

# III. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

# Глава 1. СЛУЧАЙНЫЕ СОБЫТИЯ И ВЕРОЯТНОСТИ

Серьезное предупреждение. Мы разграничиваем *содержательную задачу*, решить которую необходимо *специалисту-нематематику*, и *математическую модель*, которую исследует *математик*. Вопрос об адекватности математической модели и породившей ее содержательной задачи не только выходит за рамки нашего курса, но вообще не может быть отнесен к предмету математики.

## 1.1. Дискретное пространство элементарных событий

Мы будем рассматривать некоторый эксперимент, который может закончиться одним из *взаимоисключающих исходов*. Эти исходы эксперимента мы будем называть *элементарными событиями*.

Будем пока что предполагать, что множество всех элементарных событий *конечно* или *счетно*<sup>25</sup>. Назовем это множество *дискретным пространством элементарных событий*. Обычно его обозначают  $\Omega$ , а элементарные события –  $\omega \in \Omega$ .

Примеры. 1. Бросается правильная монета. Обычно считают, что возможны два исхода этого эксперимента, и называют их герб (Г) и решетка (Р). Таким образом, пространство элементарных событий состоит из двух элементов  $\Omega = \{Г, Р\}$ .

Замечание. Если в реальном эксперименте монета встанет *на ребро*, мы должны либо считать этот эксперимент несостоявшимся (так как подобный исход *не включен нами* в пространство элементарных событий), либо *изменить первоначальную модель* и рассматривать пространство элементарных событий из трех элементов  $\Omega = \{Г, Р, Н\}$  (Н означает "на ребро").

Этот простой пример еще раз показывает, что одна и та же *содержательная задача* может привести к различным *математическим моделям*. Именно поэтому мы и будем изучать математические модели, не задумываясь об их происхождении.

2. Эксперимент состоит в бросании правильной монеты *до первого появления герба*. Ниже выписаны несколько элементарных событий:

---

<sup>25</sup>Множество называется *счетным*, если можно установить взаимно однозначное соответствие между его элементами и натуральными числами. *Можно показать*, например, что множество всех рациональных чисел счетно, а множество всех вещественных чисел – несчетно.

Г  
 Р, Г  
 Р, Р, Г  
 ...  
 Р, Р, ..., Р, Г  
 ...

Хотя не очень верится, что десять тысяч раз подряд будет выпадать решетка, и лишь на десять тысяч первый раз выпадет герб, но исключить такой исход никто не отважится. Поэтому пространство элементарных событий в этом примере полагают счетным.

## 1.2. Вероятность

Зафиксируем некоторый исход. Каждый реальный эксперимент может закончиться либо этим исходом, либо каким-нибудь другим. Если в неизменных условиях эксперимент повторили  $n$  раз, и в  $m$  случаях он закончился отмеченным исходом, то говорят, что наблюдалась *относительная частота*  $\frac{m}{n}$  этого исхода. Очевидно, что сумма относительных частот всех возможных исходов равна единице.

Теория вероятностей пригодна для описания тех реальных экспериментов, в которых наблюдается *устойчивость относительных частот*: длинные серии экспериментов приводят, как правило, к близким значениям относительных частот.

Математической моделью реального эксперимента служит дискретное пространство элементарных событий. Математическим аналогом относительной частоты исхода эксперимента является *вероятность элементарного события*  $\omega$  – вещественное число  $\mathbf{P}(\omega)$ , удовлетворяющее двум условиям:

$$0 \leq \mathbf{P}(\omega) \leq 1; \quad \sum_{\omega \in \Omega} \mathbf{P}(\omega) = 1 \quad (1.2.1)$$

(во втором условии фигурирует *сумма* – для конечного пространства элементарных событий и *сумма ряда* – для счетного).

Очевидно, что функция  $\mathbf{P} : \Omega \rightarrow \mathbb{R}$ , удовлетворяющая условиям (1.2.1), может быть задана многими способами. Вопрос о ее выборе *относится не к математике, а к выбору математической модели*. Этот выбор делается из содержательных предположений об условиях эксперимента.

Примеры. 1. Если бросается *правильная* монета, то естественно предположить, что в длинных сериях экспериментов относительные частоты выпадения герба и решетки должны быть примерно равны. Поэтому разумно приписать элементарным событиям  $\Gamma$  и  $P$  одинаковые вероятности  $\mathbf{P}(\Gamma) = \mathbf{P}(P)$ .

Если  $\Omega = \{\Gamma, P\}$ , то  $\mathbf{P}(\Gamma) + \mathbf{P}(P) = 1$ . Отсюда  $\mathbf{P}(\Gamma) = \mathbf{P}(P) = 1/2$ .

2. *Правильная* монета бросается до первого выпадения герба. Обозначим  $\omega_n$  элементарное событие, соответствующее выпадению герба при  $n$ -м бросании монеты. Положим  $\mathbf{P}(\omega_n) = 2^{-n}$  (обсуждение причин такого выбора мы отложим до п.1.4). Вычислив

$$\sum_{n=1}^{+\infty} \mathbf{P}(\omega_n) = \sum_{n=1}^{+\infty} 2^{-n} = 1,$$

убедимся в *допустимости* такого задания вероятностей.

Серьезное предупреждение. Еще раз подчеркнем, что применение теории вероятностей для описания эксперимента возможно лишь при наличии *статистической устойчивости* относительной частоты исходов. Имеется в виду, что при *многократном* повторении *длинных* серий экспериментов в *неизменных* условиях относительные частоты, *как правило*, меняются мало. Слова "многократно" "длинные" "неизменные" "как правило" не могут быть уточнены. Они относятся к неалгоритмизируемой процедуре перехода от реального эксперимента к его математической модели, и ответственность за справедливость этих слов несет тот *специалист-нематематик*, который решается применить теорию вероятностей к описанию своей задачи.

Отметим еще, что совершенно бессмысленно употреблять слово "вероятность пока не описано пространство элементарных событий".

### 1.3. Случайные события в дискретном пространстве элементарных событий

*Случайным событием (событием)* называют всякое подмножество дискретного пространства элементарных событий. Говорят, что в результате эксперимента *событие произошло*, если эксперимент закончился одним из *взаимоисключающих* исходов, составляющих это событие.

Вероятность случайного события равна *по определению* сумме вероятностей элементарных событий, составляющих это событие:

$$A \subset \Omega \quad \Longrightarrow \quad \mathbf{P}(A) = \sum_{\omega \in A} \mathbf{P}(\omega).$$

Оправдание такого определения вероятности события – ее частотная интерпретация: относительная частота появления хотя бы одного из взаимоисключающих исходов равна, очевидно, сумме относительных частот этих исходов.

Пример. Эксперимент состоит в бросании игральной кости – правильного куба, грани которого помечены цифрами от 1 до 6. Здесь  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , и, вследствие правильности куба, разумно считать элементарные события равновероятными:

$$\mathbf{P}(n) = 1/6; \quad n = 1, \dots, 6.$$

Если  $A = \{1, 3, 5\}$  (в результате эксперимента выпала нечетная цифра), то  $\mathbf{P}(A) = \mathbf{P}(1) + \mathbf{P}(3) + \mathbf{P}(5) = 3 \cdot 1/6 = 1/2$ .

Замечания. 1. Очевидно, что вероятность *события*  $\{\omega\}$ , состоящего из одного элемента  $\omega$ , совпадает с вероятностью *элементарного события*  $\omega$ . Поэтому мы будем отождествлять  $\omega$  и  $\{\omega\}$ , несмотря на то, что это объекты разной природы.

2. Иногда в задачах необходимо суммировать вероятности большого количества элементарных событий, составляющих данное событие. При этом приходится использовать специальные методы *комбинаторики*. Мы же ограничимся рассмотрением достаточно простых примеров, в которых вычислительные сложности не возникают.

Пример. Эксперимент состоит в бросании *правильной* монеты до первого выпадения герба. Будем, как и в п.1.2, считать, что  $\omega_n$  – элементарное событие "в первый раз герб появился при  $n$ -м бросании" и  $\mathbf{P}(\omega_n) = 2^{-n}$ .

Найдем вероятности событий  $A$  (игра закончится на нечетном бросании монеты) и  $B$  (игра закончится на четном бросании монеты):

$$\mathbf{P}(A) = \sum_{n=1}^{+\infty} 2^{-(2n-1)} = \frac{1/2}{1 - 1/4} = 2/3;$$

$$\mathbf{P}(B) = \sum_{n=1}^{+\infty} 2^{-2n} = \frac{1/4}{1 - 1/4} = 1/3.$$

Содержательная интерпретация этого результата – важный для игроков факт: если два игрока бросают монету по очереди, то при *многokратном* повторении игры бросающий первым будет выигрывать примерно вдвое чаще партнера.

Замечание. Теория вероятностей обязана своим появлением потребности участников азартных игр в выработке "выигрышной стратегии". Разделом математики теория вероятностей стала сравнительно недавно. Поэтому в ней сохранилась "доматематическая" терминология.

Принято называть случайное событие  $\Omega$  ( $\Omega \subset \Omega$ ) *достоверным* событием, а случайное событие  $\emptyset$  ( $\emptyset \subset \Omega$ ) *невозможным* событием. Отметим, что  $\mathbf{P}(\Omega) = 1$ ,  $\mathbf{P}(\emptyset) = 0$ .

Пусть  $A, B \subset \Omega$ . Если  $A \cap B = \emptyset$ , то  $A$  и  $B$  называют *несовместными* событиями. В противном случае их называют *совместными* событиями.

Переведите на язык "исходов эксперимента" понятия "достоверное событие", "невозможное событие", "несовместные события".

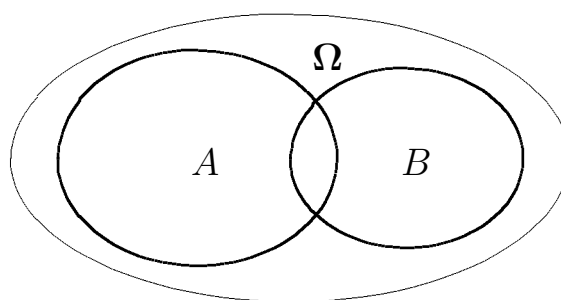


Рис.1.1

Пусть теперь  $A, B \subset \Omega$  (рис.1.1). Выделим следующие события (множества):  $A \setminus B$ ,  $B \setminus A$ ,  $A \cap B$  (напомним, что  $A \setminus B$  состоит из всех элементарных событий, которые входят в  $A$  и не входят в  $B$ ;  $B \setminus A$  – из всех элементарных событий, которые входят в  $B$  и не входят в  $A$ ;  $A \cap B$  – из всех элементарных событий, которые входят и в  $A$ , и в  $B$ ).

Любые два из трех описанных множеств не пересекаются (события  $A \setminus B$ ,  $B \setminus A$ ,  $A \cap B$  попарно несовместны).

Очевидно, что

$$A = (A \setminus B) \cup (A \cap B); \quad B = (B \setminus A) \cup (A \cap B);$$

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B).$$

Поскольку вероятность события определена как сумма вероятностей составляющих его элементарных событий, из этих равенств следует, что

$$\mathbf{P}(A) = \sum_{\omega \in A \setminus B} \mathbf{P}(\omega) + \sum_{\omega \in A \cap B} \mathbf{P}(\omega); \quad \mathbf{P}(B) = \sum_{\omega \in B \setminus A} \mathbf{P}(\omega) + \sum_{\omega \in A \cap B} \mathbf{P}(\omega);$$

$$\mathbf{P}(A \cup B) = \sum_{\omega \in A \setminus B} \mathbf{P}(\omega) + \sum_{\omega \in B \setminus A} \mathbf{P}(\omega) + \sum_{\omega \in A \cap B} \mathbf{P}(\omega).$$

Отсюда получаем утверждение, известное как *теорема сложения вероятностей*:

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$$

(если просто сложить вероятности событий  $A$  и  $B$ , то вероятности элементарных событий, содержащихся в  $A \cap B$ , войдут в сумму дважды).

В частности, если события  $A$  и  $B$  несовместны ( $A \cap B = \emptyset$ ), то

$$\mathbf{P}(A \cap B) = 0 \quad \text{и} \quad \mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B).$$

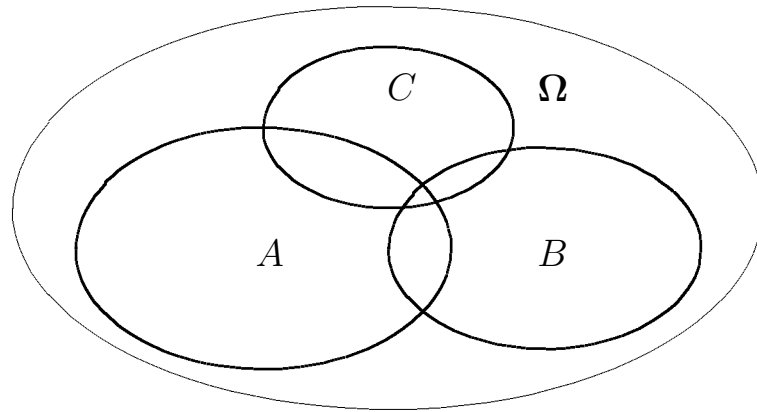


Рис.1.2

Для трех событий  $A$ ,  $B$  и  $C$  рассмотрение диаграммы Венна (рис.1.2) позволяет записать равенство

$$\begin{aligned} \mathbf{P}(A \cup B \cup C) = & \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C) - \\ & - \mathbf{P}(A \cap B) - \mathbf{P}(A \cap C) - \mathbf{P}(B \cap C) + \mathbf{P}(A \cap B \cap C). \end{aligned}$$

В частности, если события  $A, B, C$  попарно несовместны, то

$$\mathbf{P}(A \cup B \cup C) = \mathbf{P}(A) + \mathbf{P}(B) + \mathbf{P}(C).$$

Последнее утверждение легко распространяется на случай любого конечного и даже счетного множества попарно несовместных событий: если при  $i \neq j$   $A_i \cap A_j = \emptyset$ , то

$$\mathbf{P}\left(\bigcup_i A_i\right) = \sum_i \mathbf{P}(A_i).$$

Формула для вероятности объединения любого числа *совместных* событий более сложна, и мы ее не приводим.

## 1.4. Статистическая зависимость событий. Условные вероятности

Пусть  $A, B \subset \Omega$ . Будем называть случайные события  $A$  и  $B$  *статистически независимыми* (обычно говорят короче – независимыми), если

$$\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B). \quad (1.4.1)$$

Если  $\mathbf{P}(A \cap B) \neq \mathbf{P}(A) \cdot \mathbf{P}(B)$ , то события  $A$  и  $B$  называют (статистически) зависимыми.

Несложно показать (проверьте это!), что если события  $A$  и  $B$  независимы, то события  $\Omega \setminus A$  и  $B$  также независимы.

Пример. Эксперимент состоит в бросании правильной игральной кости. Исходы – выпадение граней, помеченных цифрами от 1 до 6. Здесь  $\Omega = \{1, 2, 3, 4, 5, 6\}$ , и  $\mathbf{P}(\omega) = 1/6$  для  $\omega = 1, \dots, 6$ .

Рассмотрим события:  $A$  – "результат четный"  $B$  – "результат больше двух"  $C$  – "результат меньше четырех".

$$\begin{aligned} A &= \{2, 4, 6\}, & \mathbf{P}(A) &= 3 \cdot 1/6 = 1/2; \\ B &= \{3, 4, 5, 6\}, & \mathbf{P}(B) &= 4 \cdot 1/6 = 2/3; \\ C &= \{1, 2, 3\}, & \mathbf{P}(C) &= 3 \cdot 1/6 = 1/2; \\ A \cap B &= \{4, 6\}, & \mathbf{P}(A \cap B) &= 2 \cdot 1/6 = 1/3; \\ A \cap C &= \{2\}, & \mathbf{P}(A \cap C) &= 1/6. \end{aligned}$$

События  $A$  и  $B$  независимы, так как

$$\mathbf{P}(A) \cdot \mathbf{P}(B) = 1/2 \cdot 2/3 = 1/3 = \mathbf{P}(A \cap B).$$

События  $A$  и  $C$  зависимы, так как

$$\mathbf{P}(A) \cdot \mathbf{P}(C) = 1/2 \cdot 1/2 = 1/4 \neq 1/6 = \mathbf{P}(A \cap C).$$

Замечание. Независимость (статистическая) событий  $A$  и  $B$  определена равенством (1.4.1), и не следует истолковывать ее каким-либо другим способом (например, делая из нее выводы о причинной, национальной, финансовой и т. п. независимости). Это относится и к статистической зависимости.

Пусть теперь  $A, B \subset \Omega$ , и  $\mathbf{P}(A) \neq 0$ . Тогда число  $\frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)}$  называется *условной вероятностью события  $B$  при условии, что событие  $A$  произошло*. Для условной вероятности принято обозначение

$$\mathbf{P}(B|A) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)}. \quad (1.4.2)$$



В отличие от *условной вероятности*  $\mathbf{P}(B|A)$  события  $B$  *вероятность* этого события  $\mathbf{P}(B)$  иногда называют *безусловной*.

Равенство (1.4.2) может быть переписано в виде

$$\mathbf{P}(B \cap A) = \mathbf{P}(A) \cdot \mathbf{P}(B|A).$$

Это утверждение носит название *теоремы умножения вероятностей*.

Покажем, что переход к условной вероятности события равносильен изменению пространства элементарных событий. Пусть  $A, B \subset \Omega$ , и  $\mathbf{P}(A) \neq 0$  (рис.1.1). Слова "событие  $A$  произошло" означают, что эксперимент закончился одним из исходов (неизвестно каким), составляющих событие  $A$ . Поэтому можно *исключить из рассмотрения* все элементарные события, не входящие в  $A$ , т.е. считать  $A$  *новым* пространством элементарных событий  $\Omega' = A$  (рис.1.3).

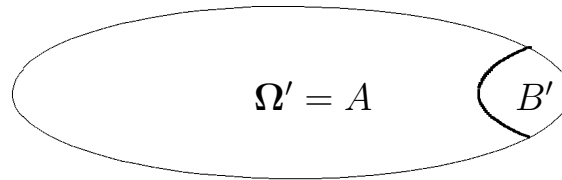


Рис.1.3

Если (что естественно) *сохранить соотношения* между вероятностями элементарных событий, т.е. изменить эти вероятности *пропорционально* ( $\mathbf{P}'(\omega) = k \cdot \mathbf{P}(\omega)$ ), то сумма вероятностей всех элементарных событий, составляющих новое пространство, должна равняться единице. Отсюда

$$\sum_{\omega \in A} \mathbf{P}'(\omega) = k \cdot \sum_{\omega \in A} \mathbf{P}(\omega) = k \cdot \mathbf{P}(A) = 1, \quad \text{т.е.} \quad k = \frac{1}{\mathbf{P}(A)}.$$

Мы построили новое пространство элементарных событий  $\Omega' = A$  с вероятностями элементарных событий

$$\mathbf{P}'(\omega) = k \cdot \mathbf{P}(\omega) = \frac{\mathbf{P}(\omega)}{\mathbf{P}(A)}.$$

В новом пространстве  $\Omega'$  событию  $B$  исходного пространства соответствует событие  $B' = B \cap A$  (рис.1.3), и

$$\begin{aligned} \mathbf{P}'(B') &= \sum_{\omega \in B'} \mathbf{P}'(\omega) = \sum_{\omega \in B'} \frac{\mathbf{P}(\omega)}{\mathbf{P}(A)} = \\ &= \frac{1}{\mathbf{P}(A)} \cdot \sum_{\omega \in B \cap A} \mathbf{P}(\omega) = \frac{\mathbf{P}(B \cap A)}{\mathbf{P}(A)} = \mathbf{P}(B|A). \end{aligned}$$

Пример. Бросается правильная игральная кость. Найдем условную вероятность события "результат больше единицы" при условии, что выпала нечетная цифра.

