокупности (номер измерения), а во второй — сам элемент (результат измерения).

Расположим элементы выборочной совокупности в порядке возрастания (в случае, если случайная величина X — одномерная): $x_{i_1} \leq x_{i_2} \leq \ldots \leq x_{i_N}$. Таблица

1	2	$\ldots$	N
x_{i_1}	x_{i_2}	$\ldots$	x_{i_N}

называется *вариационным рядом*.

Пусть K — некоторое натуральное число. Рассмотрим систему *интервалов группировки* $\{\Delta_k : \ k = 1, 2, \ldots, K\}$, где $\Delta_k = [x^*_{k-1}, x^*_k)$ $(k = 1, 2, \ldots, K)$ и $x^*_0 \leq x_{i_1} = \min x_i$, $x^*_K > x_{i_N} = \max x_i$. Пусть m_k — количество элементов выборки, принадлежащих интервалу Δ_k $(k = 1, 2, \ldots, K)$. Таблица

1	2	$\ldots$	K
m_1	m_2	$\ldots$	m_K

называется *информационной статистической таблицей*. Заметим, что величина $p_k = m_k/N$ равна частоте, с которой случайная величина X принимает значения из интервала Δ_k, и является оценкой вероятности наступления этого события:

$$p_k = \frac{m_k}{N} \approx \mathsf{Pr}\,\{X \in \Delta_k\} \qquad (k = 1, 2, \ldots, K).$$

Таблицу

1	2	$\ldots$	K
p_1	p_2	$\ldots$	p_K

также называют *информационной статистической таблицей*.

Эмпирической функцией распределения определяется соотношением

$$\widehat{P}(x) = \frac{1}{N} \sum_{x_i < x} 1,$$

где суммирование происходит по всем элементам выборочной совокупности, меньшим x. *Теорема Гливенко* утверждает, что для любого x и любого положительного ε

$$\lim_{N \to \infty} \mathsf{Pr}\left\{|\widehat{P}(x) - P(x)| < \varepsilon\right\} = 1,$$

т. е. с ростом объема выборки эмпирическая функция распределения равномерно приближается к истинной.

Гистограммой называют график кусочно-постоянной функции $\widehat{p}$, постоянной на интервалах группировки и принимающих в них значения

$$\frac{p_k}{x_k^* - x_{k-1}^*} = \frac{m_k}{N(x_k^* - x_{k-1}^*)} \qquad (k = 1, 2, \ldots, K)$$

и равной нулю во всех остальных точках. Площадь под гистограммой равна 1.

А.7. Оценка параметров распределения

Пусть известно, что случайная величина X имеет функцию распределения $P(x, \theta_1, \ldots, \theta_s)$, где $\theta_1, \ldots, \theta_s$ — некоторые параметры, которые нужно восстановить по имеющейся выборке $x_1, x_2, \ldots, x_N$. В качестве этих параметров могут выступать числовые характеристики случайной величины X: математическое ожидание, дисперсия и т. п. Напомним, что x_i является реализацией случайной величины X_i, совпадающей с X. Пусть $\widehat{\theta}_j = \psi_j(X_1, X_2, \ldots, X_N)$ $(j = 1, 2, \ldots, s)$ — некоторая неслучайная функция от случайных аргументов $X_1, X_2, \ldots, X_N$. Будем называть ее *оценкой* параметра θ_j. Чтобы оценка давала хорошее приближение, нужно, чтобы она удовлетворяла некоторым требованиям. Рассмотрим их.

Оценка $\widehat{\theta}$ параметра θ называется *несмещенной*, если $\mathsf{E}\,\widehat{\theta} = \theta$.

Оценка $\widehat{\theta}$ называется *состоятельной*, если при $N \to \infty$ последовательность $\widehat{\theta}$ сходится по вероятности к θ. *Сходимость по вероятности* означает, что для любого $\varepsilon > 0$

$$\lim_{N \to \infty} \mathsf{Pr}\left\{|\widehat{\theta} - \theta| < \varepsilon\right\} = 1.$$

Из неравенства Чебышева следует, что если $\mathsf{D}\,\widehat{\theta} \to 0$ при $N \to \infty$, то несмещенная оценка является состоятельной.

Несмещенная оценка $\widehat{\theta}$ параметра θ называется *эффективной*, если среди всех несмещенных оценок этого параметра она имеет минимальную дисперсию. Несмещенной оценки может не существовать.

Один из многих способов выразить количественно отклонение оценки $\widehat{\theta}$ от параметра θ является *средняя квадратическая ошибка*:

$$\mathrm{RSS}\,\widehat{\theta} = \mathsf{E}\,(\widehat{\theta} - \theta)^2.$$

Легко видеть, что

$$\mathrm{RSS}\,\widehat{\theta} = (\mathsf{Bias}\,\widehat{\theta})^2 + \mathsf{D}\,\widehat{\theta},$$

где

$$\mathsf{Bias}\,\theta = \mathsf{E}\,(\widehat{\theta} - \theta)$$

— смещение оценки.

