Вероятность, основные определения и примеры

§ 1. Предмет теории вероятностей

Теорией вероятностей называется раздел математики, изучающий *математические модели* экспериментов, исход которых не вполне однозначно определяется условиями опытов. Такие эксперименты называются *случайными опытами*.

Для более точного определения надо сказать, что теория вероятностей имеет дело лишь со случайными опытами, обладающими свойством *статистической устойчивости* или *устойчивости частот*. Это свойство описывается следующим образом. Рассмотрим случайный эксперимент и предположим, что идентичные условия проведения этого опыта и его повторение можно реализовать любое желаемое число раз. Такое повторение возможно представить себе не только реально физически осуществимым, а хотя бы мыслимым. Обозначим через A один из возможных исходов этого эксперимента. Повторим данный опыт n раз и обозначим через $n(\mathsf{A})$, $0 \le n(\mathsf{A}) \le n$, число наступлений исхода A при этих n повторениях. Отношение

$$\frac{n(\mathsf{A})}{n}, \qquad 0 \le \frac{n(\mathsf{A})}{n} \le 1,$$

называется *частотой* появления события (исхода) A, а свойство устойчивости частот заключается в том, что при больших значениях n частота появления события A при изменении n мало отличается от некоторого постоянного значения, которое называется *вероятностью события* A и обозначается

$$\Pr\{\mathsf{A}\} \approx \frac{n(\mathsf{A})}{n}, \qquad 0 \le \Pr\{\mathsf{A}\} \le 1.$$

Данное определение называется *частотным* или *интуитивным* определением вероятности.

Например, многократное бросание симметричной монеты с двумя возможными исходами герб - решетка при каждом бросании, дает частоты выпадений герба, близкие к $1/2$. В данном случае для события A, означающего появление герба в одном испытании, нет сомнения в том, что его вероятность $\Pr\{\mathsf{A}\} = 1/2$.

§ 2. Модель случайного опыта с конечным числом исходов

Рассмотрим опыт, N исходов которого обозначим символами $\omega_1, \omega_2, \ldots, \omega_N$. Эти символы не обязательно являются числами и их физическая природа для нас не имеет значения.

Определение 1. Всевозможные N исходов опыта называются *элементарными событиями*, а их совокупность

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\} = \{\omega\}$$

называется *пространством* элементарных событий.

Определение 2. Каждому элементарному событию ω ставится в соответствие некоторое число $\Pr\{\omega\}$, называемое *вероятностью* ω. При этом числа $\Pr\{\omega_i\}$, $i = 1, 2, \ldots, N$, удовлетворяют условиям

$$0 \le \Pr\{\omega_i\} \le 1, \qquad \sum_{i=1}^{N} \Pr\{\omega_i\} = \sum_{\omega} \Pr\{\omega\} = 1.$$

Определение 3. Любое подмножество A пространства элементарных событий Ω, т.е. $\mathsf{A} \subseteq \Omega$, называется *событием*. Событие $\mathsf{A} = \Omega$ называется *достоверным*. Кроме того, введем и обозначим символом пустого множества $\emptyset$ событие, которое не содержит элементарных событий и называется *невозможным* событием.

Определение 4. *Вероятностью* $\Pr\{\mathsf{A}\}$ *события* A называется сумма вероятностей элементарных событий, входящих в A, иначе

$$\Pr\{\mathsf{A}\} \stackrel{\text{def}}{=} \sum_{\omega \in \mathsf{A}} \Pr\{\omega\}.$$

Для невозможного события $\emptyset$ по определению полагаем $\Pr\{\emptyset\} \stackrel{\text{def}}{=} 0$.

Здесь и дальнейшем символ $\stackrel{\text{def}}{=}$ обозначает *равенство по определению*. Очевидно, что для любого события A вероятность $0 \le \Pr\{\mathsf{A}\} \le 1$, а для достоверного события Ω вероятность $\Pr\{\Omega\} = 1$.

Пример. Пусть опыт Ω состоит в проведении трех испытаний, в каждом их которых регистрируется один из двух возможных исходов: успех, обозначаемый символом 1, или неудача, обозначаемая символом 0. Математическую модель этого опыта можно описать следующим образом.

- Общее число элементарных событий $N = 2^3 = 8$. Каждое элементарное событие ω есть двоичная (из 1 и 0) последовательность $\omega = (x_1, x_2, x_3)$, где $x_i = 1,\ 0$ обозначает исход (успех или неудачу) в i-ом, $i = 1, 2, 3$, испытании. Пространство элементарных событий имеет вид

  $$\Omega = \{(0,0,0),\ (0,0,1),\ (0,1,0),\ (0,1,1),\ (1,0,0),\ (1,0,1),\ (1,1,0),\ (1,1,1)\},$$

  т.е. $\omega_1 = (0,0,0),\ \omega_2 = (0,0,1),\ \ldots,\ \omega_8 = (1,1,1)$.

- В качестве примера события рассмотрим событие A, состоящее в том, что *в трех испытаниях произойдет по крайней мере два успеха*. Это событие представляет собой совокупность из 4-х элементарных событий:

  $$\mathsf{A} = \{(0,1,1),\ (1,0,1),\ (1,1,0),\ (1,1,1)\}.$$

- Если всем элементарным событиям поставлены в соответствие одинаковые вероятности

  $$\Pr\{(0,0,0)\} = \Pr\{(0,0,1)\} = \Pr\{(0,1,0)\} = \Pr\{(0,1,1)\} =$$
  $$= \Pr\{(1,0,0)\} = \Pr\{(1,0,1)\} = \Pr\{(1,1,0)\} = \Pr\{(1,1,1)\} = \frac{1}{8},$$

  то согласно определению 4 для вероятности введенного события A имеем

  $$\Pr\{\mathsf{A}\} = \Pr\{(0,1,1)\} + \Pr\{(1,0,1)\} + \Pr\{(1,1,0)\} + \Pr\{(1,1,1)\} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}.$$

Вероятностная модель этого примера очевидным образом обобщается на случай n испытаний, в каждом их которых регистрируется один из двух возможных исходов: успех, обозначаемый символом 1, или неудача, обозначаемая символом 0. Равновероятные элементарные события ω имеют вид

$$\omega = (x_1, x_2, \ldots, x_n), \quad x_i = 1,\ 0, \quad i = 1, 2, \ldots, n,$$

где вероятность каждого ω равна

$$\Pr\{\omega\} = \Pr\{(x_1, x_2, \ldots, x_n)\} = \frac{1}{N} = \frac{1}{2^n}.$$

Такая модель называется моделью n испытаний Бернулли с $N = 2^n$ равновероятными исходами.

Рассмотрим дихотомическое (с возможными двоичными исходами 1 или 0) тестировании n *пар однородных испытуемых*, когда один из испытуемых в каждой паре становится контрольным (измерение "до"), а второй подвергается какой-либо *обработке*, например, *процедуре лечения, методике обучения* или *рекламе* (измерение "после"). Если сравнение этих измерений показывает их значимое отличие, то результатом дихотомического тестирования данной пары считается 1, в противном случае–0. Предполагается, что модель n испытаний Бернулли с $N = 2^n$ равновероятными исходами даёт адекватную вероятностную интерпретацию результатов тестирования n пар однородных испытуемых, если рассматриваемая обработка *не эффективна*. Эта модель позволяет разработать научно - обоснованные методы принятия решения об эффективности обработки на основании результатов дихотомических наблюдений. Данная задача являются предметом исследования для тесно связанного с теорией вероятностей раздела математики, называемого *математической статистикой*.

§ 3. Правила перевода

Первым этапом разработки теоретико - вероятностных моделей для применения в практических задачах является *перевод* с неформального языка, на котором обычно записывается условие задачи, на формальный язык математической модели случайного опыта. При этом следует учитывать два простых правила.

Правило 1. Пространство элементарных событий Ω есть совокупность всех мыслимых исходов опыта, при этом считается, что исходы регистрируются возможно более подробно.

Правило 2. Пусть $|\mathsf{A}|$ обозначает число исходов (элементарных событий) при которых происходит событие A. Тогда, если $N = |\Omega|$ есть число всевозможных исходов (элементарных событий), то для события A вероятность

$$\Pr\{\mathsf{A}\} = \frac{|\mathsf{A}|}{|\Omega|} = \frac{|\mathsf{A}|}{N}.$$

Если правило 2 применимо, то говорят что имеется задача на *классическую вероятность*, где вероятность события равна *отношению числа благоприятных исходов к числу всех исходов*.

Предыдущий пример модели трех испытаний Бернулли с равновероятными исходами иллюстрирует оба правила, т.е. при подсчете вероятности события A мы учли, что

$$N = |\Omega| = 2^3 = 8, \quad |\mathsf{A}| = 4, \quad \Pr\{\mathsf{A}\} = \frac{|\mathsf{A}|}{N} = \frac{4}{8} = \frac{1}{2}.$$

В разделе "Комбинаторика и вероятность" излагаются некоторые методы подсчета числа элементарных событий в типичных ситуациях, встречающихся при решении задач, для которых применимо правило 2.

§ 4. Комбинаторика и вероятность

§ 4.1. Определения и примеры

Пусть N и n – произвольные натуральные числа а $[N] = \{1, 2, \ldots, N\}$ обозначает множество всех целых чисел от 1 до N.

1) Мы выбираем *с возвращением* n элементов из множества $[N]$. Результатом выбора является N - ичная последовательность длины n, элементы которой принадлежат алфавиту $[N] = \{1, 2, \ldots, N\}$. *Число всех* N *- ичных последовательностей длины* n, *называемых* n *- последовательностями из алфавита* $[N] = \{1, 2, \ldots, N\}$, равно N^n.

   Примеры.

   а) Число всех n - последовательностей из алфавита $\{0, 1, \ldots, q-1\}$, называемых q-ичными n - последовательностями, равно q^n.

   б) Число всех двоичных n - последовательностей, т.е. n-последовательностей из алфавита $\{0, 1\}$, равно 2^n.

   в) Число комбинаций дат рождения у n людей равно 365^n.

2) Пусть $n \le N$ и мы выбираем *без возвращения* n элементов из множества $[N]$. Результатом выбора является n-последовательность, элементы которой взяты из алфавита $[N] = \{1, 2, \ldots, N\}$ и отличны друг от друга. Такая n-последовательность называется n-*перестановкой* множества $[N]$. Число всех n-перестановок обозначается символом $(N)_n$ или символом $P(N, n)$ и вычисляется по формуле

$$(N)_n = P(N, n) \stackrel{\text{def}}{=} N \cdot (N-1) \cdot (N-2) \cdot \cdots \cdot (N-n+1).$$

   Для частного случая $n = N$ используется обозначение

$$(N)_N = P(N, N) = N \cdot (N-1) \cdot (N-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1 \stackrel{\text{def}}{=} N!.$$

   Примеры

   а) Число "слов" длины 4, которые можно составить из неповторяющихся букв пятеричного алфавита $\{a, b, c, d, e\}$, равно

$$(5)_4 = P(5, 4) = 5 \cdot 4 \cdot 3 \cdot 2.$$

   б) Число *различных* дат рождения у n людей, где $2 \le n \le 365$ равно

$$(365)_n = P(365, n) = 365 \cdot 364 \cdot 363 \cdot \cdots \cdot [365 - (n-1)].$$

   в) Вероятность того, что n случайно собравшихся людей имеют различные даты рождения равна

$$\pi_n = \frac{(365)_n}{(365)^n} = \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdot \cdots \cdot \left(1 - \frac{n-1}{365}\right).$$

   В частности, для $n = 30$ число $\pi_{30} = .2937$. Следовательно, вероятность совпадения дат рождения хотя бы у двоих среди $n = 30$ людей равна $1 - \pi_{30} = .7063$.

3) Пусть $n \le N$. *Число всех n-подмножеств множества* $[N]$ обозначается через $\binom{N}{n}$. Если положить по определению $0! = 1$, то имеет место следующая формула

$$\binom{N}{n} = \binom{N}{N-n} = \frac{(N)_n}{n!} = \frac{N!}{n!(N-n)!}.$$

Примеры

а) Пусть ребёнок играет с 8 карточками, на которых написаны следующие буквы:

$$\boxed{\text{О}} \quad \boxed{\text{О}} \quad \boxed{\text{О}} \quad \boxed{\text{О}} \quad \boxed{\text{К}} \quad \boxed{\text{К}} \quad \boxed{\text{Л}} \quad \boxed{\text{Т}}.$$

Он может выложить из этих карточек всего 8! последовательностей, из которых $4! \cdot 2! = 48$ последовательностей дают слово "ОКОЛОТОК". Заметим также, что среди 8! рассматриваемых последовательностей имеется лишь $\binom{8}{4}\binom{4}{2} \cdot 2!$ "различных слов". Следовательно, вероятность получить слово "ОКОЛОТОК" можно записать следующими двумя способами

$$\Pr\{\text{"ОКОЛОТОК"}\} = \frac{4! \cdot 2!}{8!} = \frac{1}{\binom{8}{4}\binom{4}{2} \cdot 2!} = \frac{48}{8!} = 1.19 \times 10^{-3}.$$

Задача. Проведите аналогичные рассуждения для слов "ПСИХОЛОГИЯ" и "МАТЕМАТИКА"

б) Каждое n-подмножество множества $[N]$ взаимно однозначно отображается в двоичную N - последовательность, содержащую единицы на позициях, соответствующих элементам данного n-подмножества. Поэтому число *всех подмножеств* множества $[N]$ совпадает с числом всех двоичных N - последовательностей, которое равно 2^N, и справедлива формула

$$\binom{N}{0} + \binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = \sum_{n=0}^{n=N} \binom{N}{n} = \sum_{n=0}^{n=N} \frac{N!}{n!(N-n)!} = 2^N.$$

в) Пусть $0 \le i \le m \le n < N$. Имеют место следующие утверждения

– Зафиксируем произвольное n-подмножество множества $[N]$. Тогда число m-подмножеств, i элементов которых принадлежат данному n-подмножеству, а $m - i$ элементов не принадлежат этому n-подмножеству, равно

$$\binom{n}{i}\binom{N-n}{m-i} \quad \text{и} \quad \sum_{i=0}^{m} \binom{n}{i}\binom{N-n}{m-i} = \binom{N}{m}.$$

Для написанных выше формул имеется следующая вероятностная интерпретация. Пусть множество $[N]$ состоит из n "счастливых" номеров и $N - n$ "несчастливых" номеров. Мы случайно *без возвращения* выбираем m номеров из множества $[N]$. Тогда вероятность того, что в нашу выборку попадут i "счастливых" и $m - i$ "несчастливых" номеров есть

$$\frac{\binom{n}{i}\binom{N-n}{m-i}}{\binom{N}{m}} = \binom{m}{i}\frac{(n)_i (N-n)_{m-i}}{(N)_n}.$$

Пример лотереи. Пусть $m = n = 6,\ N = 49,\ i = 0, 1, \ldots, 6$. В лотерее "6 из 49" вероятность g_i угадывания i "счастливых шаров" вычисляется следующим образом

$$g_i = \frac{\binom{6}{i}\binom{43}{6-i}}{\binom{49}{6}}, \quad g_0 = .4360, \quad g_1 = .4130, \quad g_2 = .1324,$$

$$g_3 = .01765, \quad g_4 = 9.686 \times 10^{-4}, \quad g_5 = 1.845 \times 10^{-5}, \quad g_6 = 7.151 \times 10^{-8}.$$

В лотерее "6 из 49", лотерейный билет выигрывает денежный приз, если он содержит по крайней мере *три* номера "счастливых шаров". Вероятность получения такого "счастливого билета" есть

$$p = g_3 + g_4 + g_5 + g_6 = .0186.$$

– Зафиксируем произвольное m-подмножество множества $[N]$. Тогда число всех n-подмножеств, которые содержат данное m-подмножество, равно

$$\binom{N-m}{n-m} = \binom{N-m}{N-n}.$$

§ 4.2. Биномиальные вероятности

Пусть $N = N_1 + N_2$ и имеется N шаров, занумерованных числами от 1 до N. Шары, занумерованные числами от 1 до N_1, имеют *белый* цвет, а остальные N_2 шаров являются *черными*.

Предположим, что из данного N-множества шаров выбираются случайно и *с возвращением* n шаров. Для любого $k = 0, 1, 2 \ldots, n$, число n- последовательностей, содержащих k белых и $n - k$ черных шаров, равно

$$\binom{n}{k} N_1^k N_2^{n-k} \quad \text{и дробь} \quad \frac{\binom{n}{k} N_1^k N_2^{n-k}}{N^n} = \binom{n}{k} \left(\frac{N_1}{N}\right)^k \left(\frac{N_2}{N}\right)^{n-k}$$

есть *вероятность* получить при таком случайном выборе последовательность из k белых и $n - k$ черных шаров.

Определение. Зафиксируем целое число $n = 1, 2, \ldots$ и произвольное число $p,\ 0 < p < 1$. Числа

$$b_n(p, k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \ldots, n,$$

называются *биномиальными вероятностями с параметрами* (n, p).

Для случайного выбора объема n с возвращением вероятность получения k белых и $n - k$ черных шаров является биномиальной вероятностью с параметрами $(n, p = \frac{N_1}{N})$.

Примеры.