Построим пространство элементарных событий  $\Omega = \{1, 2, 3, 4, 5, 6\}$  с равными вероятностями  $\mathbf{P}(\omega) \equiv 1/6$ . Рассмотрим события  $A = \{1, 3, 5\}$ ,  $B = \{2, 3, 4, 5, 6\}$  и  $A \cap B = \{3, 5\}$ . Их вероятности

$$\mathbf{P}(A) = 3 \cdot 1/6 = 1/2, \quad \mathbf{P}(B) = 5 \cdot 1/6 = 5/6, \quad \mathbf{P}(A \cap B) = 2 \cdot 1/6 = 1/3.$$

$$\text{Отсюда } \mathbf{P}(B|A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = 2/3.$$

Эту же задачу можно решить, рассмотрев пространство элементарных событий  $\Omega' = \{1, 3, 5\}$ . В этом пространстве каждому элементарному событию следует приписать вероятность  $\mathbf{P}'(\omega) = \mathbf{P}(\omega)/\mathbf{P}(A) \equiv 1/3$ . Тогда  $B' = \{3, 5\}$ , и  $\mathbf{P}'(B') = 2 \cdot 1/3 = 2/3$ .

Если события  $A$  и  $B$  независимы, то из (1.4.1) и (1.4.2) следует, что  $\mathbf{P}(B|A) = \mathbf{P}(B)$ . Поскольку события  $\Omega \setminus A$  и  $B$  также независимы, получаем  $\mathbf{P}(B|A) = \mathbf{P}(B|(\Omega \setminus A))$ , т.е. условные вероятности события  $B$  при условиях, что событие  $A$  произошло и не произошло, равны. Этот факт обычно интерпретируется так: относительная частота события  $B$  не зависит от того, рассматриваем ли мы все проведенные эксперименты, или выбираем только те, в которых произошло (не произошло) событие  $A$ . Термин "независимые события" возник именно из этой интерпретации.

В приложениях статистическая независимость обычно *постулируется постановщиком задачи* при построении математической модели из содержательных предположений. Такой постулат дает возможность легко находить вероятности пересечения событий. Однако за адекватность этого постулата реальности отвечает *постановщик задачи*, а не теория вероятностей.

Пример. Эксперимент состоит в двух бросаниях правильной монеты. Здесь  $\Omega = \{ГГ, ГР, РГ, РР\}$ . Поскольку первый и второй броски не влияют друг на друга, разумно предположить, что события  $A$  "при первом броске выпала решетка" и  $B$  "при втором броске выпал герб" независимы. Тогда из  $\mathbf{P}(A) = \mathbf{P}(B) = 1/2$  имеем  $\mathbf{P}(РГ) = \mathbf{P}(A) \cdot \mathbf{P}(B) = 1/4$  (конечно, вероятности остальных элементарных событий тоже равны  $1/4$ ).

Аналогично, при  $n$  бросаниях правильной монеты из тех же соображений получаем  $\mathbf{P}(Р \dots РГ) = 2^{-n}$ . Именно поэтому в примере 2 п.1.2 разумно положить  $\mathbf{P}(\omega_n) = 2^{-n}$ .

## 1.5. Формула полной вероятности

Пусть пространство элементарных событий представлено в виде объединения попарно несовместных событий  $A_i$ ,  $i = 1, \dots, n$ :

$$\Omega = A_1 \cup A_2 \cup \dots \cup A_n, \quad A_i \cap A_j = \emptyset \quad \text{при} \quad i \neq j.$$

Тогда легко видеть, что для любого события  $B$

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n),$$

причем события  $(B \cap A_i)$  попарно несовместны.

Из теорем сложения и умножения вероятностей имеем

$$\mathbf{P}(B) = \sum_{i=1}^n \mathbf{P}(B \cap A_i) = \sum_{i=1}^n \mathbf{P}(A_i) \cdot \mathbf{P}(B|A_i). \quad (1.5.1)$$

Эта формула именуется *формулой полной вероятности*. Она часто применяется в тех случаях, когда из содержательных соображений легче задать условные вероятности, чем безусловные. События  $A_i$  часто называют *гипотезами*.

Пример. Имеются три кости, из них две правильных, а третья налита свинцом, так что шестерка на ней выпадает вдвое чаще, чем на правильной. На ощупь кости неразличимы. Эксперимент заключается в бросании одной наугад выбранной кости.

Определим события:  $A_1$  – "выбрана правильная кость",  $A_2$  – "выбрана неправильная кость",  $B$  – "выброшена шестерка". Тогда из данных задачи легко определить условные вероятности:  $\mathbf{P}(B|A_1) = 1/6$ ,  $\mathbf{P}(B|A_2) = 1/3$ . Далее, из условия разумно предположить, что при многократном выборе кости правильная будет попадаться примерно в два раза чаще, чем неправильная. Поэтому  $\mathbf{P}(A_1) = 2/3$ ,  $\mathbf{P}(A_2) = 1/3$ . По формуле (1.5.1) получаем  $\mathbf{P}(B) = 2/3 \cdot 1/6 + 1/3 \cdot 1/3 = 2/9$ .

## 1.6. Пространство элементарных событий с несчетным множеством исходов

Рассмотрим задачу об измерении некоторой физической величины с помощью стрелочного прибора. Если считать стрелку прибора не имеющей толщины, то результатом измерения (элементарным событием) может быть любая точка на шкале, и множество всех исходов  $\Omega$  не является счетным. *Можно показать*, что в этом случае невозможно приписать каждому элементарному событию ненулевую вероятность и обеспечить при этом конечность  $\mathbf{P}(\Omega)$ .

В такой ситуации отказываются от приписывания вероятности каждому элементарному событию. Вероятность приписывается теперь только *событиям*, к которым относят *не любые* части  $\Omega$ .

Точнее: пусть задано пространство элементарных событий  $\Omega$  с несчетным числом исходов. Строится семейство  $\mathfrak{S}$  частей  $\Omega$ , удовлетворяющее следующим условиям:

1.  $\Omega \in \mathfrak{S}$ ;
2. если  $A \in \mathfrak{S}$ , то  $\Omega \setminus A \in \mathfrak{S}$  (в частности,  $\emptyset = \Omega \setminus \Omega \in \mathfrak{S}$ );
3. если  $A_k \in \mathfrak{S}$  (набор этих множеств может быть конечным или счетным), то

$$\bigcap_k A_k \in \mathfrak{S} \quad \text{и} \quad \bigcup_k A_k \in \mathfrak{S}$$

(в частности, если  $A, B \in \mathfrak{S}$ , то  $A \setminus B = A \cap (\Omega \setminus B) \in \mathfrak{S}$ ).

Случайными событиями теперь называются не любые части  $\Omega$ , а только входящие в  $\mathfrak{S}$ , и только им приписываются вероятности – неотрицательные числа, удовлетворяющие условиям:

4.  $\mathbf{P}(\Omega) = 1$ ;
5. если  $A_k \in \mathfrak{S}$  (набор событий может быть конечным или счетным), и эти события попарно несовместны (не пересекаются), то

$$\mathbf{P}\left(\bigcup_k A_k\right) = \sum_k \mathbf{P}(A_k).$$

Замечание. Из свойств 4 и 5 следует, что для  $A \in \mathfrak{S}$

$$\mathbf{P}(\Omega \setminus A) = 1 - \mathbf{P}(A).$$

В частности,  $\mathbf{P}(\emptyset) = 0$ .

Несложно также показать (проверьте это!), что для  $A, B \in \mathfrak{S}$  выполнены теоремы сложения и умножения вероятностей. Формула полной вероятности также справедлива, если все входящие в нее множества принадлежат семейству  $\mathfrak{S}$ . Более того, формула (1.5.1) распространяется на счетный набор попарно несовместных событий  $A_i \in \mathfrak{S}$ .

Перечисленные выше пять свойств семейства  $\mathfrak{S}$  и функции  $\mathbf{P}$  называются *аксиомами теории вероятностей*. Специалист-нематематик при построении модели эксперимента, использующей теорию вероятностей, вообще говоря, ограничен только этими аксиомами. Однако при измерении физической величины естественно считать, что семейство  $\mathfrak{S}$  содержит любые промежутки (сегменты, интервалы, полуинтервалы),

целиком лежащие в множестве возможных значений этой величины. Тогда, согласно аксиоме **3**, это семейство должно содержать также каждое множество, которое можно получить из промежутков с помощью конечного или счетного числа операций объединения и пересечения. Понятно, что все множества, представляющие хоть какой-нибудь практический интерес, попадут в такое семейство.

Как и в случае дискретного пространства элементарных событий, выбор функции  $\mathbf{P} : \mathfrak{S} \rightarrow \mathbb{R}$ , удовлетворяющей условиям **4** и **5**, делается из содержательных предположений об условиях эксперимента.

Пример. В случайный момент времени измеряется мгновенное значение синусоидального напряжения

$$u = U_{max} \cdot \sin(\omega t).$$

Какова вероятность события "измеренное значение по модулю меньше половины амплитуды"?

Вследствие периодичности синусоиды *будем считать* пространством элементарных событий промежутки  $] -\frac{\pi}{\omega}, \frac{\pi}{\omega}]$  (рис.1.4). Случайность момента времени в математической модели обычно интерпретируют так: при многократном повторении эксперимента относительные частоты попадания в равные промежутки времени одинаковы. Поэтому каждому промежутку, содержащемуся в  $] -\frac{\pi}{\omega}, \frac{\pi}{\omega}]$ , *припишем* вероятность, пропорциональную длине этого промежутка. Тогда вероятность попадания в промежуток длины  $a$  будет равна  $\frac{a\omega}{2\pi}$  (независимо от того, входят граничные точки в промежуток или нет). Если событие состоит из нескольких непересекающихся промежутков, то, согласно аксиоме **3**, вероятность этого события равна сумме длин этих промежутков, умноженной на  $\frac{\omega}{2\pi}$ .

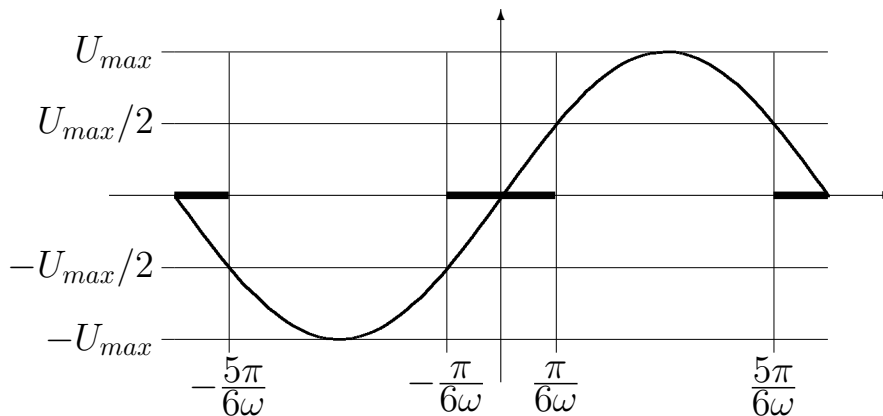


Рис.1.4

На рис.1.4 выделены промежутки оси абсцисс, на которых выполняется условие  $|u| < U_{max}/2$ . Видно, что суммарная длина этих промежутков  $4 \times \frac{\pi}{6\omega} = \frac{2\pi}{3\omega}$ . Поэтому

$$\mathbf{P}\left(|u| < \frac{U_{max}}{2}\right) = \frac{2\pi}{3\omega} \cdot \frac{\omega}{2\pi} = 1/3.$$

Замечание. Мы отождествили "событие" из содержательной задачи ( $|u| < U_{max}/2$ ) и случайное событие – элемент семейства  $\mathfrak{S}$  :

$$\left] -\frac{\pi}{\omega}, -\frac{5\pi}{6\omega} \right[ \cup \left] -\frac{\pi}{6\omega}, \frac{\pi}{6\omega} \right[ \cup \left] \frac{5\pi}{6\omega}, \frac{\pi}{\omega} \right[.$$

Так поступают часто, чтобы сэкономить на записи.

## Глава 2. СЛУЧАЙНЫЕ ПЕРЕМЕННЫЕ

### 2.1. Основные определения

Рассмотрим эксперимент, состоящий в  $n$  бросаниях правильной монеты. Построим пространство элементарных событий при  $n = 3$ :

$$\Omega = \{PPR, PRP, RPP, PRR, RRP, RPR, RRR, PRR\}.$$

Это пространство содержит 8 элементарных событий. Правильность монеты дает основание считать элементарные события равновероятными (см. пример в конце п.1.4). В таблице 2.1 представлены элементарные события, их вероятности и соответствующие этим событиям количества выпадений герба ( $X$ ).

Таблица 2.1

| $\omega$             | PPR | PRP | RPP | PRR | RRP | RPR | RRR | RRR |
|----------------------|-----|-----|-----|-----|-----|-----|-----|-----|
| $\mathbf{P}(\omega)$ | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 | 1/8 |
| $X$                  | 0   | 1   | 1   | 2   | 1   | 2   | 2   | 3   |

Видно, что  $X$  может принимать 4 различных значения: 0, 1, 2, 3. Поэтому можно рассмотреть события  $(X = 0)$ ,  $(X = 1)$ ,  $(X = 2)$ ,  $(X = 3)$  и вычислить их вероятности. Результаты сведены в таблицу 2.2.

Таблица 2.2

| $x$                 | 0   | 1   | 2   | 3   |
|---------------------|-----|-----|-----|-----|
| $\mathbf{P}(X = x)$ | 1/8 | 3/8 | 3/8 | 1/8 |

Если целью эксперимента было определение количества выпадений герба, то естественно рассмотреть новое пространство элементарных событий – множество значений  $X$ :  $\Omega' = \{0, 1, 2, 3\}$  – и заданные на нем таблицей 2.2 вероятности.

Ситуация, когда исходом эксперимента можно считать число, часто встречается в прикладных задачах.

Определение. *Дискретной случайной переменной* называется конечное или счетное *числовое* множество (пространство элементарных событий) с заданными на нем вероятностями.

Замечание. Введенное определение не исключает вырожденной ситуации, когда пространство элементарных событий состоит из одного элемента, вероятность которого равна единице. Такая случайная переменная уже будет "не случайной" и мы исключим ее из рассмотрения.

Вспомним понятие *переменная*: переменная – это буква и множество, элементы которого можно подставлять вместо этой буквы. Переменную можно интерпретировать как мешок, содержащий все возможные ее значения. Каждый имеет право выбрать в этом мешке понравившееся ему значение переменной. Аналогично, *случайной переменной* назовем букву и (числовое) множество ее значений, которое мы будем рассматривать как пространство элементарных событий. Дискретную случайную переменную также можно интерпретировать как мешок, содержащий все возможные ее значения. Однако *выбирать эти значения нельзя*. Можно запустить в этот мешок руку и вынуть из него то значение случайной переменной, которое *случайно* попадет. Если повторять этот эксперимент *многократно*, то относительные частоты появления различных значений случайной переменной будут, *как правило, мало* отличаться от их вероятностей.

Терминологическое замечание. В книгах по теории вероятностей часто вместо слов "случайная переменная" используется термин "случайная величина".

В дальнейшем мы вместо "случайная переменная" будем писать с.п.; случайные переменные обозначаются латинскими заглавными буквами, а их значения – *числа* – соответствующими строчными буквами.

Итак, дискретная с.п.  $X$  задается множеством своих значений и вероятностями этих значений, т.е. таблицей, которую называют *законом распределения* случайной переменной:

|                     |       |       |         |
|---------------------|-------|-------|---------|
| $x$                 | $x_1$ | $x_2$ | $\dots$ |
| $\mathbf{P}(X = x)$ | $p_1$ | $p_2$ | $\dots$ |

$$p_k > 0; \quad \sum_k p_k = 1.$$

Полезна следующая физическая интерпретация дискретной с.п.: значения с.п. рассматриваются как абсциссы точек на оси, а соответствующие вероятности – как массы, сосредоточенные в этих точках. Таким образом, с.п. интерпретируется как система точечных масс на оси. Существенно, что масса всей системы всегда равна единице. Можно сказать, что в нашем распоряжении есть единичная "вероятностная масса и мы должны "распределить" ее между заданными точками оси – задать *закон распределения вероятностей*.

Эта физическая аналогия помогает найти способ описания с.п. с несчетным множеством значений. Известно, что физика оперирует в таком случае понятием *плотности массы*, считая, что масса в каждой



точке равна нулю, но масса отрезка равна интегралу от плотности по этому отрезку. По аналогии мы будем рассматривать так называемую *абсолютно непрерывную* с.п.  $X$ , множество значений которой  $\Omega = \mathbb{R}$ . Мы будем считать, что задана кусочно непрерывная функция (*плотность распределения* с.п.  $X$ )  $f_X : \mathbb{R} \rightarrow \mathbb{R}$ , обладающая двумя свойствами:

$$1) f_X \geq 0; \quad 2) \int_{-\infty}^{+\infty} f_X = 1. \quad (2.1.1)$$

В этой модели вероятность попадания с.п. в промежуток  $[a, b]$  будет вычисляться по формуле

$$\mathbf{P}(X \in [a, b]) = \int_a^b f_X.$$

Замечания. 1. Мы записали  $X \in [a, b]$ , считая промежуток *сегментом*. Но если вспомнить, что значение интеграла не зависит от значений подынтегральной функции в конечном числе точек, то будет ясно, что тип промежутка (интервал, полуинтервал, сегмент) не играет роли.

2. Очевидно,  $\mathbf{P}(X = a) = \int_a^a f_X = 0$ , т.е. вероятность каждого отдельного значения для абсолютно непрерывной с.п. равна нулю.

3. В приложениях встречаются так называемые дискретно-непрерывные с.п., но их рассмотрение выходит за рамки нашего курса.

Одну из первообразных плотности распределения с.п., а именно,

$$F_X(x) = \int_{-\infty}^x f_X = \mathbf{P}(X < x), \quad (2.1.2)$$

называют *функцией распределения* с.п.  $X$ . Из свойств первообразной следует, что  $F_X$  – непрерывная, неубывающая на  $\mathbb{R}$  функция,  $F_X(-\infty) = 0$ ,  $F_X(+\infty) = 1$ . В точках, где непрерывна плотность  $f_X$ , функция распределения имеет производную, причем  $F'_X = f_X$ .

## 2.2. Числовые характеристики случайных переменных

В механике для компактного описания точечной системы масс используются так называемые *моменты*. Статический момент системы масс относительно начала координат равен, как известно,  $\sum_k x_k \cdot m_k$ . Здесь  $x_k$  – абсцисса точки, а  $m_k$  – масса, сосредоточенная в этой точке.

По аналогии, *первым начальным моментом*, или *математическим ожиданием* дискретной с.п.  $X$  называют число

$$M(X) = \sum_k x_k \cdot p_k,$$

где  $x_k$  – значение с.п.  $X$ , а  $p_k$  – вероятность этого значения.

Для абсолютно непрерывной с.п.  $X$  с плотностью  $f_X$  понятие математического ожидания вводится так:

$$M(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx.$$

Замечания. 1. Если множество значений дискретной с.п. счетно, то первый момент представляет собой *сумму ряда*. В этом случае выдвигается требование *абсолютной* сходимости ряда, так как если ряд сходится не абсолютно, то его сумма зависит от порядка слагаемых, а это невозможно интерпретировать в содержательных задачах. Для абсолютно непрерывной с.п. требуется абсолютная сходимость интеграла, определяющего первый момент. Если ряд (интеграл) не является абсолютно сходящимся, говорят, что с.п. *не имеет математического ожидания*.

2. Учитывая, что  $\sum_k p_k = 1$ , математическое ожидание дискретной с.п. можно записать так:

$$M(X) = \left( \sum_k x_k \cdot p_k \right) / \left( \sum_k p_k \right).$$

Аналогично, для абсолютно непрерывной с.п.

$$M(X) = \left( \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \right) / \left( \int_{-\infty}^{+\infty} f_X(x) dx \right)$$

(сравните с физическим понятием центра масс).