Пусть известно, что для некоторых ε и α

$$\mathsf{Pr}\left\{\widehat{\theta} - \varepsilon < \theta < \widehat{\theta} + \varepsilon\right\} = 1 - \alpha,$$

тогда $1-\alpha$ называется *доверительной вероятностью*, или *надежностью*, оценки $\widehat{\theta}$, а ε — ее *точностью*. При этом интервал $(\widehat{\theta}-\varepsilon, \widehat{\theta}+\varepsilon)$ называется (*двусторонним*) *доверительным интервалом*. Доверительный интервал может быть *односторонним*: если для некоторых θ_0 и α

$$\mathsf{Pr}\,\{\theta > \theta_0\} = 1-\alpha, \quad \text{или} \quad \mathsf{Pr}\,\{\theta < \theta_0\} = 1-\alpha,$$

то доверительный интервал это $(\theta_0,\ +\infty)$ или $(-\infty,\ \theta_0)$ соответственно.

А.7.1. Выборочные числовые характеристики

Один из методов получения оценок для числовых характеристик случайных величин — замена их *выборочными числовыми характеристиками* (метод *аналогии*). Пусть $x_1, x_2, \ldots, x_N$ — выборка. Рассмотрим дискретную случайную величину, принимающую значения $x_1, x_2, \ldots, x_N$, каждое с вероятностью $1/N$. Числовые характеристики этой случайной величины называются *выборочными* (*эмпирическими*) *характеристиками* случайной величины X. Например,

$$\widehat{\mathsf{E}}\,X = \overline{X} = \frac{1}{N}\sum_{n=1}^{N} X_n$$

называется *выборочным математическим ожиданием*, или *выборочным средним*;

$$\widehat{\mathsf{D}}\,X = \frac{1}{N}\sum_{n=1}^{N}(X_n - \widehat{\mathsf{E}}\,X)^2$$

называется *выборочной дисперсией* и т. д.

Выборочное среднее $\overline{X}$ является несмещенной, состоятельной и эффективной оценкой для $\mathsf{E}\,X$. Докажем, например, несмещенность и состоятельность. Имеем

$$\mathsf{E}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N}\left(\sum_{n=1}^{N}\mathsf{E}\,X_n\right) = \frac{1}{N} N\,\mathsf{E}\,X = \mathsf{E}\,X,$$

что означает несмещенность, и

$$\mathsf{D}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2}\left(\sum_{n=1}^{N}\mathsf{D}\,X_n\right) = \frac{1}{N^2} N\,\mathsf{D}\,X = \frac{\mathsf{D}\,X}{N} \to 0 \quad \text{при} \quad N\to\infty,$$

откуда получаем состоятельность.

Выборочная дисперсия $\widehat{\mathsf{D}}\,X$ является состоятельной, но смещенной оценкой для $\mathsf{D}\,X$. Статистика

$$s^2 = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \overline{X})^2$$

является несмещенной и состоятельной. Если математическое ожидание $\mathsf{E}\,X$ известно, то можно вычислить статистику

$$s_0^2 = \frac{1}{N}\sum_{n=1}^{N}(X_n - \mathsf{E}\,X)^2,$$

которая также является несмещенной и состоятельной оценкой для $\mathsf{D}\,X$.

Пусть

$$(x_1,\, y_1),\ (x_2,\, y_2),\ \ldots,\ (x_N,\, y_N)$$

реализации случайных величин X, Y в N испытаниях. Оценками ковариации и корреляции величин X, Y являются соответственно *выборочные ковариация* и *корреляция*:

$$\widehat{\mathsf{Cov}}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y}),$$

$$\widehat{\mathsf{Corr}}(X,Y) = \frac{\sum\limits_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum\limits_{i=1}^{N}(x_i-\overline{x})^2}\sqrt{\sum\limits_{i=1}^{N}(y_i-\overline{y})^2}}.$$

Несмещенной оценкой ковариации является

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y}).$$

Пусть $X = (X_1, X_2, \ldots, X_d)$ — d-мерная случайная величина, для которой известны N реализаций $x_1, x_2, \ldots, x_N \in \mathbf{R}^d$. Матрица

$$\widehat{\mathsf{Cov}}(X) = \widehat{\mathsf{Cov}}(X_1, X_2, \ldots, X_d) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})^\top = (c_{ij}),$$

в которой $c_{ij} = \widehat{\mathsf{Cov}}(X_i, X_j)$, называется *выборочной ковариационной матрицей*. Предполагая, что $x_1, x_2, \ldots, x_N$ — векторы-столбцы, можно записать:

$$\widehat{\mathsf{Cov}}(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})^\top.$$

Несмещенной оценкой матрицы ковариации является

$$C = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\overline{x})(x_i-\overline{x})^\top.$$