1) Задача об отношениях предпочтения. Пусть малая социальная группа объема $n + 1$ проводит (внутри себя) выборы лидера. Каждый член группы в своем протоколе голосования (ПГ), т.е. в списке из $n + 1$ участника, отмечает $m < n$ членов группы, исключая самого себя, которых он *предпочитает* остальным $n - m$ членам группы и считает кандидатами в лидеры.

Зафиксируем произвольного члена группы, которого назовем претендентом на лидерство (ПЛ). Тогда для любого из остальных n участников выборов имеем:

а) $N = \binom{n}{m}$–общее число способов заполнить ПГ;

б) $N_1 = \binom{n-1}{m-1}$–число способов заполнить ПГ, в котором отмечен ПЛ;

в) $N_2 = \binom{n-1}{m}$–число способов заполнить ПГ, в котором ПЛ не отмечен.

Легко проверить, что $N = N_1 + N_2$. Для данных параметров в качестве модели *случайных отношений предпочтения* в испытуемой группе объема $n+1$ будем использовать введенную выше модель выборки с возвращением объема n. Следовательно, если отношения предпочтения в испытуемой группе случайны, то вероятность ПЛ быть отмеченным в k ПГ и быть неотмеченным в $(n-k)$ ПГ является биномиальной вероятностью вида

$$b_n\left(\frac{N_1}{N}, k\right) = \binom{n}{k}\left(\frac{N_1}{N}\right)^k\left(\frac{N_2}{N}\right)^{n-k} =$$

$$= \binom{n}{k}\left(\frac{\binom{n-1}{m-1}}{\binom{n}{m}}\right)^k\left(\frac{\binom{n-1}{m}}{\binom{n}{m}}\right)^{n-k} = \binom{n}{k}\left(\frac{m}{n}\right)^k\left(\frac{n-m}{n}\right)^{n-k},\ k = 0, 1, \ldots, n,$$

Кроме того, для модели случайных отношений предпочтения число m можно рассматривать в качестве *среднего числа* ПГ по испытуемой группе, в которых отмечен ПЛ.

Пусть по итогам голосования ПЛ оказался отмеченным в K ПГ, где число K удовлетворяет условиям $m \ll K \leq n$. На основании такого результата выборов мы можем принять *решение* о том, что отношения предпочтения в данной малой социальной группе *неслучайны*. Тогда число

$$\alpha = \sum_{k=K}^{n}\binom{n}{k}\left(\frac{m}{n}\right)^k\left(\frac{n-m}{n}\right)^{n-k}$$

может рассматриваться в качестве *вероятности ошибки* данного решения.

2) Задача о лотерее "6 из 49". Если купить один лотерейный билет, то ранее вычисленная вероятность того, что он окажется "счастливым" равна .0186. Пусть мы покупаем n лотерейных билетов и случайно отмечаем 6 номеров в каждом из них. Обозначим через P_n вероятность того, что *хотя бы один из этих n билетов будет "счастливым"* вычисляется по формуле

$$P_n = \sum_{k=1}^{n}\binom{n}{k}(.0186)^k\,(.9814)^{n-k} = 1 - (.9814)^n.$$

Для оценки количества покупаемых билетов n, которое гарантирует вероятность $P_n \geq .95$, можем написать следующую цепочку эквивалентных неравенств

$$1 - (.9814)^n \geq .95 \Leftrightarrow (.9814)^n \leq .05 \Leftrightarrow n \ln .9814 \leq \ln .05 \Leftrightarrow n \geq \frac{\ln .05}{\ln .9814} = 159.55.$$

Таким образом, если мы хотим с надежностью $\geq 95\%$ быть уверены, что по крайней мере один из наших n лотерейных билетов будет "счастливым", то надо покупать $n \geq 160$ билетов.

§ 5. Модель (n,p) - испытаний Бернулли

Пусть $n = 1, 2, \ldots$ есть число испытаний, в каждом их которых регистрируется один из двух возможных исходов: *успех*, обозначаемый символом 1, или *неудача*, обозначаемая символом 0. Пусть p, $0 < p < 1$, - фиксированный параметр.

Модель (n,p) - *испытаний Бернулли* задается пространством $\Omega = \{\omega\}$, состоящим из $N = 2^n$ элементарных событий

$$\omega = (x_1, x_2, \ldots, x_n), \quad x_i = 1,\ 0, \quad i = 1, 2, \ldots, n,$$

где $x_i = 1,\ 0$ обозначает исход (успех или неудачу) в i-ом, $i = 1, 2, \ldots, n$, испытании. Вероятности этих элементарных событий имеют вид:

$$\Pr\{\omega\} = \Pr\{(x_1, x_2, \ldots, x_n)\} \stackrel{\text{def}}{=} p^k \cdot (1-p)^{n-k}, \quad \text{если} \quad \sum_{i=1}^{n} x_i = k, \quad 0 \le k \le n. \qquad (1)$$

Отметим, что при $p = 1 - p = 1/2$ в модели $(n, 1/2)$ - испытаний Бернулли всем 2^n элементарным событиям поставлены в соответствие одинаковые вероятности, равные $1/2^n$.

Пример. При $n = 3$ число $N = 2^3 = 8$. Из (1) получаем следующие формулы для вероятностей восьми элементарных событий в модели $(3,p)$ - испытаний Бернулли:

$$\Pr\{(0,0,0)\} = (1-p)^3 \quad \Pr\{(0,0,1)\} = \Pr\{(0,1,0)\} = \Pr\{(1,0,0)\} = p \cdot (1-p)^2,$$

$$\Pr\{(0,1,1)\} = \Pr\{(1,0,1)\} = \Pr\{(1,1,0)\} = p^2 \cdot (1-p), \quad \Pr\{(1,1,1)\} = p^3.$$

Двоичным символам $x = 1,\ 0$ сопоставим числа

$$\Pr\{x\} \stackrel{\text{def}}{=} \begin{cases} p, & \text{если } x = 1, \\ 1-p, & \text{если } x = 0, \end{cases}$$

которые удобно интерпретировать как вероятности успеха или неудачи при однократном испытании Бернулли. Используя эти обозначения, вероятности (1) при n-кратном испытании Бернулли можно представить в виде произведения вероятностей

$$\Pr\{\omega\} = \Pr\{(x_1, x_2, \ldots, x_n)\} = \Pr\{x_1\} \cdot \Pr\{x_2\} \cdot \ldots \cdot \Pr\{x_n\}.$$

Данная запись, в частности, позволяет проверить, что сумма всех 2^n чисел, задаваемых формулами (1), равна

$$\sum_{w \in \Omega} \Pr\{\omega\} = \left(\sum_{x_1=0}^{1} \Pr\{x_1\}\right) \cdot \left(\sum_{x_2=0}^{1} \Pr\{x_2\}\right) \cdot \ldots \cdot \left(\sum_{x_n=0}^{1} \Pr\{x_n\}\right) =$$

$$= [p + (1-p)] \cdot [p + (1-p)] \cdot \ldots \cdot [p + (1-p)] \ = \ 1,$$

т.е. числа (1) являются вероятностями.

Введём величину $S_n \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i$, $0 \le S_n \le n$, равную числу единичных символов в последовательности $\omega = (x_1, x_2, \ldots, x_n)$. Поскольку единицы соответствуют успехам в испытаниях, а нули – неудачам, то значение S_n представляет собой *число успехов* в элементарном событии ω, а разность $n - S_n$ есть число неудач в элементарном событии ω. Для фиксированного целого числа $k = 0, 1, \ldots, n$ рассмотрим

$$\text{событие} \quad \{S_n = k\}, \quad \text{состоящее из всех} \quad \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

элементарных событий $\omega = (x_1, x_2, \ldots, x_n)$, в которых число единиц $\sum_{i=1}^{n} x_i$ равно k. Событие $\{S_n = k\}$ означает, что в n испытаниях Бернулли произошло k успехов. Если символом $b_n(p,k)$ обозначить вероятность события $\{S_n = k\}$, то из (1) следует:

$$b_n(p,k) = \Pr\{S_n = k\} = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \; =$$

$$= \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{2}$$

Соотношения (2) описывают важную модификацию модели (n,p)- испытаний Бернулли, в которой пространство элементнарных событий $\Omega = \{0, 1, 2, \ldots, n\}$ представляет собой множество из $N = n + 1$ возможных значений для числа успехов в n испытаниях. Это соответствует типичной для приложений ситуации, когда экспериментатор регистрирует лишь общее количество успехов в n испытаниях и не интересуется в каких конкретно испытаниях происходили успехи или неудачи.

Пример. При $n = 3$ число $N = 3 + 1 = 4$ и из (2) получаем следующие формулы для вероятностей четырёх элементарных событий:

$$b_3(p,0) = \Pr\{S_3 = 0\} = (1-p)^3, \qquad b_3(p,1) = \Pr\{S_3 = 1\} = 3p \cdot (1-p)^2,$$

$$b_3(p,2) = \Pr\{S_3 = 2\} = 3p^2 \cdot (1-p), \qquad b_3(p,3) = \Pr\{S_3 = 3\} = p^3.$$

Отметим, что известная из курса математического анализа формула *бинома Ньютона*

$$(a+b)^n \; = \; \sum_{k=0}^{n} \binom{n}{k} \cdot a^k \cdot b^{n-k} \; = \; \sum_{k=0}^{n} \frac{n!}{k! \cdot (n-k)!} \cdot a^k \cdot b^{n-k}$$

позволяет проверить, что сумма $n + 1$ чисел, задаваемых (2), равна 1:

$$\sum_{k=0}^{n} b_n(p,k) = \sum_{k=0}^{n} \frac{n!}{k! \cdot (n-k)!} \cdot p^k \cdot (1-p)^{n-k} \; = \; [p + (1-p)]^n \; = \; 1,$$

Числа $b_n(p,k)$, $k = 0, 1, 2, \ldots, n$, называются *биномиальными вероятностями*. Свойства биномиальных вероятностей изучаются в Задании 1, где также анализируются необходимые для решения прикладных задач *таблицы* этих вероятностей.

§ 6. Операции над событиями, закон сложения вероятностей

С событиями, которые мы определили как подмножества пространства $\Omega = \{\omega\}$ элементарных событий ω, связывают три теоретико - множественные операции: *дополнение, пересечение* и *объединение.*

Приведём традиционную теоретико-вероятностную терминологию, связанную с этими операциями и их обозначениями в теории множеств.

- Если $\omega \in \mathsf{A}$ то говорят, что *наступило* событие A.
- Операция *дополнения* для события $\overline{\mathsf{A}}$ определяет событие, состоящее из элементарных событий $\omega \in \Omega$, не входящих в событие A. Говорят что $\overline{\mathsf{A}} \subseteq \Omega$ означает событие, состоящее в *ненаступлении* события A. Событие $\overline{\mathsf{A}}$ называют также *отрицанием* события A.
- Символ *пересечения* (A, B) событий A и B определяет событие, состоящее из элементарных событий $\omega \in \Omega$, входящих *и в* A *и в* B. Говорят, что (A, B) означает событие, состоящее в *одновременном наступлении* событий A и B.
- Символ *объединения* $\mathsf{A} \cup \mathsf{B}$ событий A и B определяет событие, состоящее из элементарных событий $\omega \in \Omega$, входящих *либо в* A *либо в* B. Говорят, что $\mathsf{A} \cup \mathsf{B}$ обозначает событие, состоящее в том, что *произошло хотя бы одно* из событий A или B.
- Символ $\emptyset$ определяет пустое множество или *невозможное* событие.
- Символ Ω определяет множество, состоящее из всех точек, которое называется *достоверным* событием. Очевидно, $\overline{\Omega} = \emptyset$.
- Символ подмножества $\mathsf{A} \subseteq \mathsf{B}$ в теории вероятностей обозначает, что из осуществления события A *необходимо следует* наступление события B.
- Если события A и B не пересекаются, т.е. $(\mathsf{A}, \mathsf{B}) = \emptyset$, то эти события называются *несовместимыми* (не могут наступить одновременно).
- Если $(\mathsf{A}, \mathsf{B}) = \emptyset$, то символ $\mathsf{A} + \mathsf{B}$ обозначает событие, состоящее в том, что *наступит хотя бы одно из несовместимых* событий A или B.

Свойства.

1) *Если события* A *и* B *несовместимы, т.е.* $(\mathsf{A}, \mathsf{B}) = \emptyset$, *то справедлива формула*

$$\Pr\{\mathsf{A} + \mathsf{B}\} = \Pr\{\mathsf{A}\} + \Pr\{\mathsf{B}\}.$$

Эта формула называется *законом сложения вероятностей.*

2) *Любое событие* A *несовместимо с* $\overline{\mathsf{A}}$. *Следовательно,* $\mathsf{A} + \overline{\mathsf{A}} = \Omega$ *и* $\Pr\{\overline{\mathsf{A}}\} = 1 - \Pr\{\mathsf{A}\}$.

3) *Для любой пары событий* A *и* B *событие* (A, B) *несовместимо с событием* $(\mathsf{A}, \overline{\mathsf{B}})$. *При этом* $(\mathsf{A}, \mathsf{B}) + (\mathsf{A}, \overline{\mathsf{B}}) = \mathsf{A}$ *и вероятность* $\Pr\{\mathsf{A}\} = \Pr\{(\mathsf{A}, \mathsf{B})\} + \Pr\{(\mathsf{A}, \overline{\mathsf{B}})\}$.

§ 7. Модель (2×2)-таблицы

Используя данные выше обозначения и их свойства, мы вводим важную для приложений математическую модель случайного опыта, называемую *моделью* (2×2)-*таблицы* и применяемую для описания *связи* (или взаимодействия) *признаков*.

Опыт состоит в *одновременном* тестировании испытуемого по двум признакам $\mathcal{A}$ и $\mathcal{B}$. Например, $\mathcal{A}$ – вес, а $\mathcal{B}$ – объём талии или $\mathcal{A}$ – рост, а $\mathcal{B}$ – объём бёдер. Предположим, что признак $\mathcal{A}$ (признак $\mathcal{B}$) может быть зарегистрирован в одном из двух состояний, интерпретируемых как событие A или его отрицание $\overline{\mathsf{A}}$ (событие B или его отрицание $\overline{\mathsf{B}}$). В итоге такого одновременного тестирования наступает одно из четырех элементарных событий (A,B), $(\overline{\mathsf{A}},\mathsf{B})$, $(\mathsf{A},\overline{\mathsf{B}})$ или $(\overline{\mathsf{A}},\overline{\mathsf{B}})$, которые вместе с их вероятностями записываются в виде (2×2)-таблиц

$$\left\| \begin{array}{c|c} (\mathsf{A},\mathsf{B}) & (\overline{\mathsf{A}},\mathsf{B}) \\ (\mathsf{A},\overline{\mathsf{B}}) & (\overline{\mathsf{A}},\overline{\mathsf{B}}) \end{array} \right\| \qquad \left\| \begin{array}{c|c} \Pr\{(\mathsf{A},\mathsf{B})\} & \Pr\{(\overline{\mathsf{A}},\mathsf{B})\} \\ \Pr\{(\mathsf{A},\overline{\mathsf{B}})\} & \Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\} \end{array} \right\|.$$

Соотношения между элементами (2×2)-таблиц, аналогичные свойству 3, имеют форму *расширенных* (2×2)-таблиц

$$\left| \begin{array}{c|c|c|c} & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\ \mathcal{B}=\mathsf{B} & (\mathsf{A},\mathsf{B}) & (\overline{\mathsf{A}},\mathsf{B}) & \mathsf{B}=(\mathsf{A},\mathsf{B})+(\overline{\mathsf{A}},\mathsf{B}) \\ \mathcal{B}=\overline{\mathsf{B}} & (\mathsf{A},\overline{\mathsf{B}}) & (\overline{\mathsf{A}},\overline{\mathsf{B}}) & \overline{\mathsf{B}}=(\mathsf{A},\overline{\mathsf{B}})+(\overline{\mathsf{A}},\overline{\mathsf{B}}) \\ \sum & \mathsf{A}=(\mathsf{A},\mathsf{B})+(\mathsf{A},\overline{\mathsf{B}}) & \overline{\mathsf{A}}=(\overline{\mathsf{A}},\mathsf{B})+(\overline{\mathsf{A}},\overline{\mathsf{B}}) & \Omega=(\mathsf{A},\mathsf{B})+(\mathsf{A},\overline{\mathsf{B}})+(\overline{\mathsf{A}},\mathsf{B})+(\overline{\mathsf{A}},\overline{\mathsf{B}}) \end{array} \right|,$$

$$\left| \begin{array}{c|c|c|c} & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\ \mathcal{B}=\mathsf{B} & \Pr\{(\mathsf{A},\mathsf{B})\} & \Pr\{(\overline{\mathsf{A}},\mathsf{B})\} & \Pr\{\mathsf{B}\} \\ \mathcal{B}=\overline{\mathsf{B}} & \Pr\{(\mathsf{A},\overline{\mathsf{B}})\} & \Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\} & \Pr\{\overline{\mathsf{B}}\} \\ \sum & \Pr\{\mathsf{A}\} & \Pr\{\overline{\mathsf{A}}\} & 1 \end{array} \right|, \qquad (1)$$

где

$$\Pr\{\mathsf{A}\}=\Pr\{(\mathsf{A},\mathsf{B})\}+\Pr\{(\mathsf{A},\overline{\mathsf{B}})\}, \quad \Pr\{\overline{\mathsf{A}}\}=\Pr\{(\overline{\mathsf{A}},\mathsf{B})\}+\Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\},$$
$$\Pr\{\mathsf{B}\}=\Pr\{(\mathsf{A},\mathsf{B})\}+\Pr\{(\overline{\mathsf{A}},\mathsf{B})\}, \quad \Pr\{\overline{\mathsf{B}}\}=\Pr\{(\mathsf{A},\overline{\mathsf{B}})\}+\Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\},$$
$$\Pr\{\mathsf{A}\}+\Pr\{\overline{\mathsf{A}}\}=\Pr\{\mathsf{B}\}+\Pr\{\overline{\mathsf{B}}\}=1.$$

Таблица (1) представляет собой удобный способ записи модели (2×2)-таблицы.

