Министерство образования и науки Российской Федерации
Саратовский государственный технический университет
имени Гагарина Ю.А.
Энгельсский технологический институт (филиал)

В.В. Новиков

ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Учебное пособие

Энгельс 2016

УДК 517.3
ББК 22.1
Н73

Н73 Новиков В.В. Основы теории вероятностей: Учебное пособие / В.В. Новиков. – Энгельс: Изд-во ЭТИ (филиал) СГТУ имени Гагарина Ю.А., 2016. – 59 с.
ISBN 978-5-9907991-1-0

Рецензенты: кафедра теории функций и приближений Саратовского национального исследовательского государственного университета имени Н.Г. Чернышевского;
главный научный сотрудник Института проблем точной механики и управления РАН, д.ф.-м.н., проф. Ольшанский В.Ю.

Учебное пособие рекомендуется студентам технических вузов и классических университетов.

УДК 517.3
ББК 22.1

Одобрено редакционно-издательским советом
ЭТИ (филиал) СГТУ имени Гагарина Ю.А.

Учебное пособие издается в авторской редакции

ISBN 978-5-9907991-1-0

ВВЕДЕНИЕ

Данное пособие содержит материал по основным разделам курса теории вероятностей. Наряду с изложением традиционных для технических вузов вопросов автор ставил целью ознакомить студентов с основными идеями аксиоматического подхода А.Н. Колмогорова. Рассмотрены такие вопросы, как алгебры и сигма-алгебры событий, вероятностная мера и ее свойства, вероятностные пространства, условная вероятность и независимость, одномерные и многомерные случайные величины, предельные теоремы теории вероятностей. Для понимания материала достаточно знакомства с математическим анализом и линейной алгеброй в рамках стандартного курса высшей математики. Основное назначение пособия – помочь студенту при освоении теоретической части курса теории вероятностей и связанных с ней дисциплин. Большая часть приведенных теорем и формул снабжена подробными доказательствами и выводами. Кроме того, практически все разделы снабжены задачами и примерами, иллюстрирующими обсуждаемые теоретические вопросы. Особенностью данного пособия является то, что значительная часть теоретического материала собрана в замечаниях, которые приводятся после формулировок и доказательств. Большая часть этих замечаний посвящена разъяснению трудных для понимания моментов, с которыми могут столкнуться учащиеся при первоначальном знакомстве с предметом. Кроме того, замечания имеют целью помочь студентам избежать некоторых характерных ошибок, которые часто возникают в процессе изучения дисциплины.

Глава I. СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

§ 1. *Математическая модель случайного эксперимента*

Определение. *Случайное событие – это событие, о котором при данных условиях нельзя заранее сказать, произойдет оно или нет. Совокупность условий, сопровождающих случайное событие, (если их можно воспроизвести неоднократно) называют случайным экспериментом или испытанием.*

Случайные события (далее – просто «события») обозначают прописными латинскими буквами $A, B, C, \ldots$

Определение. *Исход данного случайного эксперимента, не разложимый на более простые исходы, называется элементарным исходом или элементарным событием. Элементарные события обозначают как* $\omega_1, \omega_2, \ldots$ *или просто* ω. *Множество всех элементарных событий данного случайного эксперимента называют пространством элементарных событий и обозначают* Ω.

Определение. *Пусть* A *– некоторое (элементарное или нет) событие в данном испытании. Говорят, что данный элементарный исход* $\omega \in \Omega$ *благоприятствует наступлению события* A, *если* A *происходит* *одновременно с* ω.

Таким образом, всякое событие A можно рассматривать как «набор» элементарных событий $\omega \in \Omega$, благоприятствующих A.

Математической моделью случайного эксперимента служит некоторое абстрактное множество Ω, элементы которого ω (точнее, одноэлементные подмножества $\{\omega\} \subset \Omega$) изображают элементарные события, а прочие подмножества $A \subseteq \Omega$ – сложные события, состоящие из элементарных событий, благоприятствующих A.

В частности, все множество Ω (соответственно, пустое множество $\varnothing$) отождествляется с *достоверным событием* (*невозможным событием*) т.е. с тем, которое заведомо происходит (не происходит) в данном испытании.

§ 2. *Алгебры и сигма-алгебры событий*

Определение. *Произведением событий A и B называется событие $C = AB$, состоящее в том, что A и B произошли одновременно.*

Произведению событий A и B соответствует пересечение изображающих их множеств: $A \cap B$.

Определение. *Суммой событий A и B называется событие $C = A + B$, состоящее в том, что произошло хотя бы одно из этих событий.*

Сумме событий A и B соответствует объединение изображающих их множеств: $A \cup B$.

Определение. *Разностью событий A и B называется событие $C = A - B$, состоящее в том, что A произошло, а B не произошло.*

Разности событий A и B соответствует теоретико-множественная разность изображающих их множеств: $A \setminus B$.

Определение. *Говорят, что событие A влечет событие B, если B происходит всякий раз, когда произошло событие A.*

Отношению «событие A влечет событие B» соответствует отношение «A есть подмножество B» изображающих их множеств: $A \subseteq B$.

Определение. *События A и B называются несовместными, если наступление любого из них исключает наступление другого.*

Несовместным событиям A и B соответствуют не пересекающиеся множества: $A \cap B = \varnothing$.

Определение. *Если A – некоторое событие, то противоположным ему называется событие $\overline{A}$, состоящее в том, что A не произошло.*

Противоположному событию $\overline{A}$ соответствует дополнение $\Omega \setminus A$ множества A до всего пространства Ω.

Замечание. В дальнейшем мы не будем различать событие A, наступающее в рамках некоторого случайного эксперимента, и множество A, изображающее его в соответствующей математической модели. Кроме, того для обозначения операций над событиями, наряду с приведенными выше, мы будем использовать теоретико-множественные обозначения. Например, записи

$C = AB$ и $C = A \bigcap B$ означают одно и то же – произведение событий A и B.

Определение. *Совокупность* $\mathscr{A}$ *множеств* $A \subseteq \Omega$ *называется алгеброй событий, если она содержит* Ω *и из условия* $A \in \mathscr{A}$, $B \in \mathscr{A}$ *следует, что* $A + B \in \mathscr{A}$, $AB \in \mathscr{A}$ *и* $A - B \in \mathscr{A}$.

Иначе говоря, алгебра событий – это система $\mathscr{A}$ подмножеств множества Ω, содержащая само множество Ω и замкнутая относительно операций объединения, пересечения и взятия разности.

Определение. *Бесконечное множество А называется счетным, если его элементы можно занумеровать натуральными числами (представить в виде последовательности):* $A = \{a_1, a_2, \ldots, a_n, \ldots\}$ *и несчетным в противном случае.*

Примеры счетных множеств: $\mathbb{N} = \{1, 2, \ldots\}$ – множество натуральных чисел; $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ – множество целых чисел; $\mathbb{Q} = \{\frac{p}{q} \mid p \in \mathbb{Z},\ q \in \mathbb{N}\}$ – множество рациональных чисел.

Примеры несчетных множеств: $\mathbb{R}$ – множество действительных чисел (числовая прямая); $\mathbb{R}^2 = \{(x, y) \mid x \in \mathbb{R},\ y \in \mathbb{R}\}$ – множество точек координатной плоскости.

Определение. *Алгебра событий* $\mathscr{A}$, *замкнутая относительно счетного числа операций объединения и пересечения, называется* σ*-алгеброй.*

Множество всех подмножеств произвольного множества A, включая само A и $\varnothing$, называется *булеаном* A и обозначается через 2^A. Если множество $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ состоит из конечного числа n элементов, то его булеан содержит 2^n элементов и образует алгебру $\mathscr{A} = 2^\Omega$, которая одновременно будет и σ-алгеброй.

§ 3. *Вероятность события. Аксиомы вероятности*

Интуитивно вероятность события A понимается как численная мера «ожидаемости» наступления этого события, заключенная в пределах от нуля до единицы. Чем ближе вероятность события к единице, тем больше оснований ожидать, что оно наступит при данных условиях. Мы определим вероятность как числовую функцию $\mathsf{P}(A)$, заданную на некотором множестве случайных событий и удовлетворяющую определенному набору условий (аксиом).

Аксиома 1. *Пусть* Ω *– пространство элементарных событий и* $\mathscr{A}$ *– некоторая* σ*-алгебра событий. Каждому событию* $A \in \mathscr{A}$ *ставится в соответствие неотрицательное число* $\mathsf{P}(A)$, *называемое его вероятностью.*

Аксиома 2. *Вероятность достоверного события равна единице:*

$$\mathsf{P}(\Omega) = 1.$$

Аксиома 3 (σ-аддитивность вероятностной меры). *Если события* A_1, $A_2, \ldots, A_i, \ldots$ *попарно несовместны, то*

$$\mathsf{P}(A_1 + A_2 + \ldots + A_i + \ldots) = \sum_{i=1}^{\infty} \mathsf{P}(A_i).$$

В частности, для конечного числа n попарно несовместных событий $A_1, A_2, \ldots, A_n$ *имеет место равенство*

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

Свойство, выражаемое последним равенством называется аддитивностью (конечной) вероятностной меры.

Аксиома 4 (аксиома непрерывности). *Пусть дана невозрастающая последовательность* $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ *множеств из* $\mathscr{A}$, *и пусть* $A = \bigcap_{n=1}^{\infty} A_n$. *Тогда* $\mathsf{P}(A) = \lim_{n\to\infty} \mathsf{P}(A_n)$.

З а м е ч а н и е. Можно доказать, что аксиомы 3 и 4 эквивалентны друг другу.

Следствия из аксиом

I. *Вероятность невозможного события равна нулю:* $\mathsf{P}(\varnothing) = 0$.

Д о к а з а т е л ь с т в о. Запишем $\Omega = \Omega \bigcup \varnothing$. Тогда на основании аксиом 1-3 будем иметь $1 = \mathsf{P}(\Omega) = \mathsf{P}(\Omega) + \mathsf{P}(\varnothing) = 1 + \mathsf{P}(\varnothing)$, откуда $\mathsf{P}(\varnothing) = 0$.

II. *Для любого* $A \in \mathscr{A}$

$$0 \le \mathsf{P}(A) \le 1.$$

Д о к а з а т е л ь с т в о. Запишем $\Omega = A \bigcup \overline{A}$. Тогда в силу аксиом 1-3 получим

$$0 \le \mathsf{P}(A) \le \mathsf{P}(A) + \mathsf{P}(\overline{A}) = \mathsf{P}(\Omega) = 1.$$

III. *Для любого* $A \in \mathscr{A}$

$$\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A).$$

Д о к а з а т е л ь с т в о. Снова представим Ω в виде $\Omega = A \bigcup \overline{A}$. Тогда из аксиом 2 и 3 следует, что $1 = \mathsf{P}(\Omega) = \mathsf{P}(A) + \mathsf{P}(\overline{A})$, откуда $\mathsf{P}(\overline{A}) = 1 - \mathsf{P}(A)$.

IV. *Если событие A влечет за собой событие* B, *то*

$$\mathsf{P}(A) \le \mathsf{P}(B).$$

Д о к а з а т е л ь с т в о. События A и $B \setminus A$ несовместны и по аксиоме 1 $P(B \setminus A) \ge 0$. Тогда $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A) \ge \mathsf{P}(A)$.

V. (Теорема сложения вероятностей). *Если A и B – произвольные (не обязательно несовместные) события, то*

$$\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB).$$

Д о к а з а т е л ь с т в о. Запишем $A + B = A + (B \setminus A)$, откуда в силу несовместности A и $B \setminus A$ будем иметь $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$. Далее, $\mathsf{P}(B) = \mathsf{P}(B \setminus A) + \mathsf{P}(AB)$, так что $\mathsf{P}(B \setminus A) = \mathsf{P}(B) - \mathsf{P}(AB)$. Подставляя это выражение в равенство $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B \setminus A)$, получим $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(AB)$.

§ 4. *Вероятностное пространство*

Определение. *Вероятностным пространством называют тройку объектов* $(\Omega, \mathscr{A}, \mathsf{P})$, *где* Ω *– пространство элементарных событий,* $\mathscr{A}$ *–* σ*-алгебра его подмножеств, называемых случайными событиями,* $\mathsf{P}(A)$ *– вероятность, определенная на* σ*-алгебре* $\mathscr{A}$.

В современной теории вероятностей построение вероятностного пространства $(\Omega, \mathscr{A}, \mathsf{P})$ является исходным моментом для математического описания и моделирования любых случайных явлений и процессов.

Пример. Пусть случайный эксперимент состоит в однократном подбрасывании симметричной монеты. Рассмотрим события: $\omega_1 = \{0\}$ – выпадение «орла», $\omega_2 = \{1\}$ – выпадение «решки», $\varnothing$ – невозможное, $\Omega = \{0;1\}$ – достоверное. Тогда

$$\mathscr{A} = \{\varnothing;\{0\};\{1\};\{0;1\}\}$$

Зададим вероятность на событиях из $\mathscr{A}$ так:

$$\mathsf{P}(\varnothing) = 0; \mathsf{P}(\{0\}) = \mathsf{P}(\{1\}) = 1/2; \mathsf{P}(\{0;1\}) = 1.$$

Тем самым полностью построено вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$, соответствующее данному случайному эксперименту.