Сформируем матрицу $\mathbf{X}$, записывая компоненты векторов $x_1, x_2, \ldots, x_N$ по строкам:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1d} \\ x_{21} & x_{22} & \ldots & x_{2d} \\ \ldots & \ldots & \ldots & \ldots \\ x_{N1} & x_{N2} & \ldots & x_{Nd} \end{pmatrix}.$$

Предположим, что данные центрированы, т. е. среднее значение элементов каждого столбца равно 0 (этого всегда можно добиться путем замены x_{ij} на $x_{ij}-\overline{x}_j$). Легко видеть, что тогда

$$\widehat{\mathsf{Cov}}(X) = \frac{1}{N}\mathbf{X}^\top\mathbf{X}.$$

Укажем доверительные интервалы для некоторых из приведенных оценок. Будем считать, что доверительная вероятность равна $1-\alpha$.

Пусть среднеквадратическое отклонение величины X известно. По центральной предельной теореме статистика

$$U = \frac{\overline{X} - \mathsf{E}\,X}{\sigma X/\sqrt{N}}$$

близка к нормально распределенной случайной величине. Отсюда несложно получить доверительный интервал для оценки $\overline{X}$ характеристики $\mathsf{E}\,X$:

$$\overline{X} - \frac{\sigma X}{\sqrt{N}}u_{1-\alpha/2} < \mathsf{E}\,X < \overline{X} + \frac{\sigma X}{\sqrt{N}}u_{1-\alpha/2},$$

где $u_{1-\alpha/2}$ — квантиль нормального распределения $N(0,1)$. Если среднеквадратическое отклонение не известно, то доверительный интервал для той же оценки имеет вид

$$\overline{X} - \frac{s}{\sqrt{N}}t_{1-\alpha/2}(N-1) < \mathsf{E}\,X < \overline{X} + \frac{s}{\sqrt{N}}t_{1-\alpha/2}(N-1),$$

где $t_{1-\alpha/2}$ — квантиль распределения Стьюдента с $N-1$ степеню свободы.

Если математическое ожидание известно, то доверительный интервал для оценки s_0^2 дисперсии $\mathsf{D}\,X$ имеет вид

$$\frac{Ns_0^2}{\chi^2_{1-\alpha/2}(N)} < \mathsf{D}\,X < \frac{Ns_0^2}{\chi^2_{\alpha/2}(N)},$$

где $\chi^2_p(N)$ — квантиль χ^2-распределения с N степенями свободы. Если математическое ожидание не известно, то доверительный интервал для оценки s^2 дисперсии $\mathsf{D}\,X$ имеет вид

$$\frac{(N-1)s^2}{\chi^2_{1-\alpha/2}(N-1)} < \mathsf{D}\,X < \frac{(N-1)s^2}{\chi^2_{\alpha/2}(N-1)}.$$

А.7.2. Метод максимального правдоподобия

Рассмотрим еще один универсальный метод построения оценок параметров распределения.

Пусть X — непрерывная случайная величина, а $p(x;\theta)$ — ее плотность вероятности. Параметр θ необходимо определить. Пусть $X_1, X_2, \ldots, X_N$ — независимые случайные величины, совпадающие с X, тогда плотность распределения вероятности случайного вектора $(X_1, X_2, \ldots, X_N)$ равна

$$p(x_1, x_2, \ldots, x_N;\,\theta) = \prod_{i=1}^{N} p(x_i;\theta)$$

Пусть $x_1, x_2, \ldots, x_N$ — выборочная совокупность случайной величины X. Плотность вероятности $p(x_1, x_2, \ldots, x_N;\,\theta)$, рассматриваемая при фиксированных значениях $x_1, x_2, \ldots, x_N$, называется *функцией правдоподобия* и обозначается $L(\theta)$:

$$L(\theta) = \prod_{i=1}^{N} p(x_i;\theta). \tag{92}$$

Пусть теперь X — дискретная случайная величина. Функция правдоподобия вводится аналогично:

$$L(\theta) = \prod_{i=1}^{N} \Pr\left\{X = x_i\right\}.$$

Принцип максимального правдоподобия заключается в том, что в качестве оценки $\widehat{\theta}$ параметра θ выбирается значение, доставляющее максимум функции правдоподобия:

$$\widehat{\theta} = \operatorname{argmax} L(\theta).$$

Иногда вместо функции $L(\theta)$ удобно рассматривать ее логарифм. Функция

$$\ell(\theta) = \ln L(\theta)$$

называется *логарифмической функцией правдоподобия*.

Вообще говоря, оценка, полученная методом максимального правдоподобия, может быть смещенной.