Пусть для каждого из n испытуемых проведено одновременное тестирование по признакам $\mathcal{A}$ и $\mathcal{B}$. Пусть символами $n(\mathsf{A})$, $n(\overline{\mathsf{A}})$, $n(\mathsf{B})$, $n(\overline{\mathsf{B}})$, $n(\mathsf{A},\mathsf{B})$, $n(\overline{\mathsf{A}},\mathsf{B})$, $n(\mathsf{A},\overline{\mathsf{B}})$ и $n(\overline{\mathsf{A}},\overline{\mathsf{B}})$ обозначают числа наступлений соответствующих событий в n опытах. Например, число

$$n(\mathsf{A})=n(\mathsf{A},\mathsf{B})+n(\mathsf{A},\overline{\mathsf{B}}), \quad 0 \le n(\mathsf{A}) \le n,$$

обозначает количество испытуемых, для которых у признака $\mathcal{A}$ зарегистрировано событие A. Число $n(\mathsf{A},\overline{\mathsf{B}})$, $0 \le n(\mathsf{A},\overline{\mathsf{B}}) \le n(\mathsf{A})$, есть количество испытуемых, для которых у признака $\mathcal{A}$ зарегистрировано событие A и одновременно у признака $\mathcal{B}$ зарегистрировано событие $\overline{\mathsf{B}}$. Очевидно, что

$$n=n(\mathsf{A})+n(\overline{\mathsf{A}})=n(\mathsf{B})+n(\overline{\mathsf{B}})=n(\mathsf{A},\mathsf{B})+n(\overline{\mathsf{A}},\mathsf{B})+n(\mathsf{A},\overline{\mathsf{B}})+n(\overline{\mathsf{A}},\overline{\mathsf{B}}).$$

Указанные числа изображаются в виде расширенной (2×2)-таблицы

$$\left|\begin{array}{c||c|c||c} & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\ \mathcal{B}=\mathsf{B} & n(\mathsf{A},\mathsf{B}) & n(\overline{\mathsf{A}},\mathsf{B}) & n(\mathsf{B})=n(\mathsf{A},\mathsf{B})+n(\overline{\mathsf{A}},\mathsf{B}) \\ \mathcal{B}=\overline{\mathsf{B}} & n(\mathsf{A},\overline{\mathsf{B}}) & n(\overline{\mathsf{A}},\overline{\mathsf{B}}) & n(\overline{\mathsf{B}})=n(\mathsf{A},\overline{\mathsf{B}})+n(\overline{\mathsf{A}},\overline{\mathsf{B}}) \\ \sum & n(\mathsf{A})=n(\mathsf{A},\mathsf{B})+n(\mathsf{A},\overline{\mathsf{B}}) & n(\overline{\mathsf{A}})=n(\overline{\mathsf{A}},\mathsf{B})+n(\overline{\mathsf{A}},\overline{\mathsf{B}}) & n(\mathsf{A})+n(\overline{\mathsf{A}})=n(\mathsf{B})+n(\overline{\mathsf{B}}) \end{array}\right|,$$

называемой (2×2)-*таблицей сопряжённости признаков*. Числовые данные (2×2)-таблицы сопряжённости используются для решения двух нижеформулируемых задач статистического анализа о связи (взаимодействии) признаков $\mathcal{A}$ и $\mathcal{B}$.

1) Пользуясь соображениями *здравого смысла*, сделать *качественный вывод* либо о наличии прямой (обратной) связи между признаками $\mathcal{A}$ и $\mathcal{B}$, либо об отсутствии связи между $\mathcal{A}$ и $\mathcal{B}$.

2) Если в первой задаче принято решение о наличии прямой (обратной) связи между признаками $\mathcal{A}$ и $\mathcal{B}$, то постановка второй задачи состоит в том, чтобы на основе разработанной в теории вероятностей адекватной модели случайного опыта получить *количественную* оценку для ошибки этого качественного решения.

Пример. Пусть число испытуемых $n = 27$. Приведём для сравнения записанные в форме (2×2)-таблиц сопряжённости признаков *два варианта* экспериментальных данных, которые могли бы быть получены для *двух разных пар* признаков.

$$\left|\begin{array}{c||c|c||c} & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\ \mathcal{B}=\mathsf{B} & 10 & 3 & n(\mathsf{B})=10+3=13 \\ \mathcal{B}=\overline{\mathsf{B}} & 4 & 10 & n(\overline{\mathsf{B}})=4+10=14 \\ \sum & n(\mathsf{A})=10+4=14 & n(\overline{\mathsf{A}})=3+10=13 & n=n(\mathsf{A})+n(\overline{\mathsf{A}})=n(\mathsf{B})+n(\overline{\mathsf{B}})=27 \end{array}\right|$$

Таблица 1.

$$\left|\begin{array}{c||c|c||c} & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\ \mathcal{B}=\mathsf{B} & 7 & 6 & n(\mathsf{B})=7+6=13 \\ \mathcal{B}=\overline{\mathsf{B}} & 7 & 7 & n(\overline{\mathsf{B}})=7+7=14 \\ \sum & n(\mathsf{A})=7+7=14 & n(\overline{\mathsf{A}})=6+7=13 & n=n(\mathsf{A})+n(\overline{\mathsf{A}})=n(\mathsf{B})+n(\overline{\mathsf{B}})=27 \end{array}\right|$$

Таблица 2.

В следующем разделе рассматривается теоретико-вероятностный подход к вопросу о связи событий, с помощью которого в математической статистике разработаны рецепты анализа числовых данных из (2×2)-таблиц сопряжённости признаков. Опираясь на частотное определение условной вероятности, лежащее в основе этого подхода, можно будет сделать качественные выводы о том, что числовые данные таблицы 1 показывают наличие прямой связи между соответствующими признаками, а числовые данные таблицы 2 свидетельствуют об отсутствии связи.

§ 8. Связь событий, условная вероятность, независимость

В теории вероятностей характеристикой связи событий A и B служит *условная вероятность* $\Pr\{\mathsf{A}|\mathsf{B}\}$ *события* A *при условии* B, определяемая как

$$\Pr\{\mathsf{A}|\mathsf{B}\} \stackrel{\text{def}}{=} \frac{\Pr\{(\mathsf{A},\mathsf{B})\}}{\Pr\{\mathsf{B}\}}. \tag{1}$$

Это определение имеет смысл лишь в случае, если $\Pr\{\mathsf{B}\} \neq 0$.

Применим интуитивное представление о вероятности, связанное с частотами, для того, чтобы оправдать определение величины $\Pr\{\mathsf{A}|\mathsf{B}\}$ из (1) как условной вероятности осуществления события A при условии наступления события B. Используя обозначения предыдущего раздела, назовём *условной частотой* события A при условии, что событие B наступило, отношение

$$\frac{n(\mathsf{A},\mathsf{B})}{n(\mathsf{B})} = \frac{n(\mathsf{A},\mathsf{B})}{n(\mathsf{A},\mathsf{B}) + n(\overline{\mathsf{A}},\mathsf{B})} = \frac{n(\mathsf{A},\mathsf{B})/n}{n(\mathsf{B})/n}, \tag{2}$$

т.е. частоту события A, вычисленную не по совокупности всех n экспериментов, а лишь по совокупности тех $n(\mathsf{B})$ экспериментов, в которых наступило событие B. При больших значениях n левая часть равенства (2) интерпретируется как приближённое значение условной вероятности $\Pr\{\mathsf{A}|\mathsf{B}\}$ наступления события A при условии, что B наступило, а именно:

$$\text{отношение} \quad \frac{n(\mathsf{A},\mathsf{B})}{n(\mathsf{B})} \approx \Pr\{\mathsf{A}|\mathsf{B}\}.$$

При этом в правой части равенства (2)

$$\text{отношение} \quad \frac{n(\mathsf{A},\mathsf{B})}{n} \approx \Pr\{(\mathsf{A},\mathsf{B})\}, \quad \text{а отношение} \quad \frac{n(\mathsf{B})}{n} \approx \Pr\{\mathsf{B}\}.$$

Изложенные соображения мотивируют определение условной вероятности (1).

Естественно сказать, что событие A *не зависит* от события B, если условная вероятность события A при условии B равна безусловной вероятности события A, т.е.

$$\Pr\{\mathsf{A}|\mathsf{B}\} = \Pr\{\mathsf{A}\}. \tag{3}$$

Пример. Рассмотрим числа $n(\mathsf{A},\mathsf{B})$, $n(\overline{\mathsf{A}},\mathsf{B})$, $n(\mathsf{A},\overline{\mathsf{B}})$ и $n(\overline{\mathsf{A}},\overline{\mathsf{B}})$ из таблиц 1 и 2 предыдущего раздела и проведем сравнение соответствующих условных и безусловных частот, получаемых по данным числам. Например, для условной частоты события A при условии B и безусловной частоты события A таблица 1 даёт

$$\frac{n(\mathsf{A},\mathsf{B})}{n(\mathsf{B})} = \frac{10}{13} = 0.769 \approx \Pr\{\mathsf{A}|\mathsf{B}\} \quad \text{и} \quad \frac{n(\mathsf{A})}{n} = \frac{14}{27} = 0.519 \approx \Pr\{\mathsf{A}\}.$$

Эти условная и безусловная частоты значительно отличаются друг от друга. Для таблицы 2 те же самые условная и безусловная частоты есть

$$\frac{n(\mathsf{A},\mathsf{B})}{n(\mathsf{B})} = \frac{7}{13} = 0.538 \approx \Pr\{\mathsf{A}|\mathsf{B}\} \quad \text{и} \quad \frac{n(\mathsf{A})}{n} = \frac{14}{27} = 0.519 \approx \Pr\{\mathsf{A}\}.$$

Их отличие друг от друга значительно меньше, чем для таблицы 1. Аналогичные результаты сравнения справедливы и для остальных трёх условных частот, отвечающих условным вероятностям $\Pr\{\overline{\mathsf{A}}|\mathsf{B}\}$, $\Pr\{\mathsf{A}|\overline{\mathsf{B}}\}$ и $\Pr\{\overline{\mathsf{A}}|\overline{\mathsf{B}}\}$. Следовательно, по данным таблицы 1 можно сделать качественный вывод о наличии связи признаков $\mathcal{A}$ и $\mathcal{B}$, т.е. отвергнуть гипотезу об их независимости (отсутствии связи), а на основании данных таблицы 2 этого сделать нельзя.

С помощью определения условной вероятности (1) и определения независимости событий по формуле (3) получаем, что для независимых событий $\Pr\{(\mathsf{A},\mathsf{B})\} = \Pr\{\mathsf{A}\}\cdot\Pr\{\mathsf{B}\}$. Следовательно, если A не зависит от B, то и B не зависит от A, поскольку последнее равенство симметрично относительно A и B. Поэтому принимается следующее

Определение 1. События A и B называются *независимыми*, если выполнено равенство

$$\Pr\{(\mathsf{A},\mathsf{B})\} = \Pr\{\mathsf{A}\}\cdot\Pr\{\mathsf{B}\}. \tag{4}$$

Замечание 1. Формула (4) представляет несколько более общее определение независимости, поскольку не предполагает, что $\Pr\{\mathsf{B}\} > 0$.

Определение 2. Модель (2×2)-таблицы называется *моделью* (2×2)-*таблицы для независимых признаков*, если вероятность элементарного исхода $\Pr\{(\mathsf{A},\mathsf{B})\}$ задается равенством (4), а вероятности остальных трёх элементарных исходов $\Pr\{(\overline{\mathsf{A}},\mathsf{B})\}$, $\Pr\{(\mathsf{A},\overline{\mathsf{B}})\}$ и $\Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\}$ также определяются как *произведения* соответствующих вероятностей.

Замечание 2. Пусть p, $0 < p < 1$, - фиксированный параметр. Тогда модель (2×2)-таблицы для независимых признаков, в которой

$$\Pr\{\mathsf{A}\} = \Pr\{\mathsf{B}\} = p, \quad \text{и} \quad \Pr\{\overline{\mathsf{A}}\} = \Pr\{\overline{\mathsf{B}}\} = 1-p,$$

представляет собой модель $(2,p)$-испытаний Бернулли. Отметим, что используя введённое понятие независимости, мы можем сказать, что общая модель (n,p)-испытаний Бернулли является последовательностью из n *независимых испытаний*, где в каждом испытании происходит либо "успех" с одной и той же вероятностью p, либо "неудача" с одной и той же вероятностью $1-p$.

Свойства.

1) *Достоверное событие* Ω *и невозможное событие* $\emptyset$ *независимы с любым событием* A.

2) *Если события* A *и* B *независимы, то события* $\overline{\mathsf{A}}$ *и* B *также независимы.*

3) *Из независимости событий* A *и* B *следует независимость событий* $\overline{\mathsf{A}}$ *и* $\overline{\mathsf{B}}$.

4) *Модель* (2×2)*-таблицы является моделью* (2×2)*-таблицы для независимых признаков тогда и только тогда, когда выполнено условие* (4).

Первое свойство проверяется непосредственно по формуле (4), полагая в ней сначала $\mathsf{B} = \Omega$, а затем $\mathsf{B} = \emptyset$. Третье и четвёртое свойства есть очевидные следствия второго свойства, которое доказывается следующим образом:

$$\Pr\{(\overline{\mathsf{A}},\mathsf{B})\} = \Pr\{\mathsf{B}\} - \Pr\{(\mathsf{A},\mathsf{B})\} = \Pr\{\mathsf{B}\} - \Pr\{(\mathsf{A},\mathsf{B})\} =$$

$$= \Pr\{\mathsf{B}\} - \Pr\{\mathsf{A}\}\cdot\Pr\{\mathsf{B}\} = \Pr\{\mathsf{B}\}\cdot(1-\Pr\{\mathsf{A}\}) = \Pr\{\mathsf{B}\}\cdot\Pr\{\overline{\mathsf{A}}\}.$$

Рассмотрим для сравнения два примера моделей (2×2)-таблиц:

$$\left|\begin{array}{c||c|c||c|}
 & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\
\hline\hline
\mathcal{B}=\mathsf{B} & \Pr\{(\mathsf{A},\mathsf{B})\}=1/4 & \Pr\{(\overline{\mathsf{A}},\mathsf{B})\}=1/4 & \Pr\{\mathsf{B}\}=1/2 \\
\mathcal{B}=\overline{\mathsf{B}} & \Pr\{(\mathsf{A},\overline{\mathsf{B}})\}=1/4 & \Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\}=1/4 & \Pr\{\overline{\mathsf{B}}\}=1/2 \\
\hline
\sum & \Pr\{\mathsf{A}\}=1/2 & \Pr\{\overline{\mathsf{A}}\}=1/2 & 1 \\
\hline
\end{array}\right|,$$

$$\left|\begin{array}{c||c|c||c|}
 & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\
\hline\hline
\mathcal{B}=\mathsf{B} & \Pr\{(\mathsf{A},\mathsf{B})\}=1/2 & \Pr\{(\overline{\mathsf{A}},\mathsf{B})\}=0 & \Pr\{\mathsf{B}\}=1/2 \\
\mathcal{B}=\overline{\mathsf{B}} & \Pr\{(\mathsf{A},\overline{\mathsf{B}})\}=0 & \Pr\{(\overline{\mathsf{A}},\overline{\mathsf{B}})\}=1/2 & \Pr\{\overline{\mathsf{B}}\}=1/2 \\
\hline
\sum & \Pr\{\mathsf{A}\}=1/2 & \Pr\{\overline{\mathsf{A}}\}=1/2 & 1 \\
\hline
\end{array}\right|.$$

Первая модель есть частный случай модели (2×2)-таблицы для независимых признаков. Она эквивалентна модели $(2, 1/2)$-испытаний Бернулли. Вторая модель описывает сильную прямую связь признаков, когда условная вероятность

$$\Pr\{\mathsf{A}|\mathsf{B}\} \;=\; \frac{\Pr\{(\mathsf{A},\mathsf{B})\}}{\Pr\{\mathsf{B}\}} \;=\; \frac{1/2}{1/2} \;=\; 1, \quad \Pr\{\mathsf{B}|\mathsf{A}\} \;=\; \frac{\Pr\{(\mathsf{A},\mathsf{B})\}}{\Pr\{\mathsf{A}\}} \;=\; \frac{1/2}{1/2} \;=\; 1.$$

Отметим, что безусловные вероятности событий (состояний) $\Pr\{\mathsf{A}\} = \Pr\{\mathsf{B}\} = 1/2$ в обеих моделях являются одинаковыми.