4.1. Классическое определение вероятности

Предположим, что проводится случайный эксперимент с конечным числом элементарных исходов $\omega_1, \omega_2, ..., \omega_n$, и пусть имеются основания (например, соображения симметрии) считать эти исходы равновозможными:

$$\mathsf{P}(\omega_1) = \mathsf{P}(\omega_2) = ... = \mathsf{P}(\omega_n) = \frac{1}{n}.$$

Зададим вероятность на $\mathscr{A} = 2^{\Omega}$ следующим образом. Если событию $A \in 2^{\Omega}$ благоприятствуют k элементарных событий $\omega_{i_1}, \omega_{i_2}, ..., \omega_{i_k}$, то его вероятность положим равной

$$\mathsf{P}(A) = \sum_{m=1}^{k} \mathsf{P}(\omega_{i_m}) = \frac{k}{n},$$

в частности, $\mathsf{P}(\Omega) = \sum_{m=1}^{m} \mathsf{P}(\omega_m) = \frac{n}{n} = 1$. Очевидно, что все аксиомы вероятности для введенной нами функции множества выполнены. Определение вероятности события A, на основе изложенного подхода называется *классическим*. Кратко это определение можно сформулировать так:

при указанных выше условиях вероятность события A *равна отношению числа элементарных исходов, благоприятствующих событию* A *к общему числу элементарных исходов*

$$\mathsf{P}(A)=\frac{|A|}{|\Omega|},$$

где символом $|A|$ обозначено число элементов конечного множества A.

Пример. Случайный эксперимент состоит в однократном выбрасывании двух игральных костей. Найти вероятность того, что сумма выпавших на двух костях очков кратна пяти.

Решение. Каждый элементарный исход зададим упорядоченной парой чисел (i, j), где i – число очков, выпавшее на первой кости, j – на второй, $1 \le i, j \le 6$. Общее число элементарных исходов равно $n=6^2=36$. Событию A={сумма очков кратна 5} , благоприятствуют k=7 элементарных исходов

5+5=10,
6+4=10,
4+6=10,
1+4=5,
4+1=5,
2+3=5,
3+2=5.

Таким образом, $\mathsf{P}(A)=\frac{k}{n}=\frac{7}{36}$.

4.2. Геометрические вероятности

Пусть A – множество точек на прямой $\mathbb{R}$, на плоскости $\mathbb{R}^2$ или в пространстве $\mathbb{R}^3$. Символом $\mathrm{mes}(A)$, будем обозначать так называемую меру этого множества, т.е., соответственно длину, площадь или объем (от measure – англ. «мера»)

Рассмотрим следующий случайный эксперимент. Пусть на плоскости XOY задана фигура Ω с площадью $\mathrm{mes}(\Omega)$ и внутри ее фигура $A \subseteq \Omega$ с площадью $\mathrm{mes}(A)$. На плоскость «наудачу» бросают точку $(x; y)$; при этом, по условию, она не может оказаться вне фигуры Ω. Слово «наудачу» в данном случае означает, что вероятность попадания точки в пределы любой фигуры $A \subseteq \Omega$ зависит только от площади этой фигуры, но не от ее формы или расположения. Наша задача – задать вероятность события, состоящего в попадании случайной точки $(x; y)$ в пределы фигуры A. Это событие удобно обозначить той же буквой A. Множество точек фигуры Ω соответствует в этих обозначениях достоверному событию (пространству элементарных событий). Элементарному же событию в нашем случае соответствует пара координат $\omega=(x; y) \in \Omega$. Положим

$$\mathsf{P}(A)=\frac{\mathrm{mes}(A)}{\mathrm{mes}(\Omega)}.$$

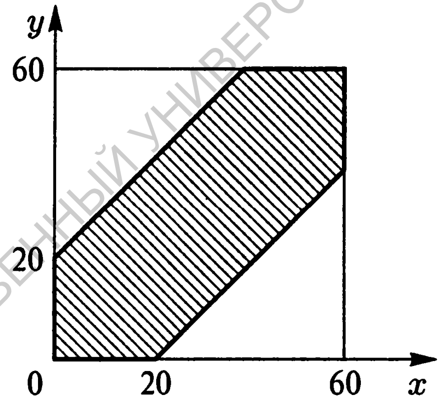
Вероятности, вычисленные по этой формуле, называют *геометрическими*. Все аксиомы вероятности при таком определении выполнены (строгая проверка этого интуитивно очевидного факта выходит за рамки настоящего пособия).

Очевидно, что $\mathsf{P}(\Omega)=1$. В то же время, нулевую вероятность будет иметь любое событие, состоящее в попадании случайной точки на множество $A \subseteq \Omega$, площадь которого равна нулю, а не только невозможное событие $\varnothing$. Итак, мы построили вероятностное пространство $(\Omega, \mathscr{A}, \mathsf{P})$, где Ω – множество точек плоскости, $\mathscr{A}$ – σ-алгебра всех его измеримых (говоря упрощенно, тех, для которых имеет смысл понятие площади) подмножеств, P – вероятность, определенная равенством $\mathsf{P}(A) = \mathrm{mes}(A)/\mathrm{mes}(\Omega)$.

Аналогичное построение можно провести для множеств точек прямой или трехмерного пространства. При этом площадь следует заменить длиной или, соответственно, объемом.

Пример (задача о встрече). Два лица, A и B, договорились о встрече в определённом месте между 18.00 и 19.00 часами. Тот, кто пришел первым, ожидает другого в течение 20 минут, затем уходит. Найти вероятность того, что эти люди встретятся, если момент прихода каждого из них в течение часа выбирается наудачу и не зависит от времени прихода другого человека.

Решение. Очевидно, временной интервал с 18.00 до 19.00 часов можно заменить промежутком [0, 60], если за единицу измерения взять минуту. Обозначим через $X \in [0, 60]$ и $Y \in [0, 60]$ моменты прихода A и B соответственно. Элементарными событиями тогда будут всевозможные точки $(X;Y)$ квадрата $\Omega = \{(X;Y): 0 \le X \le 60,\ 0 \le Y \le 60\}$.



Встреча происходит тогда и только тогда, когда точка $(X;Y)$ попадает в область $D = \{(X;Y): |X - Y| \le 20\}$ (на рисунке она заштрихована). Искомая вероятность, таким образом, равна отношению площадей фигур D и Ω

$$\mathsf{P}(D) = \frac{\mathrm{mes}(D)}{\mathrm{mes}(\Omega)} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

§ 5. *Условная вероятность. Независимость событий*

Определение. *Пусть* $(\Omega, \mathscr{A}, \mathsf{P})$ – *произвольное вероятностное пространство,* $A, B \in \mathscr{A}$ *– события, причем,* $\mathsf{P}(B) > 0$. *Условной вероятностью события A при условии, что произошло событие B, называется число*

$$\mathsf{P}(A\,|\,B)=\frac{\mathsf{P}(A\bigcap B)}{\mathsf{P}(B)},$$

Если выполнено неравенство $\mathsf{P}(A)>0$, то аналогично можно определить условную вероятность события *B* при условии *A*

$$\mathsf{P}(B\,|\,A)=\frac{\mathsf{P}(A\bigcap B)}{\mathsf{P}(A)}.$$

Пример. Студент, готовясь к экзамену, выучил из 40 билетов только половину – билеты с номерами от 1 до 20. К моменту, когда он пришел на экзамен, осталось только 25 билетов с номерами от 16 до 40. Рассмотрим события

A={студент получил выученный билет},

B={остались билеты с номерами от 16 до 40}

и найдем $\mathsf{P}(A|B)$. Элементарным исходом будет номер вытянутого студентом билета, так что Ω={1,2,…,40}. Если вычислять вероятность *A* при условии, что наступило событие *B*, то благоприятными исходами следует считать только те, которые совместимы с *B*; их число равно $|A\bigcap B|=5$. Число всех возможных исходов при дополнительной информации о наступившем *B* теперь будет равно $|B|=25$ так что

$$\mathsf{P}(A|B)=\frac{|A\bigcap B|}{|B|}=\frac{5}{25}=\frac{1}{5}.$$

Замечание. Мы видели, что при использовании классического определения условная вероятность равна $\mathsf{P}(A|B)=|AB|/|B|$. Если учесть, что безусловные вероятности событий $A\bigcap B$ и B в этом случае равны, соответственно, $\mathsf{P}(AB)=|A\bigcap B|/|\Omega|$ и $\mathsf{P}(B)=|B|/|\Omega|$, то условную вероятность $\mathsf{P}(A|B)$ можно записать в виде

$$\mathsf{P}(A|B)=\frac{|A\bigcap B|/|\Omega|}{|B|/|\Omega|}=\frac{\mathsf{P}(A\bigcap B)}{\mathsf{P}(B)}.$$

Это равенство, с учетом высказанных соображений, объясняет, почему условную вероятность естественно определить выражением $\mathsf{P}(A|B)=\mathsf{P}(A\bigcap B)/\mathsf{P}(B)$.

Определение. *События A и B называются независимыми, если*

$$\mathsf{P}(A\bigcap B)=\mathsf{P}(A)\mathsf{P}(B).$$

С содержательной точки зрения независимость событий означает отсутствие физического (причинного) влияния любого из них на вероятность наступления другого. Пусть, например, два стрелка независимо друг от друга стреляют по мишени. Тогда события *A*={мишень поражена первым стрелком} и *B*={мишень поражена вторым стрелком} будут независимыми.

При истолковании независимых событий как причинно не связанных естественно ожидать, что их условные вероятности совпадают с безусловными. Действительно, справедлива следующая

Теорема. *Пусть* $\mathsf{P}(B)>0$. *Тогда для независимости событий A и B необходимо и достаточно, чтобы* $\mathsf{P}(A)=\mathsf{P}(A|B)$.

Доказательство. Если события A и B независимы, то

$$\mathsf{P}(A\,|\,B)=\frac{\mathsf{P}(A\cap B)}{\mathsf{P}(B)}=\frac{\mathsf{P}(A)\mathsf{P}(B)}{\mathsf{P}(B)}=\mathsf{P}(A).$$

Обратно, если $\mathsf{P}(A)=\mathsf{P}(A\,|\,B)$, то из равенства

$$\mathsf{P}(A\,|\,B)=\frac{\mathsf{P}(A\cap B)}{\mathsf{P}(B)}=\mathsf{P}(A)$$

следует, что

$$\mathsf{P}(A\cap B)=\mathsf{P}(A)\mathsf{P}(B).$$

Замечание. Не следует путать понятия *независимости* и *несовместности* событий. Если $\mathsf{P}(A)>0$, $\mathsf{P}(B)>0$, то из несовместности A и B всегда следует их зависимость. Действительно, несовместность означает, что наступление одного события исключает наступление другого. Поэтому влияние одного события на другое не только имеет место, но и является строго детерминированным (неслучайным). Кроме того, нарушается и формальное требование независимости, поскольку в этом случае

$$\mathsf{P}(A\cap B)=\mathsf{P}(\varnothing)=0\neq\mathsf{P}(A)\mathsf{P}(B)>0.$$

С другой стороны, если выполнено хотя бы одно из равенств $\mathsf{P}(A)=0$ или $\mathsf{P}(B)=0$, то несовместные события A и B будут и независимыми.

Теорема. *Если события A и B независимы, то независимыми будут и пары событий A и $\overline{B}$, $\overline{A}$ и B, $\overline{A}$ и $\overline{B}$.*

Доказательство. Представим A в виде $A=(A\cap B)\cup(A\cap\overline{B})$. В силу того, что события $A\cap B$ и $A\cap\overline{B}$ несовместны, будем иметь

$$\mathsf{P}(A)=\mathsf{P}(A\cap B)+\mathsf{P}(A\cap\overline{B}),$$

откуда

$$\mathsf{P}(A\cap\overline{B})=\mathsf{P}(A)-\mathsf{P}(A\cap B)=\mathsf{P}(A)-\mathsf{P}(A)\mathsf{P}(B)=\mathsf{P}(A)(1-\mathsf{P}(B))=\mathsf{P}(A)\mathsf{P}(\overline{B})$$

Теперь, по доказанному, из независимости A и B следует независимость $\overline{A}$ и B, а из независимости $\overline{B}$ и A – независимость $\overline{B}$ и $\overline{A}$. Теорема полностью доказана.

Замечание. Перепишем формулы для условных вероятностей $\mathsf{P}(B\,|\,A)$ и $\mathsf{P}(A\,|\,B)$ в виде

$$\mathsf{P}(AB)=\mathsf{P}(A)\mathsf{P}(B\,|\,A)$$

$$\mathsf{P}(AB)=\mathsf{P}(A\,|\,B)\mathsf{P}(B).$$

Каждое из этих равенств часто называют *«теоремой умножения»* вероятностей, а равенство $\mathsf{P}(AB)=\mathsf{P}(A)\mathsf{P}(B)$, служащее определением независимости, – *«теоремой умножения»* для независимых событий.

Определение. *События $A_1, A_2, \ldots, A_n$ называются независимыми в совокупности, если для любого $1\le k\le n$ и любого набора индексов $1\le i_1<i_2<\ldots<i_k\le n$ справедливо равенство*

$$\mathsf{P}(A_{i_1}\cap A_{i_2}\cap\ldots\cap A_{i_k})=\mathsf{P}(A_{i_1})\mathsf{P}(A_{i_2})\ldots\mathsf{P}(A_{i_k})$$

Замечание. Если события независимы в совокупности, то они *попарно* *независимы*, т. е. из набора $A_1, A_2, ..., A_n$ будут независимы любые два. Чтобы убедиться в этом, достаточно положить в последнем равенстве $k = 2$. В то же время, попарная независимость, не влечет за собой независимость в совокупности.

Пример. Рассмотрим пример, принадлежащий С.Н. Бернштейну (Сергей Натанович Бернштейн – советский математик, 1880-1968). Пусть на плоскость бросают правильный однородный тетраэдр, у которого одна грань окрашена в синий цвет, другая – в красный, третья – в зеленый и , наконец, окраска четвертой грани содержит все три цвета. Введем события

$A =${грань, на которую падает тетраэдр, содержит в окраске синий цвет};

$B =${грань, на которую падает тетраэдр, содержит в окраске красный цвет};

$C =${грань, на которую падает тетраэдр, содержит в окраске зеленый цвет}.