В качестве примера использования метода рассмотрим задачу определения параметров распределения равномерной случайной дискретной величины. Пусть X равновероятно принимает случайные целочисленные значения на отрезке $[1, \theta]$. Неизвестный параметр $\theta \geq 1$ требуется восстановить по выборке $x_1, x_2, \ldots, x_N$, состоящей из N независимых реализаций случайной величины X. Легко видеть, что

$$\Pr\left\{X = x\right\} = \frac{1}{\theta} \cdot \left[x \in \{1, 2, \ldots, \theta\}\right],$$

откуда

$$L(\theta) = \prod_{i=1}^{N} \frac{1}{\theta}\left[x \in \{1, 2, \ldots, \theta\}\right].$$

Функция $L(\theta)$ принимает свое максимальное значение в точке $\widehat{\theta} = \max_i x_i$. Это смещенная оценка (она слишком занижена). Несмещенной эффективной оценкой является[3]

$$\widehat{\theta} = \frac{N+1}{N} \cdot \max_i x_i - 1.$$

А.8. Проверка статистических гипотез

Пусть X — статистически наблюдаемая случайная величина. *Статистической гипотезой* называется любое предположение о параметрах или виде распределения этой величины. Гипотеза называется *простой*, если она однозначно определяет распределение, в противном случае гипотеза называется *сложной*. Например, гипотеза о том, что X имеет нормальное распределение $N(0, 1)$ является *простой*, а гипотеза, заключающаяся в том, что X имеет распределение $N(\mu, 1)$, где $0 \leq \mu < 10$, является сложной.

[3] Рассмотренная задача близка к так называемой задаче о немецких танках (German tank problem): по серийным номерам захваченных немецких танков необходимо оценить их общее количество. Предполагается, что танки нумеруются без пропусков начиная с 1. Единственное отличие от рассмотренной задачи заключается в том, что теперь элементы выборки $x_1, x_2, \ldots, x_N$ являются реализациями *разных* случайных величин (танки изымаются из генеральной совокупности без возвращения).

Рассмотрим задачу проверки статистической гипотезы. Проверяемая гипотеза называется *нуль-гипотезой* и обозначается H_0. Наряду с гипотезой H_0 рассматривается любая *альтернативная*, или *конкурирующая*, *гипотеза* H_1. Если, например, H_0 состоит в том, что некий параметр θ распределения равен значению θ_0, то в качестве альтернативных могут выступать, например, $\theta \neq \theta_0$, $\theta > \theta_0$ (если из смысла задачи известно, что $\theta < \theta_0$ невозможно), $\theta < \theta_0$ (если известно, что $\theta \geq \theta_0$ невозможно), $\theta = \theta_1$ (если известно, что $\theta = \theta_0$ или $\theta = \theta_1$) и др. Вид альтернативной гипотезы зависит от конкретной задачи.

Правило, по которому принимается решение, принять или отклонить H_0, называется *критерием*, или *тестом*. Критерий включает в себя вычисление некоторой определенной функции от наблюдений $X_1, X_2, \ldots, X_N$, которая называется *статистикой критерия*. Статистика критерия характеризует отклонение эмпирических данных от теоретических (соответствующих гипотезе H_0). Если в результате проверки критерия гипотеза H_0 отвергается, хотя она верна (и принимается гипотеза H_1), то говорят, что произошла *ошибка* 1*-го рода*. Если гипотеза H_0 принимается, хотя она не верна, то говорят, что произошла *ошибка* 2*-го рода*.

Перед анализом выборки фиксируется α — *вероятностный уровень значимости* критерия, равный вероятности того, что гипотеза отвергается, когда на самом деле она верна. Таким образом, α — это вероятность ошибки 1-го рода. Часто используют уровни значимости α, равные 0.1, 0.05, 0.01 и т. п.

Пусть t^* — значение статистики критерия, вычисленные по наблюдениям $x_1, x_2, \ldots, x_N$. Гипотеза H_0 принимается или отвергается в зависимости от значения t^*. *Критической областью* $T_{\text{критич}} = T_{\text{критич}}(H_0, \alpha)$ называется множество тех значений t^*, при которых H_0 отвергается. Так как α есть вероятность ошибки 1-го рода, то необходимо

$$\mathsf{Pr}(t \in T_{\text{критич}} \,|\, H_0) = \alpha,$$

где $\mathsf{Pr}(\cdot \,|\, H_0)$ — вероятность при условии истинности H_0. Итак, если $t^* \in T_{\text{критич}}$, то гипотеза H_0 отвергается и принимается альтернативная гипотеза H_1. Если $t^* \notin T_{\text{критич}}$, то можно считать, что эмпирические данные согласуются с гипотезой H_0, и гипотеза H_0 принимается.

Часто в качестве $T_{\text{критич}}(H_0, \alpha)$ выбираются области вида

$$\{t : t \geq t_\alpha\}, \qquad \{t : t \leq t_\alpha\}, \qquad \{t : |t| \geq t_\alpha\}, \tag{93}$$

где величина t_α определяется по α. В первых двух случаях критерий называется *односторонним*, в последнем — *двусторонним*. См. рис. А.9.