3. В математической литературе вместо  $M(X)$  часто используется обозначение  $E(X)$  (от слова "expectation").

В физике разброс массы системы относительно ее центра масс характеризуется моментом инерции. По аналогии вводится понятие *второго центрального момента*, или *дисперсии* с.п.  $X$ :

для дискретной с.п.  $X$

$$\mu_2(X) = D(X) = \sum_k (x_k - M(X))^2 \cdot p_k, \quad (2.2.1)$$

для абсолютно непрерывной –

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx. \quad (2.2.1')$$

Очевидно, есть смысл говорить о дисперсии с.п. только при наличии у нее математического ожидания. Но даже в этом случае дисперсия не обязана существовать (ряд или интеграл может расходиться).

Теорема. Дисперсия с.п. положительна (если она существует).

Доказательство. Интеграл в формуле (2.2.1'), очевидно, неотрицателен. Более того, если он равен нулю, то подынтегральная функция равна нулю почти всюду, чего не может быть ввиду условия (2.1.1).

Аналогично, сумма в (2.2.1) может равняться нулю только в том случае, когда с.п.  $X$  принимает единственное значение  $M(X)$  с вероятностью 1. Но такую вырожденную ситуацию мы, как уже говорилось, не рассматриваем. ■

Если у с.п. есть дисперсия, то *среднеквадратическим отклонением* (иногда говорят *отклонением*) этой с.п. называют число

$$\sigma(X) = \sqrt{D(X)}.$$

Среднеквадратическое отклонение есть мера разброса значений с.п. относительно ее математического ожидания.

Докажем теперь важное неравенство. Пусть  $X$  – с.п., имеющая дисперсию. Тогда для любого  $\varepsilon > 0$

$$\mathbf{P} (|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}$$

(при малых  $\varepsilon$  ( $\varepsilon \leq \sigma(X)$ ) это неравенство становится тривиальным).

Мы ограничимся случаем абсолютно непрерывной с.п. Неравенство

$$\begin{aligned} D(X) &= \int_{-\infty}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx \geq \\ &\geq \int_{-\infty}^{M(X)-\varepsilon} (x - M(X))^2 \cdot f_X(x) dx + \int_{M(X)+\varepsilon}^{+\infty} (x - M(X))^2 \cdot f_X(x) dx \end{aligned}$$

следует из неотрицательности подынтегральной функции. Учитывая, что в подынтегральных выражениях  $|x - M(X)| \geq \varepsilon$ , можно записать

$$\begin{aligned}
D(X) &\geq \varepsilon^2 \cdot \left( \int_{-\infty}^{M(X)-\varepsilon} f_X(x) dx + \int_{M(X)+\varepsilon}^{+\infty} f_X(x) dx \right) = \\
&= \varepsilon^2 \cdot \int_{|x-M(X)| \geq \varepsilon} f_X(x) dx = \varepsilon^2 \cdot \mathbf{P}(|X - M(X)| \geq \varepsilon).
\end{aligned}$$

Доказанное неравенство именуется *неравенством Чебышева*. Его часто интерпретируют так: *большие отклонения с.п. от ее математического ожидания маловероятны*.

Математическое ожидание и дисперсия в какой-то мере характеризуют закон распределения с.п. Однако следует понимать, что существует сколько угодно законов распределения с данными математическим ожиданием и дисперсией. В то же время есть задачи, для решения которых достаточно знать только эти числовые характеристики.

Укажем еще одну часто встречающуюся числовую характеристику. *Модой* распределения абсолютно непрерывной с.п. называют абсциссу локального максимума ее плотности распределения. Если распределение имеет не одну моду, то его называют *многомодальным*.

### 2.3. Некоторые законы распределения, встречающиеся в приложениях

1. Биномиальное распределение Бернулли<sup>26</sup>. Так называется семейство дискретных распределений с двумя параметрами  $n \in \mathbb{N}$  и  $p \in ]0, 1[$ :

$$\Omega = \{0, 1, \dots, n\}; \quad \mathbf{P}(X = m) = \binom{n}{m} \cdot p^m \cdot (1-p)^{n-m}; \quad m = 0, 1, \dots, n.$$

Здесь  $\binom{n}{m} = \frac{n!}{m! \cdot (n-m)!}$  – число сочетаний<sup>27</sup> из  $n$  элементов по  $m$ .

Содержательная интерпретация биномиального распределения: эксперимент может заканчиваться одним из двух исходов: успех или неудача. Вероятность успеха равна  $p$ . Эксперимент повторяют  $n$  раз. С.п. – количество успехов.

<sup>26</sup>БЕРНУЛЛИ (Bernoulli) – семья швейцарских ученых. Распределение носит имя Якоба I БЕРНУЛЛИ (1654-1705), благодаря работам которого теория вероятностей стала играть важную роль в приложениях.

<sup>27</sup>В отечественной литературе вместо  $\binom{n}{m}$  часто пишут  $C_n^m$  (обратите внимание на расположение букв  $n$  и  $m$  в этих символах!).

Название "биномиальное" объясняется тем, что вероятности элементарных событий совпадают с членами бинома Ньютона  $(a + b)^n$  при  $a = p$ ,  $b = 1 - p$ . Отсюда следует, что

$$\sum_{m=0}^n \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = 1.$$

Для вычисления математического ожидания и дисперсии биномиального распределения нужна некоторая техника суммирования, которая не имеет отношения к теории вероятностей. Поэтому здесь, как и в дальнейшем, мы лишь приводим результаты:

$$M(X) = \sum_{m=0}^n m \cdot \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = n \cdot p,$$

$$D(X) = \sum_{m=0}^n (m - n \cdot p)^2 \cdot \binom{n}{m} \cdot p^m \cdot (1 - p)^{n-m} = n \cdot p \cdot (1 - p).$$

Мы вернемся к этому примеру в п.2.6.

Отметим очевидное неравенство  $p \cdot (1 - p) = 0.25 - (p - 0.5)^2 \leq 0.25$ , из которого следует  $D(X) \leq 0.25 \cdot n$ .

2. Геометрическое распределение – однопараметрическое семейство дискретных распределений с параметром  $p \in ]0, 1[$ :

$$\Omega = \mathbb{N}; \quad \mathbf{P}(X = n) = p \cdot (1 - p)^{n-1}.$$

Содержательная интерпретация геометрического распределения: вероятность успеха в одном эксперименте равна  $p$ ; эксперимент повторяют до первого успеха. С.п. – количество повторений эксперимента<sup>28</sup>.

Легко видеть, что

$$\sum_{n=1}^{+\infty} p \cdot (1 - p)^{n-1} = \frac{p}{1 - (1 - p)} = 1.$$

Далее, вычисление показывает, что

$$M(X) = \sum_{n=1}^{+\infty} n \cdot p \cdot (1 - p)^{n-1} = \frac{1}{p},$$

$$D(X) = \sum_{n=1}^{+\infty} (n - 1/p)^2 \cdot p \cdot (1 - p)^{n-1} = \frac{1 - p}{p^2}.$$

---

<sup>28</sup>При  $p = 1/2$  это распределение уже встречалось в главе 1.

3. Распределение Пуассона – однопараметрическое семейство дискретных распределений с параметром  $\lambda > 0$ :

$$\Omega = \mathbb{N} \cup \{0\}; \quad \mathbf{P}(X = n) = \exp(-\lambda) \cdot \frac{\lambda^n}{n!}.$$

Очевидно, что

$$\sum_{n=0}^{+\infty} \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \exp(-\lambda) \cdot \exp(\lambda) = 1.$$

Далее, вычисление показывает, что

$$M(X) = \sum_{n=0}^{+\infty} n \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \lambda;$$

$$D(X) = \sum_{n=0}^{+\infty} (n - \lambda)^2 \cdot \exp(-\lambda) \cdot \frac{\lambda^n}{n!} = \lambda.$$

4. Равномерное распределение – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами  $x_0 \in \mathbb{R}$  и  $\Delta > 0$ :

$$f_X(x) = \begin{cases} \text{const}, & \text{если } |x - x_0| \leq \Delta, \\ 0, & \text{если } |x - x_0| > \Delta. \end{cases}$$

Таким образом, "вероятностная масса" (равная 1) *равномерно* распределена по промежутку длиной  $2\Delta$  (с центром в точке  $x_0$ ). Поэтому значение константы равно  $\frac{1}{2\Delta}$ .

Одна из возможных интерпретаций: равномерно распределенной считают (по-видимому, не без основания) погрешность округления в цифровых измерительно-вычислительных устройствах.

Очевидно,

$$M(X) = \int_{x_0-\Delta}^{x_0+\Delta} x \cdot \frac{1}{2\Delta} dx = x_0; \quad D(X) = \int_{x_0-\Delta}^{x_0+\Delta} (x - x_0)^2 \cdot \frac{1}{2\Delta} dx = \frac{\Delta^2}{3}.$$

5. Нормальное распределение – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами  $m \in \mathbb{R}$  и  $\sigma > 0$  (рис.2.1):

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(x - m)^2}{2\sigma^2}\right).$$

Убедитесь в том, что  $\int_{-\infty}^{+\infty} f_X = 1$ , используя известный из курса ма-

тематического анализа интеграл Пуассона  $\int_{-\infty}^{+\infty} \exp(-y^2) dy = \sqrt{\pi}$ .

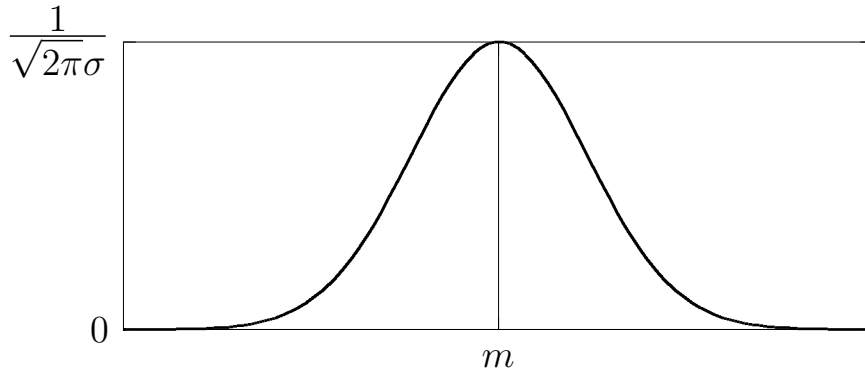


Рис.2.1. Плотность нормального распределения с параметрами  $m$  и  $\sigma$

Можно показать, что

$$M(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{+\infty} x \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = m;$$

$$D(X) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \int_{-\infty}^{+\infty} (x-m)^2 \cdot \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx = \sigma^2.$$

6. Распределение Коши – двухпараметрическое семейство абсолютно непрерывных распределений с параметрами  $x_0 \in \mathbb{R}$  и  $a > 0$  (рис.2.2):

$$f_X(x) = \frac{a}{\pi} \cdot \frac{1}{a^2 + (x - x_0)^2}.$$

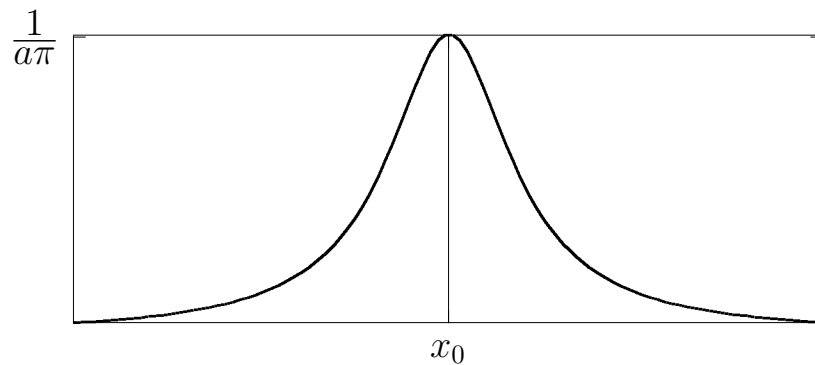


Рис.2.2. Плотность распределения Коши с параметрами  $x_0$  и  $a$

Легко видеть, что

$$\int_{-\infty}^{+\infty} f_X = \frac{1}{\pi} \cdot \arctg\left(\frac{x - x_0}{a}\right) \Big|_{x=-\infty}^{x=+\infty} = 1.$$

Сравнивая рис.2.1 и рис.2.2, можно подумать, что распределение Коши очень похоже на нормальное. Однако интеграл

$$\int_{-\infty}^{+\infty} x \cdot \frac{a}{\pi} \cdot \frac{1}{a^2 + (x - x_0)^2} dx$$

расходится, и следовательно, у распределения Коши нет ни математического ожидания, ни дисперсии. В то же время распределение Коши имеет единственную моду ( $x_0$ ).

## 2.4. Случайный вектор

Случайная переменная может использоваться как математическая модель эксперимента, в котором измеряется значение физической величины. Для описания эксперимента, в котором одновременно фиксируются значения *нескольких* величин, служит понятие *случайного вектора*<sup>29</sup>.

Дискретный случайный вектор задается таблицей, в которой перечислены все возможные значения вектора, и каждому значению приписана вероятность.

Абсолютно непрерывный случайный вектор  $X$  задается плотностью его распределения – неотрицательной кусочно непрерывной функцией  $f_X : \mathbb{R}^n \rightarrow \mathbb{R}$ , такой, что

$$\int_{\mathbb{R}^n} f_X = 1. \quad (2.4.1)$$

При этом каждой области  $G \subset \mathbb{R}^n$  с кусочно гладкой границей соответствует вероятность

$$\mathbf{P}(X \in G) = \int_G f_X.$$

Серьезное предупреждение. Обратите внимание, что задание нескольких случайных переменных, вообще говоря, *недостаточно* для задания случайного вектора.

Примеры. 1. Бросается правильная кость. Случайная переменная  $X_1$  равна единице, если результат четный, и нулю в противном случае. Легко видеть, что  $\mathbf{P}(X_1 = 0) = \mathbf{P}(X_1 = 1) = 1/2$ .

Определим еще с.п.  $X_2$ , равную единице, если на кости выпало число, большее трех, и нулю в противном случае. Тогда очевидно, что  $\mathbf{P}(X_2 = 0) = \mathbf{P}(X_2 = 1) = 1/2$ .

<sup>29</sup> в дальнейшем мы будем иногда использовать сокращение "с.в."



Если провести две различные серии экспериментов, в одной из которых определяется с.п.  $X_1$ , а в другой –  $X_2$ , то можно ожидать, что для каждой из них относительные частоты единиц и нулей будут близки к  $1/2$ . Но отличить их друг от друга или увидеть связь между ними по результатам этих экспериментов невозможно.

Если же в каждом эксперименте определять и  $X_1$ , и  $X_2$ , то (проверьте это!) мы получим двумерный дискретный случайный вектор  $X$ , заданный таблицей:

|                      |     |     |
|----------------------|-----|-----|
| $x_2 \backslash x_1$ | 0   | 1   |
| 0                    | 1/3 | 1/6 |
| 1                    | 1/6 | 1/3 |

Множество значений этого случайного вектора (пространство элементарных событий) составляют четыре числовых вектора:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Их вероятности записаны в соответствующих клетках таблицы.

Приведем два примера абсолютно непрерывных случайных векторов, часто встречающихся на практике.

Примеры. 2. Двумерный случайный вектор, равномерно распределенный в ограниченной области  $G$  с кусочно гладкой границей, задается плотностью распределения:

$$f_X : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_X(x) = \begin{cases} const, & \text{если } x \in G, \\ 0, & \text{если } x \notin G. \end{cases}$$

Таким образом, единичная "вероятностная масса" *равномерно* распределена по области  $G$ . Поэтому значение константы равно  $\frac{1}{S(G)}$ .

Аналогично определяется  $n$ -мерный случайный вектор, равномерно распределенный в  $n$ -мерной области.

3.  $n$ -мерный *нормально* распределенный случайный вектор задается плотностью распределения

$$f_X(x) = \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1}(x - m), (x - m) \rangle\right). \quad (2.4.2)$$

Здесь  $m = [m_1, \dots, m_n]^T \in \mathbb{R}^n$ ,  $B$  – симметричная, положительно определенная  $(n \times n)$ -матрица.

Проверим, что  $\int_{\mathbb{R}^n} f_X = 1$ . В курсе линейной алгебры было показано, что положительно определенная матрица допускает разложение Холецкого:

$$B = H^*H,$$

где  $H$  – верхняя треугольная матрица. Сделаем в интеграле "подстановку"  $x = m + H^*y$ . Получим

$$\begin{aligned} \int_{\mathbb{R}^n} \frac{1}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1}(x-m), (x-m) \rangle\right) &= \\ &= \int_{\mathbb{R}^n} \frac{|\det(H^*)|}{(2\pi)^{n/2} \sqrt{\det(B)}} \cdot \exp\left(-\frac{1}{2} \langle B^{-1}H^*y, H^*y \rangle\right). \end{aligned}$$

Замечая, что  $|\det(H^*)| = |\det(H)| = \sqrt{\det(B)}$ , и

$$\langle B^{-1}H^*y, H^*y \rangle = \langle H(H^*H)^{-1}H^*y, y \rangle = \langle y, y \rangle,$$

имеем

$$\begin{aligned} \int_{\mathbb{R}^n} f_X &= \frac{1}{(2\pi)^{n/2}} \cdot \int_{\mathbb{R}^n} \exp\left(-\frac{y_1^2 + \dots + y_n^2}{2}\right) dy_1 \dots dy_n = \\ &= \prod_{k=1}^n \left( \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp\left(-\frac{y_k^2}{2}\right) dy_k \right) = 1 \end{aligned}$$

(все одномерные интегралы равны 1, как следует из примера 5 п.2.3).

Если задан случайный вектор  $X$  со значениями в  $\mathbb{R}^n$ , то можно построить случайные переменные  $X_1, \dots, X_n$  – координаты  $X$ . Множество значений каждой такой с.п. известно. Требуется лишь указать распределение вероятностей на нем.

Рассмотрим вначале случай *дискретного* случайного вектора (пример 1). Множество значений с.п.  $X_1$  – это  $\Omega = \{0, 1\}$ . Событие  $X_1 = 0$  есть объединение двух элементарных событий:  $(X_1 = 0, X_2 = 0)$  и  $(X_1 = 0, X_2 = 1)$ . Поэтому  $\mathbf{P}(X_1 = 0) = 1/3 + 1/6 = 1/2$ . Аналогично,  $\mathbf{P}(X_1 = 1) = 1/6 + 1/3 = 1/2$ .

Мы нашли распределение с.п.  $X_1$  – первой координаты случайного вектора  $X$ . Естественно, оно совпадает с распределением, которое мы получили, рассматривая с.п.  $X_1$  отдельно от  $X_2$ .

Обобщая этот результат, можно сформулировать алгоритм построения закона распределения координаты *дискретного* случайного вектора, если известен весь с.в.:

$$\mathbf{P}(X_k = x_k) = \sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_n} \mathbf{P}(X = x).$$

Для каждого возможного значения координаты суммируются вероятности всех элементарных событий, на которых это значение достигается.

В случае абсолютно непрерывного случайного вектора суммирование заменяется интегрированием:

$$f_{X_k}(x_k) = \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_{k-1} \int_{-\infty}^{+\infty} dx_{k+1} \cdots \int_{-\infty}^{+\infty} f_X(x) dx_n.$$

Примеры. 4. Пусть случайный вектор  $Y$  равномерно распределен на прямоугольнике  $|y_1| \leq a$ ,  $|y_2| \leq b$  ( $a > 0$ ,  $b > 0$ ). Тогда

$$f_Y : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Y(y) = \begin{cases} \frac{1}{4ab}, & \text{если } |y_1| \leq a \text{ и } |y_2| \leq b, \\ 0, & \text{если } |y_1| > a \text{ или } |y_2| > b; \end{cases}$$

$$f_{Y_1}(y_1) = \int_{-\infty}^{+\infty} f_Y(y_1, y_2) dy_2 = \begin{cases} \frac{1}{2a}, & \text{если } |y_1| \leq a, \\ 0, & \text{если } |y_1| > a; \end{cases}$$

$$f_{Y_2}(y_2) = \int_{-\infty}^{+\infty} f_Y(y_1, y_2) dy_1 = \begin{cases} \frac{1}{2b}, & \text{если } |y_2| \leq b, \\ 0, & \text{если } |y_2| > b. \end{cases}$$

Координаты этого случайного вектора распределены равномерно на сегментах  $|y_1| \leq a$  и  $|y_2| \leq b$  соответственно.