Поскольку каждый цвет присутствует на двух гранях, а всего граней четыре, вероятности событий A, B и C одинаковы и равны

$$\mathsf{P}(A) = \mathsf{P}(B) = \mathsf{P}(C) = \frac{1}{2}.$$

Событиям AB, BC и AC соответствует только одна грань, трехцветная, поэтому

$$\mathsf{P}(AB) = \mathsf{P}(BC) = \mathsf{P}(AC) = \frac{1}{4}.$$

Следовательно,

$$\mathsf{P}(AB) = \mathsf{P}(A)\mathsf{P}(B), \mathsf{P}(BC) = \mathsf{P}(B)\mathsf{P}(C), \mathsf{P}(AC) = \mathsf{P}(A)\mathsf{P}(C),$$

поэтому события A, B и C попарно независимы. В то же время,

$$\mathsf{P}(ABC) = \frac{1}{4} \neq \frac{1}{8} = \mathsf{P}(A)\mathsf{P}(B)\mathsf{P}(C),$$

так что эти события не являются независимыми в совокупности.

5.1. Формула полной вероятности

Определение. *Говорят, что события $H_1, H_2, ..., H_n$ образуют полную группу, если они попарно несовместны и в сумме дают достоверное событие.*

В приложениях теории вероятностей события полной группы часто имеют характер предположений. Например, если из одной из двух одинаковых с виду урн с шарами наудачу извлекают шар, то события H_1={шар извлечен из первой урны} и H_2={шар извлечен из второй урны} образуют полную группу. В связи с этим события из полной группы часто называют гипотезами.

Теорема (формула полной вероятности). *Пусть события $H_1, H_2, ..., H_n$ такие, что $\mathsf{P}(H_i) > 0$, образуют полную группу. Тогда вероятность любого события А равна сумме парных произведений вероятностей каждого из этих событий на соответствующие им условные вероятности наступления события А.*

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(H_i)\mathsf{P}(A \mid H_i)$$

Доказательство. Так как события $H_1, H_2, ..., H_n$ образуют полную группу, можно записать

$$A = A\Omega = A(H_1 + H_2 + ... + H_n) = AH_1 + AH_2 + ... + AH_n = \sum_{i=1}^{n} AH_i .$$

В силу несовместности событий $H_1, H_2, ..., H_n$, события AH_i также несовместны, поэтому из аддитивности вероятностной меры следует, что:

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(AH_i)$$

Поскольку при этом $\mathsf{P}(AH_i) = \mathsf{P}(H_i)\mathsf{P}(A \mid H_i)$, окончательно получаем:

$$\mathsf{P}(A) = \sum_{i=1}^{n} \mathsf{P}(H_i)\mathsf{P}(A \mid H_i),$$

что и требовалось.

Пример. В цехе имеется два станка, штампующих однотипные детали. Первый станок производит 40% всех деталей, из которых 5% составляет брак, второй – 60% при 10% брака. Какова вероятность того, что случайно выбранная деталь окажется дефектной?

Решение. Рассмотрим следующие события: A={выбранная наугад деталь – дефектная}, H_1={выбранная деталь произведена первым станком}, H_2={ выбранная деталь произведена вторым станком}. Очевидно, события H_1 и H_2 образуют полную группу, при этом $\mathsf{P}(H_1) = 0{,}4$, $\mathsf{P}(H_2) = 0{,}6$. Далее, по условию $\mathsf{P}(A \mid H_1) = 0{,}05$, $\mathsf{P}(A \mid H_2) = 0{,}1$. Применяя формулу полной вероятности, находим

$$\mathsf{P}(A) = \mathsf{P}(A \mid H_1)\mathsf{P}(H_1) + \mathsf{P}(A \mid H_2)\mathsf{P}(H_2) = 0{,}05 \cdot 0{,}4 + 0{,}1 \cdot 0{,}6 = 0{,}08 .$$

5.2. Формула Байеса

Пусть имеется полная группа гипотез $H_1, H_2, ..., H_n$ и пусть в силу каких-либо причин (например, из соображений симметрии) заранее известны вероятности их реализации $\mathsf{P}(H_1), \mathsf{P}(H_2), ..., \mathsf{P}(H_n)$. Пусть в результате опыта произошло событие A, для которого известны условные вероятности $\mathsf{P}(A \mid H_1)$, $\mathsf{P}(A \mid H_2)$,…, $\mathsf{P}(A \mid H_n)$. Требуется определить, как изменятся вероятности гипотез $H_1, H_2, ..., H_n$ с учетом информации о наступлении события A, т.е. условные вероятности $\mathsf{P}(H_i \mid A)$. В связи с такой интерпретацией безусловные вероятности $\mathsf{P}(H_i)$ иногда называют априорными (a priori – в данном случае «до опыта», лат.), а условные вероятности $\mathsf{P}(H_i \mid A)$ – апостериорными (a posteriori – в данном случае «после опыта», лат.). Выражение для $\mathsf{P}(H_i \mid A)$ принято связывать с именем английского математика Томаса Байеса (1702-1761).

Теорема (формула Байеса). *Вероятность гипотезы при условии наступления события A такого, что $\mathsf{P}(A)>0$, равна*

$$\mathsf{P}(H_i \mid A)=\frac{\mathsf{P}(H_i)\mathsf{P}(A|H_i)}{\sum_{i=1}^{n}\mathsf{P}(H_i)\mathsf{P}(A|H_i)}, i=1,2,...n.$$

Доказательство. По теореме умножения вероятностей

$$\mathsf{P}(A)\mathsf{P}(H_i \mid A)=\mathsf{P}(H_i)\mathsf{P}(A|H_i),$$

откуда, с учетом того, что $\mathsf{P}(A)\neq 0$, получаем

$$\mathsf{P}(H_i \mid A)=\frac{\mathsf{P}(H_i)\mathsf{P}(A|H_i)}{\mathsf{P}(A)}.$$

Применив формулу полной вероятности для $\mathsf{P}(A)$, окончательно найдем

$$\mathsf{P}(H_i \mid A)=\frac{\mathsf{P}(H_i)\mathsf{P}(A|H_i)}{\sum_{i=1}^{n}\mathsf{P}(H_i)\mathsf{P}(A|H_i)}.$$

Пример. В условиях задачи из предыдущего примера найти вероятность того, что случайно выбранная деталь изготовлена на втором станке, при условии, что она оказалась бракованной.

Решение. Требуется найти условную вероятность $\mathsf{P}(H_2 \mid A)$. Очевидно, применима формула Байеса, причем все необходимые числовые значения уже известны. Находим

$$\mathsf{P}(H_2 \mid A)=\frac{\mathsf{P}(H_2)\mathsf{P}(A|H_2)}{\mathsf{P}(A)}=\frac{0{,}6\cdot 0{,}1}{0{,}08}=0{,}75.$$

§ 6. *Последовательности независимых испытаний. Схема Бернулли*

Пусть проводится серия из n однотипных независимых испытаний. Независимость понимается в том смысле, что на результаты любого из испытаний не влияет исход любого другого испытания. Пусть, далее, в ходе каждого испытания может наступить событие A с известной вероятностью $P(A)=p>0$, которое мы для краткости будем условно называть «успехом». Противоположное ему событие $\overline{A}$ назовем «неудачей». Его вероятность равна $P(\overline{A})=1-p=q$. Описанный тип случайного эксперимента принято называть «схемой Бернулли» в честь выдающегося швейцарского математика Якоба Бернулли (1654-1705).

Рассмотрим задачу: в условиях схемы Бернулли найти вероятность $P_n(k)$ события $A_{n,k}$, состоящего в том, что в серии из n испытаний «успех» наступил ровно k раз. Элементарными событиями в данном случае будут исходы серий из n испытаний вида

$$\omega_{\alpha_1\alpha_2...\alpha_n}=A^{(1)}_{\alpha_1}A^{(2)}_{\alpha_2}...A^{(n)}_{\alpha_n}, \tag{1}$$

где на i-ом месте ($i=1,2,\ldots,n$) расположен значок A, если индекс $\alpha_i=1$ и значок $\overline{A}$, если $\alpha_i = 0$. При этом символ A встречается ровно k раз, а символ $\overline{A}$ – ровно $n-k$ раз, т. е. среди индексов α_i имеется k единиц и $(n-k)$ нулей. Вероятность каждого события $\omega_{\alpha_1\alpha_2\ldots\alpha_n}$ в силу независимости отдельных испытаний естественно положить равной произведению соответствующих вероятностей

$$P\left(\omega_{\alpha_1\alpha_2\ldots\alpha_n}\right) = p^k(1-p)^{n-k} = p^k q^{n-k}. \tag{2}$$

Число элементарных событий, благоприятствующих интересующему нас событию $A_{n,k}$ равно числу способов, которыми можно выбрать k объектов среди имеющихся n объектов, т. е. числу сочетаний

$$C_n^k = \frac{n!}{k!(n-k)!}.$$

Таким образом, искомая вероятность равна

$$P_n(k) = C_n^k p^k q^{n-k}, k = 0,1,\ldots,n. \tag{3}$$

Полученная формула называется *формулой Бернулли*. Говорят также, что формула (3) задает биномиальное распределение вероятностей. Это название связано с тем, что правая часть этой формулы представляет собой $(k+1)$-й член бинома Ньютона

$$(p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k}.$$

Отметим также, что сумма всех вероятностей, заданных формулой (3) равна единице:

$$1 = (p+q)^n = \sum_{k=0}^{n} P_n(k). \tag{4}$$

Это равенство соответствует тому факту, что события $\{A_{n,k}\}_{k=0}^{n}$ образуют так называемую полную группу, т. е. они попарно несовместны и в сумме дают достоверное событие $\sum_{k=0}^{n} A_{n,k} = \Omega$.

Найдем теперь вероятность $P_n(m_1;m_2)$ того, что «успех» в n испытаниях наступил не менее, чем m_1 раз, но не более, чем m_2 раз. Интересующее нас событие есть сумма попарно несовместных событий $A_{n,k}$, $m_1 \le k \le m_2$, значит по аксиоме сложения вероятностей $P_n(m_1;m_2) = \sum_{k=m_1}^{m_2} P\left(A_{n,k}\right)$, так что

$$P_n(m_1;m_2) = \sum_{k=m_1}^{m_2} P(k), \tag{5}$$

где $P_n(k)$ определяются формулами (3).

З а м е ч а н и е Важно понимать, что вероятности элементарных событий $\omega_{\alpha_1\alpha_2\ldots\alpha_n}$ были нами не вычислены, а взяты «равными по определению» значениям $p^k q^{n-k}$. При этом оказывается, что соотношения (2) действительно задают вероятностную меру на конечном вероятностном пространстве $\Omega = \{\omega_{\alpha_1\alpha_2\ldots\alpha_n}\}$, поскольку в силу (4)

$$\sum P\left(\omega_{\alpha_1\alpha_2\ldots\alpha_n}\right)=\sum_{k=0}^{n}P\left(A_{n,k}\right)=\sum_{k=0}^{n}P_n(k)=1,$$

т.е. сумма вероятностей всех элементарных событий равна 1.

Строгое обоснование равенства (2) связано с так называемыми прямыми произведениями вероятностных пространств и здесь рассматриваться не будет.

Пример. Найти вероятность того что, при десятикратном подбрасывании «правильной» (т. е. симметричной и однородной) монеты «орел» выпадет ровно 7 раз.

Решение. Для правильной монеты вероятность выпадения «орла» равна 1/2, поэтому в силу формулы Бернулли (3) будем иметь

$$P_{10}(7)=C_{10}^{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7}=\frac{10!}{7!3!}\cdot\frac{1}{2^{10}}\approx 0{,}1172.$$

Пример. Правильная монета подбрасывается 10 раз. Найти вероятность того, что герб выпадет не менее 5, но не более 7 раз.

Решение. Воспользуемся формулами (5) и (3). Имеем

$$P_{10}(5)=C_{10}^{5}\left(\frac{1}{2}\right)^{5}\left(\frac{1}{2}\right)^{10-5}=\frac{10!}{5!5!}\cdot\frac{1}{2^{10}}\approx 0{,}246,$$

$$P_{10}(6)=C_{10}^{6}\left(\frac{1}{2}\right)^{6}\left(\frac{1}{2}\right)^{10-6}=\frac{10!}{6!4!}\cdot\frac{1}{2^{10}}\approx 0{,}205,$$

$$P_{10}(7)=C_{10}^{7}\left(\frac{1}{2}\right)^{7}\left(\frac{1}{2}\right)^{10-7}=\frac{10!}{7!3!}\cdot\frac{1}{2^{10}}\approx 0{,}117,$$

$$P_{10}(5;7)=P_{10}(5)+P_{10}(6)+P_{10}(7)\approx 0{,}246+0{,}205+0{,}117=0{,}568.$$

§ 7. *Предельные теоремы в схеме Бернулли*

7.1. Локальная теорема Муавра-Лапласа

Формула Бернулли позволяет вычислить вероятность наступления ровно k «успехов» в серии из n однотипных независимых испытаний при любых k и n. Однако при больших значениях k и n вычисление вероятностей $P_n(k)$ становится технически весьма сложной задачей. Поэтому возникает необходимость в асимптотических формулах, которые позволяли бы находить приближенные значения этих вероятностей. Первый результат такого рода был получен Муавром (Абрахам Муавр – французский математик, 1667-1754) в 1730 г. и значительно позже (в 1812 г.) обобщен Лапласом (Пьер-Симон Лаплас

– выдающийся французский математик, 1749-1827).