Итак, проверка гипотезы по некоторому критерию состоит из следующих шагов:

1. выбирается вероятностный уровень значимости α;
2. по данным $x_1, x_2, \ldots, x_N$ вычисляется значение t^* статистики теста;
3. если $t^* \in T_{\text{критич}}$, то гипотезу H_0 отвергаем и принимаем гипотезу H_1; иначе принимаем гипотезу H_0.

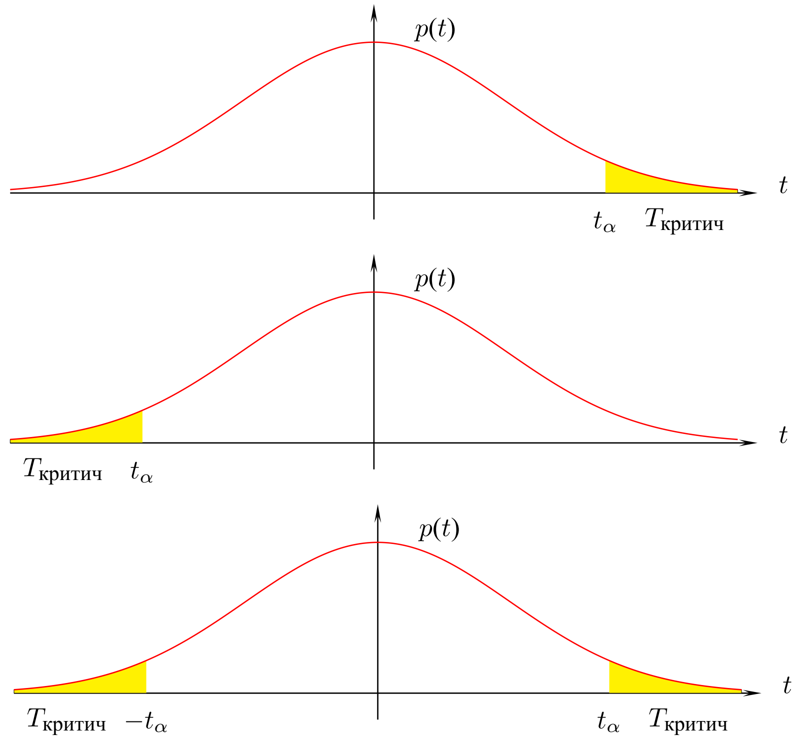

Рис. А.8. Часто используемые критические области $T_{\text{критич}}$. Площадь желтой области на каждом графике равна вероятностному уровню значимости α.

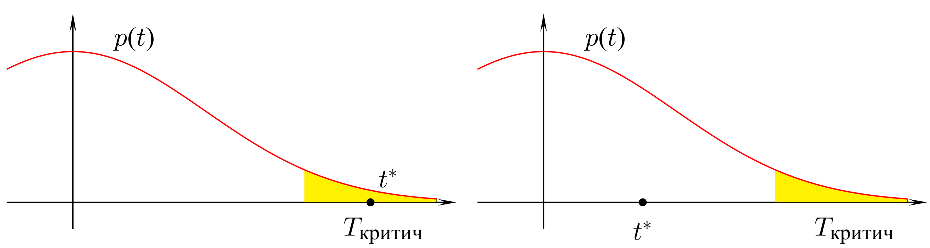

Рис. А.9. Критическая область $T_{\text{критич}}$. Если $t^* \in T_{\text{критич}}$, то гипотезу H_0 отвергаем, иначе гипотезу H_0 принимаем.

Степень согласия данных с гипотезой, или p-value, — это минимальное значение уровня значимости α, при котором данное значение t^* статистики $t(X)$ принадлежит критической области $T_{\text{критич}}(H_0, \alpha)$:

$$\text{p-value}(t^*, H_0) = \inf\left\{\alpha: \ t^* \in T_{\text{критич}}(H_0, \alpha)\right\}$$

Заметим, что p-value вычисляется по значению t^* и зависит также от критерия и гипотезы H_0, но не зависит от α. Из определение p-value вытекает следующее