§ 9. Гипергеометрические вероятности для (2×2)-таблиц сопряжённости в случае независимых признаков

Приведём список всех 14 вариантов (2×2)-таблицы сопряжённости

$$\left|\begin{array}{c||c|c||c|}
 & \mathcal{A}=\mathsf{A} & \mathcal{A}=\overline{\mathsf{A}} & \sum \\
\hline\hline
\mathcal{B}=\mathsf{B} & n(\mathsf{A},\mathsf{B}) & n(\overline{\mathsf{A}},\mathsf{B}) & n(\mathsf{B})=n(\mathsf{A},\mathsf{B})+n(\overline{\mathsf{A}},\mathsf{B}) \\
\mathcal{B}=\overline{\mathsf{B}} & n(\mathsf{A},\overline{\mathsf{B}}) & n(\overline{\mathsf{A}},\overline{\mathsf{B}}) & n(\overline{\mathsf{B}})=n(\mathsf{A},\overline{\mathsf{B}})+n(\overline{\mathsf{A}},\overline{\mathsf{B}}) \\
\hline
\sum & n(\mathsf{A})=n(\mathsf{A},\mathsf{B})+n(\mathsf{A},\overline{\mathsf{B}}) & n(\overline{\mathsf{A}})=n(\overline{\mathsf{A}},\mathsf{B})+n(\overline{\mathsf{A}},\overline{\mathsf{B}}) & n(\mathsf{A})+n(\overline{\mathsf{A}})=n(\mathsf{B})+n(\overline{\mathsf{B}}) \\
\hline
\end{array}\right|,$$

которые возможны в опытах с $n = 27$ испытуемыми, если фиксированы значения $n(\mathsf{A}) = 14$ и $n(\mathsf{B}) = 13$:

$$\left|\begin{array}{c|c||c|}
13-i & i=n(\overline{\mathsf{A}},\mathsf{B}) & 13 \\
\hline
i+1 & 13-i & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right| = \left|\begin{array}{c|c||c|}
13 & 0 & 13 \\
\hline
1 & 13 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
12 & 1 & 13 \\
\hline
2 & 12 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
11 & 2 & 13 \\
\hline
3 & 11 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|,$$

$$\left|\begin{array}{c|c||c|}
10 & 3 & 13 \\
\hline
4 & 10 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
9 & 4 & 13 \\
\hline
5 & 9 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
8 & 5 & 13 \\
\hline
6 & 8 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
7 & 6 & 13 \\
\hline
7 & 7 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|,$$

$$\left|\begin{array}{c|c||c|}
6 & 7 & 13 \\
\hline
8 & 6 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
5 & 8 & 13 \\
\hline
9 & 5 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
4 & 9 & 13 \\
\hline
10 & 4 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|, \quad \left|\begin{array}{c|c||c|}
3 & 10 & 13 \\
\hline
11 & 3 & 14 \\
\hline\hline
14 & 13 & 27 \\
\hline
\end{array}\right|,$$

$$\left|\begin{array}{c|c||c} 2 & 11 & 13 \\ 12 & 2 & 14 \\ \hline 14 & 13 & 27 \end{array}\right|, \qquad \left|\begin{array}{c|c||c} 1 & 12 & 13 \\ 13 & 1 & 14 \\ \hline 14 & 13 & 27 \end{array}\right|, \qquad \left|\begin{array}{c|c||c} 0 & 13 & 13 \\ 14 & 0 & 14 \\ \hline 14 & 13 & 27 \end{array}\right|.$$

Каждый вариант идентифицируется числом $n(\overline{\mathsf{A}}, \mathsf{B}) = i$, где $i = 0, 1, \ldots, 13$.

Здравый смысл подсказывает, что если в результате опытов с $n = 27$ испытуемыми будет получена одна из первых четырёх-пяти (или одна из последних четырёх-пяти) таблиц данного списка, то можно принять более или менее надёжное решение о наличии прямой (обратной) связи между признаками, т.е. сделать вывод о том, что отсутствии связи не подтверждается. В остальных случаях, видимо, можно считать, что эксперименты подтверждают отсутствие связи. Количественную оценку ошибки решения здравого смысла можно получить на основе теории вероятностей, а именно: *если признаки независимы, то* можно показать, что *условная вероятность*

$$\Pr\{n(\overline{\mathsf{A}}, \mathsf{B}) = i \,|\, n(\mathsf{A}) = 14, n(\mathsf{B}) = 13\} \;=\; \frac{\binom{13}{i}\binom{14}{13-i}}{\binom{27}{13}}, \quad i = 0, 1, \ldots, 13.$$

Эти числа называются *гипергеометрическими вероятностями*, которые в общем случае задаются формулами:

$$Q_i(M_1, M_2, m) \stackrel{\text{def}}{=} \binom{M_1}{i} \cdot \binom{M_2}{m-i} \Big/ \binom{M_1 + M_2}{m}, \quad 0 \le i \le m \le M_1 \le M_2.$$

При этом $\sum_{i=1}^{m} Q_i(M_1, M_2, m) = 1$.

Непосредственно или по таблицам вычислим нижние α_k^-, $k = 2, 3, 4$, и верхние α_k^+, $k = 8, 9, 10$, "хвосты" гипергеометрических вероятностей:

$$\alpha_k^- = \Pr\{n(\overline{\mathsf{A}}, \mathsf{B}) \le k \,|\, n(\mathsf{A}) = 14, n(\mathsf{B}) = 13\} = \sum_{i=0}^{k} \frac{\binom{13}{i}\binom{14}{13-i}}{\binom{27}{13}} = \begin{cases} .0014, & \text{если } k = 2, \\ .016, & \text{если } k = 3, \\ .087, & \text{если } k = 4, \end{cases}$$

$$\alpha_k^+ = \Pr\{n(\overline{\mathsf{A}}, \mathsf{B}) \ge k \,|\, n(\mathsf{A}) = 14, n(\mathsf{B}) = 13\} = \sum_{i=k}^{13} \frac{\binom{13}{i}\binom{14}{13-i}}{\binom{27}{13}} = \begin{cases} .006, & \text{если } k = 10, \\ .041, & \text{если } k = 9, \\ .169, & \text{если } k = 8. \end{cases}$$

Выводы о связи между признаками могут быть сделаны в следующей форме.

Выводы. Предположим, что в результате опытов с $n = 27$ испытуемыми получена (2×2)-таблица сопряжённости, в которой

$$n(\mathsf{A}) = 14, \quad n(\mathsf{B}) = 13, \qquad n(\overline{\mathsf{A}}, \mathsf{B}) = k, \quad \text{где} \quad k = 2, 3, 4 \quad \text{или} \quad k = 10, 9, 8.$$

При $k = 2, 3, 4$ ($k = 10, 9, 8$) принимается опирающееся на здравый смысл решение о наличии прямой (обратной) связи между признаками. Тогда количественной оценкой *ошибки* решения здравого смысла о наличии прямой (обратной) связи между признаками является число α_k^-, (α_k^+), которое можно интерпретировать как вероятность *напрасного отказа* от гипотезы о независимости признаков.

§ 10. Случайные величины и распределения вероятностей

§ 10.1. Дискретная модель

Рассмотрим случайный опыт, вероятностная модель которого задается пространством элементарных событий $\Omega = \{\omega\}$ и набором чисел

$$\Pr\{\omega\}, \quad 0 \le \Pr\{\omega\} \le 1, \quad \sum_{\omega} \Pr\{\omega\} = 1,$$

являющихся вероятностями этих элементарных событий.

Определение 1. *Случайной величиной* называется числовая функция $\xi = \xi(\omega)$, определённая на элементарных событиях ω пространства $\Omega = \{\omega\}$.

Пример. (Случайные величины в модели (n, p)-испытаний Бернулли.) Пусть двоичные (из 1 и 0) последовательности $\omega = (x_1, x_2, \ldots, x_n)$ являются элементарными событиями для модели (n, p)-испытаний Бернулли. Этот случайный опыт представляет собой последовательность из n независимых испытаний с двумя возможными исходами в i-том, $i = 1, 2, \ldots, n$, испытании: успехом или неудачей. Успех в каждом испытании происходит с одной и той же вероятностью p, $0 < p < 1$, и обозначается символом 1, а неудача происходит с вероятностью $1-p$ и обозначается символом 0. Согласно определению независимости, вероятность элементарного события $\omega = (x_1, x_2, \ldots, x_n)$ задаётся как произведение вероятностей соответствующих исходов, относящихся к разным испытаниям:

$$\Pr\{\omega\} = \Pr\{(x_1, x_2, \ldots, x_n)\} = \Pr\{x_1\} \cdot \Pr\{x_2\} \cdot \ldots \cdot \Pr\{x_n\},$$

где $\Pr\{1\} = p$, а $\Pr\{0\} = 1 - p$. Введём двоичные случайные величины $\xi_i = \xi_i(\omega)$, каждая из которых может принять одно из двух значений 1 или 0:

$$\xi_i = \xi_i(\omega) = \begin{cases} 1, & \text{если в } i\text{-ом испытании был успех,} \\ 0, & \text{если в } i\text{-ом испытании была неудача,} \end{cases} \quad i = 1, 2, \ldots, n$$

а также случайную величину $S_n = S_n(\omega)$, называемую *числом успехов в n испытаниях*, которая может принять одно из $n + 1$ целочисленных значений $\{0, 1, \ldots, n\}$:

$$S_n = S_n(\omega) = k, \text{ если в } n \text{ испытаниях произошло } k \text{ успехов, } k = 0, 1, \ldots, n.$$

Для частного случая $n = 3$ эти определения иллюстрируется таблицей:

$\omega = (x_1, x_2, x_3)$	$\Pr\{\omega\}$	$\xi_1 = \xi_1(\omega)$	$\xi_2 = \xi_2(\omega)$	$\xi_3 = \xi_3(\omega)$	$S_3 = S_3(\omega)$
$\omega = (0, 0, 0)$	$(1-p)^3$	0	0	0	0
$\omega = (0, 0, 1)$	$p(1-p)^2$	0	0	1	1
$\omega = (0, 1, 0)$	$p(1-p)^2$	0	1	0	1
$\omega = (0, 1, 1)$	$p^2(1-p)$	0	1	1	2
$\omega = (1, 0, 0)$	$p(1-p)^2$	1	0	0	1
$\omega = (1, 0, 1)$	$p^2(1-p)$	1	0	1	2
$\omega = (1, 1, 0)$	$p^2(1-p)$	1	1	0	2
$\omega = (1, 1, 1)$	p^3	1	1	1	3

,

Формулы для записи случайных величин как функций от $\omega = (x_1, x_2, \ldots, x_n)$ имеют вид:

$$\xi_i = \xi_i(\omega) = \xi_i((x_1, x_2, \ldots, x_n)) = x_i, \quad i = 1, 2, \ldots, n,$$

$$S_n \;=\; S_n(\omega) = S_n((x_1, x_2, \ldots, x_n)) = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \xi_i.$$

Обозначим различные значения случайной величины ξ через $a_1, a_2, \ldots, a_K$, $K = 1, 2, \ldots$. Отметим, что вырожденный случай $K = 1$ означает, что $\xi = c$, где $c = a_1$, является *постоянной* (*неслучайной*) величиной. Значения $a_1, a_2, \ldots, a_K$ могут быть любыми числами, как положительными, так и отрицательными. При $K \geq 2$ для определённости, не нарушая общности, будем считать, что $a_1 < a_2 < \cdots < a_K$. Пусть q_k, $k = 1, 2, \ldots, K$, есть вероятность события $\{\xi = a_k\}$, т.е.

$$q_k = \Pr\{\xi = a_k\}, \quad \text{где} \quad \sum_{k=1}^{K} q_k \;=\; 1.$$

Определение 2. *Распределением вероятностей* дискретной случайной величины ξ называется функция $q(x)$ действительного аргумента x, $-\infty \leq x \leq \infty$, вида:

$$q(x) \stackrel{\text{def}}{=} \begin{cases} q_k, & \text{если } x = a_k, \quad k = 1, 2, \ldots, K, \\ 0, & \text{для остальных значений } x, \; -\infty \leq x \leq \infty, \end{cases}$$

Такую функцию $q(x)$ назовем *вероятностной*.

Мы будем использовать следующую краткую запись данного определения:

$$\xi \;\sim\; q(x) = \begin{cases} q_k, & \text{если } x = a_k, \quad k = 1, 2, \ldots, K, \\ 0, & \text{для остальных значений } x. \end{cases} \tag{1}$$

Отметим, что для неслучайной (детерминированной) величины $\xi = c$ такая запись имеет вид

$$c \;\sim\; q(x) = \begin{cases} 1, & \text{если } x = c, \\ 0, & \text{для остальных значений } x. \end{cases}$$

Для любого x, $-\infty \leq x \leq \infty$, очевидно, что вероятностная функция

$$q(x) = \Pr\{\xi = x\} \geq 0 \quad \text{и} \quad \sum_{x} q(x) \;=\; \sum_{k=1}^{K} q_k \;=\; 1. \tag{2}$$

Кроме того, для любых действительных чисел $-\infty \leq a < b \leq \infty$ вероятность события

$$\Pr\{a \leq \xi \leq b\} \;=\; \sum_{x\,:\,a \leq x \leq b} q(x). \tag{3}$$

Запись под знаком $\sum$ означает, что суммирование берётся по всем действительным числам x таким, что $a \leq x \leq b$.

Определение 3. Функция $F(t)$ действительного аргумента t, $-\infty \leq t \leq \infty$, называется *функцией распределения* случайной величины ξ (или, кратко, $\xi \;\sim\; F(t)$), если

$$F(t) \stackrel{\text{def}}{=} \Pr\{\xi < t\} = \sum_{x\,:\,x<t} q(x) = \begin{cases} 0, & \text{если} \quad t \le a_1, \\ q_1, & \text{если} \quad a_1 < t \le a_2, \\ q_1 + q_2, & \text{если} \quad a_2 < t \le a_3, \\ \dots & \dots \quad \dots \\ \sum_{i=1}^{k} q_i, & \text{если} \quad a_k < t \le a_{k+1}, \quad k = 1, 2, \dots, K-1, \\ \dots & \dots \quad \dots \\ \sum_{i=1}^{K-1} q_i, & \text{если} \quad a_{K-1} < t \le a_K, \\ 1, & \text{если} \quad t > a_K. \end{cases} \tag{4}$$

Функция распределения $F(t)$ есть ступенчатая функция, меняющая свои значения в точках, совпадающих со значениями $a_1 < a_2 < \ldots < a_K$ случайной величины ξ.

В реальных научных опытах целью экспериментатора является получение информации о *неизвестном* [1] распределении вероятностей (1) случайной величины (признака) ξ на основании результатов опытов с n однородными испытуемыми, для каждого из которых он регистрирует одно из $\{a_1, a_2, \dots, a_K\}$ возможных значений ξ.

Пусть $n(a_k)$, – зарегистрированное в n опытах число испытуемых, у которых признак $\xi = a_k$. При этом, очевидно, $0 \le n(a_k) \le n$ и $\sum_{k=1}^{K} n(a_k) = n$. Если $K \ll n$, то в принципе можно пользоваться частотным вычислением вероятности

$$q_k \;=\; \Pr\{\xi = a_k\} \;\approx\; \frac{n(a_k)}{n}, \qquad k = 1, 2, \dots, K.$$

Для многих других практически интересных ситуаций (например, когда $K > n$) такое вычисление невозможно. Желательно поэтому охарактеризовать распределении вероятностей (1) несколькими числовыми параметрами, которые можно было бы *оценивать экспериментально*. Сейчас мы займёмся формальным введением двух важнейших параметров, свойства которых будут исследованы в дальнейшем.

Определение 4. Числа

$$\mathbf{M}\xi \stackrel{\text{def}}{=} \sum_{x} x \cdot q(x) \;=\; \sum_{k=1}^{K} a_k \cdot q_k, \quad \mathbf{D}\xi \stackrel{\text{def}}{=} \sum_{x} (x - \mathbf{M}\xi)^2 \cdot q(x) \;=\; \sum_{k=1}^{K} (a_k - \mathbf{M}\xi)^2 \cdot q_k \tag{5}$$

называются соответственно *математическим ожиданием* (*средним значением*) и *дисперсией* (средним значением *квадрата отклонений* от математического ожидания) случайной величины ξ с распределением вероятностей (1).

Очевидно, что дисперсия $\mathbf{D}\xi \ge 0$. Число $\sigma_\xi \stackrel{\text{def}}{=} \sqrt{\mathbf{D}\xi}$, размерность которого совпадает с размерностью ξ, называется *среднеквадратичным отклонением* случайной величины ξ.

Замечание. Если числа $a_1 < a_2 < \cdots < a_K$ являются возможными значениями случайной величины ξ, то интерпретация математического ожидания $\mathbf{M}\xi$ как среднего значения ξ связана с очевидными неравенствами $a_1 \le \mathbf{M}\xi \le a_K$.

[1] Например, обычно экспериментатор знает множество $\{a_1, a_2, \dots, a_K\}$ возможных значений ξ, но не знает вероятности $q_k = \Pr\{\xi = a_k\}$, $k = 1, 2, \dots, K$, с которыми признак ξ принимает эти значения.