Теорема (локальная теорема Муавра-Лапласа). *Если вероятность «успеха» A в n испытаниях по схеме Бернулли равна $p>0$, то вероятность $P_n(k)$ того, что событие A в этих испытаниях наступит ровно k раз, удовлетворяет при $n\to\infty$ следующему предельному соотношению*

$$\sqrt{npq}P_n(k):\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\to 1. \qquad (6)$$

Это соотношение выполняется равномерно для всех k, для которых значения

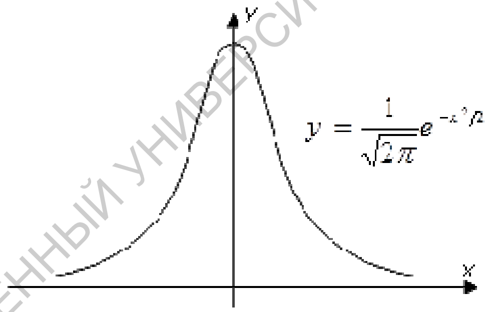
$$x=x_{k,n}=\frac{k-np}{\sqrt{npq}}$$

находятся в каком-либо конечном промежутке.

Из соотношения (7) следует приближённая формула

$$P_n(k)\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2}. \qquad (7)$$

Для функции $\varphi_0(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ составлены подробные таблицы, используя которые можно выполнять вычисления по формуле (7). На рисунке представлен график этой функции.



Однозначных рекомендаций, при каких значениях стоит применять формулу (7) не существует. Некоторые авторы рекомендуют ее использование при $n>100$ и $\sqrt{npq}>20$ (см., например, [3]).

Замечание. Равномерную сходимость, речь о которой идет в теореме 1, следует понимать следующим образом. Пусть дан произвольный фиксированный отрезок $[a,b]$. Тогда для любого $\varepsilon>0$ найдется номер $N=N(\varepsilon,a,b)$ такой, что для всех $n>N$ и для всех k, удовлетворяющих условию

$$x_{k,n}=\frac{k-np}{\sqrt{npq}}\in[a,b]$$

будет выполняться неравенство

$$\left|\sqrt{npq}P_n(k):\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_{k,n}^2}-1\right|<\varepsilon.$$

Пример. Используя приближенную формулу (7), найти вероятность

$P_n(k)$ при $n = 10000$, $k = 15$, $p = 0{,}001$.

Решение. В нашем случае

$$\sqrt{npq} = \sqrt{10000 \cdot 0{,}001 \cdot 0{,}999} = \sqrt{9{,}99} \approx 3{,}1607,$$

$$\frac{k - np}{\sqrt{npq}} \approx \frac{15 - 10}{3{,}1607} \approx 1{,}58.$$

Поэтому

$$P_n(k) \approx \frac{1}{3{,}1607\sqrt{2\pi}} e^{-\frac{1}{2}(1{,}58)^2}$$

В таблице значений функции $\varphi_0(x)$ находим $\varphi_0(1{,}58) = 0{,}1145$. Таким образом, искомая вероятность равна

$$P_n(k) \approx \frac{0{,}1145}{3{,}1607} = 0{,}03623.$$

7.2. Интегральная теорема Муавра-Лапласа

Снова рассмотрим вопрос: какова вероятность $P_n(k_1;k_2)$ того, что событие A, имеющее вероятность $p > 0$ при n испытаниях наступит не менее, чем k_1 раз, но не более, чем k_2 раза. Точное значение дает полученная ранее формула

$$P_n(k_1;k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k} \tag{8}$$

Однако, при больших значениях k_1, k_2 и n подсчет $P_n(k_1;k_2)$ связан с теми же трудностями, что и подсчет вероятностей $P_n(k)$. Попытаемся получить приближённое выражение для $P_n(k_1;k_2)$, воспользовавшись формулой (7). Заменяя каждое слагаемое в (8) выражением (7), будем иметь

$$P_n(k_1;k_2) \approx \sum_{k=k_1}^{k_2} \frac{1}{\sqrt{npq}} \varphi(x_{k,n}) \tag{9}$$

где, как и раньше,

$$x_{k,n} := \frac{k - np}{\sqrt{npq}}, \ k_1 \le k \le k_2,$$

$$\varphi_0(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Далее,

$$\Delta x_{k,n} = x_{k+1,n} - x_{k,n} = \frac{(k+1)-np}{\sqrt{npq}} - \frac{k-np}{\sqrt{npq}} = \frac{1}{\sqrt{npq}},$$

так что

$$P_n(k_1;k_2) \approx \sum_{k=k_1}^{k_2} \varphi_0(x_{k,n})\Delta x_{k,n}. \qquad (10)$$

Выражение в правой части (10) является интегральной суммой для функции $\varphi_0(x)$ на отрезке $[x_{k_1,n}, x_{k_2,n}]$. При $n \to \infty$ будем иметь $\Delta x_{k,n} \to 0$, поэтому пределом суммы будет соответствующий определенный интеграл. Таким образом, считая n достаточно большим, получаем приближенную формулу

$$P_n(k_1;k_2) \approx \int_{x_{k_1,n}}^{x_{k_2,n}} \varphi_0(t)dt = \frac{1}{\sqrt{2\pi}} \int_{x_{k_1,n}}^{x_{k_2,n}} e^{-\frac{1}{2}t^2}\,dt, \qquad (11)$$

где

$$x_{k_1,n} = \frac{k_1-np}{\sqrt{npq}}, \quad x_{k_2,n} = \frac{k_2-np}{\sqrt{npq}}.$$

Проведенное рассуждение, позволяет предположить (не являясь строгим доказательством), что интеграл в правой части (6) служит пределом вероятностей $P_n(k_1;k_2)$ при $n \to \infty$. Сформулируем без доказательства точное утверждение.

Теорема (интегральная теорема Муавра-Лапласа). *Если k есть число наступлений события A в n испытаниях по схеме Бернулли, то равномерно относительно a и b, $(-\infty \le a < b \le +\infty)$ при $n \to \infty$ имеет место соотношение*

$$\mathrm{P}\left(a \le \frac{k-np}{\sqrt{npq}} \le b\right) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}t^2}\,dt.$$

Введем функцию (*стандартный интеграл вероятностей* или *функция Лапласа*)

$$\Phi_o(x) = \int_0^x \varphi_0(t)dt = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2}\,dt. \qquad (12)$$

Поскольку функция $\Phi_o(x)$ является первообразной для $\varphi_0(x)$, на основании формулы Ньютона-Лейбница будем иметь

$$P_n(k_1;k_2) \approx \Phi_o(x_{k_2,n}) - \Phi_o(x_{k_1,n}). \qquad (13)$$

Формула (8) дает приближенное значение вероятности того, что число k наступлений события A при n испытаниях удовлетворяет неравенству $k_1 \le k \le k_2$. Это событие равносильно тому, что случайная величина $X = x_{k,n}$ удовлетворяет неравенству $x_{k_1,n} \le X \le x_{k_2,n}$. Поэтому формулу (11) часто записывают в виде

$$P(k_1 \le k \le k_2) = P(x_{k_1,n} \le X \le x_{k_2,n}) \approx \Phi_o(x_{k_2,n}) - \Phi_o(x_{k_1,n})$$

(*интегральная формула Лапласа.*)

Пример. Вероятность поражения цели при одиночном выстреле одного орудия равна $p = 0{,}04$. Какова вероятность того, что при залпе из 100 орудий цель будет поражена не менее 30 раз?

Решение. Здесь $n = 100$, $30 \le k \le 100$. Имеем

$$\sqrt{npq} = \sqrt{100 \cdot 0{,}3 \cdot 0{,}7} \approx 4{,}6.$$

Отсюда

$$x_{30,100} = \frac{30 - 100 \cdot 0{,}3}{4{,}6} \approx 0$$

$$x_{100,100} = \frac{100 - 100 \cdot 0{,}3}{4{,}6} \approx 15{,}22.$$

На основании формулы (8) получаем

$$P_n(30 \le k \le 100) \approx \Phi_o(15{,}22) - \Phi_o(0) \approx 0{,}5000 - 0 = 0{,}5.$$

7.3. Теорема Пуассона

Пусть последовательно производятся серии из n $(n = 1, 2, \ldots)$ испытаний по схеме Бернулли, причем вероятность p_n наступления «успеха» в данной серии зависит от ее номера n и стремится к нулю при $n \to \infty$ (т. н. последовательность «редких событий»). Обозначим $\lambda_n = np_n$ и будем считать, что существует предел

$$\lim_{n\to\infty} \lambda_n = \lambda > 0. \qquad (14)$$

Зафиксируем число k и изучим предельное поведение вероятности $P_n(k)$ при условии, что длина серии $n \to \infty$. По формуле Бернулли имеем

$$P_n(k) = C_n^k p_n^k (1 - p_n)^{n-k} = C_n^k \left(\frac{\lambda_n}{n}\right)^k \left(1 - \frac{\lambda_n}{n}\right)^{n-k}. \qquad (15)$$

Преобразуем произведение первых двух сомножителей выражения (15)

$$C_n^k \left(\frac{\lambda_n}{n}\right)^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!\,n^k}\lambda_n^k =$$

$$= \frac{\lambda_n^k}{k!}\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\ldots\left(1 - \frac{k-1}{n}\right). \qquad (16)$$

Так как число скобок в (16) фиксировано (оно равно $k-1$) и каждая скобка стремится к 1 при $n \to \infty$, получим, с учетом (14)

$$\lim_{n\to\infty} C_n^k \left(\frac{\lambda_n}{n}\right)^k = \frac{\lambda^k}{k!}. \qquad (17)$$

Используя второй замечательный предел

$$\lim_{t\to\infty}(1+t)^{\frac{1}{t}} = e,$$

и учитывая, что

$$\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{-k} = 1,$$

найдем

$$\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k} = \lim_{n\to\infty}\left[\left(1-\frac{\lambda_n}{n}\right)^{\frac{n}{\lambda_n}}\right]^{\lambda_n}\left(1-\frac{\lambda_n}{n}\right)^{-k} = e\cdot 1 = e^{-\lambda}. \tag{18}$$

Наконец, принимая во внимание (15), (17) и (18), получаем

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!}\,e^{-\lambda}. \tag{19}$$

Утверждение, содержащееся в формуле (19) называется *теоремой Пуассона*. Из предельного соотношения (19) при больших значениях n получаем приближённую *формулу Пуассона*

$$P_n(k) \approx \frac{\lambda^k}{k!}\,e^{-\lambda}. \tag{20}$$

З а м е ч а н и е. Формула Пуассона применяется в случаях, когда число испытаний «велико» , вероятность p_n события A «мала», а произведение $\lambda_n = np_n$ «не мало, и не велико» . Формула Пуассона находит применение в теории массового обслуживания.

П р и м е р. При производстве некоторой массовой продукции вероятность появления одного нестандартного изделия составляет 0,001. Какова вероятность того, что в партии из 1000 изделий продукции этого вида 3 изделия будут нестандартными?

Р е ш е н и е. Здесь вероятность $p = 0{,}001$ мала, число $n = 1000$ велико, причем $\lambda = np = 1000\cdot 0{,}001 = 1$. Используя формулу Пуассона для искомой вероятности, получаем следующее значение:

$$P_{1000}(3) \approx \frac{\lambda^2}{3!}e^{-\lambda} = \frac{1}{6}e^{-1} \approx 0{,}0613.$$

Глава II. СЛУЧАЙНЫЕ ВЕЛИЧИНЫ

§ 1. Случайная величина и ее распределение

В приложениях теории вероятностей под случайной величиной понимают скалярную или векторную величину, принимающую свои значения в зависимости от исхода некоторого случайного эксперимента. Например:

1) число очков, выпадающее при однократном бросании игральной кости;

2) число клиентов, обратившихся на фирму в течение рабочего дня;

3) число испытаний в схеме Бернулли до наступления первого «успеха» при неограниченной длине серии;

4) величина уклонения точки падения снаряда от центра цели.

Рассмотрим теперь математическое определение случайной величины. Сначала сформулируем понятие борелевского множества (Эмиль Борель – французский математик, 1871-1956).

Определение. *Системой борелевских множеств* $\mathscr{B}(\mathbb{R})$ *на прямой* $\mathbb{R}$ *называют минимальную сигма-алгебру (т.е. содержащуюся во всех остальных сигма-алгебрах с указанным свойством), включающую в себя все открытые и замкнутые множества из* $\mathbb{R}$.

Упрощенно можно считать, что борелевские множества – это множества, которые можно построить с помощью не более чем счетного числа операций объединения и пересечения открытых или замкнутых множеств.

Определение. *Пусть дано вероятностное пространство* $(\Omega, \mathscr{A}, \mathsf{P})$ *и* $X = X(\omega)$ *– однозначная числовая функция элементарного события* ω, *определенная на множестве* Ω. *Функцию* X *называют измеримой (относительно сигма-алгебры* $\mathscr{A}$*), если для любого борелевского множества* $B \subset \mathscr{B}(\mathbb{R})$ *множество* $A = X^{-1}(B) = \{\omega : X(\omega) \in B\}$ *является элементом нашей сигма-алгебры.*

Иными словами, прообраз любого $B \in \mathscr{B}$ должен быть элементом $\mathscr{A}$, т.е. событием. С содержательной точки зрения измеримость означает, что факт попадания случайной точки в пределы любого борелевского множества должен рассматриваться нами как событие.

Замечание. Можно доказать, что если вместо всевозможных борелевских множеств B рассматривать только лучи вида $(-\infty, x)$, то при этом получается определение измеримости, эквивалентное только что сформулированному.