5. Пусть случайный вектор  $Z$  равномерно распределен на круге  $\|z\| \leq r$ . Тогда

$$f_Z : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Z(z) = \begin{cases} \frac{1}{\pi r^2}, & \text{если } \|z\| \leq r, \\ 0, & \text{если } \|z\| > r; \end{cases}$$

$$f_{Z_1}(z_1) = \int_{-\infty}^{+\infty} f_Z(z_1, z_2) dz_2 = \begin{cases} \frac{2}{\pi r^2} \sqrt{r^2 - z_1^2}, & \text{если } |z_1| \leq r, \\ 0, & \text{если } |z_1| > r; \end{cases}$$

$$f_{Z_2}(z_2) = \int_{-\infty}^{+\infty} f_Z(z_1, z_2) dz_1 = \begin{cases} \frac{2}{\pi r^2} \sqrt{r^2 - z_2^2}, & \text{если } |z_2| \leq r, \\ 0, & \text{если } |z_2| > r. \end{cases}$$

Координаты этого случайного вектора распределены одинаково (но неравномерно!).

Еще раз подчеркнем, что знание распределений координат не дает, вообще говоря, возможности построить распределение случайного вектора. Исключение составляет случай, для описания которого мы введем важное новое понятие.

Координаты дискретного с.в. называются (*статистически*) *независимыми в совокупности*, если

$$\mathbf{P}(X_1 = x_1, \dots, X_n = x_n) = \prod_{k=1}^n \mathbf{P}(X_k = x_k)$$

для всех возможных значений  $x = (x_1, \dots, x_n)$  этого вектора.

Возвращаясь к примеру **1**, видим, что координаты вектора  $X$  статистически зависимы, так как, например,  $\mathbf{P}(X_1 = 0, X_2 = 1) = 1/6$ , но  $\mathbf{P}(X_1 = 0) \cdot \mathbf{P}(X_2 = 1) = 1/4$ .

Координаты абсолютно непрерывного с.в. называются *независимыми в совокупности*, если

$$f_X(x) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \dots \cdot f_{X_n}(x_n),$$

т.е. если плотность распределения случайного вектора равна произведению плотностей распределения его координат.

Легко видеть, что в примере **4** это условие выполнено, т.е. координаты случайного вектора статистически независимы. А вот в примере **5** координаты оказываются зависимыми.

Очевидно, если координаты случайного вектора  $X$  независимы в совокупности, то распределение с.в. полностью определяется распределениями его координат. В этом случае можно изучать с.п.  $X_1, \dots, X_n$  по отдельности.

В приложениях эксперимент часто стараются планировать так, чтобы "из физических соображений" координаты случайного вектора (математической модели эксперимента) можно было считать независимыми в совокупности. Такое предположение существенно упрощает задачу и удешевляет ее решение. Подчеркнем, что ответственность за правильность этих "физических соображений" несет экспериментатор.

## 2.5. Числовые характеристики случайного вектора

Определение. Математическим ожиданием  $n$ -мерного с.в.  $X$  называется числовой вектор  $m \in \mathbb{R}^n$ , вычисляемый по правилу:

в дискретном случае

$$m = M(X) = \sum_x x \cdot \mathbf{P}(X = x)$$

(суммирование ведется по всем значениям случайного вектора);

в абсолютно непрерывном случае

$$m = M(X) = \int_{\mathbb{R}^n} x \cdot f_X(x)$$

(напомним, что интегрирование вектора выполняется по координатам).

Вычислим первую координату вектора  $M(X)$  для дискретного с.в.:

$$m_1 = \sum_{x_1} x_1 \cdot \sum_{x_2, \dots, x_n} \mathbf{P}(X = x) = \sum_{x_1} x_1 \cdot \mathbf{P}(X_1 = x_1) = M(X_1)$$

(суммирование по  $x_2, \dots, x_n$  дает распределение  $X_1$ ).

В случае абсолютно непрерывного с.в.

$$m_1 = \int_{-\infty}^{+\infty} x_1 dx_1 \int_{\mathbb{R}^{n-1}} f_X(x) dx_2 \dots dx_n = \int_{-\infty}^{+\infty} x_1 \cdot f_{X_1}(x_1) dx_1 = M(X_1)$$

(интегрирование по  $x_2, \dots, x_n$  дает плотность распределения  $X_1$ ).

Аналогично,  $m_k = M(X_k)$ ,  $k = 1, \dots, n$ .

Итак, координаты математического ожидания случайного вектора равны математическим ожиданиям соответствующих координат этого вектора. При этом, конечно, предполагается *абсолютная* сходимость всех числовых рядов или несобственных интегралов.

Определение. Ковариационной матрицей  $n$ -мерного с.в.  $X$  называется числовая матрица, вычисляемая по правилу:

$$B(X) = cov(X) = \sum_x (x - M(X)) \cdot (x - M(X))^T \cdot \mathbf{P}(X = x)$$

в дискретном случае (суммирование ведется по всем значениям с.в.);

$$B(X) = cov(X) = \int_{\mathbb{R}^n} (x - M(X)) \cdot (x - M(X))^T \cdot f_X(x)$$

в абсолютно непрерывном случае (матрица интегрируется поэлементно).

Поскольку  $(x - M(X)) \cdot (x - M(X))^T$  – матрица порядка  $n$ , то и  $B(X)$  – матрица порядка  $n$ . Вычислим ее первый *диагональный* элемент для дискретного с.в.:

$$\begin{aligned} b_{11} &= \sum_{x_1} (x_1 - m_1)^2 \cdot \sum_{x_2, \dots, x_n} \mathbf{P}(X = x) = \\ &= \sum_{x_1} (x_1 - m_1)^2 \cdot \mathbf{P}(X_1 = x_1) = D(X_1). \end{aligned}$$

В случае абсолютно непрерывного с.в.

$$\begin{aligned} b_{11} &= \int_{-\infty}^{+\infty} (x_1 - m_1)^2 dx_1 \int_{\mathbb{R}^{n-1}} f_X(x) dx_2 \dots dx_n = \\ &= \int_{-\infty}^{+\infty} (x_1 - m_1)^2 \cdot f_{X_1}(x_1) dx_1 = D(X_1). \end{aligned}$$

Аналогично,  $b_{kk} = D(X_k)$ ,  $k = 1, \dots, n$ .

Таким образом, диагональные элементы ковариационной матрицы равны дисперсиям соответствующих координат случайного вектора.

Рассмотрим теперь внедиагональные элементы матрицы  $B(X)$ :

$$b_{kj} = b_{jk} = \sum_x (x_k - m_k)(x_j - m_j) \cdot \mathbf{P}(X = x)$$

в дискретном случае;

$$b_{kj} = b_{jk} = \int_{\mathbb{R}^n} (x_k - m_k)(x_j - m_j) \cdot f_X(x)$$

в абсолютно непрерывном случае.

Число  $b_{kj}$  называется *ковариацией* координат  $X_k$  и  $X_j$ , или их *вторым смешанным центральным моментом* (читателю, знакомому с механикой, рекомендуем вспомнить центробежный момент инерции).

Напоминаем, что для существования матрицы ковариаций требуется существование вектора математического ожидания и абсолютная сходимость всех определяющих ее элементы рядов (интегралов).

Важное свойство матрицы ковариаций доказывает

Теорема. Матрица  $B(X)$  неотрицательно определена, т.е.

$$\langle B(X) \alpha, \alpha \rangle \geq 0 \quad \text{для любого } \alpha \in \mathbb{R}^n.$$

Доказательство. Пусть  $\alpha \in \mathbb{R}^n$ . Для дискретного случайного вектора  $X$  имеем

$$\begin{aligned} \langle B(X) \alpha, \alpha \rangle &= \alpha^T B(X) \alpha = \sum_x \alpha^T (x - m) \cdot (x - m)^T \alpha \cdot \mathbf{P}(X = x) = \\ &= \sum_x |\langle (x - m), \alpha \rangle|^2 \cdot \mathbf{P}(X = x); \end{aligned} \quad (2.5.1)$$

аналогично, для абсолютно непрерывного случайного вектора

$$\langle B(X) \alpha, \alpha \rangle = \int_{\mathbb{R}^n} |\langle (x - m), \alpha \rangle|^2 \cdot f_X(x). \quad (2.5.1')$$

Интеграл в (2.5.1'), очевидно, неотрицателен. Более того, если  $\alpha \neq \theta_n$ , то он не может равняться нулю, ибо в силу (2.4.1) и кусочной непрерывности  $f_X$  найдется область, в которой оба сомножителя положительны.

Аналогично, если  $\alpha \neq \theta_n$ , то сумма в (2.5.1) может равняться нулю только в случае, когда все точки  $x$ , для которых  $\mathbf{P}(X = x) > 0$ , расположены в плоскости  $\langle (x - m), \alpha \rangle = 0$ . А это равенство, как известно из курса линейной алгебры, означает, что одна из координат вектора  $x - m$  есть линейная комбинация остальных:

$$x_k = m_k - \frac{1}{\alpha_k} \sum_{j \neq k} \alpha_j \cdot (x_j - m_j). \quad (2.5.2)$$

Таким образом, матрица ковариации является даже положительно определенной, за исключением вырожденного случая, когда одна из координат с.в. – полином первой степени от остальных координат. ■

Примеры. Найдем числовые характеристики случайных векторов, рассмотренных в предыдущем пункте.

1. Дискретный случайный вектор из примера 1.

|                      |     |     |
|----------------------|-----|-----|
| $x_2 \backslash x_1$ | 0   | 1   |
| 0                    | 1/3 | 1/6 |
| 1                    | 1/6 | 1/3 |

$$m_1 = M(X_1) = 0 \cdot (1/3 + 1/6) + 1 \cdot (1/6 + 1/3) = 1/2.$$

Аналогичные подсчеты дают

$$M(X) = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}; \quad cov(X) = \begin{bmatrix} \frac{1}{4} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{4} \end{bmatrix}.$$

2. Абсолютно непрерывный случайный вектор из примера 4:

$$f_Y : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Y(y) = \begin{cases} \frac{1}{4ab}, & \text{если } |y_1| \leq a \text{ и } |y_2| \leq b, \\ 0, & \text{если } |y_1| > a \text{ или } |y_2| > b. \end{cases}$$

$$m_1 = M(Y_1) = \int_{-\infty}^{+\infty} y_1 \cdot f_{Y_1}(y_1) dy_1 = \frac{1}{2a} \cdot \int_{-a}^a y_1 dy_1 = 0$$

(интеграл от нечетной функции по симметричному относительно нуля промежутку равен нулю). Аналогично,  $m_2 = 0$ , т.е.  $M(Y) = [0, 0]^T$ .

$$b_{11} = D(Y_1) = \int_{-\infty}^{+\infty} (y_1 - m_1)^2 \cdot f_{Y_1}(y_1) dy_1 = \frac{1}{2a} \cdot \int_{-a}^a (y_1 - 0)^2 dy_1 = a^2/3.$$

Аналогично,  $b_{22} = D(Y_2) = b^2/3$ . Далее, вычислим  $b_{12} = b_{21}$ :

$$\int_{\mathbb{R}^2} (y_1 - m_1)(y_2 - m_2) \cdot f_Y(y_1, y_2) dy_1 dy_2 = \frac{1}{4ab} \cdot \int_{-a}^a y_1 dy_1 \int_{-b}^b y_2 dy_2 = 0$$

(интеграл от нечетной функции по симметричному относительно нуля промежутку равен нулю). Итак,

$$\text{cov}(Y) = \begin{bmatrix} \frac{a^2}{3} & 0 \\ 0 & \frac{b^2}{3} \end{bmatrix}.$$

3. Для абсолютно непрерывного случайного вектора из примера 5

$$f_Z : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad f_Z(z) = \begin{cases} \frac{1}{\pi r^2}, & \text{если } \|z\| \leq r, \\ 0, & \text{если } \|z\| > r \end{cases}$$

аналогичные выкладки дают

$$M(Z) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T; \quad \text{cov}(Z) = \begin{bmatrix} \frac{r^2}{4} & 0 \\ 0 & \frac{r^2}{4} \end{bmatrix}.$$

4. Для нормально распределенного случайного вектора с параметрами  $m$  и  $B$  (пример 3), не приводя достаточно громоздких выкладок, укажем, что  $M(X) = m$ ,  $\text{cov}(X) = B$ .

Иногда вместо ковариационной матрицы используют *корреляционную матрицу*, элементы которой определяются формулой

$$r_{jk} = \frac{b_{jk}}{\sqrt{b_{jj} \cdot b_{kk}}}.$$



Очевидно, что  $r_{ii} = 1$ . Покажем, что  $|r_{ik}| \leq 1$ . Действительно, применив доказанную выше теорему к вектору

$$\alpha = \lambda e^{(j)} - e^{(k)}$$

(здесь  $\lambda \in \mathbb{R}$ , а  $e^{(j)}$  и  $e^{(k)}$  – векторы стандартного базиса), получим

$$b_{jj}^2 \lambda^2 - 2b_{jk} \lambda + b_{kk}^2 \geq 0.$$

Известно, что квадратный трехчлен *всюду* неотрицателен, только если его дискриминант неположителен. Поэтому  $b_{jk}^2 - (b_{jj}b_{kk})^2 \leq 0$ , или  $r_{jk}^2 \leq 1$ , т.е.  $|r_{jk}| \leq 1$ .

Замечание. Равенство  $|r_{jk}| = 1$  означает, что координаты случайного вектора  $X_j - M(X_j)$  и  $X_k - M(X_k)$  линейно зависимы<sup>30</sup> (см. (2.5.2)).

Число  $r_{jk}$  называется *коэффициентом корреляции* между случайными переменными  $X_j$  и  $X_k$ . Если  $r_{jk} = 0$ , то с.п.  $X_j$  и  $X_k$  называются *некоррелированными*.

Покажем, что независимые случайные переменные некоррелированы. Пусть  $f_X(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$ . Тогда

$$\begin{aligned} b_{12} &= \int_{\mathbb{R}^2} (x_1 - m_1)(x_2 - m_2) f_{X_1}(x_1) f_{X_2}(x_2) dx_1 dx_2 = \\ &= \left( \int_{-\infty}^{+\infty} (x_1 - m_1) f_{X_1}(x_1) dx_1 \right) \times \left( \int_{-\infty}^{+\infty} (x_2 - m_2) f_{X_2}(x_2) dx_2 \right). \end{aligned}$$

Легко видеть, что каждый из полученных интегралов равен нулю. Поэтому  $b_{12} = 0$ .

Серьезное предупреждение. В рассмотренных выше примерах (случайные векторы, равномерно распределенные на прямоугольнике и на круге) координаты случайных векторов оказались *некоррелированными*. В то же время в предыдущем пункте было показано, что координаты первого случайного вектора независимы, а второго – зависимы. Следовательно, равенство нулю коэффициента корреляции между двумя с.п. не несет никакой информации о наличии статистической зависимости между ними. К сожалению, слово "корреляция" иногда употребляется в смысле "статистическая зависимость". Исторические причины этой путаницы будут изложены ниже.

<sup>30</sup>Не путайте линейную зависимость со статистической!

Замечание. Можно показать, что координаты нормального случайного вектора – нормально распределенные с.п.:

$$f_{X_k}(x_k) = \frac{1}{\sqrt{2\pi} \cdot \sigma_k} \cdot \exp\left(-\frac{(x_k - m_k)^2}{2\sigma_k^2}\right) \quad (\sigma_k = \sqrt{b_{kk}}).$$

Если координаты *нормального* случайного вектора попарно некоррелированы, то  $B = \text{diag} [\sigma_1^2, \dots, \sigma_n^2]$  и

$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sigma_1 \dots \sigma_n} \times \\ \times \exp\left(-\frac{(x_1 - m_1)^2}{2\sigma_1^2} - \dots - \frac{(x_n - m_n)^2}{2\sigma_n^2}\right) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \dots f_{X_n}(x_n).$$

Таким образом (в исключение из общего правила), некоррелированность координат *нормального* случайного вектора равносильна их независимости в совокупности.

## 2.6. Преобразования случайных переменных

Пусть задан случайный вектор  $X$ , т.е. задано пространство элементарных событий  $\Omega_X \subset \mathbb{R}^n$ , и на нем заданы либо вероятности (для дискретного случайного вектора), либо плотность распределения вероятностей (для абсолютно непрерывного случайного вектора).

Зададим теперь функцию  $\phi : \Omega_X \rightarrow \mathbb{R}^m$  и обозначим множество ее значений  $\Omega_Y$ . Тогда на  $\Omega_Y$  естественным образом задается распределение вероятностей, порождаемое распределением вероятностей на  $\Omega_X$ : вероятность каждого события  $A \subset \Omega_Y$  равна вероятности его полного прообраза, т.е. множества всех точек из  $\Omega_X$ , которые переводятся в  $A$  функцией  $\phi$ . Таким образом определяется новый случайный вектор, обозначаемый  $Y = \phi(X)$ .

Легко видеть, что распределение вероятностей образа *дискретного* случайного вектора  $Y = \phi(X)$  вычисляется по формуле

$$\mathbf{P}(Y = y) = \sum_{\{x|\phi(x)=y\}} \mathbf{P}(X = x).$$

Пример. Пусть  $\Omega_X = \{-3, -2, -1, 0, 1, 2, 3\}$ , и все значения *равновероятны*. Зададим функцию  $g : \Omega_X \rightarrow \mathbb{R}$  формулой  $g(x) = x^2$ . Видно, что множество значений с.п.  $Y = X^2$   $\Omega_Y = \{0, 1, 4, 9\}$ . При этом у значения 0 один прообраз, а у остальных значений – по два. Теперь легко сосчитать вероятности всех значений. С.п.  $Y$  задана таблицей:

|                 |     |     |     |     |
|-----------------|-----|-----|-----|-----|
| $y$             | 0   | 1   | 4   | 9   |
| $\mathbf{P}(y)$ | 1/7 | 2/7 | 2/7 | 2/7 |

Вопрос о преобразовании плотности распределения абсолютно непрерывного случайного вектора легко решается, если  $\phi$  – гладкая функция, взаимно однозначно отображающая  $\Omega_X \subset \mathbb{R}^n$  на  $\Omega_Y \subset \mathbb{R}^n$ , причем  $\det(\phi') \neq 0$ . В этой ситуации  $X \in G \iff Y \in \phi(G)$ . Поэтому  $\mathbf{P}(X \in G) = \mathbf{P}(Y \in \phi(G))$ , и по теореме о преобразовании интеграла

$$\int_G f_X = \int_{\phi(G)} f_Y = \int_G (f_Y \circ \phi) \cdot |\det(\phi')|. \quad (2.6.1)$$

Поскольку это равенство должно быть выполнено для любой области  $G$  с кусочно гладкой границей, подынтегральные функции слева и справа в (2.6.1) совпадают, откуда

$$f_Y(\phi(x)) = \frac{f_X(x)}{|\det(\phi'(x))|}.$$

Если отображение  $\phi$  не является взаимно однозначным, то вопрос о распределении  $Y = \phi(X)$  более сложен. Остановимся подробнее на случае, когда  $\phi : \Omega_X \rightarrow \mathbb{R}$  – функционал.