Пример. (Модель (n,p)-испытаний Бернулли.) *Различные* случайные величины ξ_i, $i = 1,2,\ldots,n$, которые в этой модели регистрируют успехи - неудачи в n независимых испытыниях, очевидно, имеют *одно и то же* распределение вероятностей:

$$\xi_i \sim q(x) = \begin{cases} 1-p, & \text{если } x = 0, \\ p, & \text{если } x = 1, \\ 0, & \text{для остальных значений } x,\ -\infty \le x \le \infty, \end{cases}$$

или

$$\xi_i \sim F(t) = \Pr\{\xi_i < t\} = \begin{cases} 0, & \text{если } t \le 0, \\ 1-p, & \text{если } 0 < t \le 1, \\ 1, & \text{если } t > 1. \end{cases}$$

График функции распределения $F(t)$ является монотонно неубывающей ступенчатой функцией, которая меняет свои значения при $t = 0$ и при $t = 1$. Математическое ожидание

$$\mathbf{M}\xi_i = \sum_{k=1}^{K} a_k \cdot q_k = 0 \cdot (1-p) + 1 \cdot p = p,$$

а дисперсия

$$\mathbf{D}\xi_i = \sum_{k=1}^{K} (a_k - \mathbf{M}\xi_i)^2 \cdot q_k = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p)[p + (1-p)] = p(1-p).$$

Число успехов в n испытаниях

$$S_n \sim q(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{если } x = 0,1,\ldots,n, \\ 0, & \text{для остальных значений } x,\ -\infty \le x \le \infty. \end{cases}$$

Можно также написать, что $S_n \sim F(t)$, где функция распределения

$$F(t) = \Pr\{S_n < t\} = \begin{cases} 0, & \text{если } t \le 0, \\ \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}, & \text{если } k < t \le k+1, \quad k = 0,1,\ldots,n-1, \\ 1, & \text{если } t > n. \end{cases}$$

Для любых целых чисел $0 \le a \le b \le n$ вероятность события

$$\Pr\{a \le S_n \le b\} = \sum_{i=a}^{b} \binom{n}{i} p^i (1-p)^{n-i}.$$

Справедливы следующие формулы для вычисления математического ожидания и дисперсии числа успехов в модели (n,p)-испытаний Бернулли:

$$\mathbf{M}S_n = \sum_{i=0}^{n} i \cdot \binom{n}{i} p^i (1-p)^{n-i} = np,$$

$$\mathbf{D}S_n = \sum_{i=0}^{n} (i-np)^2 \cdot \binom{n}{i} p^i (1-p)^{n-i} = np(1-p).$$

§ 10.2. Непрерывная модель

Формально для приложений можно ограничиться изучением вышеописанной дискретной модели распределения вероятностей случайной величины ξ, принимающей лишь конечное число N значений. Это объясняется тем, что любой реальный прибор имеет ограниченную точность, т.е. число знаков, с которыми наблюдаются (измеряются) результаты случайного эксперимента ξ всегда ограничено и, следовательно, число возможных исходов N реального опыта всегда конечно. Тем не менее, для проведения различных расчётов в ситуации, когда $N \to \infty$, оказывается очень полезным вместо дискретного распределения рассматривать описываемый ниже его непрерывный аналог, который можно интерпретировать как математическую модель случайного опыта ξ в указанном предельном случае.

Пусть $p(x) \geq 0$ – произвольная неотрицательная функция аргумента x, $-\infty < x < \infty$, для которой сходящийся несобственный интеграл

$$\int_{-\infty}^{\infty} p(x)\,dx \;=\; 1, \quad p(x) \geq 0. \tag{2'}$$

Определение 2'. Пусть $-\infty \leq a \leq b \leq \infty$ – произвольные действительные числа. Функция $p(x)$ называется *плотностью распределения* случайной величины ξ (или, кратко, $\xi \sim p(x)$), если вероятность события $\{a \leq \xi \leq b\}$, обозначаемая символом $\Pr\{a \leq \xi \leq b\}$, задается как определенный интеграл от функции $p(x)$ по промежутку $[a; b]$:

$$\Pr\{a \leq \xi \leq b\} \;\stackrel{\text{def}}{=}\; \int_a^b p(t)\,dt. \tag{3'}$$

Плотность распределения $p(x)$ является аналогом вероятностной функции $q(x)$ в определении дискретного распределения.

Как и в дискретном случаем, даются следующие определения.

Определение 3'. Функция $F(x)$ действительного аргумента x, $-\infty < x < \infty$, определяемая формулой

$$F(x) \;\stackrel{\text{def}}{=}\; \Pr\{\xi < x\} \;=\; \Pr\{-\infty < \xi < x\} \;=\; \int_{-\infty}^{x} p(t)\,dt, \quad -\infty < x < \infty, \tag{4'}$$

называется *функцией распределения* случайной величины ξ (или, кратко, $\xi \sim F(x)$).

Определение 4'. Числа

$$\mathbf{M}\xi \;\stackrel{\text{def}}{=}\; \int_{-\infty}^{\infty} x \cdot p(x)\,dx, \quad \mathbf{D}\xi \;\stackrel{\text{def}}{=}\; \int_{-\infty}^{\infty} (x - \mathbf{M}\xi)^2 \cdot p(x)\,dx, \tag{5'}$$

называются соответственно *математическим ожиданием* (*средним значением*) и *дисперсией* (средним значением *квадрата отклонений* от математического ожидания) случайной величины ξ с плотностью распределения $p(x)$.

Очевидно, что дисперсия $\mathbf{D}\xi \geq 0$. Число $\sigma_\xi = \sqrt{\mathbf{D}\xi}$, размерность которого совпадает с размерностью ξ, называется *среднеквадратичным отклонением* случайной величины ξ.

Применяя известную из курса математического анализа основную теорему интегрального исчисления (теорему Ньютона - Лейбница) о производной определённого интеграла по переменному верхнему пределу, получаем

Свойство 1. *Если плотность распределения $p(x)$ есть непрерывная функция на всей числовой оси $-\infty < x < \infty$, то для функции распредения $F(x)$ имеют место следующие утверждения.*

1) *Производная $F'(x) = p(x)$ для любого x, $-\infty < x < \infty$, т.е. производная функции распределения равна плотности этого распределения.*

2) *Если $a < b$, то* $\Pr\{a \le \xi \le b\} = \int_a^b p(t)\,dt = F(b) - F(a)$.

Пример. (Распределение вероятностей $\mathcal{N}(a,\sigma)$). Согласно определению, непрерывная случайная величина ξ имеет *стандартное нормальное распределение вероятностей* $\mathcal{N}(0,1)$ (или, кратко, $\xi \sim \mathcal{N}(0,1)$), если её плотность распределения $p(x) = g(x)$, а функция распределения $F(x) = G(x)$, где функции $g(x)$ и $G(x)$ задаются соотношениями

$$g(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = G'(x), \quad G(x) \stackrel{\text{def}}{=} \int_{-\infty}^{x} g(t)\,dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt.$$

Отметим, что $g(x)$ –четная положительная функция, а функция $G(x)$ монотонно возрастает от 0 до 1, причем

$$G(-\infty) = 0, \quad G(0) = 1/2, \quad G(\infty) = 1.$$

Пусть a и $\sigma > 0$ – произвольные фиксированные числа. Согласно определению, непрерывная случайная величина η имеет *нормальное распределение с параметрами a и σ* (или, кратко, $\eta \sim \mathcal{N}(a,\sigma)$), если её плотность распределения $p(x) = f_\eta(x)$, а функция распределения $F(x) = F_\eta(x)$, где функции $f_\eta(x)$ и $F_\eta(x)$ задаются соотношениями

$$f_\eta(x) \stackrel{\text{def}}{=} \sigma^{-1} g\left(\frac{x-a}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/2\sigma^2}, \quad F_\eta(x) \stackrel{\text{def}}{=} G\left(\frac{x-a}{\sigma}\right) = \int_{-\infty}^{\frac{x-a}{\sigma}} g(t)\,dt,$$

которые равносильны тому, что

$$\eta \stackrel{\text{def}}{=} a + \sigma \cdot \xi \quad \text{или} \quad \mathcal{N}(a,\sigma) \stackrel{\text{def}}{=} a + \sigma \cdot \mathcal{N}(0,1).$$

Для математического ожидания $\mathbf{M}\eta$ и дисперсии $\mathbf{D}\eta$ справедливы равенства

$$\mathbf{M}\eta = \int_{-\infty}^{\infty} t f_\eta(t)\,dt = a, \qquad \mathbf{D}\eta = \int_{-\infty}^{\infty} (t-a)^2 f_\eta(t)\,dt = \sigma^2.$$

Отметим, что параметр $\sigma = \sqrt{\mathbf{D}\eta}$ является среднеквадратичным отклонением распределения $\mathcal{N}(a,\sigma)$. Пусть $-\infty \le A < B \le \infty$ – произвольные заданные числа. В силу свойства 1, вероятность попадания наблюдения $\eta \sim \mathcal{N}(a,\sigma)$ в заданный промежуток $[A;B]$ вычисляется по формуле

$$\Pr\{A \le \eta \le B\} = \int_A^B f_\eta(t)\,dt = \int_{(A-a)/\sigma}^{(B-a)/\sigma} g(t)\,dt = G\left(\frac{B-a}{\sigma}\right) - G\left(\frac{A-a}{\sigma}\right).$$

Данная формула сводит вычисление вероятностей для $\eta \sim \mathcal{N}(a, \sigma)$ к вычислению по таблицам для стандартного нормального распределения $\xi \sim \mathcal{N}(0, 1)$. Эти таблицы называются таблицами *интеграла вероятностей.*

Методы математической статистики, в которых при обработке результатов наблюдений используется предположение об их нормальном распределении $\mathcal{N}(a, \sigma)$ с неизвестными параметрами a и σ, называются *параметрическими* методами.

Более общие методы математической статистики, в которых не используется предположение об $\mathcal{N}(a, \sigma)$ для обрабатываемых наблюдений, называются *непараметрическими* или *ранговыми* методами. Для обоснования применения ранговых методов важное значение имеет следующее *общее свойство* функции распределения $F(x)$.

Свойство 2. *Пусть случайная величина ξ имеет функцию распредения $F(x)$, которая является непрерывной и монотонно возрастающей функцией при всех x, $-\infty < x < \infty$, таких что $0 < F(x) < 1$. Тогда случайная величина $\eta \stackrel{def}{=} F(\xi)$ удовлетворяет неравенствам $0 < \eta < 1$ и имеет функцию распределения*

$$U(t) = \Pr\{\eta < t\} = \begin{cases} 0, & \text{если } t \le 0, \\ t, & \text{если } 0 < t < 1, \\ 1, & \text{если } t \ge 1. \end{cases}$$

которая не зависит от исходной функции $F(x)$.

Данная стандартная функция $U(t)$ называется *равномерной* функцией распределения на интервале $(0\,;\,1)$. Ей соответствует *плотность равномерного распределения* на интервале $(0\,;\,1)$:

$$u(t) = U'(t) = \begin{cases} 1, & \text{если } 0 < t < 1, \\ 0, & \text{если } t \le 0 \text{ или } t \ge 1, \end{cases}$$

Отметим, что если случайная величина (наблюдение) η имеет равномерное распределение на интервале $(0\,;\,1)$, то для любых чисел $0 \le a < b \le 1$ вероятность попадания наблюдения в интервал $(a\,;\,b)$ равна длине этого интервала $b - a$, т.е. $\Pr\{a < \eta < b\} = b - a$.

Доказательство свойства 2. Поскольку $t = F(x)$ есть непрерывная и монотонно возрастающая от $t = 0$ до $t = 1$ функция аргумента x, $-\infty < x < \infty$, то согласно теореме математического анализа об обратной функции, для функции $t = F(x)$ существует непрерывная и монотонно возрастающая от $x = -\infty$ до $x = \infty$ обратная функция $x = F^{-1}(t)$, аргумента t, $0 < t < 1$. При этом, в силу строгой монотонности и непрерывности прямой и обратной функций,

$$\text{для любого} \quad t,\ 0 < t < 1, \quad \text{множество} \quad \{x\ :\ F(x) < t\} = \{x\ :\ x < F^{-1}(t)\}$$

и, кроме того,

$$\text{для любого} \quad t,\ 0 < t < 1, \quad \text{справедливо тождество} \quad F\left(F^{-1}(t)\right) = t.$$

Следовательно, для любого t, $0 < t < 1$, функция распределения

$$U(t) = \Pr\{\eta < t\} = \Pr\{F(\xi) < t\} = \Pr\{\xi < F^{-1}(t)\} = F\left(F^{-1}(t)\right) = t.$$

Свойство 2 доказано.

§ 11. Совместное распределение случайных величин, независимость случайных величин

§ 11.1. Дискретная модель

Пусть $\{a_1, a_2, \dots, a_K\}$ – набор различных значений случайной величины $\xi = \xi(\omega)$ и индекс $k = 1, 2, \dots K$. Пусть $\{b_1, b_2, \dots, b_J\}$ – набор различных значений случайной величины $\eta = \eta(\omega)$, и индекс $j = 1, 2, \dots J$.

Определение 1. Произвольный набор (матрица) $\{q_{kj}\} = \|q_{kj}\|$ из $K \cdot J$ чисел, удовлетворяющих условиям

$$q_{kj} \ \geq 0, \qquad \sum_{k=1}^{K} \sum_{j=1}^{J} q_{kj} \ = \ 1 \tag{1}$$

и задающих вероятности вида

$$q_{kj} \stackrel{\text{def}}{=} \Pr\{\xi = a_k, \ \eta = b_j\}, \qquad k = 1, 2, \dots K, \quad j = 1, 2, \dots J, \tag{2}$$

называется *совместным распределением пары случайных величин* $(\xi, \ \eta)$.

Очевидно, что события

$$\{\xi = a_k, \ \eta = b_j\} \qquad k = 1, 2, \dots K, \quad j = 1, 2, \dots J,$$

несовместимы, т.е. попарно непересекаются. Кроме того, событие

$$\{\xi = a_k\} = \sum_{j=1}^{J} \{\xi = a_k, \ \eta = b_j\}, \quad k = 1, 2, \dots K,$$

а событие

$$\{\eta = b_j\} = \sum_{k=1}^{K} \{\xi = a_k, \ \eta = b_j\}, \quad j = 1, 2, \dots J.$$

Отсюда вытекает важное

Свойство 1. *Справедливы формулы*

$$\Pr\{\xi = a_k\} = \sum_{j=1}^{J} \Pr\{\xi = a_k, \ \eta = b_j\} = \sum_{j=1}^{J} q_{kj} \stackrel{def}{=} q_{k\cdot} \quad k = 1, 2, \dots K,$$

$$\Pr\{\eta = b_j\} = \sum_{k=1}^{K} \Pr\{\xi = a_k, \ \eta = b_j\} = \sum_{k=1}^{K} q_{kj} \stackrel{def}{=} q_{\cdot j} \quad j = 1, 2, \dots J$$

с помощью которых по совместному распределению $\{q_{kj}\}$ *пары случайных величин* $(\xi, \ \eta)$ *вычисляются распределения самих случайных величин* ξ *и* η.

Сопоставим совместному распределению $\{q_{kj}\} = \|q_{kj}\|$ функцию двух переменных x, $-\infty \leq x \leq \infty$, и y, $-\infty \leq y \leq \infty$:

$$q(x, y) \stackrel{\text{def}}{=} \begin{cases} q_{kj}, & \text{если } x = a_k, \text{ а } y = b_j, \\ 0, & \text{в остальных случаях.} \end{cases}$$

С помощью этой функции совместное распределение пары (ξ, η) и получаемые по формулам свойства 1 распределения самих случайных величин ξ и η можно записать в виде

$$\Pr\{\xi = x,\ \eta = y\} = q(x, y),$$

$$\Pr\{\xi = x\} = \sum_y q(x, y) \stackrel{\text{def}}{=} q(x, .), \quad \Pr\{\eta = y\} = \sum_x q(x, y) \stackrel{\text{def}}{=} q(., y). \tag{3}$$

Определение 2. Случайные величины ξ и η называются *независимыми*, если для любой пары чисел $(x,\ y)$ события $\{\xi = x\}$ и $\{\eta = y\}$ независимы между собой, т.е.

$$q(x, y) = \Pr\{\xi = x,\ \eta = y\} \stackrel{\text{def}}{=} \Pr\{\xi = x\} \cdot \Pr\{\eta = y\} = q(x, .) \cdot q(., y). \tag{4}$$

Рассмотрим вопрос о распределении случайной величины $\xi + \eta$, т.е. задачу вычисления *распределения суммы случайных величин*. Как для произвольного числа x, $-\infty \le x \le \infty$, вычислить вероятность $\Pr\{\xi + \eta = x\}$?