Определение. *Случайной величиной* $X = X(\omega)$, *определенной на вероятностном пространстве* $(\Omega, \mathscr{A}, \mathsf{P})$ *называется любая измеримая функция элементарного события.*

Пусть, X – случайная величина, B – произвольное борелевское множество и $A = X^{-1}(B)$ – прообраз B при отображении X. Поскольку A принадлежит $\mathscr{A}$, для него определена вероятность $\mathsf{P}(A)$. Если каждому $B \in \mathscr{B}(\mathbb{R})$ сопоставить указанным образом число $\mathsf{P}(A)$, то мы получим меру $\mathsf{P}_X(B)$, заданную на множестве $\mathscr{B}(\mathbb{R})$ всех борелевских подмножеств числовой прямой. Она называется *распределением* случайной величины X.

Определение. *Функцией распределения случайной величины* X *называется функция*

$$F_X(x) = \mathsf{P}_X(-\infty, x) = \mathsf{P}\{\omega : X(\omega) < x\}.$$

Обычно, в обозначении $F_X(x)$ опускают нижний индекс, если ясно, о какой случайной величине идет речь. Функция распределения $F(x)$ однозначно задает распределение случайной величины X (т.е. меру $\mathsf{P}_X(B)$, $B \in \mathscr{B}(\mathbb{R})$). В то же время, следует иметь в виду, что исходную вероятностную меру $\mathsf{P}(A)$, заданную на $\mathscr{A}$, в общем случае по $F(x)$ однозначно восстановить нельзя.

§ 2. Дискретные случайные величины

Определение. *Случайная величина X называется дискретной, если она принимает конечное или счетное число попарно различных значений*

$$x_1, x_2, \dots, x_n, \dots \tag{1}$$

с вероятностями

$$p_1, p_2, \dots, p_n, \dots, \quad p_n = \mathsf{P}(\omega : X(\omega) = x_n).$$

Так как события $\omega : X(\omega) = x_n$, $n = 1, 2, \dots$, образуют полную группу, имеем

$$\sum_n p_n = 1. \tag{2}$$

Здесь подразумевается, что суммирование поизводится по всем индексам n, которыми снабжены значения (1). Если значений конечное число, выражение (2) представляет собой конечную сумму, если счетное, то – бесконечный числовой ряд. Подобного соглашения мы будем придерживаться и в дальнейшем, не указывая явно пределы суммирования. Распределение вероятностей дискретной случайной величины имеет вид

$$\mathsf{P}_X(B) = \sum_{x_n \in B} p_n.$$

Удобнее всего задавать распределение дискретной величины с помощью таблицы. Например, если величина X принимает значения $x_1, x_2, \dots, x_n$ с вероятностями $p_i = \mathsf{P}(X = x_i)$, то соответствующая таблица выглядит следующим образом

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

Если случайная величина принимает счетное число значений, то таблица будет бесконечной.

Примерами дискретных случайных величин могут служить первые три из перечисленных в начале параграфа. При этом первые две принимают конечное, третья – счетное множество значений. Подробно мы будем рассматривать примеры распределений дискретных случайных величин в гл. V.

§ 3. Абсолютно непрерывные случайные величины

Определение. *Случайная величина* X *называется абсолютно непрерывной, если существует неотрицательная, интегрируемая на* $\mathbb{R}$ *функция* $p_X(x)$ *(плотность вероятности) такая, что*

$$\mathsf{P}_X(B) = \int_B p_X(x)dx \tag{3}$$

для любого множества $B \in \mathscr{B}(\mathbb{R})$. *Распределение вероятностей* $\mathsf{P}_X(B)$, *определяемое формулой* (3) *также называют абсолютно непрерывным.*

Если в (3) положить $B = (-\infty, +\infty)$, то получим равенство

$$\int_{-\infty}^{+\infty} p_X(x)dx = 1. \tag{4}$$

которое является аналогом формулы (2) для случая абсолютно непрерывного распределения. Как и для функции $F(x)$ в дальнейшем будем писать $p(x)$ вместо $p_X(x)$.

Очевидно, примером абсолютно непрерывной величины может служить четвертая из перечисленных в начале параграфа величин. Подробное исследование некоторых абсолютно непрерывных величин содержится в гл. V .

Замечание. С качественной точки зрения основное различие дискретных и непрерывных величин состоит в том, что первые могут принимать значения только из некоторых дискретных (т.е. состоящих из изолированных точек) множеств, тогда как значения вторых сплошь заполняют некоторые промежутки на числовой прямой.

Замечание. Распределение вероятностей допускает следующую физическую (механическую) аналогию. Аналогом дискретного распределения вероятностей $\mathsf{P}(X = x_k) = p_k$, $k = 1,2,...$, служит единичная масса, сосредоточенная в точках $x_1, x_2, ...$, числовой прямой $\mathbb{R}$ (бесконечно длинного невесомого стержня). Абсолютно непрерывному расспределению вероятностей с плотностью $p(x)$ соответствует единичная масса, «размазанная» вдоль бесконечного стержня непрерывным образом. При этом линейная плотность стержня в точке x, определяемая как предел $\rho(x) = \lim\limits_{\Delta x \to 0} \dfrac{m(x + \Delta x) - m(x)}{\Delta x}$, где $m(x)$ – масса промежутка $(-\infty; x]$, совпадает с плотностью распределения $p(x)$.

Замечание. Существуют случайные величины, которые не относятся ни к одному из двум рассмотренных выше типов. Их называют величинами смешанного типа. В подробных курсах теории вероятностей доказывается, что распределение любой случайной величины сводится к смеси распределений не более, чем трех основных типов: дискретного, абсолютно непрерывного и так называемого сингулярного.

§ 4. Свойства функции распределения и плотности вероятности

Сначала перечислим свойства, присущие функции распределения $F(x)$ случайной величины X любого типа.

1. *Значения функции распределения принадлежат отрезку* $[0,1]$.

Это очевидно, т.к. $F(x)$ представляет собой некоторую вероятность.

2. $F(x)$ *– неубывающая на* $(-\infty,+\infty)$ *функция.*

Действительно, пусть $x_1 < x_2$. Тогда событие $\{X < x_1\}$ влечет за собой событие $\{X < x_2\}$, так что по свойству вероятности $F(x_1) = \mathsf{P}(X < x_1) \le \mathsf{P}(X < x_2) = F(x_2)$.

3. *Вероятность того, что случайная величина* X *примет значение, заключенное в полуинтервале* $[a,b)$, равна

$$\mathsf{P}(a \le X < b) = F(b) - F(a).$$

В самом деле, обозначим $A = X < a$, $B = X < b$ и $C = \{a \le X < b\}$. Тогда $B = A + C$, откуда в силу несовместности A и C получим $\mathsf{P}(B) = \mathsf{P}(A) + \mathsf{P}(C)$ или $\mathsf{P}(C) = \mathsf{P}(B) - \mathsf{P}(A)$, т.е. $\mathsf{P}(a \le X < b) = F(b) - F(a)$.

4. *Справедливы равенства*

$$\lim_{x\to+\infty} F(x) = 0\,,\ \lim_{x\to+\infty} F(x) = 1\,.$$

Действительно, событие $\{-\infty < X < -\infty\}$ является невозможным, а событие $\{-\infty < X < +\infty\}$ – достоверным.

5. *Функция распределения непрерывна слева во всех точках, т.е.* $\forall x_0 \in \mathbb{R}$

$$F(x_0 - 0) = \lim_{x\to x_0-0} F(x) = F(x_0)\ . \tag{5}$$

Докажем это свойство. Пусть $\{x_i\}$ – неубывающая последовательность точек, сходящаяся к x_0 и пусть $A_i = \{x_i \le X < x_0\}$. Тогда $A_j \subset A_i$ при $j > i$, $\bigcap_{i=1}^{\infty} A_i = \varnothing$ и по аксиоме непрерывности получаем

$$\lim_{i\to\infty}\mathsf{P}(A_i) = \lim_{i\to\infty}[F(x_0) - F(x_i)] = F(x_0) - \lim_{i\to\infty}F(x_i) == F(x_0) - F(x_0 - 0) = 0.$$

6. *Функция распределения имеет разрывы только первого рода, и множество ее точек разрыва не более, чем счетно.*

Это свойство приведем без доказательства.

Теперь перечислим свойства функции распределения случайных величин специального вида. Доказательства следующих утверждений непосредственно следуют из соответствующих определений.

7. *Функция распределения дискретной случайной величины является ступенчатой, то есть является постоянной всюду за исключением точек* (1), *а в самих этих точках имеет скачки величины* $p_1, p_2, \ldots, p_n, \ldots$ *соответственно. Она выражаеся формулой*

$$F(x) = \sum_{x_n < x} p_n\ . \tag{6}$$

8. *Функция распределения абсолютно непрерывной случайной величины*

есть интеграл от плотности вероятности

$$F(x)=\int\limits_{-\infty}^{x}p(t)dt. \tag{7}$$

9. *Функция распределения абсолютно непрерывной случайной величины непрерывна.*

Это следует из формулы (7) и соответствующего свойства определенного интеграла.

Замечание. В отличие от функции распределения $F(x)$ абсолютно непрерывной случайной величины, ее плотность вероятности $p(x)$ может быть разрывной функцией. Кроме того она может быть не определена в одной или нескольких изолированных точках (и даже в бесконечном их числе).

10. Если x есть точка непрерывности функции $p(x)$, то справедливо равенство

$$p(x)=F'(x).$$

Данное свойство также следует из соответствующего свойства определенного интеграла.

11. *Вероятность того, что абсолютно непрерывная случайная величина* X *примет заранее указанное определенное значение* x_0, *равна нулю*

$$\mathsf{P}(X=x_0)=0,\ \forall x_0\in\mathbb{R}.$$

Доказательство. Пусть x_i – невозрастающая последовательность точек, сходящаяся к x_0 и пусть $A_i=\{x_0\le X<x_i\}$. Тогда

$$\mathsf{P}(x_0\le X<x_i)=F(x_i)-F(x_0),$$

$A_j\subset A_i$ при $j>i$, $\bigcap\limits_{i=1}^{\infty}A_i=x_0$ и по аксиоме непрерывности получаем

$$\mathsf{P}(X=x_0)=\lim_{i\to\infty}[F(x_i)-F(x_0)]=\lim_{i\to\infty}F(x_i)-F(x_0)=F(x_0)-F(x_0)=0.$$

Здесь мы воспользовались непрерывностью функции $F(x)$.

12. *Вероятность того, что абсолютно непрерывная случайная величина* X *примет значение в любом из конечных промежутков* $[a,b)$, $(a,b]$, $[a,b]$, (a,b) *одна и та же и вычисляется по формуле*

$$\mathsf{P}(a\le X<b)=\int\limits_{a}^{b}p(t)dt=F(b)-F(a). \tag{8}$$

Доказательство. Равенство указанных вероятностей следует из свойства 10. Действительно, рассмотрим, для определенности, события $\{a\le X<b\}$ и $\{X=b\}$. Они несовместны, при этом $\{a\le X\le b\}=\{a\le X<b\}\cup\{X=b\}$. Поскольку $\mathsf{P}(X=b)=0$, получаем

$$\mathsf{P}(a\le X\le b)=\mathsf{P}(a\le X<b)+0=F(b)-F(a).$$

Равенство $\mathsf{P}(a\le X<b)=\int\limits_{a}^{b}p(t)dt=F(b)-F(a)$ верно в силу свойства 8 и формулы Ньютона-Лейбница.

Глава III. СИСТЕМЫ СЛУЧАЙНЫХ ВЕЛИЧИН (СЛУЧАЙНЫЕ ВЕКТОРЫ)

§ 1. Распределение системы случайных величин

Определение. *Пусть на вероятностном пространстве* $(\Omega, \mathscr{A}, \mathsf{P})$ *определены случайные величины* $X_1, X_2, \dots, X_n$. *Набор* $\vec{X} = (X_1, X_2, \dots, X_n)$ *называется системой случайных величин, или* n*-мерным случайным вектором, или многомерной случайной величиной.*

Определение. *Функцией распределения случайного вектора* $\vec{X} = (X_1, X_2, \dots, X_n)$ *называется функция от* n *переменных, определенная на* $\mathbb{R}^n$ *вида*

$$F(x_1, x_2, \dots, x_n) = \mathsf{P}(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n). \qquad (1)$$

Функцию (1) часто называют также функцией совместного распределения величин $X_1, X_2, \dots, X_n$ или многомерной функцией распределения.

Основные свойства функции совместного распределения мы рассмотрим для двумерного случая $\boldsymbol{X} = (X, Y)$, $F(x, y) = \mathsf{P}(X < x, Y < y)$.

1. Функция $F(x, y)$ не убывает по каждой из переменных:

$$F(x_2; y) \geq F(x_1; y) \text{ при } x_2 > x_1; \qquad F(x; y_2) \geq F(x; y_1) \text{ при } y_2 > y_1.$$

2. Функция $F(x, y)$ непрерывна слева по каждой переменной.

3. Индивидуальные (маргинальные) функции распределения $F_1(x)$, $F_2(y)$ величин X и Y соответственно вычисляются по формулам

$$\lim_{y \to +\infty} F(x; y) = F_1(x), \qquad \lim_{x \to +\infty} F(x; y) = F_2(y).$$

4. При одновременном стремлении обеих переменных к $+\infty$ имеем

$$\lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x; y) = 1. \qquad (2)$$

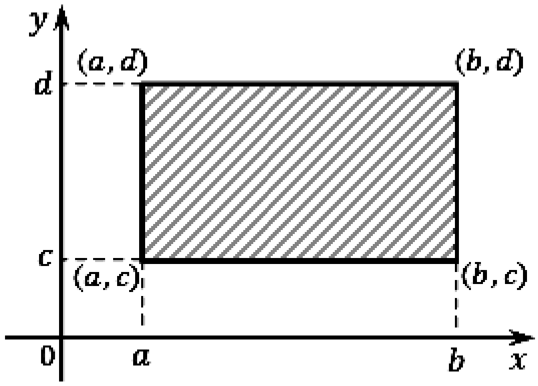
5. При стремлении к $-\infty$ хотя бы одной из переменных справедливо равенство

$$\lim_{x \to -\infty} F(x; y) = \lim_{y \to -\infty} F(x; y) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x; y) = 0. \qquad (3)$$

Замечание. Обратим внимание, что в предельном соотношении (3) равенство нулю достигается при стремлении к $-\infty$ хотя бы одной из переменных, тогда как в (2) требуется, чтобы к $+\infty$ стремились *обе* переменные.