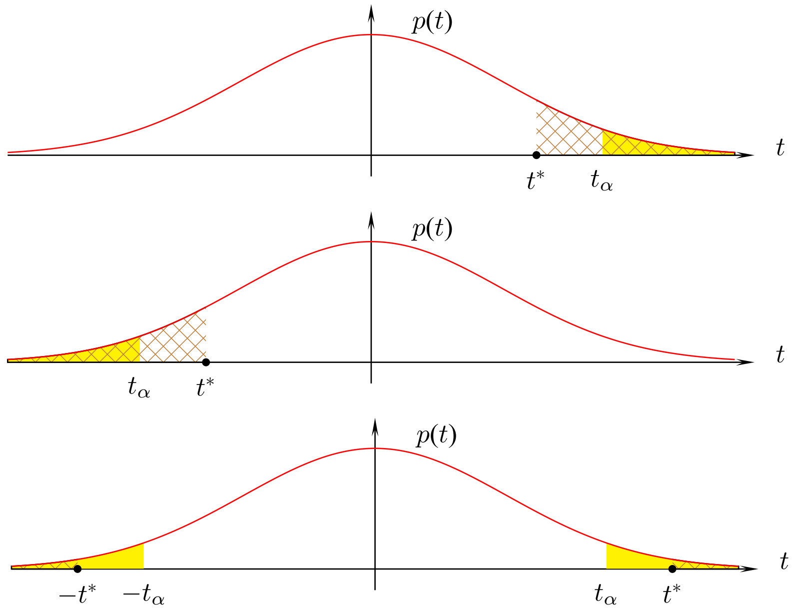

Рис. А.10. p-value для разных способов задания $T_{\text{критич}}$. На каждом рисунке p-value равно площади заштрихованной фигуры, а вероятностный уровень α равен шлощади желтой фигуры.

основное свойство p-value: если p-value меньше α, то гипотеза H_0 отвергается, иначе гипотезу H_0 отвергнуть нельзя.

Проиллюстрируем понятие p-value на каждом $T_{\text{критич}}(H_0, \alpha)$ из (93):

$$\begin{aligned}
&\text{если } T_{\text{критич}} = \{t : t \geq t_\alpha\}, && \text{то p-value} = \Pr\{t(X) \geq t^*\},\\
&\text{если } T_{\text{критич}} = \{t : t \leq t_\alpha\}, && \text{то p-value} = \Pr\{t(X) \leq t^*\},\\
&\text{если } T_{\text{критич}} = \{t : |t| \geq t_\alpha\}, && \text{то p-value} = \Pr\{|t(X)| \geq |t^*|\}.
\end{aligned}$$

См. рис. А.10.

А.8.1. Критерий согласия χ^2

Рассмотрим *критерий согласия* χ^2, принадлежащий Пирсону, для проверки простой гипотезы H_0. Пусть H_0 утверждает, что статистически наблюдаемая случайная величина X имеет закон распределения $P(x)$. Числовая ось разбивается на r областей:

$$d_1 = (c_0, c_1), \quad d_2 = [c_1, c_2), \quad \ldots, \quad d_r = [c_{r-1}, c_r),$$

где $c_0 = -\infty$, $c_r = +\infty$, и вычисляются величины m_j равные количеству элементов выборки $x_1, x_2, \ldots, x_N$, попавших в i-ю область ($i = 1, 2, \ldots, r$). Области нужно выбирать так, чтобы каждая из них содержала по крайней мере 1 точку (а лучше больше). Статистика критерия определяется формулой

$$t^* = \sum_{i=1}^{r} \frac{(m_i - Np_i)^2}{Np_i}, \tag{94}$$

где p_i — «теоретическая» вероятность попадания элемента выборки в i-ю область, т. е.

$$p_i = \mathsf{Pr}\left\{X \in d_i\right\} = P(c_r) - P(c_{r-1}) \qquad (i = 1, 2, \ldots, r).$$

По теореме Пирсона–Фишера, если гипотезы H_0 верна, то статистика критерия не зависит от распределения $P(x)$ и при $N \to \infty$ стремится по распределению к $\chi^2(r-1)$ и поэтому при больших N близка к ней (рекомендованное значение на практике $N \geq 30$). В качестве критической области используется

$$T_{\text{критич}} = \left\{t: \ t \geq \chi^2_{1-\alpha}(r-1)\right\},$$

где $\chi^2_{1-\alpha}(r-1)$ — квантиль случайной величины $\chi^2(r-1)$ порядка $1-\alpha$.

Критерий использует тот факт, что случайная величина $\frac{(m_i - Np_i)}{\sqrt{Np_i}}$ близка к нормальной. Для этого необходимо, чтобы величина Np_i была бы достаточно большой, например, $Np_i \geq 5$. Если это неравенство не выполнено, то соседние области следует объединить.

Фишер предложил модифицировать критерий, чтобы его можно было использовать для проверки сложных гипотез. Пусть H_0 заключается в том, что функция распределения случайной величины X равна $P(x, \theta_1, \theta_2, \ldots, \theta_s)$, где $\theta_1, \theta_2, \ldots, \theta_s$ — неизвестные параметры. Теперь мы не можем вычислить значение статистики t^* по формуле (94), так как не знаем θ_j $(j = 1, 2, \ldots, s)$. Оказывается, в качестве θ_j можно выбрать значения, доставляющие минимум функции t^*. В частности, можно доказать, что эти параметры могут быть найдены из системы уравнений

$$\sum_{i=1}^{r} \frac{m_i}{p_i(\theta_1, \ldots, \theta_s)} \cdot \frac{\partial p_i(\theta_1, \ldots, \theta_s)}{\partial \theta_j} \qquad (j = 1, 2, \ldots, s) = 0,$$