Воспользуемся свойством (2.1.2) функции распределения:

$$F_Y(y) = \mathbf{P}(\phi(X) < y) = \int_{\{x|\phi(x)<y\}} f_X. \quad (2.6.2)$$

Если удастся вычислить функцию распределения, то плотность  $f_Y$  находится из нее дифференцированием.

Примеры. 1. Мгновенное значение синусоидального напряжения определяется по формуле  $u = U_{max} \cdot \sin(\varphi)$ . Измерение мгновенного значения производится в *случайный* момент времени, т.е. предполагается, что с.п.  $\Phi$  (фаза) распределена равномерно на  $]-\pi, \pi[$ :

$$f_\Phi(\varphi) = \begin{cases} \frac{1}{2\pi}, & \text{если } -\pi < \varphi \leq \pi, \\ 0, & \text{если } \varphi \leq -\pi \text{ или } \varphi > \pi. \end{cases}$$

При этих условиях определена с.п.  $U$ , значения которой заполняют сегмент  $\Omega_U = [-U_{max}, U_{max}]$ . Найдем плотность ее распределения, положив для упрощения, что  $U_{max} = 1$ .

Начнем с построения  $F_U(u) = \mathbf{P}(U < u)$ . Очевидно, что  $F_U(u) = 0$  при  $u \leq -1$  и  $F_U(u) = 1$  при  $u > 1$ .

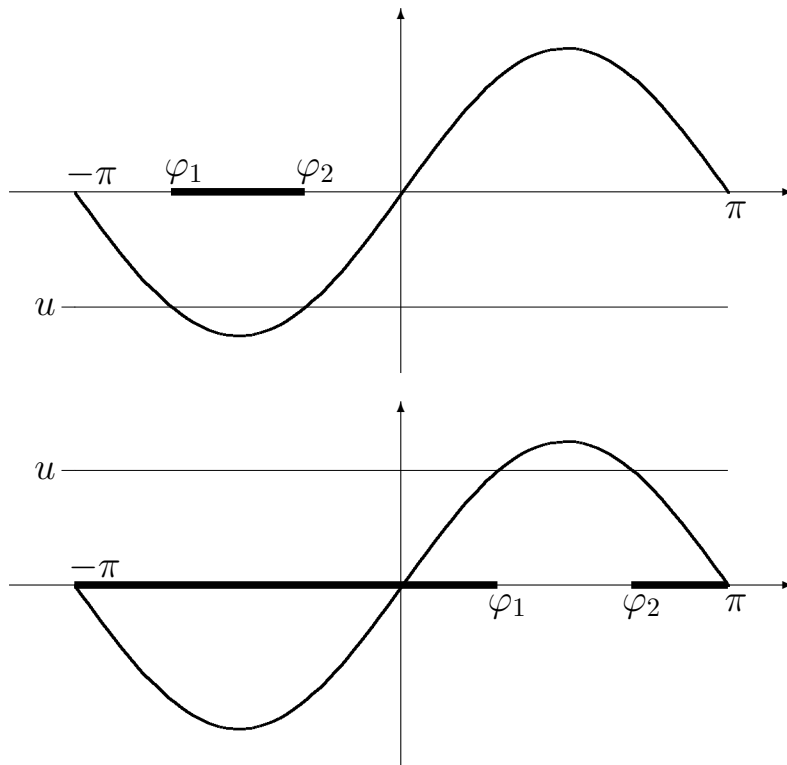


Рис.2.3

На рис.2.3 выделен участок оси абсцисс, на котором  $U < u$ : вверху для  $u \in ]-1, 0[$ , а внизу – для  $u \in [0, 1]$ . Видно, что при  $-1 < u < 0$

$$F_U(u) = \mathbf{P}(U < u) = \frac{\varphi_2 - \varphi_1}{2\pi} = \frac{\pi + 2\arcsin(u)}{2\pi},$$

а при  $0 \leq u \leq 1$

$$F_U(u) = \mathbf{P}(U < u) = \frac{\varphi_1 + 2\pi - \varphi_2}{2\pi} = \frac{\pi + 2\arcsin(u)}{2\pi}.$$

Итак,

$$F_U(u) = \begin{cases} 0 & \text{при } u \leq -1, \\ \frac{\pi + 2\arcsin(u)}{2\pi} & \text{при } -1 \leq u \leq 1, \\ 1 & \text{при } u \geq 1. \end{cases}$$

Дифференцируя, найдем плотность распределения

$$f_U(u) = F'_U(u) = \begin{cases} \frac{1}{\pi\sqrt{1-u^2}} & \text{при } |u| < 1, \\ 0 & \text{при } |u| > 1 \end{cases}$$

(при  $u = \pm 1$  плотность не определена).

Мы получили так называемый *закон арксинуса*.

2. Задан двумерный случайный вектор  $X$  с плотностью распределения  $f_X$ . Найдем распределение с.п.  $Y = X_1 + X_2$ .

Формула (2.6.2) дает

$$F_Y(y) = \int_{\{x|x_1+x_2 < y\}} f_X(x_1, x_2) dx_1 dx_2.$$

Как известно, множество  $\{x \in \mathbb{R}^2 | x_1 + x_2 < y\}$  – полуплоскость (рис.2.4).

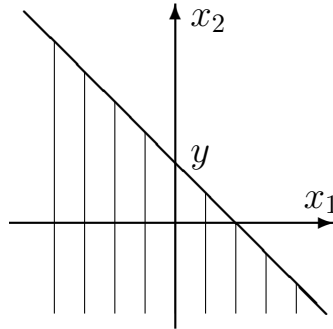


Рис.2.4

Преобразуем двойной интеграл в повторный:

$$F_Y(y) = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{y-x_1} f_X(x_1, x_2) dx_2 = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^y f_X(x_1, z - x_1) dz$$

(во внутреннем интеграле мы сделали подстановку  $x_2 = z - x_1$ ).

Теперь преобразуем повторный интеграл обратно в двойной, а затем – в повторный в другом порядке ("поменяем порядок интегрирования"):

$$F_Y(y) = \int_{-\infty}^y dz \int_{-\infty}^{+\infty} f_X(x_1, z - x_1) dx_1.$$

Дифференцируя по  $y$ , получим плотность распределения:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_X(x_1, y - x_1) dx_1. \quad (2.6.3)$$

Замечание. Если координаты  $x_1$  и  $x_2$  *независимы*, то формула (2.6.3) перепишется так:

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X_1}(x_1) \cdot f_{X_2}(y - x_1) dx_1 = (f_{X_1} * f_{X_2})(y). \quad (2.6.4)$$

Плотность распределения суммы двух независимых с.п.  
равна свертке плотностей распределения слагаемых.

Примеры. 3. Найдем распределение суммы координат двумерного *нормального* случайного вектора.

Подставляя в формулу (2.4.2)

$$m = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}; \quad B = \begin{bmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

( $r$  – коэффициент корреляции), имеем

$$f_X(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \\ \times \exp\left(-\frac{\left(\frac{x_1-m_1}{\sigma_1}\right)^2 - 2r \cdot \frac{x_1-m_1}{\sigma_1} \cdot \frac{x_2-m_2}{\sigma_2} + \left(\frac{x_2-m_2}{\sigma_2}\right)^2}{2(1-r^2)}\right).$$

Если подставить  $f_X$  в (2.6.3) и вычислить интеграл, то получим

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{(y-m)^2}{2\sigma^2}\right),$$

где  $m = m_1 + m_2$ ;  $\sigma^2 = \sigma_1^2 + \sigma_2^2 + 2r \cdot \sigma_1 \cdot \sigma_2$ .

*Можно показать, что* и в  $n$ -мерном случае сумма координат *нормального* случайного вектора имеет нормальное распределение.

4. Если двумерный случайный вектор  $X$  равномерно распределен на квадрате  $[-1, 1] \times [-1, 1]$ , то, как было показано в п.2.4, его координаты независимы и равномерно распределены на сегменте  $[-1, 1]$  :

$$f_{X_1}(t) = f_{X_2}(t) = \begin{cases} \frac{1}{2} & \text{при } |t| \leq 1, \\ 0 & \text{при } |t| > 1. \end{cases}$$

По формуле (2.6.4)

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_{X_1}(t) \cdot f_{X_2}(y-t) dt.$$

Подынтегральная функция здесь отлична от нуля только при выполнении одновременно двух неравенств:  $-1 \leq t \leq 1$  и  $-1 \leq y-t \leq 1$  (в этом случае она равна  $1/4$ ). Поэтому  $f_{X_1+X_2}(y) = 1/4 \cdot \Delta$ , где  $\Delta$  – длина общей части промежутков  $[-1, 1]$  и  $[y-1, y+1]$ .

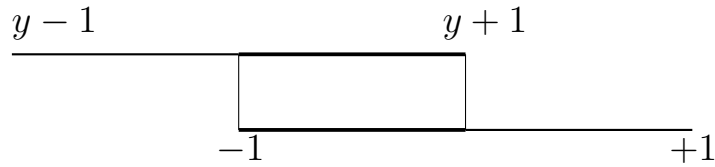


Рис.2.5

Из рис.2.5 видно, что при  $y \in [-2, 0]$   $\Delta = (y + 1) - (-1) = y + 2$ . Аналогично, при  $y \in ]0, 2]$   $\Delta = 2 - y$ . Наконец, очевидно, что  $\Delta = 0$  при  $|y| \geq 2$ . График плотности распределения  $X_1 + X_2$  (так называемое *треугольное* распределение) изображен на рис.2.6.

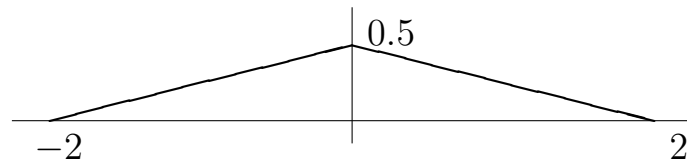


Рис.2.6

Вычислим теперь математическое ожидание и ковариационную матрицу образа случайного вектора.

Пусть  $X$  –  $n$ -мерный дискретный случайный вектор,  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $Y = \phi(X)$ . Тогда из определения видно, что

$$M(Y) = \sum_x \phi(x) \cdot \mathbf{P}(X = x). \quad (2.6.5)$$

Можно показать, что и для абсолютно непрерывного случайного вектора  $X$  справедлива аналогичная формула

$$M(Y) = \int_{\mathbb{R}^n} \phi(x) \cdot f_X(x) \quad (2.6.5')$$

(как всегда, предполагаем, что ряд или интеграл абсолютно сходится).

Замечание. Формулы (2.2.1), (2.2.1') можно теперь переписать так:

$$D(X) = M \left( (X - M(X))^2 \right). \quad (2.6.6)$$

Дисперсия с.п. – это математическое ожидание квадрата отклонения с.п. от ее математического ожидания.

Отметим, что формулу (2.6.6) можно преобразовать так:

$$D(X) = M(X^2 - 2M(X) \cdot X + (M(X))^2) = M(X^2) - (M(X))^2. \quad (2.6.6')$$

Аналогично формуле (2.6.5), из определения ковариационной матрицы для дискретного случайного вектора видно, что

$$B(Y) = \sum_x (\phi(x) - M(Y)) \cdot (\phi(x) - M(Y))^T \cdot \mathbf{P}(X = x), \quad (2.6.7)$$

и можно показать, что для абсолютно непрерывного случайного вектора

$$B(Y) = \int_{\mathbb{R}^n} (\phi(x) - M(Y)) \cdot (\phi(x) - M(Y))^T \cdot f_X(x). \quad (2.6.7')$$

Пример. В приложениях часто приходится вычислять математическое ожидание и дисперсию линейной комбинации случайных переменных.

Пусть  $X$  –  $n$ -мерный случайный вектор,  $Y = \alpha_1 X_1 + \dots + \alpha_n X_n$ . В этом случае  $\phi(x) = \langle x, \alpha \rangle$ , где  $\alpha = [\alpha_1, \dots, \alpha_n]^T$ . Поэтому, вынося в формулах (2.6.5), (2.6.5') постоянный множитель  $\alpha$  за знак суммы (интеграла), получаем

$$M(Y) = \langle M(X), \alpha \rangle \quad (2.6.8)$$

(в предположении, что  $M(X)$  существует).

Аналогично, подставив  $\phi(x) = \langle x, \alpha \rangle$  и  $M(Y) = \langle M(X), \alpha \rangle$  в формулы (2.6.7), (2.6.7'), можно вынести постоянные множители за знак суммы (интеграла) и получить

$$D(Y) = \langle B(X) \alpha, \alpha \rangle \quad (2.6.9)$$

(если существует  $B(X)$ ).

В важном частном случае, когда  $Y = \sum_{k=1}^n X_k$ , имеем  $\alpha = [1, \dots, 1]^T$ .

Поэтому

$$M(Y) = \sum_{k=1}^n M(X_k); \quad D(Y) = \sum_{j=1}^n \sum_{k=1}^n b_{jk}.$$

Если координаты случайного вектора имеют математические ожидания, то математическое ожидание суммы координат равно сумме математических ожиданий слагаемых.

Если координаты случайного вектора имеют дисперсии, то дисперсия суммы координат равна сумме всех элементов ковариационной матрицы.



Отметим очень важное обстоятельство: *дисперсия суммы координат, вообще говоря, не равна сумме дисперсий этих координат*. Она может быть и больше, и меньше.

Если считать дисперсию с.п. мерой "неупорядоченности" этой с.п., то при суммировании случайных переменных эта "неупорядоченность" может как увеличиваться, так и уменьшаться.

Если же координаты случайного вектора *попарно некоррелированы*, то дисперсия суммы координат равна сумме дисперсий слагаемых.

Пример. Рассмотрим биномиальное распределение (пример 1 п.2.3). Если обозначить  $X_j$  число успехов в  $j$ -м испытании, то, очевидно,  $M(X_j) = 1 \cdot p + 0 \cdot (1 - p) = p$ . Далее,  $X_j^2 = X_j$  (поскольку значение  $X_j$  всегда равно нулю или единице). Поэтому формула (2.6.6') дает

$$D(X_j) = M(X_j^2) - (M(X_j))^2 = p - p^2 = p(1 - p).$$

Заметим теперь, что по смыслу задачи с.п.  $X_j, j = 1, \dots, n$ , независимы в совокупности, и  $X = \sum_{j=1}^n X_j$ . Поэтому

$$M(X) = \sum_{j=1}^n M(X_j) = n \cdot p; \quad D(X) = \sum_{j=1}^n D(X_j) = n \cdot p \cdot (1 - p),$$

что согласуется с результатом, объявленным в п.2.3.

## 2.7. Датчик псевдослучайных чисел

Задача о преобразовании случайной переменной часто встречается в приложениях (например, задача о преобразовании "случайной помехи" при ее прохождении через приемно-усилительное устройство). В силу сложности структуры физических преобразователей "аналитические" методы (некоторые из них рассматривались в предыдущем пункте) нередко оказываются неприменимыми. Хорошие результаты дает в этом случае машинное моделирование: алгоритм преобразования, осуществляемого физическим устройством, реализуется в виде машинной программы, на вход которой подается достаточно длинный набор чисел, имитирующих значения случайной переменной ("помехи"). Вопрос о качестве имитации будет рассмотрен в главе 3. Числа эти именуются *псевдослучайными*.

В библиотеках стандартных подпрограмм и в средах конечного пользователя имеются программные датчики (генераторы) псевдослучайных чисел. В их основе лежат алгоритмы, вырабатывающие псевдослучайные числа, равномерно распределенные на  $[0, 1]$ . Устройство таких

алгоритмов весьма сложно и не может рассматриваться в нашем курсе. Взяв достаточно большую выборку, можно убедиться, что относительные частоты попадания псевдослучайных чисел в части интервала  $[0, 1]$  примерно пропорциональны длинам этих частей.

Имея равномерный датчик, можно конструировать датчики с заданными законами распределения. Покажем, как это делается, на примере абсолютно непрерывной с.п.

Пусть  $X$  – с.п., равномерно распределенная на  $[0, 1]$ , и  $\Psi$  – заданная функция распределения ( $\psi$  – соответствующая плотность распределения). Рассмотрим объединение промежутков, на которых  $\psi > 0$ . На этом множестве (обозначим его  $Q$ ) функция  $\Psi$  возрастает и, следовательно, имеет обратную. Поскольку  $0 \leq \Psi \leq 1$ , то  $\Psi^{-1}$  определена почти всюду на  $[0, 1]$ , т.е. на множестве значений с.п.  $X$ . Докажем, что с.п.  $Y = \Psi^{-1}(X)$  имеет функцию распределения  $\Psi$ .

Действительно, из возрастания  $\Psi$  следует возрастание  $\Psi^{-1}$ , и

$$\Psi^{-1}(X) < y \iff X < \Psi(y).$$

Отсюда

$$\begin{aligned} F_Y(y) &= \mathbf{P}(Y < y) = \mathbf{P}(\Psi^{-1}(X) < y) = \mathbf{P}(X < \Psi(y)) = \\ &= \int_{-\infty}^{\Psi(y)} f_X(x) dx = \int_0^{\Psi(y)} 1 \cdot dx = \Psi(y). \end{aligned}$$

Пример. Построим датчик *экспоненциального* распределения с параметром  $\alpha > 0$ :

$$F_Y(y) = \begin{cases} 1 - \exp(-\alpha \cdot y) & \text{при } y > 0, \\ 0 & \text{при } y \leq 0. \end{cases}$$

Здесь  $Q = ]0, +\infty[$ ,

$$F_Y^{-1}(x) = -\frac{\ln(1-x)}{\alpha} \quad \text{при } 0 < x < 1.$$

Следовательно, преобразуя числа  $x_k$ , полученные от равномерного датчика, по формуле  $y_k = -\frac{1}{\alpha} \cdot \ln(1-x_k)$ , мы построим набор значений экспоненциальной с.п.

Отметим, что у описанного метода есть трудность: необходимо уметь вычислять значения функции, обратной к  $F_Y$ .

## 2.8. Центральная предельная теорема

Рассмотрим  $n$ -мерный абсолютно непрерывный случайный вектор  $X = [X_1, \dots, X_n]^T$  с одинаково распределенными и независимыми в совокупности координатами. Будем предполагать, что координаты с.в. имеют математическое ожидание  $m$ , дисперсию  $\sigma^2$  и *абсолютный третий центральный момент*, определяемый равенством

$$\mu_3(X_k) = M(|X_k - M(X_k)|^3).$$

Введем новые случайные переменные  $Y_k = \frac{X_k - m}{\sigma\sqrt{n}}$ . Из формул, полученных в п.2.6, видно, что

$$M(Y_k) = 0, \quad D(Y_k) = \frac{1}{n}, \quad \mu_3(Y_k) = \frac{\mu_3(X_k)}{\sigma^3 n \sqrt{n}}. \quad (2.8.1)$$

Обозначим  $f_Y$  общую плотность распределения с.п.  $Y_k$  и рассмотрим ее преобразование Фурье:

$$\tilde{f}_Y(\omega) = \int_{-\infty}^{+\infty} \exp(-i\omega y) \cdot f_Y(y) dy. \quad (2.8.2)$$

Формула Тейлора третьего порядка дает

$$\exp(-i\omega y) = 1 - i\omega y - \frac{(\omega y)^2}{2} + i \frac{(\omega y)^3}{6} \cdot g(\omega y),$$

где  $g(\omega y) = \cos(\gamma_1) - i \cdot \sin(\gamma_2)$ ,  $\gamma_1$  и  $\gamma_2$  – некоторые точки между 0 и  $\omega y$ .

Подставим это выражение в (2.8.2):

$$\begin{aligned} \tilde{f}_Y(\omega) = & \int_{-\infty}^{+\infty} f_Y(y) dy - i\omega \int_{-\infty}^{+\infty} y f_Y(y) dy - \\ & - \frac{\omega^2}{2} \int_{-\infty}^{+\infty} y^2 f_Y(y) dy + i \frac{\omega^3}{6} \int_{-\infty}^{+\infty} g(\omega y) y^3 f_Y(y) dy. \end{aligned} \quad (2.8.3)$$

Очевидно, что первый интеграл в этой сумме равен единице. Далее, в силу (2.8.1) второй интеграл равен нулю, а третий –  $\frac{1}{n}$ .