6. Вероятность попадания случайной точки (X, Y) в произвольный прямоугольник со сторонами, параллельными координатным осям (см. рис.), вычисляется, по формуле

$$P\{a \leq X < b, c \leq Y < d\} = F(b, d) - F(a, d) - F(b, c) + F(a, c).$$



§ 2. Дискретные случайные векторы

Определение. *Случайный вектор называется дискретным, если его координаты являются дискретными случайными величинами.*

Совместное распределение пары дискретных величин удобно задавать таблицей. Пусть X и Y – дискретные случайные величины, возможные значения которых (x_i, y_j), где $i = 1,2,\ldots,n;\ j = 1,2,\ldots,m$. Тогда распределение системы таких случайных величин может быть охарактеризовано указанием вероятностей $p_{ij} = \mathsf{P}\{X = x_i, Y = y_j\}$ того, что случайная величина X примет значение x_i и одновременно с этим случайная величина Y примет значение y_j. Вероятности p_{ij} записывают в таблицу.

$x_i \backslash y_j$	y_1	y_2	…	y_m
x_1	p_{11}	p_{12}	…	p_{1m}
…	…	…	…	…
x_n	p_{n1}	p_{n2}	…	p_{nm}

Все возможные события $\{X = x_i, Y = y_j\}$ при $i = 1,2,\ldots,n,\ j = 1,2,\ldots,m$ составляют полную группу, поэтому

$$\sum_{i,j} p_{i,j} = \sum_{i,j} \mathsf{P}\{X = x_i, Y = y_j\} = 1.$$

Суммируя вероятности по строкам и столбцам, получим маргинальные распределения величин X и Y:

$$\sum_{j} p_{i,j} = \sum_{j} \mathsf{P}\{X = x_i, Y = y_j\} = P\{X = x_i\} = p_i;$$
$$\sum_{i} p_{i,j} = \sum_{i} \mathsf{P}\{X = x_i, Y = y_j\} = P\{Y = y_j\} = q_j.$$

§ 3. Абсолютно непрерывные случайные векторы

Определение. *Говорят, что случайный вектор* $\vec{X} = (X_1, X_2, \ldots, X_n)$ *имеет абсолютно непрерывное распределение, если существует* *неотрицательная интегрируемая функция* $p(x_1, x_2, \ldots, x_n)$ *(называемая плотностью совместного распределения) такая, что для любого борелевского множества* $D \subset \mathbb{R}^n$ выполняется равенство

$$\mathsf{P}(X \in D) = \int\int\limits_D \ldots \int p(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n. \qquad (4)$$

В дальнейшем снова будем для простоты рассматривать двумерный вектор $\vec{X} = (X, Y)$. Для него равенство (4) примет вид

$$\mathsf{P}\{(X, Y) \in D\} = \iint\limits_D p(x, y) dxdy.$$

Если функция совместного распределения известна и дважды дифференцируема в точке (x, y), то по ней можно вычислить совместную плотность

$$p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}.$$

Наоборот, если плотность задана, то функция совместного распределения получается интегрированием совместной плотности

$$F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} p(u, v) dudv.$$

Поскольку событие $(X, Y) \in \mathbb{R}^2$ является достоверным, имеет место равенство

$$\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} p(x, y) dxdy = 1. \qquad (5)$$

Можно доказать, что всякая неотрицательная интегрируемая функция $p(x, y)$ такая, что верно (5) является плотностью совместного распределения некоторого абсолютно непрерывного случайного вектора (X, Y). Это замечание верно для векторов любой размерности.

По заданной совместной плотности можно найти маргинальные плотности $p_1(x)$ и $p_2(y)$ величин X и Y соответственно

$$p_1(x) = \int\limits_{-\infty}^{+\infty} p(x, y) dy,$$

$$p_2(y) = \int\limits_{-\infty}^{+\infty} p(x, y) dx.$$

П р и м е р (равномерное распределение на плоскости). *Пусть* D *– плоское множество с конечной площадью* $\mathrm{mes}(D)$. *Говорят, что вектор* (X, Y) *имеет равномерное распределение в области* D, *если плотность совместного распределения* $p(x, y)$ *постоянна в области* D *и равна нулю вне этой области*

$$p(x, y) = \begin{cases} \dfrac{1}{\mathrm{mes}(D)}, \text{ если } (x, y) \in D; \\ 0, \text{ если } (x, y) \notin D. \end{cases}$$

Объем подграфика этой функции равен объему цилиндра с основанием D и высотой $1/\operatorname{mes}(D)$, т.е. равен единице. Таким образом функция $p(x,y)$ может служить плотностью вероятности, т. к. она неотрицательна и двойной интеграл от нее по всей плоскости равен единице. Равномерное распределение имеет случайная точка (X,Y), которую наудачу бросают в пределы фигуры D. При этом вероятность попадания (X,Y) в пределы любой другой измеримой фигуры $F \subset \mathbb{R}^2$ равна отношению площадей $\operatorname{mes}(F \cap D)/\operatorname{mes}(D)$, в частности,

$$\mathsf{P}\{(X,Y) \in F\} = \begin{cases} \dfrac{\operatorname{mes}(F)}{\operatorname{mes}(D)}, \text{ если } F \subset D; \\ \quad 1, \text{ если } F = D. \end{cases}$$

Последние соотношения полностью согласуются с введенным ранее понятием геометрической вероятности.

Замечание. Дискретный случайный вектор определялся как вектор с дискретными компонентами. Для абсолютно непрерывных векторов такой подход недопустим. Можно привести примеры векторов с абсолютно непрерывными компонентами, которые не имеют совместной плотности. В то же время, доказывается, что если вектор является абсолютно непрерывным в смысле приведенного выше определения, то его компонетны также абсолютно непрерывны.

Замечание. Отметим следующее свойство случайных векторов, которое справедливо для любых их типов. Если известно совместное распределение, то всегда можно найти индивидуальные. Обратное верно только для тех случаев, когда совместное распределение полностью характеризуется индивидуальными. Такая ситуация имеет местно, например, для так называемых *независимых* случайных величин. В общем случае, зная только частные распределения, восстановить однозначно совместное распределение невозможно.

§ 4. Независимость случайных величин

Определение. *Случайные величины $X_1,\ldots,X_n$ называются независимыми (в совокупности), если для любой системы борелевских множеств $B_i \in \mathscr{B}(\mathbb{R})$, $i=1,\ldots,n$* выполняется равенство

$$\mathsf{P}(X_1 \in B_1,\ldots,X_n \in B_n) = \mathsf{P}(X_1 \in B_1)\mathsf{P}(X_2 \in B_2)\cdots\mathsf{P}(X_n \in B_n).$$

Можно доказать, что это определение эквивалентно следующему.

Определение. *Случайные величины $X_1,\ldots,X_n$ называются независимыми (в совокупности), если для любых $x_1,\ldots,x_n$ выполняется равенство*

$$F(x_1,\ldots,x_n) = F_1(x_1)\cdots F_n(x_n),$$

где F – совместная функция распределения, а $F_1(x_1),\ldots,F_n(x_n)$ – маргинальные функции распределения величин $X_1,\ldots,X_n$ соответственно.

В случае, когда система случайных величин имеет абсолютно непрерывное совместное распределение, условие независимости допускает еще одну эквивалентную формулировку.

Т е о р е м а. *Случайные величины* $X_1,\ldots,X_n$ *являются независимыми в совокупности тогда и только тогда, когда для любых* $x_1,\ldots,x_n$ *выполняется равенство*

$$p(x_1,\ldots,x_n) = p_1(x_1)\cdots p_n(x_n),$$

где p *– совместная плотность вероятности, а* $p_1(x_1),\ldots,p_n(x_n)$ *– маргинальные плотности величин* $X_1,\ldots,X_n$ *соответственно.*

Для системы двух дискретных случайных величин $\vec{X} = (X,Y)$ с совместным распределением $p_{ij} = \mathsf{P}\{X = x_i, Y = y_j\}$ и маргинальными распределениями $\mathsf{P}\{X = x_i\} = p_i$ и $\mathsf{P}\{Y = y_j\} = q_j$, аналогом предыдущей теоремы будет следующая.

Т е о р е м а. *Случайные величины* $\vec{X} = (X,Y)$ *являются независимыми тогда и только тогда, когда выполняются равенства*

$$p_{i,j} = p_i q_j. \tag{6}$$

§ 5. Распределение функции от случайных величин

Рассмотрим следующую задачу: пусть дан случайный вектор $(X_1,\ldots,X_n)$ с известным распределением и набор из k неслучайных, измеримых по Борелю функций от n переменных $g_i(x_1,x_2,\ldots,x_n)$, $i=1,\ldots,k$. Требуется найти распределение случайного вектора $(Y_1,Y_2,\ldots,Y_k)$, где $Y_i = g_i(X_1,X_2,\ldots,X_n)$, $i=1,\ldots,k$.

Пусть сначала $(X_1,\ldots,X_n)$ – абсолютно непрерывный вектор с многомерной плотностью $p(x_1,x_2,\ldots,x_n)$ и $\Phi(y_1,y_2,\ldots,y_k)$ – неизвестная многомерная функция распределения вектора $(Y_1,Y_2,\ldots,Y_k)$. Тогда из определения совместной плотности и функции распределения получим для Φ следующее выражение

$$\Phi(y_1,y_2,\ldots,y_k) = \int\limits_{D(y_1,y_2,\ldots,y_k)}\!\!\!\ldots\int p(x_1,x_2,\ldots,x_n)\,dx_1dx_2\ldots dx_n, \tag{7}$$

где переменная область интегрирования $D(y_1,y_2,\ldots,y_k)\subset\mathbb{R}^n$ определяется условиями

$$g_i(x_1,x_2,\ldots,x_n) < y_i,\ i=1,2,\ldots k.$$

Для функции распределения дискретного вектора имеет место аналогичное выражение, где вместо интеграла фигурирует n-мерная сумма, распространенная на область $D(y_1,y_2,\ldots,y_k)$.

П р и м е р. Абсолютно непрерывная величина X имеет плотность вероятности $p_X(x)$. Требуется найти выражение для плотности $p_Y(y)$ величины $Y = CX$, где $C = \text{const} \neq 0$. В данном случае $n=1$, $k=1$, $g(x)=Cx$. Пусть сначала $C>0$. Тогда

$$F_Y(y) = \mathsf{P}(CX < y) = \mathsf{P}\left(X < \frac{y}{C}\right) = F_X\left(\frac{y}{C}\right),$$

откуда

$$p_Y(y) = \frac{d}{dy} F_X\left(\frac{y}{C}\right) = F'_X\left(\frac{y}{C}\right) \cdot \frac{1}{C} = \frac{1}{C} \cdot p_X\left(\frac{y}{C}\right), \tag{8}$$

Пусть теперь $C<0$. Тогда

$$F_Y(y) = \mathsf{P}(CX < y) = \mathsf{P}\left(X > \frac{y}{C}\right) = 1 - \mathsf{P}\left(X < \frac{y}{C}\right) = 1 - F_X\left(\frac{y}{C}\right),$$

так что

$$p_Y(y) = \frac{d}{dy}\left(1 - F_X\left(\frac{y}{C}\right)\right) = -F'_X\left(\frac{y}{C}\right) \cdot \frac{1}{C} = -\frac{1}{C} \cdot p_X\left(\frac{y}{C}\right), \tag{9}$$

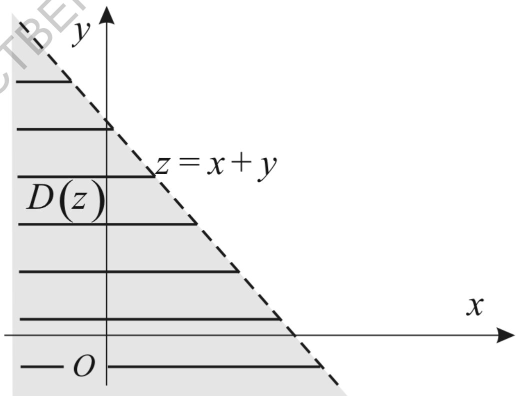
Объединяя (8) и (9), запишем

$$p_Y(y) = \frac{1}{|C|} \cdot p_X\left(\frac{y}{C}\right). \tag{10}$$

П р и м е р. Абсолютно непрерывный вектор (X,Y) имеет известную плотность вероятности $p(x,y)$. Требуется найти распределение суммы его компонент $Z = X + Y$. В нашем случае $n=2$, $k=1$, $g(x,y) = x+y$ и общая формула (7) примет вид

$$\Phi(z) = \iint\limits_{D(z)} p(x,y)\,dx\,dy, \tag{11}$$

где $D(z) = \{(x,y) : x + y < z\}$ (см. рис.).