где m_i имеет такой же смысл, как и в случае простой гипотезы, и

$$p_i(\theta_1, \theta_2, \ldots, \theta_s) = P(c_r, \theta_1, \theta_2, \ldots, \theta_s) - P(c_{r-1}, \theta_1, \theta_2, \ldots, \theta_s).$$

После того, как θ_j найдены, статистика t^* вычисляется по формуле (94). По теореме Пирсона–Фишера, если гипотезы H_0 верна, то статистика критерия при $N \to \infty$ стремится по распределению к $\chi^2(r-1-s)$. Обращаем внимание, что количество степеней свободы уменьшилось на число параметров. В качестве критической области используется

$$T_{\text{критич}} = \left\{t: \ t \geq \chi^2_{1-\alpha}(r-1-s)\right\}.$$

A.8.2. Критерии согласия Колмогорова и Колмогорова–Смирнова

Критерий Колмогорова применяется для проверки простых гипотез. Пусть нулевая гипотеза заключается в том, что случайная переменная X имеет непрерывную функцию распределения $P(x)$. По выборке $x_1 \leq x_2 \leq \ldots \leq x_N$ найдем эмпирическую функцию распределения:

$$P^*(x) = \begin{cases} 0, & x < x_1, \\ n/N, & x_n \le x < x_{n+1} \quad (n = 1, 2, \dots, N-1), \\ 1, & x \ge x_N. \end{cases}$$

Статистика в критерии Колмогорова вычисляется по формуле

$$t^* = \sqrt{N} \sup_x |P^*(x) - P(x)|.$$

Эта статистика не зависит от распределения $P(x)$ и имеет предельное (при $N \to \infty$) распределение

$$\sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2} \tag{95}$$

В качестве критической области рассматривается

$$T_{\text{критич}} = \{t : \ t \ge \lambda_{1-\alpha}\},$$

где $\lambda_{1-\alpha}$ — квантиль распределения (95). На практике критерий Колмогорова используют при $N > 50$.

В тестах Колмогорова–Смирнова используются похожие статистики:

$$t_+^* = \sqrt{N} \sup_x \big(P^*(x) - P(x)\big), \qquad t_-^* = -\sqrt{N} \inf_x \big(P^*(x) - P(x)\big).$$

А.8.3. Двухвыборочный критерий Колмогорова–Смирнова

Пусть имеется две выборки, соответсвующие двум случайным величинам X и Y:

$$X : \ x_1, x_2, \dots x_{N_1}, \qquad Y : \ y_1, y_2, \dots y_{N_2}.$$

Требуется определить, совпадают ли X и Y. Таким образом, нуль-гипотеза H_0 заключается в том, что $P_X = P_Y$. Используемые статистики определяются разностью эмпирических функций распределения $P_X^*(x)$ и $P_Y^*(x)$:

$$t^* = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \max_x \big|P_X^*(x) - P_Y^*(x)\big|, \qquad t_+^* = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \max_x \big(P_X^*(x) - P_Y^*(x)\big).$$

Предельное распределение статистик известно. Критерием рекомендуется пользоваться, если $\min\{n_1, n_2\} \ge 100$. В этом случае справедливо приближенное равенство

$$\Pr(t_+^* < z) \approx 1 - e^{-2z^2}.$$

А.8.4. Проверка гипотез о параметрах распределений

Критерии для проверки гипотез о равенстве параметров распределения заданному значению могут быть построены на основе доверительных интервалов. При этом одностороннему доверительному интервалу соответсвует односторонний критерий, а двустороннему доверительному интервалу — двусторонний критерий. Гипотеза $H_0 : \theta = \theta_0$ принимается, если значение θ_0 накрывается доверительным интервалом, в противном случае гипотеза отклоняется.

Для проверки гипотез вида $H_0 : \theta_1 = \theta_2$, где θ_1, θ_2 — значения параметра θ двух выборок из двух генеральных совокупностей, рассматривают доверительные интервалы для разности $\theta_1 - \theta_2$ или частного θ_1/θ_2. В таблице A.1 приведена информация о некоторых тестах для проверки гипотез о математическом ожидании μ и дисперсии σ^2 нормально распределенных генеральных совокупностей. Эти тесты можно применять и в случае, когда закон распределения отличается от нормального, тогда критические области несколько изменятся.