Оценим модуль последнего слагаемого в (2.8.3):

$$\left| i \frac{\omega^3}{6} \int_{-\infty}^{+\infty} g(\omega y) y^3 f_Y(y) dy \right| \leq \left| \frac{\omega^3}{3} \int_{-\infty}^{+\infty} |y|^3 f_Y(y) dy \right| = \left| \frac{\mu_3(Y_k) \omega^3}{3} \right|.$$

Поэтому (2.8.3) можно переписать так:

$$\tilde{f}_Y(\omega) = 1 - \frac{\omega^2}{2n} + \frac{\omega^3 h(\omega)}{n\sqrt{n}}, \quad (2.8.4)$$

где  $|h(\omega)| \leq \frac{\mu_3(X_k)}{3\sigma^3}$ .

Пусть теперь  $Z_n = \sum_{k=1}^n Y_k$ . Плотность распределения случайной переменной  $Z_n$  – суммы (независимых в совокупности) с.п.  $Y_k$  – есть, как показано в п.2.6, свертка их плотностей распределения, а образ Фурье свертки, как указано в п.2.4 части "Математическая физика есть произведение образов Фурье этих плотностей, т.е.

$$\tilde{f}_{Z_n}(\omega) = (\tilde{f}_Y(\omega))^n \quad (2.8.5)$$

(образы Фурье всех с.п.  $Y_k$  одинаковы).

Подставим (2.8.4) в (2.8.5):

$$\tilde{f}_{Z_n}(\omega) = \left(1 - \frac{\omega^2}{2n} + \frac{\omega^3 h(\omega)}{n\sqrt{n}}\right)^n.$$

Выясним теперь, что произойдет при неограниченном увеличении  $n$ . Логарифмируя  $\tilde{f}_{Z_n}$  и вспоминая степенной ряд для логарифма, получим

$$\ln(\tilde{f}_{Z_n}) = n \cdot \ln\left(1 - \frac{\omega^2}{2n} + \dots\right) = n \cdot \left(-\frac{\omega^2}{2n} + \dots\right) \xrightarrow{n \rightarrow +\infty} -\frac{\omega^2}{2}$$

(многоточие обозначает совокупность слагаемых, содержащих степени  $1/n$  выше первой).

Отсюда находим предел образов Фурье с.п.  $Z_n$ :

$$\lim_{n \rightarrow +\infty} (\tilde{f}_{Z_n}) = \exp\left(-\frac{\omega^2}{2}\right).$$

Таким образом, предел  $\tilde{f}_{Z_n}$  не зависит от распределения компонент исходного вектора  $X$ .

Терминологическое замечание. Поскольку

$$Z_n = \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - m), \quad M(Z_n) = 0, \quad D(Z_n) = 1,$$

случайную переменную  $Z_n$  называют *центрированной и нормированной суммой* координат с.в.  $X$ .

Можно показать, что  $\exp\left(-\frac{\omega^2}{2}\right)$  – преобразование Фурье плотности стандартного нормального распределения, т.е. нормального распределения с нулевым математическим ожиданием и единичной дисперсией.

Итак, если координаты абсолютно непрерывного с.в.

– одинаково распределены,

– независимы в совокупности,

– имеют дисперсию и абсолютный третий центральный момент,

то при неограниченном увеличении размерности вектора образ Фурье плотности центрированной и нормированной суммы координат (*при любом, но одинаковом их распределении*) стремится к образу Фурье плотности стандартного нормального распределения.

Этот факт не случаен. *Можно показать, что* имеет место

Центральная предельная теорема. Пусть  $(X_n)$  – последовательность с.п., имеющих дисперсии и абсолютные третьи центральные моменты. Пусть при любом  $n$  координаты случайного вектора  $[X_1, \dots, X_n]^T$  независимы в совокупности. Обозначим

$$\mathcal{D}_n = \sum_{k=1}^n D(X_k), \quad \mathcal{M}_n = \sum_{k=1}^n \mu_3(X_k)$$

и предположим, что

$$\lim_{n \rightarrow +\infty} \mathcal{M}_n / (\mathcal{D}_n)^{3/2} = 0. \quad (2.8.6)$$

Тогда распределение последовательности центрированных и нормированных сумм

$$\left( \frac{1}{\sqrt{\mathcal{D}_n}} \sum_{k=1}^n (X_k - M(X_k)) \right)_{n=1}^{+\infty}$$

сходится к стандартному нормальному распределению.

Слово "сходится" здесь означает, что для любого сегмента  $[a, b]$

$$\mathbf{P}\left(\frac{1}{\sqrt{\mathcal{D}_n}} \sum_{k=1}^n (X_k - M(X_k)) \in [a, b]\right) \xrightarrow{n \rightarrow +\infty} \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{t^2}{2}\right) dt.$$

Эта теорема была доказана А.М. Ляпуновым.

Замечание. В рассмотренном нами случае, когда все  $X_k$  имеют одинаковые распределения, имеем  $\mathcal{D}_n = nD(X_k)$ ,  $\mathcal{M}_n = n\mu_3(X_k)$ , и условие (2.8.6) выполнено автоматически.

Подчеркнем еще раз, что в центральной предельной теореме случайные переменные  $X_k$  могут быть любыми (не обязательно абсолютно непрерывными); более того, их распределения *могут быть различными*, лишь бы наборы  $X_1, \dots, X_n$  при всех  $n$  были независимы в совокупности и выполнялось условие (2.8.6).

Из центральной предельной теоремы следует, что если бы всегда можно было иметь дело с суммой *большого* числа таких случайных переменных, то единственным нужным распределением стало бы нормальное. Именно поэтому нормальное распределение играет исключительную роль в теории вероятностей и ее приложениях.

Серьезное предупреждение. Возможно, под влиянием центральной предельной теоремы возникло распространенное суеверие, суть которого сводится к тому, что в прикладных задачах *все* случайные переменные обязаны иметь нормальное распределение. В частности, бытует мнение, что случайные погрешности измерительных приборов всегда имеют нормальное распределение. Эта точка зрения опровергается экспериментами П.В. Новицкого<sup>31</sup>, обработавшего большое количество протоколов поверки приборов и не обнаружившего ожидавшейся нормальности погрешностей.

Упомянутая выше ошибка – отождествление независимости и некоррелированности – также проистекает, по-видимому, из убеждения, что все случайные переменные – нормальные.

---

<sup>31</sup>Петр Васильевич НОВИЦКИЙ (1922-2000) – профессор Санкт-Петербургского технического университета, основатель и разработчик информационных критериев качества измерительных устройств.

## Глава 3. ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ

Математическая статистика изучает методы построения вероятностных моделей, не противоречащих результатам эксперимента. При этом априори предполагается, что эти результаты *могут быть описаны в вероятностных терминах*<sup>32</sup>.

Мы ограничимся в этом курсе рассмотрением двух задач:

- 1) оценивание параметров распределения случайной переменной;
- 2) проверка статистических гипотез.

### 3.1. Одна содержательная задача

С помощью измерительного прибора проделаны  $n$  измерений некоторой физической величины. *Предполагается*, что за время измерений значение этой величины не изменяется. В то же время полученные числа  $x_1, \dots, x_n$  будут, вообще говоря, различны. *Предполагается*, что это различие объясняется наличием у измерительного прибора аддитивной случайной погрешности, т.е.

$$\xi = a + \eta,$$

где  $a$  – неизвестное значение измеряемой физической величины;  $\eta$  – случайная погрешность прибора;  $\xi$  – наблюдаемая физическая величина.

Задача состоит в оценивании значения  $a$  на основании полученных данных  $x_1, \dots, x_n$ .

В математической модели физическим величинам  $\xi$  и  $\eta$  соответствуют случайные переменные, которые мы также обозначим  $\xi$  и  $\eta$ .

При построении математической модели *удобно* считать, что имеется не  $n$  отсчетов, снятых с одного прибора, а  $n$  однотипных приборов, с каждого из которых снимается один отсчет. При этом числовой вектор  $x = [x_1, \dots, x_n]^T$  оказывается значением случайного вектора  $X = [X_1, \dots, X_n]^T$ , координаты которого *независимы в совокупности* (это еще одно предположение!) и имеют одинаковое распределение, совпадающее с распределением случайной переменной  $\xi$ .

*Предположим* еще, что у прибора отсутствует систематическая погрешность, т.е.  $M(\eta) = 0$ . Тогда  $a = M(\xi)$ . *Предположим*, наконец, как делают довольно часто, что случайная погрешность имеет нормальное

---

<sup>32</sup>Еще раз подчеркнем, что ответственность за применение математической модели (за вероятностную трактовку содержательной задачи) несет постановщик задачи!

распределение с известным (из паспорта измерительного прибора) среднеквадратическим уклонением  $\sigma$ .

Сделанные *предположения* позволяют сказать, что  $\xi$  принадлежит однопараметрическому семейству с.п. с плотностью распределения

$$f_{\xi}(t, a) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right),$$

параметр которого  $a$  подлежит оцениванию.

Оценкой параметра  $a$  распределения с.п.  $\xi$  мы будем, как обычно, называть интервал  $]\hat{a} - \Delta, \hat{a} + \Delta[$ , где  $\Delta$  – число, задаваемое постановщиком задачи,  $\hat{a}$  – число, определяемое по полученным результатам измерений<sup>33</sup> ( $\hat{a} = \varphi(x)$ ).

Серьезное предупреждение. Ранее, в курсе математического анализа, когда речь шла об оценивании (корня уравнения, суммы ряда, интеграла и т. д.), интервал-оценка накрывал оцениваемое число *гарантированно*. В математической статистике интервал  $]\hat{a} - \Delta, \hat{a} + \Delta[$  в единичном эксперименте может содержать оцениваемый параметр  $a$ , а может не содержать. Однако если эксперимент в *неизменных* (еще одно предположение!) условиях повторять многократно, то можно говорить об *относительной частоте* накрытия интервалом  $]\hat{a} - \Delta, \hat{a} + \Delta[$  оцениваемого параметра, и, следовательно, о *вероятности* накрытия. Таким образом, сказав, что моделью эксперимента является случайная переменная, мы тем самым предполагаем, что эксперимент будет *многократно* повторяться. Без этого предположения статистическая обработка результатов измерений бессмысленна.

При оценивании математического ожидания в качестве центра интервала обычно берут среднее арифметическое результатов измерений

$$\hat{a} = \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k.$$

Соответствующая случайная переменная  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ , согласно результатам п.2.6, имеет *при сделанных предположениях* нормальное распределение. Его параметры вычисляются по формулам (2.6.8), (2.6.9):

$$M(\bar{X}) = \frac{1}{n} \sum_{k=1}^n a = a; \quad D(\bar{X}) = \frac{1}{n^2} \sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n},$$

---

<sup>33</sup>В математической статистике оценку именуют *интервальной оценкой*, а ее центр  $\hat{a}$  – *точечной оценкой* параметра распределения.



т.е.

$$f_{\bar{X}}(t) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(t-a)^2 n}{2\sigma^2}\right).$$

Итак,  $M(\bar{X}) = a$ , т.е. математическое ожидание среднего арифметического совпадает с оцениваемым параметром. В силу неравенства Чебышева *случайные* значения среднего арифметического, полученные в различных сериях измерений, будут группироваться около его математического ожидания, т.е. около оцениваемого параметра. Разброс *случайных* значений среднего арифметического относительно  $M(\bar{X}) = a$  характеризуется дисперсией  $D(\bar{X}) = \frac{\sigma^2}{n}$ , которую можно сделать как угодно малой за счет увеличения длины обрабатываемой серии измерений.

При заданном числе  $\Delta$  можно вычислить вероятность  $\beta$  накрытия оцениваемого параметра  $a$  интервалом  $]\hat{a} - \Delta, \hat{a} + \Delta[$ :

$$\beta = \mathbf{P}(|\bar{X} - a| < \Delta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \int_{-\Delta}^{\Delta} \exp\left(-\frac{nx^2}{2\sigma^2}\right) dx = \operatorname{erf}\left(\sqrt{n} \frac{\Delta}{\sigma\sqrt{2}}\right), \quad (3.1.1)$$

где  $\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-x^2) dx$  – уже упоминавшаяся в курсе математического анализа "функция ошибок" имеющаяся во всех средах конечного пользователя и библиотеках Фортрана.

В таблице 3.1 приведены значения переменной  $\sqrt{n} \frac{\Delta}{\sigma}$  для некоторых часто используемых значений вероятности  $\beta$ .

Таблица 3.1

|                                  |      |      |      |       |       |
|----------------------------------|------|------|------|-------|-------|
| $\beta$                          | 0.9  | 0.95 | 0.99 | 0.995 | 0.999 |
| $\sqrt{n} \frac{\Delta}{\sigma}$ | 1.64 | 1.96 | 2.58 | 2.81  | 3.29  |

Проанализируем формулу (3.1.1). Она связывает три числа:  $n$  – количество обрабатываемых измерений ("стоимость" эксперимента),  $\Delta$  – полуширина интервала-оценки ("качество" оценки) и  $\beta$  – вероятность накрытия оцениваемого параметра<sup>34</sup> ("надежность" оценки). Видно, что при фиксированной стоимости эксперимента ( $n$ ) повышение качества оценки (уменьшение  $\Delta$ ) приводит к уменьшению ее надежности ( $\beta$ ): интеграл от положительной функции убывает при уменьшении длины промежутка интегрирования. Увеличить надежность оценки при сохранении ее качества можно лишь за счет увеличения затрат и т. д.

<sup>34</sup>В математической статистике  $\beta$  именуется *доверительной вероятностью*, а интервал-оценка с соответствующей  $\Delta$  – *доверительным интервалом*.

### 3.2. Оценивание параметров распределения случайной переменной

Обобщим задачу, рассмотренную в предыдущем пункте. Единственные *объективные* данные, имеющиеся в нашем распоряжении – это экспериментально полученные  $n$  чисел  $w_1, \dots, w_n$ . Числовой вектор  $w = [w_1, \dots, w_n]^T$  называется *выборкой*, а число  $n$  – *объемом выборки*.

Возможность статистической обработки выборки обеспечивается следующими *предположениями*:

- 1)  $w$  является значением случайного вектора  $X = [X_1 \dots X_n]^T$ ;
- 2) координаты случайного вектора  $X$  независимы в совокупности;
- 3) координаты случайного вектора  $X$  являются копиями наблюдаемой случайной переменной  $\xi$ ;
- 4) наблюдаемая с.п.  $\xi$  принадлежит известному семейству случайных переменных с  $k$  параметрами  $\vartheta_1, \dots, \vartheta_k$ .

*Ответственность за справедливость этих предположений несет постановщик задачи (экспериментатор).*

Таким образом, задача свелась к оцениванию по имеющейся выборке  $w = [w_1 \dots w_n]^T$  параметра (векторного)  $\vartheta = [\vartheta_1 \dots \vartheta_k]^T$ . Оценкой параметра  $\vartheta$  мы будем называть  $k$ -мерный параллелепипед  $J$  с центром в точке  $\hat{\vartheta} = [\hat{\vartheta}_1 \dots \hat{\vartheta}_k]^T$  и "ребрами"

$$] \hat{\vartheta}_1 - \Delta_1, \hat{\vartheta}_1 + \Delta_1[ , \dots , ] \hat{\vartheta}_k - \Delta_k, \hat{\vartheta}_k + \Delta_k[.$$

Центр параллелепипеда-оценки находится по имеющейся выборке  $w$ :

$$\hat{\vartheta} = \varphi(w),$$

где  $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^k$  – заданная функция, именуемая *статистикой*. Так как предполагается, что процедура оценивания будет повторяться многократно, рассмотрим случайную переменную  $T = \varphi(X)$ , которую также называют статистикой.

Предположим для определенности, что случайная переменная  $\xi$  абсолютно непрерывна. Поскольку ее плотность распределения  $f_\xi(x; \vartheta)$  известна с точностью до оцениваемого параметра, можно построить плотность распределения случайного вектора  $X$ :

$$f_X(x; \vartheta) = f_\xi(x_1; \vartheta) \cdot f_\xi(x_2; \vartheta) \dots f_\xi(x_n; \vartheta).$$

Отсюда следует *принципиальная* возможность найти плотность распределения  $f_T(t, \vartheta)$  статистики  $T$ .

Поскольку большие отклонения случайного вектора от его математического ожидания маловероятны, можно надеяться, что случайные значения статистики  $T = \varphi(X)$  будут концентрироваться около ее математического ожидания. Отсюда вытекает естественное требование: математическое ожидание статистики должно совпадать с оцениваемым параметром ( $M(T) = \vartheta$ ) или хотя бы приближаться к нему при увеличении объема выборки ( $\lim_{n \rightarrow +\infty} M(T) = \vartheta$ ).

Статистика, удовлетворяющая условию  $M(T) = \vartheta$ , называется *несмещенной*; использование такой статистики при обработке эксперимента не дает дополнительной *систематической* ошибки.

Статистика, удовлетворяющая условию  $\lim_{n \rightarrow +\infty} M(T) = \vartheta$ , называется *асимптотически несмещенной*. Практическое значение асимптотической несмещенности состоит в возможности получения пренебрежимо малой дополнительной систематической ошибки за счет увеличения объема выборки.

Серьезное предупреждение. Обратите внимание, что мы изначально предполагаем отсутствие *систематической погрешности наблюдения*. Если такая погрешность имеется, то ее невозможно устранить никакой математической обработкой!

Предположим, что статистика является несмещенной. Тогда для всех компонент  $M(T_j) = \vartheta_j$ , и мы можем представить случайное событие "оцениваемый параметр  $\vartheta$  накрывается параллелепипедом  $J$ " так:

$$(\vartheta \in J) = \bigcap_{j=1}^k A_j,$$

где  $A_j$  – случайное событие ( $|T_j - M(T_j)| < \Delta_j$ ).

Оценим вероятность события  $\vartheta \in J$  в предположении, что все с.п.  $T_j$  имеют дисперсии. В силу неравенства Чебышева

$$\mathbf{P}(|T_j - M(T_j)| < \Delta_j) \geq 1 - \frac{D(T_j)}{\Delta_j^2}, \quad j = 1, \dots, k. \quad (3.2.1)$$

Если  $k = 2$ , то по теореме сложения вероятностей

$$\mathbf{P}(\vartheta \in J) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cup A_2) \geq \mathbf{P}(A_1) + \mathbf{P}(A_2) - 1,$$

откуда с учетом (3.2.1) получаем

$$\mathbf{P}(\vartheta \in J) \geq 1 - \frac{D(T_1)}{\Delta_1^2} - \frac{D(T_2)}{\Delta_2^2}.$$

Аналогично при любом  $k$  имеем

$$\beta = \mathbf{P}(\vartheta \in J) \geq 1 - \sum_{j=1}^k \frac{D(T_j)}{\Delta_j^2}. \quad (3.2.2)$$

Предположим вдобавок, что дисперсии всех компонент статистики  $T$  стремятся к нулю при неограниченном росте объема выборки. Тогда неравенство (3.2.2) показывает, что вероятность накрытия оцениваемого параметра параллелепипедом-оценкой с заданными длинами ребер может быть сделана сколь угодно близкой к единице за счет увеличения объема выборки<sup>35</sup>. Такая статистика называется *состоятельной*.

Очевидно, "хорошая" статистика должна быть несмещенной (хотя бы асимптотически) и состоятельной. В следующем параграфе мы рассмотрим один из приемов построения таких "хороших" статистик.

Естественно считать, как и в п.3.1, что размеры *доверительного параллелепипеда*  $J$  характеризуют *качество* оценки: чем эти размеры меньше, тем оценка точнее.