Сводя двойной интеграл (11) к повторному, получим

$$\Phi(z) = \iint\limits_{D(z)} p(x,y)\,dx\,dy = \int\limits_{-\infty}^{+\infty} dx \int\limits_{-\infty}^{z-x} p(x,y)\,dy. \tag{12}$$

Дифференцируя интеграл (12) по параметру z, окончательно найдем

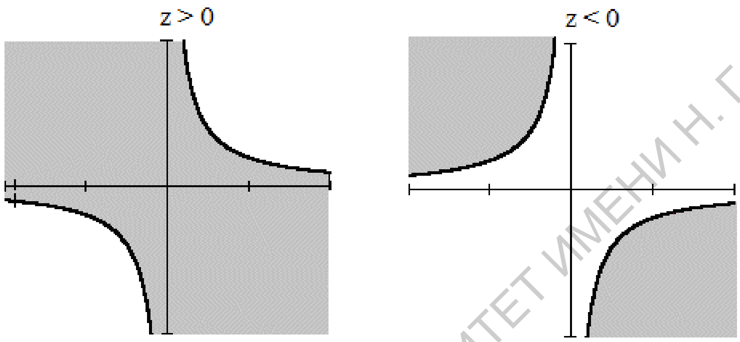
$$p_Z(z) = \Phi'(z) = \int\limits_{-\infty}^{+\infty} p(x, z-x)\,dx. \tag{13}$$

В частном случае, когда величины X и Y независимы, их совместная плотность равна произведению маргинальных плотностей $p(x,y)=p_1(x)p_2(y)$, и формула (13) запишется в виде

$$p_Z(z)=\int\limits_{-\infty}^{+\infty} p_X(x)p_Y(z-x)dx=\int\limits_{-\infty}^{+\infty} p_Y(x)p_X(z-x)dx. \qquad (14)$$

Каждый из интегралов (14) называется *сверткой* функций $p_X(x)$ и $p_Y(x)$. Таким образом, плотность суммы независимых случайных величин равна свертке их плотностей.

Пример. Абсолютно непрерывный вектор (X,Y) имеет плотность вероятности $p(x,y)$. Требуется найти распределение произведения его компонент $Z=XY$. Теперь область интегрирования $D(z)=\{(x,y): xy<z\}$ в (11) имеет разную форму в зависимости от знака переменной z (см. рис.).



Однако при сведении двойного интеграла по области $D(z)$ к повторному последний будет иметь один и тот же вид вне зависимости от знака z

$$\Phi(z)=\iint\limits_{D(z)} p(x,y)dx\,dy=\int\limits_{-\infty}^{0} dx\int\limits_{\frac{z}{x}}^{+\infty} p(x,y)dy+\int\limits_{0}^{+\infty} dx\int\limits_{-\infty}^{\frac{z}{x}} p(x,y)dy, \qquad (15)$$

Дифференцирование (15) по параметру z дает

$$p_Z(z)=\Phi'(z)=-\int\limits_{-\infty}^{0} p\left(x,\frac{z}{x}\right)\frac{dx}{x}+\int\limits_{0}^{+\infty} p\left(x,\frac{z}{x}\right)\frac{dx}{x}=\int\limits_{-\infty}^{+\infty} p\left(x,\frac{z}{x}\right)\frac{dx}{|x|}, \qquad (16)$$

причем выражение (16) справедливо для всех $z\neq 0$. При $z=0$ функцию $p_Z(z)$ можно определить произвольно, т.к. абсолютно непрерывное распределение не зависит от величины плотности в отдельно взятой точке. В частности, можно положить

$$p_Z(0)=\int\limits_{-\infty}^{+\infty} p(x,0)\frac{dx}{|x|},$$

если этот несобственный интеграл сходится.

Таким образом,

$$p_Z(z)=\int\limits_{-\infty}^{+\infty} p\left(x,\frac{z}{x}\right)\frac{dx}{|x|} \qquad (17)$$

для всех $z\in\mathbb{R}$.

Глава IV. ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ РАСПРЕДЕЛЕНИЙ

§ 1. Математическое ожидание

Определение. *Математическим ожиданием дискретной случайной величины X с конечным множеством значений называется сумма произведений всех ее возможных значений x_i на их вероятности $p_i = \mathsf{P}(X = x_i)$*

$$\mathsf{M}X = \sum_{i=1}^{n} x_i p_i \,. \quad (1)$$

Если величина X принимает счетное множество значений, то ее математическое ожидание определяется как сумма ряда

$$\mathsf{M}X = \sum_{i=1}^{\infty} x_i p_i \,, \quad (2)$$

в предположении, что этот ряд сходится абсолютно.

Определение. *Математическим ожиданием абсолютно непрерывной случайной величины X с плотностью вероятности $p(x)$ называется интеграл*

$$\mathsf{M}X = \int_{-\infty}^{\infty} xp(x)dx \,, \quad (3)$$

в предположении, что он сходится абсолютно.

Если ряд (2) или интеграл (3) не сходятся абсолютно, то говорят, что математическое ожидание случайной величины X не существует.

Замечание. Математическое ожидание соответствует наиболее характерному, «среднему» значению случайной величины. Оно имеет следующий физический (механический) смысл. Пусть на бесконечном стержне (числовой прямой) дискретным или непрерывным образом распределена единичная масса (см. замечание из §3 гл. II). Тогда величина $\mathsf{M}X$, вычисленная по любой из формул (1)-(3), представляет собой координату центра тяжести этого стержня.

1.1. Свойства математического ожидания

Пусть $\vec{X} = (X_1, X_2, \ldots X_n)$ – дискретный случайный вектор с возможными значениями $\vec{x}_{\vec{i}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$, где каждая компонента i_s мультииндекса $\vec{i} = (i_1, i_2, \ldots, i_n)$ принимает значения $i_s = 1,2,\ldots,m_s$, $s = 1,\ldots,n$. Здесь m_s – либо натуральное число, либо символ $+\infty$, в зависимости от того, конечное или счетное множество значений $\{x_{i_s}\}_{i_s=1}^{m_s}$ может принимать величина X_s.

Пусть теперь имеется некоторое множество чисел $\{a_{i_1,i_2,\ldots,i_n}\}$, снабженных индексами $i_1, i_2, \ldots, i_n$. Тогда в целях сокращения записей по определению положим

$$\sum_{\bar{i}} a_{\bar{i}} = \sum_{i_1, i_2, \ldots, i_n} a_{i_1, i_2, \ldots, i_n},$$

где суммирование производится по всевозможным наборам значений индексов $i_1, i_2, \ldots, i_n$ так, что каждый индекс изменяется в указанных выше пределах.

Теорема. *Пусть* $\overrightarrow{X} = (X_1, X_2, \ldots X_n)$ *– дискретный случайный вектор с распределением*

$$\mathsf{P}(\vec{X} = \vec{x}_{\bar{i}}) = p_{\bar{i}} > 0, \quad \sum_{\bar{i}} p_{\bar{i}} == 1.$$

Пусть, далее, $y = g(\vec{x})$, $\vec{x} = (x_1, x_2, \ldots, x_n)$, *– числовая функция от* n *переменных такая, что ряд* $\sum_{\bar{i}} \left| g(\vec{x}_{\bar{i}}) \right| p_{\bar{i}}$ *сходится. Тогда математическое ожидание случайной величины* $Y = g(\vec{X})$ *существует, причем*

$$\mathsf{M}[g(\vec{X})] = \sum_{\bar{i}} g(\vec{x}_{\bar{i}}) p_{\bar{i}}. \tag{4}$$

Доказательство. Пусть $y_1, y_2, \ldots, y_k, \ldots$ – попарно различные значения функции g на всевозможных векторах $\vec{x}_{\bar{i}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$. Поскольку функция g может принимать одинаковые значения при разных значениях аргумента: $g(\vec{x}_{\bar{i}}) = g(\vec{x}_{\bar{j}})$, $(x_{i_1}, x_{i_2}, \ldots, x_{i_n}) \neq (x_{j_1}, x_{j_2}, \ldots, x_{j_n})$, распределение случайной величины $Y = g(\vec{X})$ запишется в виде

$$p_k = \mathsf{P}(Y = y_k) = \sum_{\bar{i} \in A_k} p_{\bar{i}},$$

где $A_k = \{\vec{i} \in \mathbb{N}^n : g(\vec{x}_i) = y_k\}$, $k = 1, 2, \ldots$ Множества мультииндексов $\{A_k\}$ образуют разбиение множества $I = \{\vec{i}\}$ *всех* мультииндексов нашей суммы (т.е. $A_k \bigcap A_j = \varnothing$ при $k \neq j$, и $\bigcup_k A_k = I$). Кроме того, ряд $\sum_{\bar{i}} g(\vec{x}_{\bar{i}}) p_{\bar{i}}$ абсолютно сходится, так что суммирование его членов можно производить в любом порядке. С учетом этого имеем

$$\mathsf{M}Y = \sum_k y_k p_k = \sum_k y_k \sum_{\bar{i} \in A_k} p_{\bar{i}} = \sum_k \sum_{\bar{i} \in A_k} y_k p_{\bar{i}} = \sum_k \sum_{\bar{i} \in A_k} g(\vec{x}_{\bar{i}}) p_{\bar{i}} = \sum_{\bar{i}} g(\vec{x}_{\bar{i}}) p_{\bar{i}}.$$

Теорема доказана.

Замечание. В силу только что упомянутого свойства абсолютно сходящихся рядов, суммирование в выражениях типа (4) можно производить в любом порядке. В частности, если $\vec{X} = (X, Y)$ – двумерный дискретный вектор с распределением $\mathsf{P}(X = x_i, Y = y_j) = p_{ij} > 0$, а $y = g(x, y)$ – функция двух переменных, то формулу (4) можно записать в виде

$$\mathsf{M}[g(X, Y)] = \sum_i \sum_j g(x_i, y_j) p_{ij}. \tag{5}$$

Теорема. *Пусть* $\overrightarrow{X} = (X_1, X_2, \ldots X_n)$ *– абсолютно непрерывный случайный вектор с многомерной плотностью вероятности* $p(x_1, x_2, \ldots, x_n)$ *и* $y = g(\vec{x})$, $\vec{x} = (x_1, x_2, \ldots, x_n)$, *– числовая функция от* n *переменных такая, что несобствен-*

ный n-кратный интеграл $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}...\int_{-\infty}^{+\infty} |g(x_1, x_2, ..., x_n)|\, p(x_1, x_2, ..., x_n)\, dx_1 dx_2 ... dx_n$ *сходится. Тогда математическое ожидание случайной величины* $Y = g(\vec{X})$ *существует, причем*

$$\mathsf{M}Y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}...\int_{-\infty}^{+\infty} g(x_1, x_2, ..., x_n)\, p(x_1, x_2, ..., x_n)\, dx_1 dx_2 ... dx_n\,. \tag{6}$$

Доказательство. В общем случае доказательство требует средств, выходящих за рамки стандартного курса математического анализа. Поскольку в дальнейшем нам понадобится только один частный случай этой теоремы, мы ограничимся доказательством именно этого частного случая.

Итак, предположим, что $n = 1$, т.е. имеется одна случайная величина X с плотностью вероятности $p_X(x)$ и неслучайная функция $y = g(x)$, $x \in \mathbb{R}$, – дифференцируемая и строго монотонная, причем

$$\int_{-\infty}^{+\infty} |g(x)|\, p_X(x)\, dx < +\infty\,. \tag{7}$$

Тогда существует обратная функция $x = g^{-1}(y)$, которая также дифференцируема и строго монотонна. Докажем, что для случайной величины $Y = g(X)$ математическое ожидание $\mathsf{M}Y$ существует и равно

$$\mathsf{M}[g(X)] = \int_{-\infty}^{+\infty} g(x)\, p_X(x)\, dx\,. \tag{8}$$

Найдем плотность вероятности величины Y. Ее функция распределения в силу непрерывности и строгой монотонности $g(x)$ равна

$$F_Y(y) = \mathsf{P}(Y < y) = \mathsf{P}(g(X) < y) = \mathsf{P}(X < g^{-1}(y)) = F_X[g^{-1}(y)].$$

Тогда

$$p_Y(y) = F_Y'(y) = \frac{d}{dy} F_X[g^{-1}(y)] = F_X'[g^{-1}(y)] \frac{d}{dy} g^{-1}(y) = p_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y).$$

Предполагая, что $\mathsf{M}Y$ существует (это пока не доказано), запишем

$$\mathsf{M}Y = \int_{-\infty}^{+\infty} y p_Y(y)\, dy = \int_{-\infty}^{+\infty} y p_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y)\, dy\,. \tag{9}$$

Выполнив в последнем интеграле замену переменной $x = g^{-1}(y)$, $dx = \frac{d}{dy} g^{-1}(y)\, dy$, получим

$$\mathsf{M}Y = \int_{-\infty}^{+\infty} y p_X[g^{-1}(y)] \frac{d}{dy} g^{-1}(y)\, dy = \int_{-\infty}^{+\infty} g(x)\, p_X(x)\, dx\,.$$

Таким образом, формула (8) верна, если только интеграл (9) сходится абсолютно. Но его абсолютная сходимость следует из условия (7) и теоремы о замене переменной в несобственном интеграле.

Теперь мы приведем важнейшие свойства математического ожидания. Отметим, что все они будут получены как следствия полностью доказанных ранее утверждений.

1) *Математическое ожидание постоянной величины равно самой постоянной*

$$\mathsf{M}(C) = C.$$

Доказательство. Постоянную можно рассматривать как случайную величину, которая имеет *вырожденное распределение*, сосредоточенное в точке $C \in \mathbb{R}$, т.е., если

$$\mathsf{P}\{X = C\} = 1.$$

Для величины X такого типа математическое ожидание, по определению, равно $\mathsf{M}X = C \cdot 1 = C$.