Свойство 2. *Для распределения суммы случайных величин ξ и η справедливы формулы*

$$\Pr\{\xi + \eta = x\} = \sum_y \Pr\{\xi = x - y,\ \eta = y\} = \sum_y q(x - y, y), \quad -\infty \le x \le \infty, \tag{5}$$

которые для частного случая независимых слагаемых имеют вид

$$\Pr\{\xi + \eta = x\} = \sum_y \Pr\{\xi = x - y\} \cdot \Pr\{\eta = y\} = \sum_y q(x - y, .) \cdot q(., y), \quad -\infty \le x \le \infty. \tag{6}$$

§ 11.2. Непрерывная модель

Пусть $p(x, y) \ge 0$ – произвольная неотрицательная функция двух действительных аргументов x, $-\infty < x < \infty$, и y, $-\infty < y < \infty$, для которой сходящийся несобственный интеграл

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y)\, dx\, dy \ = \ 1, \quad p(x, y) \ge 0, \tag{1'}$$

Определение 1'. Пусть $-\infty \le a \le b \le \infty$ и $-\infty \le c \le d \le \infty$ – произвольные действительные числа. Функция $p(x, y)$ называется *плотностью совместного распределения* пары случайных величин $(\xi,\ \eta)$, если вероятность события $\{a \le \xi \le b,\ c \le \eta \le d\}$ задается как определённый интеграл от функции $p(x, y)$ вида:

$$\Pr\{a \le \xi \le b,\ c \le \eta \le d\} \stackrel{\text{def}}{=} \int_a^b \int_c^d p(x, y)\, dx\, dy. \tag{2'}$$

Формулы для вычисление плотностей распределений вероятностей величин ξ и η по плотности их совместного распределения даёт

Свойство 1'. *Если пара случайных величин* $(\xi,\ \eta)$ *имеет плотность совместного распределения* $p(x,y)$*, то функции*

$$p(x,\ .)\ \stackrel{def}{=}\ \int_{-\infty}^{\infty} p(x,y)\,dy\ \stackrel{def}{=}\ p_\xi(x) \quad и \quad p(.,\ y)\ \stackrel{def}{=}\ \int_{-\infty}^{\infty} p(x,y)\,dx\ \stackrel{def}{=}\ p_\eta(y) \qquad (3')$$

являются соответственно плотностями распределения величин ξ *и* η*, т.е. для любых чисел* $-\infty \le a \le b \le \infty$ *и любых чисел* $-\infty \le c \le d \le \infty$

$$\Pr\{a \le \xi \le b\} = \int_a^b p(x,.)\,dx \quad и \quad \Pr\{c \le \eta \le d\} = \int_c^d p(.,y)\,dy.$$

Определение независимости случайных величин в общем случае формулируется следующим образом.

Определение 2'. Случайные величины ξ и η называются *независимыми*, если для любых чисел $-\infty \le a \le b \le \infty$ и любых чисел $-\infty \le c \le d \le \infty$ события $\{a \le \xi \le b\}$ и $\{c \le \eta \le d\}$ независимы между собой, т.е. для вероятности события $\{a \le \xi \le b,\ c \le \eta \le d\}$ имеет место соотношение

$$\Pr\{a \le \xi \le b,\ c \le \eta \le d\}\ \stackrel{\text{def}}{=}\ \Pr\{a \le \xi \le b\} \cdot \Pr\{c \le \eta \le d\}.$$

Можно доказать, что справедливо

Свойство 3. *Случайные величины* ξ *и* η *с плотностью совместного распределения* $p(x,y)$ *являются независимыми тогда и только тогда, когда для любой пары чисел* $(x,\ y)$ *плотность* $p(x,y)$ *равна произведению плотностей распределения случайных величин* ξ *и* η*, т.е.*

$$p(x,y)\ =\ p(x,.) \cdot p(.,y), \qquad -\infty < x < \infty, \quad -\infty < y < \infty. \qquad (4')$$

Для непрерывной модели аналогом свойства 2 является

Свойство 2'. *Если пара случайных величин* $(\xi,\ \eta)$ *имеет плотность совместного распределения* $p(x,y)$*, то функция* $p_{\xi+\eta}(x)$ *действительного аргумента* x*,* $-\infty \le x \le \infty$*, определяемая формулой*

$$p_{\xi+\eta}(x)\ \stackrel{def}{=}\ \int_{-\infty}^{\infty} p(x-y,y)\,dy, \quad -\infty \le x \le \infty, \qquad (5')$$

является плотностью распределения суммы случайных величин ξ *и* η*, т.е. для любых чисел* $-\infty \le a \le b \le \infty$ *вероятность события*

$$\Pr\{a \le \xi + \eta \le b\} = \int_a^b p_{\xi+\eta}(x)\,dx = \int_a^b \left\{\int_{-\infty}^{\infty} p(x-y,y)\,dy\right\} dx.$$

Для частного случая независимых слагаемых ξ *и* η*, имеющих соответственно плотности распределения* $p_\xi(x)$ *и* $p_\eta(x)$*, плотность распределения суммы* $\xi + \eta$ *имеет вид*

$$p_{\xi+\eta}(x)\ =\ \int_{-\infty}^{\infty} p(x-y,.) \cdot p(.,y)\,dy\ =\ \int_{-\infty}^{\infty} p_\xi(x-y) \cdot p_\eta(y)\,dy, \quad -\infty \le x \le \infty. \qquad (6')$$

Пример. (Суммирование независимых нормальных величин). Применим свойство 2' для вычисления плотности распределения суммы независимых нормальных случайных величин

$$\xi \sim \mathcal{N}(a_1, \sigma_1),\ \mathbf{M}\xi = a_1,\ \mathbf{D}\xi = \sigma_1^2 \quad \text{и} \quad \eta \sim \mathcal{N}(a_2, \sigma_2),\ \mathbf{M}\eta = a_2,\ \mathbf{D}\eta = \sigma_2^2.$$

Согласно данному выше определению, плотности рассматриваемых нормальных распределений имеют вид

$$p_\xi(x-y) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left\{\frac{-[(x-y)-a_1]^2}{2\sigma_1^2}\right\}, \quad p_\eta(y) = \frac{1}{\sigma_2\sqrt{2\pi}} \exp\left\{\frac{-[y-a_2]^2}{2\sigma_2^2}\right\},$$

где для удобства записи используется стандартное обозначение $\exp\{z\} = e^z$. Можно показать, что если эти формулы подставить в правую часть $(6')$ и вычислить полученный определённый интеграл, то функция $p_{\xi+\eta}(x)$ примет вид:

$$p_{\xi+\eta}(x) = \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}\cdot\sqrt{2\pi}} \exp\left\{\frac{-[x-(a_1+a_2)]^2}{2(\sigma_1^2+\sigma_2^2)}\right\}.$$

Это означает, что имеет место следующее важное свойство нормальных случайных величин.

Свойство 4. *Сумма $\xi+\eta$ независимых нормальных случайных величин $\xi \sim \mathcal{N}(a_1, \sigma_1)$ и $\eta \sim \mathcal{N}(a_2, \sigma_2)$ является нормальной случайной величиной:*

$$\xi + \eta \ \sim\ \mathcal{N}(a_1, \sigma_1) + \mathcal{N}(a_2, \sigma_2) \ \sim\ \mathcal{N}\left(a_1 + a_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right),$$

математическое ожидание и дисперсия которой вычисляются по формулам

$$\mathbf{M}(\xi+\eta) = a_1 + a_2 = \mathbf{M}\xi + \mathbf{M}\eta, \quad \mathbf{D}(\xi+\eta) = \sigma_1^2 + \sigma_2^2 = \mathbf{D}\xi + \mathbf{D}\eta.$$

§ 11.3. Совместное распределение n случайных величин $(\xi_1, \xi_2, \ldots, \xi_n)$

Совместное распределение n случайных величин $(\xi_1, \xi_2, \ldots, \xi_n)$ задается *вероятностной функцией от n переменных.*

- В случае дискретной модели эта функция $\mathsf{q} = \mathsf{q}(x_1, x_2, \ldots, x_n)$, называемая *совместным распределением вероятностей*, удовлетворяет условиям

$$\mathsf{q}(x_1, x_2, \ldots, x_n) \ \geq\ 0, \qquad \sum_{i=1}^{n} \sum_{x_i} \mathsf{q}(x_1, x_2, \ldots, x_n) \ =\ 1$$

  и определяет вероятности событий

$$\Pr\{\xi_1 = x_1, \xi_2 = x_2, \ldots, \xi_n = x_n\} \stackrel{\text{def}}{=} \mathsf{q}(x_1, x_2, \ldots, x_n).$$

- В случае непрерывной модели эта функция $\mathsf{p} = \mathsf{p}(x_1, x_2, \ldots, x_n)$, называемая *плотностью совместного распределения вероятностей*, удовлетворяет условиям

$$\mathsf{p}(x_1, x_2, \ldots, x_n) \geq 0, \qquad \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \ldots \int\limits_{-\infty}^{\infty} \mathsf{p}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n \;=\; 1$$

и определяет вероятности событий

$$\Pr\{a_1 \leq \xi_1 \leq b_1;\; a_2 \leq \xi_2 \leq b_2;\; \ldots;\; a_n \leq \xi_n \leq b_n\} \stackrel{\text{def}}{=}$$

$$\stackrel{\text{def}}{=} \int\limits_{a_1}^{b_1} \int\limits_{a_2}^{b_2} \ldots \int\limits_{a_n}^{b_n} \mathsf{p}(x_1, x_2, \ldots, x_n)\, dx_1\, dx_2 \cdots dx_n.$$

При определении независимости n случайных величин $(\xi_1, \xi_2, \ldots, \xi_n)$ мы ограничимся наиболее важным для приложений случаем n независимых одинаково распределённых (однородных) случайных величин. В математической статистике эта модель наблюдений называется *выборкой объёма* n. Рассмотрим произвольную вероятностную функцию аргумента x, $-\infty < x < \infty$, т.е. распределение вероятностей $q(x)$ (дискретная модель) или плотность распределения вероятностей $p(x)$ (непрерывная модель).

Определение 3. Будем говорить, что случайные величины $(\xi_1, \xi_2, \ldots, \xi_n)$ являются *независимыми одинаково распределёнными величинами* $\xi_i \sim q(x)$ (дискретная модель) или $\xi_i \sim p(x)$ (непрерывная модель), если совместная вероятность $\mathsf{q} = \mathsf{q}(x_1, x_2, \ldots, x_n)$ (дискретная модель) или совместная плотность $\mathsf{p} = \mathsf{p}(x_1, x_2, \ldots, x_n)$ (непрерывная модель) задаются в виде *произведения* n вероятностных функций:

$$\mathsf{q} = \mathsf{q}(x_1, x_2, \ldots, x_n) \stackrel{\text{def}}{=} q(x_1) \cdot q(x_2) \cdots \cdot q(x_n),$$

$$\mathsf{p} = \mathsf{p}(x_1, x_2, \ldots, x_n) \stackrel{\text{def}}{=} p(x_1) \cdot p(x_2) \cdots \cdot p(x_n).$$

Используя это определение и формулу $(6')$, можно доказать следующее обобщение свойства 4 о распределении суммы n независимых нормальных случайных величин.

Свойство 5. *Пусть* $(\xi_1, \xi_2, \ldots, \xi_n)$ *– независимые одинаково распределённые величины, где* $\xi_i \sim \mathcal{N}(a, \sigma)$ *для любого* $i = 1, 2, \ldots, n$. *Тогда их сумма* $S_n = \sum\limits_{i=1}^{n} \xi_i$ *также является нормальной случайной величиной*

$$S_n \;=\; \sum_{i=1}^{n} \xi_i \;\sim\; \mathcal{N}(na, \sigma\sqrt{n}),$$

математическое ожидание и дисперсия которой вычисляются по формулам

$$\mathbf{M}\, S_n \;=\; \mathbf{M} \left\{ \sum_{i=1}^{n} \xi_i \right\} \;=\; \sum_{i=1}^{n} \mathbf{M}\xi_i \;=\; na, \qquad \mathbf{D}\, S_n \;=\; \mathbf{D} \left\{ \sum_{i=1}^{n} \xi_i \right\} \;=\; \sum_{i=1}^{n} \mathbf{D}\xi_i \;=\; n\sigma^2.$$

§ 12. Числовые характеристики распределения вероятностей

§ 12.1. Математическое ожидание, дисперсия и их свойства

1. Пусть ξ – произвольная случайная величина, которая имеет распределение вероятностей, задаваемое вероятностной функцией $q(x)$ (дискретная модель) или плотностью распределения вероятностей $p(x)$ (непрерывная модель). Символом $f(x)$ обозначим произвольную функцию действительного аргумента x, $-\infty \le x \le \infty$, а символом c, $-\infty \le c \le \infty$, – произвольную постоянную.

Рассмотрим случайную величину $\eta = f(\xi)$, которую можно называть *функцией от случайной величины* ξ. Математическое ожидание функции от случайной величины ξ обозначается символом $\mathbf{M}f(\xi)$ и определяется формулой

$$\mathbf{M}f(\xi) \stackrel{\text{def}}{=} \begin{cases} \sum\limits_{x} f(x)q(x) & \text{для дискретной модели,} \\ \int\limits_{-\infty}^{\infty} f(x)p(x)\,dx & \text{для непрерывной модели.} \end{cases}$$

Если для любого x, $-\infty \le x \le \infty$, положим функцию $f(x) = c$, то очевидно

$$\mathbf{M}c \stackrel{\text{def}}{=} \begin{cases} \sum\limits_{x} c\,q(x) = c \cdot \sum\limits_{x} q(x) = c & \text{для дискретной модели,} \\ \int\limits_{-\infty}^{\infty} c\,p(x)\,dx = c \cdot \int\limits_{-\infty}^{\infty} p(x)\,dx = c & \text{для непрерывной модели.} \end{cases}$$

Если положить функцию $f(x) = (x - \mathbf{M}\xi)^2$, то дисперсия $\mathbf{D}\xi$ представляется как математическое ожидание этой функции от случайной величины ξ:

$$\mathbf{D}\xi \;=\; \begin{cases} \sum\limits_{x} (x - \mathbf{M}\xi)^2\, q(x) = \mathbf{M}(\xi - \mathbf{M}\xi)^2 & \text{для дискретной модели,} \\ \int\limits_{-\infty}^{\infty} (x - \mathbf{M}\xi)^2\, p(x)\,dx = \mathbf{M}(\xi - \mathbf{M}\xi)^2 & \text{для непрерывной модели.} \end{cases}$$

В частности, $\mathbf{D}c = \mathbf{M}(c - \mathbf{M}c)^2 = \mathbf{M}(c-c)^2 = \mathbf{M}\,0 = 0$. Следовательно, имеет место

Теорема 1.

1) *Математическое ожидание постоянной величины равно этой постоянной величине, т.е.* $\mathbf{M}c = c$.

2) *Дисперсия* $\mathbf{D}\xi \ge 0$, *где знак равенства, если и только если случайная величина* $\xi = c = \mathbf{M}\xi$ – *постоянная величина.*

Аналогичным образом можно проверить и другие важные формулы для математического ожидания и дисперсии, которые описывает

Теорема 2.

1) $\mathbf{M}(c \cdot \xi) = c \cdot \mathbf{M}\xi, \qquad \mathbf{M}(\xi + c) = \mathbf{M}\xi + c, \qquad \mathbf{M}(\xi - \mathbf{M}\xi) = 0;$

2) $\mathbf{D}(\xi + c) = \mathbf{D}\xi, \quad \mathbf{D}(c \cdot \xi) = c^2 \cdot \mathbf{D}\xi, \quad \mathbf{D}\xi \;=\; \mathbf{M}(\xi - \mathbf{M}\xi)^2 \;=\; \mathbf{M}(\xi^2) \;-\; (\mathbf{M}\xi)^2;$

3) *Если* $\mathbf{M}\xi = 0$, *то* $\mathbf{D}\xi \;=\; \mathbf{M}(\xi^2)$.

Доказательство.

1)
$$\mathbf{M}(c\cdot\xi)=\sum_x c\cdot x\,q(x)=c\cdot\sum_x x\,q(x)=c\cdot\mathbf{M}\xi,$$
$$\mathbf{M}(\xi+c)=\int\limits_{-\infty}^{\infty}(x+c)\,p(x)\,dx=\int\limits_{-\infty}^{\infty}x\,p(x)\,dx+\int\limits_{-\infty}^{\infty}c\,p(x)\,dx=\mathbf{M}\xi+c,$$
а последнее равенство $\mathbf{M}(\xi-\mathbf{M}\xi)=0$ является частным случаем предыдущего, если положить постоянную $c=-\mathbf{M}\xi$.

2)
$$\mathbf{D}(\xi+c)\;=\;\int\limits_{-\infty}^{\infty}\left[(x+c)-\mathbf{M}(\xi+c)\right]^2p(x)\,dx=\int\limits_{-\infty}^{\infty}(x-\mathbf{M}\xi)^2\,p(x)\,dx=\mathbf{D}\xi,$$
$$\mathbf{D}(c\cdot\xi)\;=\;\sum_x\left[c\cdot x-\mathbf{M}(c\cdot\xi)\right]^2q(x)=\sum_x c^2\left[x-\mathbf{M}(\xi)\right]^2q(x)=c^2\cdot\mathbf{D}\xi.$$
Для непрерывного случая, применяя известные свойства определённого интеграла, проверяем
$$\mathbf{D}\xi\;=\;\int\limits_{-\infty}^{\infty}(x-\mathbf{M}\xi)^2\,p(x)\,dx=\int\limits_{-\infty}^{\infty}\left[x^2-2x\mathbf{M}\xi+(\mathbf{M}\xi)^2\right]p(x)\,dx=$$
$$=\int\limits_{-\infty}^{\infty}x^2\,p(x)\,dx-2\mathbf{M}\xi\int\limits_{-\infty}^{\infty}x\,p(x)\,dx+(\mathbf{M}\xi)^2\int\limits_{-\infty}^{\infty}p(x)\,dx=$$
$$=\mathbf{M}(\xi^2)-2(\mathbf{M}\xi)^2+(\mathbf{M}\xi)^2=\mathbf{M}(\xi^2)-(\mathbf{M}\xi)^2.$$

3) Если $\mathbf{M}\xi=0$, то $\mathbf{D}\xi=\mathbf{M}(\xi^2)-(\mathbf{M}\xi)^2=\mathbf{M}(\xi^2)-0=\mathbf{M}(\xi^2)$.