Точно так же *доверительная вероятность*  $\beta$  (вероятность накрытия оцениваемого параметра  $\vartheta$  доверительным параллелепипедом) характеризует *надежность* оценки: чем больше  $\beta$ , тем реже доверительный параллелепипед *не будет* накрывать оцениваемый параметр. Аналогично п.3.1, уменьшение размеров параллелепипеда  $J$  приводит к уменьшению доверительной вероятности  $\beta$ .

И, наконец, объем выборки  $n$  (количество экспериментов) характеризует *стоимость* получения оценки.

Понятно, что лозунг "лучше, больше, дешевле!" не может быть реализован. Если фиксировать затраты ( $n$ ), то *увеличение точности* оценки неизбежно приведет к *уменьшению ее надежности*. Желая увеличить надежность, мы будем вынуждены либо пожертвовать точностью, либо заплатить большую цену и т. д.

### 3.3. Метод максимального правдоподобия

Вообще говоря, статистики изобретаются (и поэтому обычно носят имена их авторов). Метод, изобретенный Р.А. Фишером<sup>36</sup>, является одним из регулярных методов построения "хороших" статистик.

<sup>35</sup> Можно показать, что то же самое справедливо, если заменить несмещенность статистики ее *асимптотической несмещенностью*.

<sup>36</sup> Рональд Айлмер ФИШЕР (R.A. Fisher, 1890-1962) – английский биолог, математик и статистик. С его именем связаны многие понятия и утверждения в математической статистике.

Если  $f_X(x; \vartheta)$  – семейство плотностей распределения с векторным параметром  $\vartheta$ , то, подставив вместо переменного вектора  $x$  полученную в эксперименте выборку  $w$ , получим новую функцию, заданную на множестве допустимых значений параметра  $\vartheta$ . Эта функция обозначается  $lik$  и называется *функцией правдоподобия*<sup>37</sup> выборки:

$$lik(\vartheta) = f_X(x; \vartheta)|_{x=w}.$$

Поскольку координаты случайного вектора  $X$  независимы в совокупности и имеют одинаковые плотности распределения, то

$$lik(\vartheta) = \prod_{j=1}^n f_{\xi}(w_j; \vartheta).$$

Аналогично, в дискретном случае имеем

$$lik(\vartheta) = \prod_{j=1}^n \mathbf{P}(\xi = w_j; \vartheta).$$

*Метод максимального правдоподобия предлагает принять в качестве центра доверительного параллелепипеда точку глобального максимума функции правдоподобия.*

Примеры. 1. При оценивании математического ожидания нормальной наблюдаемой с.п.  $\xi$  с известной дисперсией  $\sigma^2$  плотность распределения принадлежит однопараметрическому семейству

$$f_{\xi}(t; a) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot \exp\left(-\frac{(t-a)^2}{2\sigma^2}\right).$$

Функция правдоподобия равна

$$lik(a) = f_X(w; a) = \left(\frac{1}{\sqrt{2\pi} \cdot \sigma}\right)^n \cdot \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^n (w_j - a)^2\right).$$

В силу *возрастания* логарифма точка максимума функции правдоподобия совпадает с точкой максимума ее логарифма

$$l(a) = \ln(lik(a)) = -n \cdot \ln(\sqrt{2\pi} \cdot \sigma) - \frac{1}{2\sigma^2} \sum_{j=1}^n (w_j - a)^2.$$

---

<sup>37</sup>likelihood (англ.) – правдоподобие. Мы настоятельно рекомендуем считать название "функция правдоподобия" *одним* словом, чтобы избежать соблазна найти связь с интуитивным представлением о правдоподобии.

Решив уравнение

$$l'(a) \equiv \frac{1}{\sigma^2} \sum_{j=1}^n (w_j - a) = 0,$$

найдем единственную стационарную точку функции правдоподобия

$$\bar{w} = \hat{a} = \frac{1}{n} \sum_{j=1}^n w_j.$$

Поскольку  $l''(a) \equiv -\frac{n}{\sigma^2} < 0$ , то  $\hat{a} = \bar{w}$  — точка *локального максимума*, который, очевидно, является и *глобальным*.

2. Если семейство нормальных плотностей распределения содержит два *оцениваемых* параметра ( $a$  и  $\mathcal{D} = \sigma^2$ ), то

$$f_{\xi}(t; a, \mathcal{D}) = \frac{1}{\sqrt{2\pi\mathcal{D}}} \cdot \exp\left(-\frac{(t-a)^2}{2\mathcal{D}}\right);$$

$$lik(a, \mathcal{D}) = f_X(w; a, \mathcal{D}) = \left(\frac{1}{\sqrt{2\pi\mathcal{D}}}\right)^n \cdot \exp\left(-\frac{1}{2\mathcal{D}} \sum_{j=1}^n (w_j - a)^2\right);$$

$$l(a, \mathcal{D}) = \ln(lik(a, \mathcal{D})) = -\frac{n}{2} \cdot \ln(2\pi\mathcal{D}) - \frac{1}{2\mathcal{D}} \sum_{j=1}^n (w_j - a)^2.$$

Решая уравнение  $\nabla l = \theta$ , т.е. систему

$$\begin{cases} \frac{1}{\mathcal{D}} \sum_{j=1}^n (w_j - a) = 0 \\ -\frac{n}{2\mathcal{D}} + \frac{1}{2\mathcal{D}^2} \sum_{j=1}^n (w_j - a)^2 = 0 \end{cases},$$

найдем *единственную* стационарную точку:

$$\bar{w} = \hat{a} = \frac{1}{n} \sum_{j=1}^n w_j; \quad s^2 = \hat{\mathcal{D}} = \frac{1}{n} \sum_{j=1}^n (w_j - \bar{w})^2.$$

Вычислив в этой точке матрицу Гессе:

$$l''(\hat{a}, \hat{\mathcal{D}}) = \begin{bmatrix} -\frac{n}{\hat{\mathcal{D}}} & 0 \\ 0 & -\frac{n}{2\hat{\mathcal{D}}} \end{bmatrix},$$

убедимся в ее *отрицательной определенности*. Отсюда следует, что стационарная точка — точка *локального максимума*, который, очевидно, является и *глобальным*.

3. Наблюдаемая с.п.  $\xi$  принадлежит семейству равномерных распределений с двумя параметрами  $a$  и  $b$  ( $b > a$ ):

$$f_{\xi}(t; a, b) = \begin{cases} \frac{1}{b-a} & \text{при } t \in [a, b]; \\ 0 & \text{при } t \notin [a, b]. \end{cases}$$

Функция правдоподобия выборки

$$lik(a, b) = \begin{cases} \frac{1}{(b-a)^n} & \text{при } a \leq \min_j(w_j) \text{ и } \max_j(w_j) \leq b; \\ 0 & \text{при } a > \min_j(w_j) \text{ или } \max_j(w_j) > b. \end{cases}$$

Очевидно, дробь  $\frac{1}{b-a}$  будет максимальна при минимальном значении ее знаменателя. Но  $a$  нельзя сделать больше, чем  $\min_j(w_j)$ , а  $b$  — меньше, чем  $\max_j(w_j)$ . Итак,

$$\hat{a} = \min_j(w_j), \quad \hat{b} = \max_j(w_j).$$

4. Дискретная случайная переменная  $\xi$  принадлежит семейству распределений с одним параметром  $p$  ( $0 < p < 1$ ):

$$\mathbf{P}(\xi = t; p) = \begin{cases} p, & t = 1; \\ 1 - p, & t = 0. \end{cases}$$

Обозначив количество экспериментов  $n$ , а количество единиц ("успехов") в этих экспериментах  $m$ , получим

$$lik(p) = \mathbf{P}(X = w; p) = \prod_{j=1}^n \mathbf{P}(X_j = w_j; p) = p^m \cdot (1-p)^{n-m}$$

(среди  $n$  сомножителей  $m$  равны  $p$ , а остальные равны  $1-p$ ). Поэтому

$$l(p) = \ln(lik(p)) = m \cdot \ln(p) + (n-m) \cdot \ln(1-p).$$

Производная  $l'(p) = \frac{m}{p} - \frac{n-m}{1-p}$  обращается в нуль только в одной точке, и эта единственная стационарная точка  $\hat{p} = \frac{m}{n}$  дает, очевидно, глобальный максимум функции правдоподобия. Таким образом, *относительная частота успеха* является статистикой максимального правдоподобия для вероятности в случае двух исходов эксперимента.

Найдем числовые характеристики статистик максимального правдоподобия, полученных в этих примерах.

Пример 1. Статистика  $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$ .

Как показано в п.3.1,

$$M(\bar{X}) = a, \quad D(\bar{X}) = \frac{\sigma^2}{n}.$$

Пример 2. Статистики

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad S^2 = \frac{1}{n} \sum_{j=1}^n (X_j - \bar{X})^2.$$

Аналогично примеру 1 получаем

$$M(\bar{X}) = a, \quad D(\bar{X}) = \frac{\mathcal{D}}{n}.$$

Можно показать, что

$$M(S^2) = \frac{n-1}{n} \mathcal{D}; \quad D(S^2) = \frac{(n-1)^2}{n^3} \cdot \mu_4(\xi) - \frac{(n-1)(n-3)}{n^3} \cdot \mathcal{D}^2$$

(здесь  $\mu_4(\xi) = M((\xi - a)^4)$  – четвертый центральный момент с.п.  $\xi$ ).

Пример 3. Статистики

$$X_{max} = \max_j(X_j); \quad X_{min} = \min_j(X_j).$$

Построим функцию распределения

$$F_{X_{max}}(t) = \mathbf{P}(X_{max} < t).$$

Неравенство  $X_{max} < t$  означает, что все координаты вектора  $X$  меньше, чем  $t$ . Поскольку все  $X_j$  независимы в совокупности, имеем

$$\mathbf{P}(X_{max} < t) = \prod_{j=1}^n \mathbf{P}(X_j < t) = (F_{\xi}(t))^n = \begin{cases} 0 & \text{при } t < a; \\ \left(\frac{t-a}{b-a}\right)^n & \text{при } a \leq t \leq b; \\ 1 & \text{при } t > b. \end{cases}$$

Несложные вычисления дают

$$M(X_{max}) = \frac{nb + a}{n + 1}; \quad D(X_{max}) = \frac{n(b-a)^2}{(n+1)(n+2)}.$$

Аналогично, для статистики  $X_{min}$  получаем

$$M(X_{min}) = \frac{na + b}{n + 1}; \quad D(X_{min}) = \frac{n(b-a)^2}{(n+1)(n+2)}.$$



Пример 4. Статистика "относительная частота успеха".

Обозначим с.п. "количество успехов"  $\nu$ . Тогда, очевидно, относительная частота успеха равна  $\frac{\nu}{n}$ .

Как известно из п.2.3, с.п.  $\nu$  имеет биномиальное распределение, причем  $M(\nu) = n \cdot p$ ,  $D(\nu) = n \cdot p \cdot (1 - p)$ . Отсюда

$$M\left(\frac{\nu}{n}\right) = p, \quad D\left(\frac{\nu}{n}\right) = \frac{p(1-p)}{n}.$$

Отметим, что для статистики  $\bar{X}$  в примерах **1** и **2** математическое ожидание совпадает с оцениваемым параметром, т.е.  $\bar{X}$  – *несмещенная* статистика. То же можно сказать и об относительной частоте в примере **4**. В то же время в примерах **2** и **3**

$$M(S^2) = \frac{n-1}{n} \mathcal{D} \neq \mathcal{D},$$

$$M(X_{max}) = \frac{nb+a}{n+1} \neq b, \quad M(X_{min}) = \frac{na+b}{n+1} \neq a,$$

т.е. эти три статистики – *смещенные*. Однако они являются *асимптотически несмещенными*, так как

$$\lim_{n \rightarrow +\infty} M(S^2) = \mathcal{D},$$

$$\lim_{n \rightarrow +\infty} M(X_{max}) = b, \quad \lim_{n \rightarrow +\infty} M(X_{min}) = a.$$

Отметим также, что дисперсии всех рассмотренных статистик стремятся к нулю при  $n \rightarrow \infty$ . В силу неравенства (3.2.2) все эти статистики *состоятельны*.

*Можно показать, что при достаточно общих допущениях статистики максимального правдоподобия являются состоятельными и асимптотически несмещенными.*

## Глава 4. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

В предыдущей главе рассматривалась ситуация, когда закон распределения случайной переменной *известен* с точностью до нескольких числовых параметров. Сейчас мы рассмотрим другую задачу: пусть имеется выборка, порожденная с.п., закон распределения которой *неизвестен*. Формулируется некоторая гипотеза об этом законе распределения, и ставится вопрос: согласуется ли полученная выборка с этой гипотезой? Точнее: дает ли эта выборка основание отвергнуть гипотезу?

Наиболее простая ситуация возникает, если появление выборки *невозможно* при справедливости проверяемой гипотезы. Так, гипотеза о равномерном на  $[0, 1]$  распределении наблюдаемой с.п. очевидно *отвергается* при наличии в выборке, например, числа 2 или числа  $-1$ . Подобные тривиальные случаи в дальнейшем не рассматриваются.

Если выборка *не противоречит* гипотезе, то эта гипотеза не может быть отвергнута безоговорочно. Речь может идти лишь о признании полученной выборки *маловероятной* при справедливости данной гипотезы.

Если установлено, что полученная выборка *плохо согласуется* с гипотезой, то гипотезу следует *отклонить*. В противном случае гипотеза *сохраняется* (для дальнейшей проверки).

При этом следует помнить, что в обоих случаях мы можем ошибиться: отклонить верную гипотезу (так называемая *ошибка первого рода*) или сохранить неверную гипотезу (*ошибка второго рода*).

В предположении справедливости проверяемой гипотезы можно оценить вероятность появления наблюдаемой выборки, а, следовательно, и вероятность ошибки первого рода. В то же время оценить вероятность ошибки второго рода невозможно, так как для ее оценки потребовалось бы исследовать все альтернативные гипотезы. Поэтому обычно употребляют выражение "гипотеза сохраняется а не "гипотеза принимается".

Вероятность сохранения верной гипотезы (доверительную вероятность) принято обозначать  $\beta$ . Постановщик задачи обычно назначает допустимую (с его точки зрения) вероятность ошибки первого рода  $(1 - \beta)$ .

Серьезное предупреждение. Доверять *доверительной вероятности* в задаче проверки гипотезы, как и в задаче оценивания параметров, можно только при *достаточно большом* количестве проверок. При *однократной* проверке мы либо примем правильное решение, либо ошибемся. И ни о каких вероятностях в этом случае говорить нельзя!

## 4.1. Проверка гипотезы о согласии с заданным законом распределения

Простейшая возможная гипотеза заключается в том, что имеющаяся выборка порождена случайной переменной (случайным вектором) с *заданным* законом распределения. Алгоритмы проверки этой гипотезы обычно называются *критериями согласия*. Мы опишем два таких критерия – *критерий хи-квадрат*, изобретенный К. Пирсоном<sup>38</sup>, и *критерий Колмогорова*<sup>39</sup>.

### Критерий хи-квадрат

Этот алгоритм проверки гипотезы состоит в следующем:

1) пространство элементарных событий, т.е. гипотетическое множество значений случайной переменной (случайного вектора), разбивают на  $m$  непересекающихся частей (попарно несовместных событий)  $J_1, \dots, J_m$ ;

2) в предположении справедливости проверяемой гипотезы находят вероятности этих событий  $p_1, \dots, p_m$  ( $p_1 + \dots + p_m = 1$ ) и "ожидаемые частоты"  $\nu_1 = p_1 n, \dots, \nu_m = p_m n$  (здесь  $n$  – объем выборки), т.е. количества элементов выборки, которые *должны попасть* в эти части;

3) находят "наблюдённые частоты"  $n_1, \dots, n_m$  ( $n_1 + \dots + n_m = n$ ), т.е. количества элементов выборки, *фактически попавших* в эти части.

4) Если гипотеза верна, наблюдённые частоты должны *мало* отличаться от ожидаемых. Расстояние между векторами наблюдённых и ожидаемых частот определяется формулой

$$\hat{\chi}^2 = \sum_{k=1}^m \frac{(n_k - \nu_k)^2}{\nu_k}.$$

Число  $\hat{\chi}^2$  может рассматриваться как значение с.п. (статистики)  $\chi^2$ . К. Пирсон доказал, что для *больших* объемов выборок плотность распределения этой с.п. (независимо от распределения породившего выборку случайного вектора) хорошо аппроксимируется функцией

---

<sup>38</sup>Карл ПИРСОН (C. Pearson, 1857-1936) – английский математик и биолог, профессор Лондонского университета, основатель журнала "Биометрика".

<sup>39</sup>Андрей Николаевич КОЛМОГОРОВ (1903-1987) – один из крупнейших математиков XX века. Действительный член АН СССР, член практически всех наиболее авторитетных научных сообществ мира. Создатель одной из крупнейших в стране научных школ. Автор основополагающих работ по теории функций, теории вероятностей, теории информации, теории алгоритмов...

$$f_{\chi^2}(t, \varkappa) = \frac{t^{\varkappa/2-1}}{2^{\varkappa/2}\Gamma(\varkappa/2)} \cdot \exp(-t/2) \cdot \delta_1(t), \quad (4.1.1)$$

где  $\delta_1$  – функция Хевисайда,  $\varkappa = m - 1$ .

Функция  $f_{\chi^2}(t, \varkappa)$  именуется *плотностью распределения хи-квадрат с  $\varkappa$  степенями свободы*. Ее график при  $\varkappa = 4$  изображен на рис.4.1.

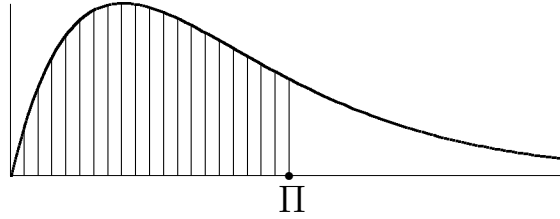


Рис.4.1

5) Назначив допустимую вероятность ошибки первого рода  $(1 - \beta)$ , можно определить *пороговое значение* статистики  $\chi^2$  (точка  $\Pi$  на рис.4.1) как решение уравнения

$$\mathbf{P}(\chi^2 < \Pi) = \int_{-\infty}^{\Pi} f_{\chi^2}(t, \varkappa) dt = \beta,$$

т.е. как значение в точке  $\beta$  функции, *обратной* к функции распределения с.п.  $\chi^2$ .

Замечание. Функция распределения хи-квадрат и обратная ей функция табулированы. Их значения "умеют" вычислять среды конечного пользователя и библиотеки Фортрана.

6) Проверяемая гипотеза отклоняется, если вычисленное по выборке значение статистики *больше* порогового ( $\hat{\chi}^2 > \Pi$ ), и сохраняется – в противном случае.

При многократном повторении процедуры проверки гипотезы по критерию хи-квадрат можно ожидать, что относительная частота ошибки первого рода будет близка к заданной вероятности  $(1 - \beta)$ .

Как видно из рис.4.1, статистика хи-квадрат может принимать на выборках как угодно большие значения, но чем больше значение статистики, тем реже оно будет встречаться. Площадь заштрихованной части фигуры на рис.4.1 равна доверительной вероятности  $\beta$ . Площадь незаштрихованной части  $(1 - \beta)$  – это вероятность ошибки первого рода.

Замечания. 1. Каково бы ни было гипотетическое распределение с.п.  $X$ , применение критерия хи-квадрат начинается с замены этого распределения на дискретное распределение с вероятностями элементарных событий  $p_k = \mathbf{P}(X \in J_k)$ ,  $k = 1, \dots, m$ .

Если  $m$  мало, то будут потеряны характерные особенности гипотетического распределения, и сделанные выводы будут ненадежными. Если же  $m$  слишком велико, то наблюдаемые частоты могут оказаться очень малыми, что тоже снижает достоверность выводов.