Таблица А.1. Критерии для проверки гипотез о средних и дисперсии нормально распределенной генеральной совокупности

H_0	*Предположения*	*Статистика критерия*	*Распределение статистики*	*Двусторонний критерий*	*Правосторонний критерий*	
				Критическая область	*Критическая область*	H_1
$\mu = \mu_0$	σ^2 известна	$\dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$N(0,1)$	$\dfrac{\lvert\overline{x} - \mu_0\rvert}{\sigma/\sqrt{N}} \geq u_{1-\alpha/2}$	$\dfrac{\lvert\overline{x} - \mu_0\rvert}{\sigma/\sqrt{N}} \geq u_{1-\alpha}$	$\mu > \mu_0$
	σ^2 не известна	$\dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$T(N-1)$	$\dfrac{\lvert\overline{x} - \mu_0\rvert}{s/\sqrt{N}} \geq t_{1-\alpha/2}(N-1)$	$\dfrac{\lvert\overline{x} - \mu_0\rvert}{S/\sqrt{N}} \geq t_{1-\alpha}(N-1)$	$\mu > \mu_0$
$\mu_1 = \mu_2$	σ_1^2, σ_2^2 известны	$\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_1^2}{N_1}}}$	$N(0,1)$	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_1^2}{N_1}}} \geq u_{1-\alpha/2}$	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_1^2}{N_1}}} \geq u_{1-\alpha}$	$\mu_1 > \mu_2$
	σ_1^2, σ_2^2 не известны; гип. $\sigma_1^2 = \sigma_2^2$ принимается	$\dfrac{\overline{X}_1 - \overline{X}_2}{S\sqrt{1/N_1 + 1/N_2}}$, где $S = \dfrac{(N_1-1)S_1^2 + (N_2-1)S_2^2}{N_1 + N_2 - 2}$	$T(N_1 + N_2 - 2)$	$\dfrac{\overline{x}_1 - \overline{x}_2}{s\sqrt{1/N_1 + 1/N_2}} \geq t_{1-\alpha/2}(N_1 + N_2 - 2)$	$\dfrac{\overline{x}_1 - \overline{x}_2}{s\sqrt{1/N_1 + 1/N_2}} \geq t_{1-\alpha}(N_1 + N_2 - 2)$	$\mu_1 > \mu_2$
	σ_1^2, σ_2^2 не известны; гип. $\sigma_1^2 = \sigma_2^2$ отклоняется	$\dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_1^2/N_1 + S_2^2/N_2}}$	$T(k)$, где $k = \dfrac{s_1^2/N_1 + s_2^2/N_2}{\dfrac{s_1^2/N_1}{N_1 - 1} + \dfrac{s_2^2/N_2}{N_2 - 1}}$	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}} \geq t_{1-\alpha/2}(k)$	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{s_1^2/N_1 + s_2^2/N_2}} \geq t_{1-\alpha}(k)$	$\mu_1 > \mu_2$
$\sigma^2 = \sigma_0^2$	μ известно	$\dfrac{NS_0^2}{\sigma_0^2}$	$\chi^2(N)$	$\chi^2_{\alpha/2}(N) < \dfrac{NS_0^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(N)$	$\dfrac{NS_0^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(N)$	$\sigma^2 > \sigma_0^2$
	μ не известно	$\dfrac{(N-1)s^2}{\sigma_0^2}$	$\chi^2(N-1)$	$\chi^2_{\alpha/2}(N-1) < \dfrac{(N-1)S_0^2}{\sigma_0^2} < \chi^2_{1-\alpha/2}(N-1)$	$\dfrac{(N-1)s^2}{\sigma_0^2} < F^2_{1-\alpha}(N_1, N_2)$	$\sigma^2 > \sigma_0^2$
$\sigma_1^2 = \sigma_2^2$	μ_1, μ_2 известны	S_{01}^2/S_{02}^2 $s_{01}^2 > s_{02}^2$	$F(N_1, N_2)$	$\dfrac{s_{01}^2}{s_{02}^2} < F_{1-\alpha/2}(N_1, N_2)$	$\dfrac{s_{01}^2}{s_{02}^2} < F_{1-\alpha/2}(N_1, N_2)$	$\sigma_1^2 > \sigma_2^2$
	μ_1, μ_2 не известны	S_{01}^2/S_{02}^2 $s_{01}^2 > s_{02}^2$	$F(N_1 - 1, N_2 - 1)$	$\dfrac{s_{01}^2}{s_{02}^2} < F_{1-\alpha/2}(N_1 - 1, N_2 - 1)$	$\dfrac{s_{01}^2}{s_{02}^2} < F_{1-\alpha/2}(N_1 - 1, N_2 - 1)$	$\sigma_1^2 > \sigma_2^2$

Литература

[1] *Hastie T., Tibshirani R., Friedman J.* The elements of statistical learning. Springer, 2001.

[2] *Ripley B.D.* Pattern recognition and neural networks. Cambridge University Press, 1996.

[3] *Воронцов К.В.* Математические методы обучения по прецедентам. Курс лекций. Москва, ВЦ РАН, 2005.
http://www.ccas.ru/voron/teaching.html

[4] *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики, 1999.

[5] *Николенко С.* Машинное обучение. Курс лекций. СПб.: ПОМИ РАН, 2006.
http://logic.pdmi.ras.ru/~sergey/

Предметный указатель