Теорема 2 доказана.

Будем говорить, что случайная величина ξ, имеющая "стандартные" параметры
$$\mathbf{M}\xi=0\quad\text{и}\quad\mathbf{D}\xi=\mathbf{M}(\xi^2)=1,$$
является *нормированной*. Для произвольной случайной величины ξ с математическим ожиданием $a=\mathbf{M}\xi$ и дисперсией $\sigma^2=\mathbf{D}\xi=\mathbf{M}(\xi-a)^2$ введём *безразмерную* величину
$$\xi^*\;\stackrel{\text{def}}{=}\;\frac{\xi-\mathbf{M}\xi}{\sqrt{\mathbf{D}\xi}}\;=\;\frac{\xi-a}{\sigma},$$
называемую *нормировкой* ξ. Из теоремы 2 вытекает

Следствие. *Математическое ожидание* $\mathbf{M}\xi^*=0$, *а дисперсия* $\mathbf{D}\xi^*=\mathbf{M}\,(\xi^*)^2=1$.

Доказательство. Применяя соответствующие утверждения теоремы 2, проверяем
$$\mathbf{M}\xi^*=\frac{\mathbf{M}(\xi-\mathbf{M}\xi)}{\sqrt{\mathbf{D}\xi}}=0,\qquad\mathbf{M}\,(\xi^*)^2=\mathbf{D}\xi^*=\frac{\mathbf{D}(\xi-\mathbf{M}\xi)}{\mathbf{D}\xi}=\frac{\mathbf{D}\xi}{\mathbf{D}\xi}=1.$$

Следствие доказано.

2. Пусть (ξ, η) – произвольная пара случайных величин с совместным распределением $q(x, y)$ (дискретная модель) или плотностью совместного распределения $p(x, y)$ (непрерывная модель). Символом $f(x, y)$ обозначим произвольную функцию двух действительных аргументов x, $-\infty \le x \le \infty$ и y, $-\infty \le y \le \infty$.

Рассмотрим случайную величину $\theta = f(\xi, \eta)$, которую можно называть *функцией от пары* случайных величин (ξ, η). Очевидно, что распределение вероятностей $\theta = f(\xi, \eta)$ однозначно вычисляется по совместному распределению пары (ξ, η). Математическое ожидание функции от пары случайных величин (ξ, η) обозначается символом $\mathbf{M} f(\xi, \eta)$ и определяется формулой

$$\mathbf{M} f(\xi, \eta) \stackrel{\text{def}}{=} \begin{cases} \sum\limits_x \sum\limits_y f(x, y)\, q(x, y) & \text{для дискретной модели,} \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} f(x, y)\, p(x, y)\, dx\, dy & \text{для непрерывной модели.} \end{cases} \tag{1}$$

Если в (1) положить $f(x, y) = x + y$, то для дискретного (аналогично и для непрерывного) случая имеем

$$\mathbf{M}(\xi + \eta) \;=\; \sum_x \sum_y (x + y)\, q(x, y) \;=\; \sum_x \sum_y x\, q(x, y) \;+ \sum_x \sum_y y\, q(x, y) \;=$$

$$= \sum_x x \sum_y q(x, y) \;+ \sum_y y \sum_x q(x, y) \;= \sum_x x\, q(x, .) \;+ \sum_y y\, q(., y) \;= \mathbf{M}\xi + \mathbf{M}\eta.$$

Следовательно, математическое ожидание суммы двух случайных величин равно сумме их математических ожиданий, т.е.

$$\mathbf{M}(\xi + \eta) \;=\; \mathbf{M}\xi + \mathbf{M}\eta. \tag{2}$$

Этот результат по индукции, очевидным образом, обобщается на сумму любого числа слагаемых. Например, для случая трех слагаемых, применив два раза формулу (2), получим

$$\mathbf{M}(\xi_1 + \xi_2 + \xi_3) = \mathbf{M}[(\xi_1 + \xi_2) + \xi_3] = \mathbf{M}[(\xi_1 + \xi_2)] + \mathbf{M}\xi_3 = \mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \mathbf{M}\xi_3.$$

Таким образом, справедлива

Теорема 3. *Математическое ожидание суммы случайных величин равно сумме их математических ожиданий, иначе*

$$\mathbf{M}(\xi_1 + \xi_2 + \cdots + \xi_n) \;=\; \mathbf{M}\xi_1 + \mathbf{M}\xi_2 + \cdots + \mathbf{M}\xi_n.$$

Для *независимых* случайных величин также верна

Теорема 4. *Математическое ожидание произведения двух независимых случайных величин ξ и η равно произведению их математических ожиданий, иначе*

$$\mathbf{M}(\xi \cdot \eta) \;=\; \mathbf{M}\xi \cdot \mathbf{M}\eta. \tag{2}$$

Доказательство. Если в определении (1) положить $f(x,y) = x \cdot y$, то для дискретного (аналогично и для непрерывного) случая имеем

$$\mathbf{M}(\xi \cdot \eta) \;=\; \sum_x \sum_y (x \cdot y) \cdot q(x,y).$$

В силу определения независимости, совместное распределение $q(x,y) = q(x,.) \cdot q(.,y)$. Поэтому

$$\mathbf{M}(\xi \cdot \eta) \;=\; \sum_x \sum_y \, [x \cdot q(x,.)] \cdot [y \cdot q(.,y)] = \left[\sum_x \, x \cdot q(x,.)\right] \cdot \left[\sum_y (y \cdot q(.,y)\right] = \mathbf{M}\xi \cdot \mathbf{M}\eta$$

Теорема 4 доказана.

Рассмотрим теперь вопрос о дисперсии суммы двух случайных величин ξ и η. Введём, для краткости, обозначения математических ожиданий $a = \mathbf{M}\xi$ и $b = \mathbf{M}\eta$. Применяя определение дисперсии и теоремы 1-3, можем написать

$$\mathbf{D}(\xi+\eta) = \mathbf{M}[(\xi+\eta) - \mathbf{M}(\xi+\eta)]^2 = \mathbf{M}[(\xi+\eta) - (a+b)]^2 = \mathbf{M}[(\xi-a) + (\eta-b)]^2 =$$

$$= \mathbf{M}[(\xi-a)^2 + (\eta-b)^2 + 2(\xi-a)(\eta-b)] = \mathbf{M}(\xi-a)^2 + \mathbf{M}(\eta-b)^2 + 2\mathbf{M}[(\xi-a)(\eta-b)] =$$

$$= \mathbf{M}(\xi - \mathbf{M}\xi)^2 + \mathbf{M}(\eta - \mathbf{M}\eta)^2 + 2\mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)] = \mathbf{D}\xi + \mathbf{D}\eta + 2\mathrm{cov}(\xi,\eta),$$

где ввели число

$$\mathrm{cov}(\xi,\eta) \stackrel{\mathrm{def}}{=} \mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)] = \sum_x \sum_y (x-a)(y-b)\, q(x,y),$$

называемое *ковариацией* пары случайных величин (ξ,η). Таким образом, в общем случае дисперсия суммы двух случайных величин связана с дисперсиями слагаемых следующей формулой

$$\mathbf{D}(\xi+\eta) \;=\; \mathbf{D}\xi + \mathbf{D}\eta + 2\mathrm{cov}(\xi,\eta). \tag{3}$$

Поскольку

$$\mathrm{cov}(\xi,\eta) = \mathbf{M}[(\xi - \mathbf{M}\xi)(\eta - \mathbf{M}\eta)] = \mathbf{M}[(\xi-a)(\eta-b)] = \mathbf{M}[\xi\eta - b\xi - a\eta + ab],$$

то используя теоремы 1-3, получаем формулу

$$\mathrm{cov}(\xi,\eta) = \mathbf{M}(\xi\cdot\eta) - b\mathbf{M}\xi - a\mathbf{M}\eta + ab =$$

$$= \mathbf{M}(\xi\cdot\eta) - 2ab + ab = \mathbf{M}(\xi\cdot\eta) - ab = \mathbf{M}(\xi\cdot\eta) - \mathbf{M}\xi\cdot\mathbf{M}\eta,$$

из которой, в силу теоремы 4 следует, что для пары (ξ,η) независимых случайных величин $\mathrm{cov}(\xi,\eta) = 0$. Поэтому из формулы (3) вытекает, что

$$\text{если случайные величины} \quad \xi \quad \text{и} \quad \eta \quad \text{независимы, то} \quad \mathbf{D}(\xi+\eta) \;=\; \mathbf{D}\xi + \mathbf{D}\eta. \tag{4}$$

Этот результат по индукции распространяется на сумму любого числа независимых слагаемых. Например, для случая трех независимых слагаемых, применив два раза свойство (4), получим

$$\mathbf{D}(\xi_1+\xi_2+\xi_3)=\mathbf{D}[(\xi_1+\xi_2)+\xi_3]=\mathbf{D}[(\xi_1+\xi_2)]+\mathbf{D}\xi_3=\mathbf{D}\xi_1+\mathbf{D}\xi_2+\mathbf{D}\xi_3.$$

Таким образом, справедлива

Теорема 5. *Дисперсия суммы независимых случайных величин равна сумме дисперсий слагаемых, иначе*

$$\mathbf{D}(\xi_1+\xi_2+\cdots+\xi_n)\;=\;\mathbf{D}\xi_1+\mathbf{D}\xi_2+\cdots+\mathbf{D}\xi_n.$$

Замечание. Теорема 5 описывает очень полезное свойство дисперсии, которое во многом объясняет выбор последней, т.е. числа $\mathbf{D}\xi=\mathbf{M}(\xi-\mathbf{M}\xi)^2$, в качестве численной меры отклонения от математического ожидания. Можно показать, что для других столь же естественных мер отклонения, например

$$\mathbf{M}|\xi-\mathbf{M}\xi|,\qquad \mathbf{M}|\xi-\mathbf{M}\xi|^3,\qquad \mathbf{M}(\xi-\mathbf{M}\xi)^4,$$

свойство, состоящее в том, что отклонение суммы независимых случайных величин равно сумме отдельных отклонений, не имеет места.

Пример. (Модель (n,p) испытаний Бернулли). Введённое в примерах из § 9.1 число успехов S_n в n испытаниях Бернулли представляется в виде суммы n независимых одинаково распределённых случайных величин $\xi_1,\xi_2,\ldots,\xi_n$ где ξ_i, $i=1,2,\ldots,n$, принимает значения 1 или 0, регистрируя "успех" или "неудачу" в i-ом испытании. Следовательно,

$$S_n\;=\;\xi_1+\xi_2+\cdots+\xi_n,\quad \text{где}\quad \Pr\{\xi_i=1\}=p,\quad \Pr\{\xi_i=0\}=1-p.$$

В § 10.1 были вычислены математическое ожидание $\mathbf{M}\xi_i=p$ и дисперсия $\mathbf{D}\xi_i=p(1-p)$. Поэтому, в силу теорем 3 и 5, математическое ожидание числа успехов в n испытаниях Бернулли $\mathbf{M}\,S_n=np$, а дисперсия числа успехов $\mathbf{D}\,S_n=np(1-p)$.

§ 12.2. Выборка, выборочные характеристики, закон больших чисел

В математической статистике *выборкой* объёма n с теоретическими (неизвестными) характеристиками: математическим ожиданием (средним значением) a, дисперсией σ^2 и функцией распределения $F(t)$, $-\infty<t<\infty$, называется совокупность из n наблюдений $(x_1,x_2,\ldots,x_n)$ (случайных величин), удовлетворяющих двум условиям.

1) Наблюдения $(x_1,x_2,\ldots,x_n)$ *независимы.*

2) Наблюдения $(x_1,x_2,\ldots,x_n)$ *однородны*, т.е. имеют *одинаковые* теоретические характеристики

$$a=\mathbf{M}x_i,\quad \sigma^2=\mathbf{D}x_i=\mathbf{M}(x_i-a)^2,\quad F(t)=\Pr\{x_i<t\},\qquad i=1,2,\ldots,n,$$

где параметр $\sigma=\sqrt{\mathbf{D}x_i}$ называтся теоретическим (неизвестным) среднеквадратичным отклонением выборки.

1. Для среднего арифметического значения наблюдений $(x_1, x_2, \ldots, x_n)$, называемого *выборочным* (или *эмпирическим*) *средним*, введём стандартное обозначение

$$\bar{x} \stackrel{\text{def}}{=} \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum\limits_{i=1}^{n} x_i}{n}.$$

Отметим, что выборочное среднее является случайной величиной, значение которой вычисляется экспериментатором после проведения n опытов.

С помощью теорем 2, 3 и 5 находим

$$\mathbf{M}\bar{x} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{M}x_i = \frac{na}{n} = a, \quad \mathbf{D}\bar{x} = \frac{1}{n^2}\sum_{i=1}^{n} \mathbf{D}x_i = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \quad \sqrt{\mathbf{D}\bar{x}} = \frac{\sigma}{\sqrt{n}}. \tag{5}$$

Другими словами, математическое ожидание $\mathbf{M}\bar{x}$ выборочного среднего совпадает с теоретическим средним выборки $a = \mathbf{M}x_i$, а среднеквадратичное отклонение $\sqrt{\mathbf{D}\bar{x}}$ выборочного среднего в $\sqrt{n}$ раз меньше теоретического $\sigma = \sqrt{\mathbf{D}x_i}$.

При $n \to \infty$ имеем

$$\lim_{n\to\infty} \mathbf{D}\bar{x} = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0. \tag{6}$$

Согласно теореме 1, $\mathbf{D}\bar{x} = 0$ тогда и только тогда, когда $\bar{x} = \mathbf{M}\bar{x} = a$. Поэтому, опираясь на (6), мы можем сформулировать *связь* вычисляемого экспериментатором выборочного среднего $\bar{x}$ и неизвестного экспериментатору теоретического среднего a в виде следующей теоремы, называемой в теории вероятностей *законом больших чисел*.

Теорема 6. (Закон больших чисел). *Если случайные величины $x_1, x_2, \ldots, x_n$ независимы, имеют одинаковые теоретические средние $\mathbf{M}x_i = a$ и одинаковые теоретические дисперсии $\mathbf{D}x_i = \sigma^2$, то при $n \to \infty$ их среднее арифметическое значение $\bar{x}$ стремится к теоретическому среднему a, иначе*

$$\lim_{n\to\infty} \bar{x} = \lim_{n\to\infty} \frac{\sum\limits_{i=1}^{n} x_i}{n} = a \quad \textit{или} \quad \bar{x} \approx a \quad \textit{при больших значениях} \quad n. \tag{7}$$

Закон больших чисел (6)-(7) отражает давно отмеченное свойство реальных опытов, состоящее в том, что *в то время как результаты отдельных измерений $x_1, x_2, \ldots, x_n$ могут колебаться сильно, их средние арифметические $\bar{x}$ обнаруживают гораздо большую устойчивость, т.е. мало меняются в различных сериях опытов.*

Пример. (Модель (n, p) испытаний Бернулли). Пусть при $i = 1, 2, \ldots, n$ измерение $x_i =$ 0, 1 регистрирует "успех ~ 1" или "неудачу ~ 0" в i-ом испытании Бернулли. Тогда число успехов $S_n = \sum\limits_{i=1}^{n} x_i$, теоретическое среднее $a = \mathbf{M}\, x_i = p$ совпадает с вероятностью успеха в одном испытании, а выборочное среднее $\bar{x} = S_n/n$ есть *доля* числа успехов в n испытаниях. Таким образом, для модели (n, p) испытаний Бернулли закон больших чисел означает, что при $n \to \infty$ вычисляемая экспериментатором доля числа успехов в n испытаниях стремится к теоретической (неизвестной) вероятности успеха p в одном испытании:

$$\lim_{n\to\infty} \frac{S_n}{n} = \lim_{n\to\infty} \frac{\sum\limits_{i=1}^{n} x_i}{n} = p.$$

Приближенное равенство (7) можно уточнить для частного случая *нормальной выборки*, в которой предполагается, что наблюдения $x_1, x_2, \ldots, x_n$ имеют вид

$$x_i \sim \mathcal{N}(a,\sigma) \quad \text{или} \quad x_i = a + \sigma\cdot\xi_i, \quad \text{где} \quad \xi_i \sim \mathcal{N}(0,1), \quad i = 1,2,\ldots,n.$$

В силу свойства 5 из § 11.3., среднее арифметическое $\bar{x}$ независимых величин x_i, имеющих нормальное распределение $\mathcal{N}(a,\sigma)$, также имеет нормальное распределение, т.е.