Обсуждение оптимального значения  $m$  выходит за рамки нашего курса. Однако очевидно, что объем выборки должен быть достаточно велик, чтобы обеспечить и достаточно большое число  $m$ , и достаточно большую среднюю наблюдаемую частоту  $\frac{n}{m}$ .

2. Если  $m$  – количество частей, на которые разбивается гипотетическое пространство элементарных событий, – задано, то возникает вопрос: а как выбирать эти части? По-видимому, не будет хуже, если брать их *равновероятными*:

$$p_k = \mathbf{P}(X \in J_k) \equiv \frac{1}{m}; \quad k = 1, \dots, m.$$

3. В приложениях часто встречается ситуация, когда гипотетическое распределение не задано полностью, а содержит несколько неизвестных параметров. В этом случае можно сначала оценить эти параметры по выборке методом максимального правдоподобия, описанным в п.3.3, а затем применить критерий хи-квадрат, заменив неизвестные параметры на их оценки. Р. Фишер показал, что алгоритм применения критерия хи-квадрат в этом случае сохраняется, но число степеней свободы уменьшается на число оцениваемых параметров, т.е. в формуле (4.1.1) следует положить  $\varkappa = m - 1 - r$ , где  $r$  – количество оцениваемых параметров.

### Критерий Колмогорова

Пусть  $F_\xi$  – гипотетическая функция распределения наблюдаемой *абсолютно непрерывной* с.п., а  $x = [x_1, \dots, x_n]$  – *вариационный ряд*, т.е. *упорядоченная по возрастанию* выборка. Тогда гипотетическая вероятность события ( $\xi < t$ ) равна  $F_\xi(t)$ , а наблюдаемая относительная частота этого события равна количеству элементов вариационного ряда, лежащих левее точки  $t$ . Пусть  $\Delta$  – наибольшее отклонение гипотетической вероятности от наблюдаемой относительной частоты. Легко видеть, что

$$\Delta = \max_{1 \leq k \leq n} \left\{ F_\xi(x_k) - \frac{k-1}{n}, \frac{k}{n} - F_\xi(x_k) \right\}.$$

А.Н. Колмогоров показал, что при большом объеме выборки  $n$  функция распределения статистики  $\Lambda = \sqrt{n} \cdot \Delta$  хорошо аппроксимируется функцией

$$K(\lambda) = \left[ 1 - 2 \sum_{m=1}^{+\infty} (-1)^{m-1} \exp(-2m^2 \lambda^2) \right] \cdot \delta_1(\lambda).$$

Эта функция *не зависит* от гипотетической функции распределения  $F_\xi$ . Ее называют *функцией распределения Колмогорова*. Функция  $K(\lambda)$  и обратная ей функция табулированы, их "умеют" вычислять среды конечного пользователя и библиотеки Фортрана.

Сформулируем алгоритм критерия Колмогорова.

- 1) Выборка упорядочивается по возрастанию.
- 2) Вычисляется  $\hat{\lambda}$  – значение статистики  $\Lambda$  на выборке.
- 3) Назначается доверительная вероятность  $\beta$  (или, что равносильно, допустимая вероятность ошибки первого рода  $(1 - \beta)$ ).
- 4) Вычисляется пороговое значение статистики Колмогорова  $\Pi$ , т.е. решение уравнения  $K(\Pi) = \beta$ .
- 5) Проверяемая гипотеза отклоняется, если вычисленное по выборке значение статистики *больше* порогового ( $\hat{\lambda} > \Pi$ ), и сохраняется – в противном случае.

## 4.2. Пример: проверка датчика псевдослучайных чисел

По утверждению разработчика программного датчика псевдослучайных чисел (п.2.7), этот датчик генерирует случайную переменную со стандартным нормальным распределением ( $M(X) = 0, \sigma(X) = 1$ ). Имеется выборка из  $n = 100$  чисел, полученных от этого датчика:

|       |       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| -0.64 | 1.78  | 0.24  | 1.07  | -1.74 | 0.18  | 0.09  | 1.57  | 0.08  | -1.76 |
| -0.83 | -0.15 | -0.43 | 0.79  | -1.60 | -0.69 | 0.17  | 0.86  | 1.60  | -0.80 |
| -0.15 | -1.99 | 0.97  | 1.32  | 0.85  | -0.49 | 1.00  | 0.44  | 0.81  | -0.35 |
| -0.43 | -0.34 | 1.04  | 0.67  | -0.72 | -0.76 | 1.23  | -0.10 | -0.85 | -0.84 |
| 0.29  | 0.41  | 1.08  | 1.82  | 0.81  | 0.59  | 0.28  | -0.21 | -0.35 | 0.15  |
| -0.84 | 1.84  | -1.71 | 0.56  | 0.90  | -1.69 | 0.93  | -0.32 | 0.07  | -1.39 |
| 1.68  | -1.80 | 0.45  | 0.82  | -1.09 | 1.73  | -0.08 | -0.25 | 0.06  | 0.15  |
| -0.14 | -0.04 | 1.22  | 0.01  | 0.20  | -0.58 | -1.72 | 2.25  | 1.93  | -1.82 |
| -0.35 | 0.29  | -2.15 | -1.21 | -1.20 | -0.98 | -1.25 | -1.25 | 0.93  | -1.48 |
| -0.89 | 2.10  | 0.76  | -0.25 | -0.93 | 0.06  | -0.38 | -0.34 | -2.39 | 0.27  |

Проверим, согласуется ли полученная выборка с декларированным законом распределения, с помощью критерия хи-квадрат и критерия Колмогорова.

1. **Критерий хи-квадрат.** Согласно алгоритму, изложенному в п.4.1, разделим множество значений с.п. ( $\mathbb{R}$ ) на  $m = 10$  равновероятных (для стандартного нормального распределения) промежутков.

Решая уравнения

$$\frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{x_k} \exp\left(-\frac{t^2}{2}\right) dt = \frac{1}{2} \cdot \left(1 + \operatorname{erf}\left(\frac{x_k}{\sqrt{2}}\right)\right) = \frac{k}{10}; \quad k = 1, \dots, 9,$$

найдем границы этих промежутков:

|       |       |       |       |       |      |      |      |      |      |
|-------|-------|-------|-------|-------|------|------|------|------|------|
| $k$   | 1     | 2     | 3     | 4     | 5    | 6    | 7    | 8    | 9    |
| $x_k$ | -1.28 | -0.84 | -0.52 | -0.25 | 0.00 | 0.25 | 0.52 | 0.84 | 1.28 |

Очевидно, *ожидаемое* количество чисел в каждом из равновероятных промежутков  $]-\infty, x_1[$ ,  $[x_1, x_2[$ ,  $\dots$ ,  $[x_9, +\infty[$ , равно 10. Количества фактически наблюдаемых чисел в промежутках приведены в следующей таблице:

|       |    |    |   |    |   |    |   |   |    |    |
|-------|----|----|---|----|---|----|---|---|----|----|
| $k$   | 1  | 2  | 3 | 4  | 5 | 6  | 7 | 8 | 9  | 10 |
| $n_k$ | 12 | 10 | 9 | 11 | 8 | 12 | 7 | 8 | 12 | 11 |

Вычислим значение статистики хи-квадрат на выборке:

$$\widehat{\chi^2} = 0.1 \sum_{k=1}^{10} (n_k - 10)^2 = 3.2$$

Число степеней свободы  $\nu = 9$ . В таблице 4.1 приведены пороговые значения статистики хи-квадрат с 9 степенями свободы для некоторых часто используемых значений доверительной вероятности  $\beta$ :

Таблица 4.1

|         |      |      |      |       |       |
|---------|------|------|------|-------|-------|
| $\beta$ | 0.9  | 0.95 | 0.99 | 0.995 | 0.999 |
| $\Pi$   | 14.7 | 16.9 | 21.7 | 23.6  | 27.9  |

Выбрав  $\beta = 0.95$ , т.е. допуская ошибочное отклонение гипотезы в пяти случаях из ста, находим  $\Pi = 16.9$ . Поскольку  $\widehat{\chi^2} = 3.2 < 16.9$ , мы не имеем основания предъявить претензии разработчику датчика.

**2. Критерий Колмогорова.** Упорядочив выборку по возрастанию, найдем  $\hat{\lambda} = 10\Delta = 0.572$ .

В таблице 4.2 приведены пороговые значения статистики Колмогорова для некоторых часто используемых значений доверительной вероятности  $\beta$ .

Таблица 4.2

|         |      |      |      |       |       |
|---------|------|------|------|-------|-------|
| $\beta$ | 0.9  | 0.95 | 0.99 | 0.995 | 0.999 |
| $\Pi$   | 1.22 | 1.36 | 1.63 | 1.73  | 2.23  |

Зададим (как и для критерия хи-квадрат)  $\beta = 0.95$ . Этой доверительной вероятности соответствует пороговое значение статистики Колмогорова  $\Pi = 1.36$ . Поскольку  $\hat{\lambda} < \Pi$ , критерий Колмогорова также не дает оснований забраковать датчик псевдослучайных чисел.

Замечание. Повторим еще раз: мы не можем утверждать, что гипотеза о нормальности верна. Имеющаяся выборка лишь *не дает оснований отвергнуть* ее.

Серьезное предупреждение. Мы применили два статистических критерия для проверки качества программного датчика псевдослучайных чисел.

Иногда эти критерии применяются в другой ситуации. Пусть требуется оценить вероятность некоторого события. Следовало бы это оценивание провести так: проделать серию экспериментов и найти относительную частоту появления события. Однако если событие редкое, то требуется очень большой объем выборки (не встретив событие ни разу при трех испытаниях, не стоит говорить, что вероятность его появления равна нулю!). Экспериментатор, желая сэкономить на эксперименте, делит множество значений с.п. на *небольшое* число частей  $m$ , выдвигает некоторую гипотезу о законе распределения, а затем, получив удовлетворительное (*за счет малости  $m$* ) согласие выборки с гипотезой, применяет гипотетическое распределение для оценки вероятности *редкого* события. Нетрудно убедиться в том, что таким "методом" можно получить любое заданное наперед значение искомой вероятности.

### 4.3. Некоторые другие применения критерия хи-квадрат

В этом пункте мы кратко рассмотрим еще две часто возникающие задачи проверки гипотез. В обоих случаях мы ограничимся алгоритмом применения критерия хи-квадрат.



## Проверка гипотезы о независимости в совокупности координат случайного вектора

Рассмотрим *двумерный* случайный вектор с координатами  $X$  и  $Y$ . Имея выборку объема  $N$ , разделим координатную ось  $OX$  на непересекающиеся промежутки  $J_{x1}, \dots, J_{xK}$ , а координатную ось  $OY$  – на непересекающиеся промежутки  $J_{y1}, \dots, J_{yR}$ . Если верна гипотеза о независимости координат, то вероятность события  $(X \in J_{xk}) \cap (Y \in J_{yr})$  равна произведению вероятностей событий  $X \in J_{xk}$  и  $Y \in J_{yr}$ .

Построим так называемую *таблицу сопряженности признаков*, т.е.  $(K \times R)$ -матрицу  $T$ , элемент  $t_{kr}$  которой – количество элементов выборки, попавших в прямоугольник  $J_{xk} \times J_{yr}$ .

Вычислим частоты попадания с.п.  $X$  в промежутки  $J_{xk}$ :

$$n_{xk} = \sum_{r=1}^R t_{kr}, \quad k = 1, \dots, K$$

и частоты попадания с.п.  $Y$  в промежутки  $J_{yr}$ :

$$n_{yr} = \sum_{k=1}^K t_{kr}, \quad r = 1, \dots, R.$$

Поскольку *вероятности* событий  $X \in J_{xk}$  и  $Y \in J_{yr}$  неизвестны, заменим их точечными оценками – *относительными частотами* этих событий. Получим, что при *независимых* координатах ожидаемая частота попадания случайного вектора в прямоугольник  $J_{xk} \times J_{yr}$  равна

$$\nu_{kr} = \frac{n_{xk} \cdot n_{yr}}{N}.$$

Вычислим значение статистики хи-квадрат:

$$\hat{\chi}^2 = \sum_{k=1}^K \sum_{r=1}^R \frac{(t_{kr} - \nu_{kr})^2}{\nu_{kr}} = N \cdot \left( \sum_{k=1}^K \sum_{r=1}^R \frac{t_{kr}^2}{n_{xk} \cdot n_{yr}} - 1 \right).$$

Определим число степеней свободы. Гипотетическое пространство элементарных событий мы разбили на  $K \cdot R$  частей. При этом мы оценивали по выборке параметры дискретного распределения – вероятности

$$p_{xk} = \mathbf{P}(X \in J_{xk}), \quad k = 1, \dots, K; \quad p_{yr} = \mathbf{P}(Y \in J_{yr}), \quad r = 1, \dots, R.$$

Но из этих  $K + R$  параметров лишь  $K + R - 2$  свободных, так как  $\sum_{k=1}^K p_{xk} = 1$  и  $\sum_{r=1}^R p_{yr} = 1$ . Согласно замечанию **3** из п.4.1,

$$\varkappa = K \cdot R - 1 - (K + R - 2) = (K - 1) \cdot (R - 1). \quad (4.2.1)$$

Пример. Имея выборку объема  $N = 20$ ,

|   |    |    |    |    |    |    |    |    |    |    |
|---|----|----|----|----|----|----|----|----|----|----|
| X | 62 | 72 | 88 | 3  | 51 | 84 | 84 | 41 | 46 | 17 |
| Y | 73 | 99 | 23 | 35 | 59 | 41 | 26 | 53 | 28 | 15 |
| X | 57 | 30 | 49 | 74 | 89 | 84 | 21 | 13 | 27 | 41 |
| Y | 80 | 53 | 95 | 55 | 62 | 15 | 71 | 9  | 0  | 2  |

разделим координатные оси на непересекающиеся части (мы взяли на каждой оси по четыре *равночастотных* промежутка):

$$J_{x1} = ] - \infty, 28[, \quad J_{x2} = [28, 50[, \quad J_{x3} = [50, 80[, \quad J_{x4} = [80, +\infty[;$$

$$J_{y1} = ] - \infty, 20[, \quad J_{y2} = [20, 50[, \quad J_{y3} = [50, 70[, \quad J_{y4} = [70, +\infty[.$$

Построим таблицу сопряженности признаков –  $(4 \times 4)$ -матрицу  $T$ :

|          | $J_{y1}$ | $J_{y2}$ | $J_{y3}$ | $J_{y4}$ |
|----------|----------|----------|----------|----------|
| $J_{x1}$ | 3        | 1        | 0        | 1        |
| $J_{x2}$ | 1        | 1        | 2        | 1        |
| $J_{x3}$ | 0        | 0        | 2        | 3        |
| $J_{x4}$ | 1        | 3        | 1        | 0        |

Так как промежутки на обеих осях взяты равночастотные, имеем  $n_{xk} = n_{yr} \equiv 5$ .

Находим значение статистики хи-квадрат на имеющейся выборке:

$$\hat{\chi}^2 = 20 \cdot \left( \sum_{k=1}^4 \sum_{r=1}^4 \frac{t_{kr}^2}{5 \cdot 5} - 1 \right) = 13.6.$$

Число степеней свободы, согласно формуле (4.2.1), равно 9. Назначив  $\beta = 0.9$ , т.е. допуская ошибочное отклонение гипотезы в одном случае из десяти, из таблицы 4.1 находим  $\Pi = 14.7$ . Поскольку  $\hat{\chi}^2 < \Pi$ , гипотеза о независимости сохраняется.

Замечания. 1. Рассмотренный пример носит методический характер: имея выборку столь малого объема, конечно, не следует строить таблицу сопряженности признаков из 16 клеток.

2. Нетрудно распространить описанный алгоритм на случай проверки гипотезы о независимости в совокупности координат случайного вектора, размерность которого больше двух.

## Проверка гипотезы о том, что выборки порождены одной и той же случайной переменной

Имеются  $s$  выборок:

$$x^{(1)}, \dots, x^{(s)},$$

где  $x^{(j)} = \{x_1^{(j)}, \dots, x_{n_j}^{(j)}\}$ . Проверяется гипотеза об однородности: все выборки представляют одну и ту же случайную переменную. Опишем алгоритм применения критерия хи-квадрат в этой задаче.

1) Выборки объединяются, образуя новую выборку  $z$  объема  $\sum_{j=1}^s n_j$ .

2) Выборка  $z$  упорядочивается по возрастанию.

3) Назначается число  $m$ , и числовая ось делится на  $m$  непересекающихся промежутков  $J_1, \dots, J_m$ , в каждый из которых попадает одинаковое число элементов выборки  $z$ . Границы этих промежутков служат точечными оценками границ *равновероятных* промежутков, вероятность попадания с.п. в каждый из которых равна  $\frac{1}{m}$ .

В предположении, что гипотеза об однородности верна, относительная частота попадания с.п. в любой из этих промежутков должна быть близка к  $\frac{1}{m}$  для *каждой* выборки  $x^{(j)}$ ,  $j = 1, \dots, s$ . Поэтому ожидаемое количество элементов этой выборки в любом промежутке равно  $\frac{n_j}{m}$ .

4) Находятся  $t_{jk}$  – наблюдаемые количества элементов выборки  $x^{(j)}$  в промежутке  $J_k$ .

5) Вычисляется значение статистики хи-квадрат:

$$\hat{\chi}^2 = \sum_{j=1}^s \sum_{k=1}^m \frac{\left(t_{jk} - \frac{n_j}{m}\right)^2}{\frac{n_j}{m}}.$$

6) Можно показать, что число степеней свободы в этой задаче  $\varkappa = (s - 1) \cdot (m - 1)$  (по  $m - 1$  степени свободы на каждую из  $s$  выборок, причем по выборке оценивается  $m - 1$  параметров гипотетического распределения – границы равновероятных промежутков).

7) Задается доверительная вероятность  $\beta$  и находится  $\Pi$  – соответствующее пороговое значение статистики хи-квадрат.

8) Если  $\hat{\chi}^2 > \Pi$ , гипотеза отвергается.

## ЗАКЛЮЧЕНИЕ

Напомним еще раз, что практическое применение теории вероятностей и математической статистики возможно только при описании *массовых* явлений при условии *статистической устойчивости* относительных частот. Это условие часто постулируется без достаточных на то оснований. Приведем в связи с этим цитату из учебника В.Н. Тутубалина<sup>40</sup> (Теория вероятностей. – М.: МГУ, 1972; второе издание – М.: МГУ, 1993): "Чрезвычайно важно искоренить заблуждение, встречающееся иногда у недостаточно знакомых с теорией вероятностей инженеров и естествоиспытателей, что результат любого эксперимента можно рассматривать как случайную величину. В особо тяжелых случаях к этому присоединяется вера в нормальный закон распределения".

Из-за этого заблуждения теория вероятностей очень часто оказывается объектом вульгаризаций и некорректного применения; встречаются и откровенные спекуляции (достаточно упомянуть псевдоисторические труды академика А.Т. Фоменко).

Мы рекомендуем читателю три брошюры В.Н. Тутубалина (Теория вероятностей в естествознании. – М.: Знание, 1972; Статистическая обработка рядов наблюдений. – М.: Знание, 1973; Границы применимости. – М.: Знание, 1977), в которых обсуждаются "идеологические" проблемы, связанные с применением вероятностно-статистических методов.

Тем, кому необходимо более основательное знакомство с теорией вероятностей и математической статистикой, мы рекомендуем цитированный выше учебник, особенно вторую его часть – "Научные и методические замечания". Дело в том, что учебники по теории вероятностей и математической статистике можно разделить на две группы: первая – учебники для математиков, недоступные прикладнику, и вторая – "учебники где предлагаются определения типа "случайной величиной называется величина, принимающая случайные значения". Курс В.Н. Тутубалина – единственное известное нам исключение.

---

<sup>40</sup>Валерий Николаевич ТУТУБАЛИН (род. 1936) – российский математик, профессор Московского университета.