$$\bar{x} \sim \mathbf{M}\bar{x} + \sqrt{\mathbf{D}\bar{x}}\cdot\mathcal{N}(0,1) = a + \frac{\sigma}{\sqrt{n}}\cdot\mathcal{N}(0,1) \quad \text{или} \quad \frac{\bar{x}-a}{\sigma/\sqrt{n}} = \frac{(\bar{x}-a)\sqrt{n}}{\sigma} \sim \mathcal{N}(0,1). \qquad (8)$$

Пусть α, $1/2 < \alpha < 1$, – фиксированное число, называемое *уровнем значимости*. Введём число $x'_\alpha > 0$, определяемое как решение уравнения

$$\Pr\{-x'_\alpha < \mathcal{N}(0,1) < x'_\alpha\} = \int\limits_{-x'_\alpha}^{x'_\alpha} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt = 1-\alpha$$

и называемое *двусторонним критическим значением* стандартного нормального распределения $\mathcal{N}(0,1)$ для уровня значимости α. В силу (8) можем написать, что

$$\Pr\left\{-x'_\alpha < \frac{(\bar{x}-a)\sqrt{n}}{\sigma} < x'_\alpha\right\} = 1-\alpha \quad \text{или} \quad \Pr\left\{\bar{x} - \frac{x'_\alpha\cdot\sigma}{\sqrt{n}} < a < \bar{x} + \frac{x'_\alpha\cdot\sigma}{\sqrt{n}}\right\} = 1-\alpha. \quad (9)$$

Предположим, что среднеквадратичное отклонение σ известно экспериментатору. Тогда (9) показывает, что вычисляемый на основании среднего арифметического значения наблюдений $x_1, x_2, \ldots, x_n$ интервал

$$\left(a_\alpha^-;\, a_\alpha^+\right), \quad \text{где} \quad a_\alpha^- \stackrel{\text{def}}{=} \bar{x} - \frac{x'_\alpha\cdot\sigma}{\sqrt{n}}, \quad a_\alpha^+ \stackrel{\text{def}}{=} \bar{x} + \frac{x'_\alpha\cdot\sigma}{\sqrt{n}}, \quad \bar{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n},$$

накрывает неизвестное математическое ожидание a с вероятностью $1-\alpha$. Такой интервал в математической статистике называется *доверительным интервалом* или *интервальной оценкой* параметра a с *коэффициентом доверия* (*надёжностью*), равным $1-\alpha$. Отметим, что *при фиксированном коэффициенте доверия* $1-\alpha$ *длина доверительного интервала с ростом объёма наблюдений* n *уменьшается пропорционально* $1/\sqrt{n}$.

2. Наряду со средним арифметическим значением $\bar{x}$, для выборки $(x_1, x_2, \ldots, x_n)$ вычисляют *выборочную* (*эмпирическую*) *функцию распределения*:

$$F_n(t) \stackrel{\text{def}}{=} \frac{\text{число наблюдений} \quad x_i \quad \text{таких, что} \quad x_i < t}{n}, \qquad -\infty < t < \infty,$$

и *выборочную* (*эмпирическую*) *дисперсию*:

$$s^2 \stackrel{\text{def}}{=} \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \cdots + (x_n-\bar{x})^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2.$$

Эти характеристики представляют собой *выборочные аналоги* теоретической функции распределения $F(t) = \Pr\{x_i < t\}$, $0 \le F(t) \le 1$, и теоретической дисперсии $\sigma^2 = \mathbf{D}x_i$. Как функция аргумента t, $-\infty < t < \infty$, $F_n(t)$ является ступенчатой функцией, меняющей свои значения в точках $x_1, x_2, \ldots, x_n$.

Зафиксируем произвольное t, $-\infty < t < \infty$, и введём независимые одинаково распределённые случайные величины

$$\eta_i \stackrel{\text{def}}{=} \begin{cases} 1, & \text{если } x_i < t, \\ 0, & \text{если } x_i \ge t, \end{cases} \quad i = 1, 2, \ldots, n,$$

среднее арифметическое значение которых равно $F_n(t)$, т.е.

$$F_n(t) = \frac{\eta_1 + \eta_2 + \cdots + \eta_n}{n} = \frac{1}{n} \sum_{i=1}^{n} \eta_i. \tag{10}$$

Поскольку

$$\Pr\{x_i < t\} = F(t), \qquad \Pr\{x_i \ge t\} = 1 - F(t),$$

то вероятностная функция $q_t(x)$, $-\infty < x < \infty$, задающая распределение вероятностей η_i, и её числовые характеристики имеют вид:

$$q_t(x) \stackrel{\text{def}}{=} \Pr\{\eta_i = x\} = \begin{cases} 1 - F(t), & \text{если } x = 0, \\ F(t), & \text{если } x = 1, \\ 0, & \text{для остальных } x, \end{cases}$$

$$\mathbf{M}\eta_i = \sum_x x\, q_t(x) = 0 \cdot (1 - F(t)) + 1 \cdot F(t) = F(t),$$

$$\mathbf{M}\eta_i^2 = \sum_x x^2\, q_t(x) = 0^2 \cdot (1 - F(t)) + 1^2 \cdot F(t) = F(t),$$

$$\mathbf{D}\eta_i = \mathbf{M}\eta_i^2 - (\mathbf{M}\eta_i)^2 = F(t) - F(t)^2 = F(t)(1 - F(t)).$$

Теперь, учитывая (10) и закон больших чисел теоремы 6, мы можем описать *связь* между вычисляемой экспериментатором выборочной функцией распределения $F_n(t)$ и неизвестной экспериментатору теоретической функцией распределения $F(t)$:

$$\text{если} \quad n \to \infty, \quad \text{то} \quad \lim_{n\to\infty} F_n(t) = F(t) \quad \text{или} \quad F_n(t) \approx F(t) \quad \text{при больших значениях} \quad n.$$

Можно проверить, что математическое ожидание $\mathbf{M}s^2 = \sigma^2$, а дисперсия $\mathbf{D}s^2 \to 0$, если $n \to \infty$. Следовательно, связь между выборочной дисперсией s^2 и теоретической дисперсией σ^2 имеет вид:

$$\lim_{n\to\infty} s^2 = \lim_{n\to\infty} \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma^2$$

или $s^2 \approx \sigma^2$ при больших значениях n.

§ 12.3. Коэффициент корреляции и его свойства

Пусть (ξ, η) – произвольная пара случайных величин с совместным распределением $q(x, y)$ (дискретная модель) или плотностью совместного распределения $p(x, y)$ (непрерывная модель). Символами

$$a_\xi = \mathbf{M}\xi, \quad \sigma_\xi = \sqrt{\mathbf{D}\xi}, \qquad a_\eta = \mathbf{M}\eta, \quad \sigma_\eta = \sqrt{\mathbf{D}\eta}$$

обозначим определённые в § 9 и рассматриваемые в § 11.1 математические ожидания и среднеквадратичные отклонения, характеризующие отдельные распределения случайных величин ξ и η. Введём безразмерные нормированные случайные величины

$$\xi^* = \frac{\xi - a_\xi}{\sigma_\xi}, \qquad \eta^* = \frac{\eta - a_\eta}{\sigma_\eta},$$

для которых, согласно следствию из теоремы 2,

$$\mathbf{M}\xi^* = \mathbf{M}\eta^* = 0, \qquad \mathbf{D}\xi^* = \mathbf{M}(\xi^*)^2 = \mathbf{D}\eta^* = \mathbf{M}(\eta^*)^2 = 1. \tag{11}$$

Определение. Количественной *мерой связи* случайных величин ξ и η служит числовая характеристика совместного распределения пары (ξ, η), определяемая формулой

$$\rho = \rho(\xi, \eta) \stackrel{\text{def}}{=} \mathbf{M}(\xi^* \cdot \eta^*) = \frac{\mathbf{M}[(\xi - a_\xi)(\eta - a_\eta)]}{\sigma_\xi \cdot \sigma_\eta} = \frac{\mathbf{M}(\xi \cdot \eta) - a_\xi \cdot a_\eta}{\sigma_\xi \cdot \sigma_\eta} =$$

$$= \begin{cases} \sum\limits_x \sum\limits_y \frac{(x-a_\xi)(y-a_\eta)}{\sigma_\xi \cdot \sigma_\eta}\, q(x,y) & \text{для дискретной модели,} \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \frac{(x-a_\xi)(y-a_\eta)}{\sigma_\xi \cdot \sigma_\eta}\, p(x,y)\, dx\, dy & \text{для непрерывной модели.} \end{cases} \tag{12}$$

и называемая *коэффициетом корреляции* (или *нормированной ковариацией*) пары (ξ, η).

Свойства коэффициета корреляции ρ описывает

Теорема 7. *Справедливы следующие утверждения.*

1) *Если случайные величины ξ и η независимы, то коэффициет корреляции $\rho = 0$.*

2) *Имеют место неравенства $-1 \le \rho \le 1$.*

3) *Коэффициет корреляции $\rho = 1$ тогда и только тогда, когда*

$$\eta^* = \xi^* \quad \textit{или} \quad \frac{\eta - a_\eta}{\sigma_\eta} = \frac{\xi - a_\xi}{\sigma_\xi} \quad \textit{или} \quad \eta = a_\eta + \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - a_\xi),$$

т.е. если случайная величина η связана со случайной величиной ξ прямо пропорциональной линейной зависимостью.

4) *Коэффициет корреляции $\rho = -1$ тогда и только тогда, когда*

$$\eta^* = -\xi^* \quad \textit{или} \quad \frac{\eta - a_\eta}{\sigma_\eta} = -\frac{\xi - a_\xi}{\sigma_\xi} \quad \textit{или} \quad \eta = a_\eta - \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - a_\xi),$$

т.е. если случайная величина η связана со случайной величиной ξ обратно пропорциональной линейной зависимостью.

Доказательство.

1) Это утверждение является следствием теоремы 4.

2) Для любого значения x, $-\infty < x < \infty$, число

$$f(x) \stackrel{\text{def}}{=} \mathbf{M}(\eta^* - x \cdot \xi^*)^2,$$

определяемое как математическое ожидание от неотрицательной случайной величины, также является неотрицательным, т.е. $f(x) \geq 0$. Поэтому, применяя свойства математического ожидания из теорем 1-3, равенства (11) и определение коэффициента корреляции $\rho = \mathbf{M}(\xi^* \cdot \eta^*)$ из (12), можем написать

$$0 \leq f(x) = \mathbf{M}(\eta^* - x \cdot \xi^*)^2 = \mathbf{M}\left[(\eta^*)^2 - 2x \cdot \xi^* \cdot \eta^* + x^2 \cdot (\eta^*)^2\right] =$$

$$= \mathbf{M}(\eta^*)^2 - 2x \cdot \mathbf{M}(\xi^* \cdot \eta^*) + x^2 \cdot \mathbf{M}(\eta^*)^2 = 1 - 2x\rho + x^2.$$

Следовательно, для любого действительного значения x, $-\infty < x < \infty$, квадратный трёхчлен $1 - 2x\rho + x^2 \geq 0$. Отсюда вытекает, что

$$\text{корни} \quad (x_1, x_2) \quad \text{уравнения} \quad 1 - 2x\rho + x^2 = 0, \quad \text{имеющие вид} \quad x_{1,2} = \rho \pm \sqrt{\rho^2 - 1},$$

либо являются мнимыми, если $\rho^2 - 1 < 0$, либо совпадают между собой, если $\rho^2 - 1 = 0$. Это возможно лишь при условии $\rho^2 - 1 \leq 0$, которое равносильно доказываемому утверждению.

3) Случай $\rho = 1$ равносилен тому, что $x_1 = x_2 = 1$. Иначе,

$$f(1) = 0 \quad \text{или} \quad \mathbf{M}(\eta^* - \xi^*)^2 \quad \text{или} \quad \eta^* = \xi^*.$$

4) Случай $\rho = -1$ равносилен тому, что $x_1 = x_2 = -1$. Иначе,

$$f(-1) = 0 \quad \text{или} \quad \mathbf{M}(\eta^* + \xi^*)^2 \quad \text{или} \quad \eta^* = -\xi^*.$$

Теорема доказана.

§ 12.4. Линейная модель корреляции признаков

Пусть случайная величина (признак) η и случайная величина (признак) ξ имеют числовые характеристики

$$a_\xi = \mathbf{M}\xi, \quad \sigma_\xi = \sqrt{\mathbf{D}\xi}, \qquad a_\eta = \mathbf{M}\eta, \quad \sigma_\eta = \sqrt{\mathbf{D}\eta},$$

а ρ, $-1 \leq \rho \leq 1$, –произвольное фиксированное число. Для статистического анализа связи признаков η и ξ, измеряемой коэффициентом корреляции

$$\rho = \mathbf{M}\left\{\frac{\eta - a_\eta}{\sigma_\eta} \cdot \frac{\xi - a_\xi}{\sigma_\xi}\right\} = \mathbf{M}(\xi^* \cdot \eta^*), \quad \text{где} \quad \xi^* = \frac{\xi - a_\xi}{\sigma_\xi}, \quad \eta^* = \frac{\eta - a_\eta}{\sigma_\eta},$$

удобно применять следующую запись:

$$\frac{\eta - a_\eta}{\sigma_\eta} \;=\; A \cdot \frac{\xi - a_\xi}{\sigma_\xi} \;+\; B \cdot \delta \quad \text{или} \quad \eta^* \;=\; A \cdot \xi^* + \; B \cdot \delta, \tag{13}$$

в которой символы A и B обозначают постоянные (неслучайные) величины, а символ δ обозначает случайную величину, удовлетворяющую условиям:

- δ – безразмерная случайная величина, которая *не зависит от* ξ;
- δ – *нормированная* случайная величина, т.е. $\mathbf{M}\delta = 0$ и $\mathbf{D}\delta = \mathbf{M}\delta^2 = 1$.

"Линейная" форма связи (13) между признаками η и ξ, называется *линейной корреляцией признаков*. Отметим, что линейной корреляции признаков η и ξ соответствует специальный *частный случай* совместного распределения $q(x, y)$ (дискретная модель) или плотности совместного распределения $p(x, y)$ (непрерывная модель) пары признаков (ξ, η).

Чтобы найти в соотношениях (13) зависимость чисел A и B от коэффициента корреляции ρ, представим с помощью (13) случайные величины $(\eta^*)^2$ и $\eta^* \cdot \xi^*$ в виде

$$(\eta^*)^2 \;=\; A^2 \cdot (\xi^*)^2 + \; B^2 \cdot \delta^2 \; + 2\, A\, B \cdot \xi^* \cdot \delta, \qquad \eta^* \cdot \xi^* \;=\; A \cdot (\xi^*)^2 + \; B \cdot \delta \cdot \xi^*.$$

Рассмотрим математические ожидания этих величин. Применяя свойства математического ожидания из теорем 1-3, равенства (11), определение коэффициента корреляции $\rho = \mathbf{M}(\xi^* \cdot \eta^*)$ из (12) и свойства случайной величины δ, можем написать

$$1 = \mathbf{M}(\eta^*)^2 \;=\; A^2 \cdot \mathbf{M}(\xi^*)^2 + \; B^2 \cdot \mathbf{M}\delta^2 \; + 2\, A\, B \cdot \mathbf{M}\left\{\xi^* \cdot \delta\right\} \;=$$

$$= \; A^2 \cdot 1 \; + \; B^2 \cdot 1 \; + 2\, A\, B \cdot \mathbf{M}\xi^* \cdot \mathbf{M}\delta \;=\; A^2 \; + \; B^2 \; + 2\, A\, B \cdot 0 \cdot 0 = A^2 \; + \; B^2,$$

$$\rho \;=\; \mathbf{M}(\eta^* \cdot \xi^*) \;=\; A \cdot \mathbf{M}(\xi^*)^2 + \; B \cdot \mathbf{M}\left\{\xi^* \cdot \delta\right\} \;=\; A \cdot 1 + \; B \cdot 0 \;=\; A.$$

Из доказанных равенств $A^2 + B^2 \;=\; 1$ и $A = \rho$ получаем, что $B = \sqrt{1 - \rho^2}$. Следовательно, в соотношениях (13) зависимость чисел A и B от коэффициента корреляции ρ имеет вид

$$A = \rho, \qquad B = \sqrt{1 - \rho^2}.$$

Другими словами, линейная корреляция признаков η и ξ с коэффициентом корреляции ρ записывается в виде:

$$\frac{\eta - a_\eta}{\sigma_\eta} \;=\; \rho \cdot \frac{\xi - a_\xi}{\sigma_\xi} \;+\; \sqrt{1 - \rho^2} \cdot \delta, \quad \text{или} \quad \eta^* \;=\; \rho \cdot \xi^* + \; \sqrt{1 - \rho^2} \cdot \delta,$$

где символ δ обозначает случайную величину, удовлетворяющую условиям:

- δ – безразмерная случайная величина, которая *не зависит от* ξ;
- δ – *нормированная* случайная величина, т.е. $\mathbf{M}\delta = 0$ и $\mathbf{D}\delta = \mathbf{M}(\delta)^2 = 1$.

В математической статистике данная форма описания связи наблюдений (случайных величин) η и ξ называется *линейной моделью корреляционного анализа.*

Оглавление