2) *Постоянный множитель можно выносить за знак математического ожидания*

$$\mathsf{M}(CX) = C\mathsf{M}(X).$$

Доказательство. Проведем доказательство сначала для абсолютно непрерывной величины X с плотностью распределения $p(x)$. При $C = 0$ имеет место предыдущий случай. Пусть $C \neq 0$. Величина $\mathsf{M}(CX)$ есть матожидание от случайной величины $Y = CX$, плотность которой равна $p_Y(y) = \frac{1}{|C|} \cdot p_X\left(\frac{y}{C}\right)$ (см. §5 гл. V). Тогда

$$\mathsf{M}Y = \int_{-\infty}^{\infty} y p_Y(y) dy = \frac{1}{|C|} \int_{-\infty}^{\infty} y p_X\left(\frac{y}{C}\right) dy = C \operatorname{sgn}(C) \int_{-\infty}^{\infty} \frac{y}{C} p_X\left(\frac{y}{C}\right) d\left(\frac{y}{C}\right) =$$

$$= C \int_{-\infty}^{\infty} x p_X(x) dt = C\mathsf{M}(X).$$

Для дискретной величины X доказательство проще, т.к. распределение величины $Y = CX$ в этом случае имеет вид

$\mathsf{P}\{X = x_i\} = \mathsf{P}\{CX = Cx_i\} = \mathsf{P}\{Y = y_i\} = p_i$, так что

$$\mathsf{M}Y = \sum_i y_i \mathsf{P}\{Y = y_i\} = \sum_i y_i p_i = \sum_i Cx_i p_i = C \sum_i x_i p_i = C\mathsf{M}(X).$$

3) *Математическое ожидание суммы двух случайных величин* X и Y *равно сумме математических ожиданий слагаемых, при условии, что* $\mathsf{M}X$ *и* $\mathsf{M}Y$ *существуют*

$$\mathsf{M}(X + Y) = \mathsf{M}(X) + \mathsf{M}(Y).$$

Доказательство. Сначала докажем это свойство для дискретных величин X и Y с совместным распределением

$$\mathsf{P}(X = x_i, Y = y_j) = p_{ij} > 0, \quad \sum_{i,j} p_{ij} = 1,$$

и маргинальными распределениями $\mathsf{P}(X = x_i) = p_i$, $\mathsf{P}(Y = y_j) = q_j$, где $p_i = \sum_j p_{ij}$, $q_j = \sum_i p_{ij}$. Для этого положим в формуле (5) $g(x, y) = x + y$ и воспользуемся абсолютной сходимостью соответствующих рядов. Получаем

$$\mathsf{M}(X+Y)=\sum_{i,j}(x_i+y_j)p_{ij}=\sum_i x_i\sum_j p_{ij}+\sum_j y_j\sum_i p_{ij}=\sum_i x_i p_i+\sum_j y_j q_j=\mathsf{M}(X)+\mathsf{M}(Y).$$

Теперь проведем доказательство для абсолютно непрерывных величин X и Y с совместной плотностью $p(x,y)$ и маргинальными плотностями $p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$ и $p_Y(y)=\int_{-\infty}^{+\infty}p(x,y)dx$. Используя формулу (6) при $n=2$, можно было бы доказать нужное свойство по аналогии с предыдущим пунктом. Но, поскольку эта формула была выведена лишь для $n=1$, мы проведем доказательство, основанное на другой идее. В §5 гл. V было показано, что плотность вероятности суммы $Z=X+Y$ равна $p_Z(z)=\int_{-\infty}^{+\infty}p(x,z-x)dx$. Поэтому, теорема будет доказана, если мы убедимся, что из существования $\mathsf{M}X$ и $\mathsf{M}Y$ следует абсолютная сходимость интеграла

$$\mathsf{M}(X+Y)=\int_{-\infty}^{+\infty}zp_Z(z)dz=\int_{-\infty}^{+\infty}z\left(\int_{-\infty}^{+\infty}p(x,z-x)dx\right)dz,$$

и установим равенство

$$\mathsf{M}(X)+\mathsf{M}(Y)=\int_{-\infty}^{+\infty}z\left(\int_{-\infty}^{+\infty}p(x,z-x)dx\right)dz.$$

Действительно, учитывая абсолютную сходимость интегралов для $\mathsf{M}X$ и $\mathsf{M}Y$, меняя порядок интегрирования и выполняя замену переменной $y=z-x$, $dz=dy$, найдем

$$\mathsf{M}(X)+\mathsf{M}(Y)=\int_{-\infty}^{+\infty}xp_X(x)dx+\int_{-\infty}^{+\infty}yp_Y(y)dy=\int_{-\infty}^{+\infty}x\left(\int_{-\infty}^{+\infty}p(x,y)dy\right)dx+\int_{-\infty}^{+\infty}y\left(\int_{-\infty}^{+\infty}p(x,y)dx\right)dy=$$

$$=\int_{-\infty}^{+\infty}dx\left(\int_{-\infty}^{+\infty}xp(x,y)dy+\int_{-\infty}^{+\infty}yp(x,y)dy\right)=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}(x+y)p(x,y)dy=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}zp(x,z-x)dz.$$

Теорема полностью доказана.

З а м е ч а н и е. Свойства 2) и 3) можно объединить в одно следующим образом: если математические ожидания компонент случайного вектора (X,Y) существуют, то для любых постоянных чисел a и b справедливо равенство

$$\mathsf{M}(aX+bY)=a\mathsf{M}(X)+b\mathsf{M}(Y).$$

Свойство, выражаемое этим равенством, называют *линейностью* математического ожидания.

4) *Математическое ожидание произведения двух независимых случайных величин равно произведению их математических ожиданий*

$$\mathsf{M}(XY)=\mathsf{M}(X)\mathsf{M}(Y).$$

Д о к а з а т е л ь с т в о. Снова будем доказывать свойство сначала для дискретных величин X и Y с описанным в п. 3) совместным распределением. Полагая в формуле (5) $g(x,y)=xy$ и учитывая независимость X и Y, найдем

$$\mathsf{M}(XY)=\sum_{i,j}x_i y_j p_{ij}=\sum_{i,j}x_i y_j p_i q_j=\sum_i x_i p_i\sum_j y_j q_j=\mathsf{M}(X)\mathsf{M}(Y).$$

Пусть теперь X и Y – абсолютно непрерывные величины с совместной плотностью $p(x,y)$ и маргинальными плотностями $p_X(x)=\int_{-\infty}^{+\infty}p(x,y)dy$ и

$p_Y(y)=\int_{-\infty}^{+\infty} p(x,y)dx$. Для плотности произведения $Z = XY$ была получена формула

$$p_Z(z)=\int_{-\infty}^{+\infty} p\left(x,\frac{z}{x}\right)\frac{dx}{|x|}.$$

В силу независимости X и Y совместная плотность равна произведению индивидуальных плотностей $p(x,y)=p_X(x)p_Y(y)$. Находим

$$\mathsf{M}(XY)=\int_{-\infty}^{+\infty} zp_Z(z)dz=\int_{-\infty}^{+\infty} zdz\int_{-\infty}^{+\infty} p\left(x,\frac{z}{x}\right)\frac{dx}{|x|}=\int_{-\infty}^{+\infty} zdz\int_{-\infty}^{+\infty} p_X(x)p_Y\left(\frac{z}{x}\right)\frac{dx}{|x|}=$$

$$=\int_{-\infty}^{+\infty} p_X(x)\frac{dx}{|x|}\int_{-\infty}^{+\infty} zp_Y\left(\frac{z}{x}\right)dz=\int_{-\infty}^{+\infty} p_X(x)\frac{dx}{|x|}x^2\int_{-\infty}^{+\infty}\frac{z}{x}p_Y\left(\frac{z}{x}\right)d\left(\frac{z}{x}\right)=$$

$$=\int_{-\infty}^{+\infty} xp_X(x)dx\operatorname{sgn}x\int_{-\infty}^{+\infty}\frac{z}{x}p_Y\left(\frac{z}{x}\right)d\left(\frac{z}{x}\right)=\int_{-\infty}^{+\infty} xp_X(x)dx\int_{-\infty}^{+\infty} yp_Y(y)dy=\mathsf{M}X\,\mathsf{M}Y.$$

Здесь мы использовали соотношение $x/|x|=\operatorname{sgn}x$, $x\neq 0$, где, как обычно,

$$\operatorname{sgn}x=\begin{cases}1, \text{если } x>0;\\ 0, \text{если } x=0;\\ -1, \text{если } x<0.\end{cases}$$

Заметим, что свойства 3) и 4) справедливы для произвольного конечного числа случайных величин (в последнем случае величины должны быть независимы в совокупности).

§ 2. Дисперсия

О п р е д е л е н и е. *Дисперсией случайной величины X называется математическое ожидание квадрата отклонения этой величины от ее математического ожидания*

$$\mathsf{D}(X)=\mathsf{M}[X-\mathsf{M}(X)]^2, \tag{10}$$

при условии, что указанные матожидания существуют.

Дисперсия служит мерой рассеяния значений случайной величины X относительно ее среднего значения $\mathsf{M}X$. Если дисперсия мала, то значения величины X с высокой вероятностью группируются вблизи ее среднего значения. При больших значениях дисперсии величина X с большой вероятностью может далеко отклоняться от $\mathsf{M}X$.

В некоторых случаях дисперсию удобнее вычислять по другой формуле, эквивалентной, формуле (10)

$$\mathsf{D}(X)=\mathsf{M}(X^2)-[\mathsf{M}(X)]^2, \tag{11}$$

Выведем формулу (11). Учитывая линейность математического ожидания и тот факт, что математическое ожидание $\mathsf{M}X$ и его квадрат $[\mathsf{M}(X)]^2$ – суть величины постоянные, можно записать:

$$\mathsf{D}(X)=\mathsf{M}[X-\mathsf{M}(X)]^2=\mathsf{M}\left[X^2-2X\mathsf{M}(X)+\mathsf{M}^2(X)\right]=$$

$$=\mathsf{M}(X^2)-2\mathsf{M}(X)\mathsf{M}(X)+\mathsf{M}^2(X)=\mathsf{M}(X^2)-[\mathsf{M}(X)]^2.$$

Учитывая одномерный вариант формулы (5) при $g(x)=(x-\mathsf{M}(x))^2$ или $g(x)=x^2$, равенства (10) и (11) для дискретной случайной величины можно записать в виде

$$\mathsf{D}X=\sum_i (x_i-\mathsf{M}X)^2 p_i\,,$$

$$\mathsf{D}X=\sum_i x_i^2 p_i-[\mathsf{M}(X)]^2.$$

В случае абсолютно непрерывной случайной величины формулы (10) и (11) примут вид

$$\mathsf{D}X=\int_{-\infty}^{\infty}(x-\mathsf{M}X)^2 p(x)dx\,,$$

$$\mathsf{D}X=\int_{-\infty}^{\infty}x^2 p(x)dx-[\mathsf{M}(X)]^2.$$

Для обоснования этих соотношений достаточно применить формулу (8), где снова $g(x)=(x-\mathsf{M}(x))^2$ или, соответственно, $g(x)=x^2$.

2.1. Свойства дисперсии

1) *Дисперсия постоянной величины равна нулю*

$$\mathsf{D}(C)=0.$$

2) *Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат*

$$\mathsf{D}(CX)=C^2\mathsf{D}(X).$$

Эти свойства следуют из определения дисперсии и линейности математического ожидания.

3) *Прибавление к случайной величине X любой постоянной C не меняет значение дисперсии*

$$\mathsf{D}(X+C)=\mathsf{D}(X).$$

Доказательство. Действительно,

$$\mathsf{D}(X+C)=\mathsf{M}[X+C-\mathsf{M}(X+C)]^2=\mathsf{M}[X+C-\mathsf{M}(X)-C]^2=\mathsf{M}[X-\mathsf{M}(X)]^2=\mathsf{D}(X).$$

4) *Дисперсия суммы двух независимых случайных величин с конечными дисперсиями равна сумме дисперсий этих величин*

$$\mathsf{D}(X+Y)=\mathsf{D}(X)+\mathsf{D}(Y).$$

Для доказательства свойства 4) достаточно воспользоваться определением дисперсии и свойствами 4) и 5) математического ожидания.

5) *Дисперсия разности двух независимых случайных величин с конечными дисперсиями равна сумме дисперсий этих величин*

$$\mathsf{D}(X-Y)=\mathsf{D}(X)+\mathsf{D}(Y)$$

Справедливость этого равенства вытекает из свойств дисперсии 2) и 4).

Определение. *Средним квадратическим отклонением случайной величины X с конечной дисперсией* $\mathsf{D}(X)$ *называется число*

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

6) *Среднее квадратическое отклонение суммы независимых случайных величин равно квадратному корню из суммы квадратов средних квадратических отклонений этих величин*

$$\sigma(X+Y)=\sqrt{\sigma^2(X)+\sigma^2(Y)}\,.$$

Для доказательства достаточно учесть, что $\sigma^2(X)=\mathsf{D}(X)$ и применить свойство 4).

7) *Дисперсия суммы двух произвольных случайных величин с конечными дисперсиями равна*

$$\mathsf{D}(X+Y)=\mathsf{D}(X)+\mathsf{D}(Y)+2\operatorname{cov}(X,Y),$$

где $\operatorname{cov}(X,Y)=\mathsf{M}[(X-\mathsf{M}X)(Y-\mathsf{M}Y)]$ *– т.н. ковариация величин* X *и* Y.

Доказательство. Имеем

$$\mathsf{D}(X+Y)=\mathsf{M}[(X+Y)-\mathsf{M}(X+Y)]^2=$$
$$=\mathsf{M}\left[(X+Y)^2-2(X+Y)(\mathsf{M}X+\mathsf{M}Y)-(\mathsf{M}X+\mathsf{M}Y)^2\right]=$$
$$=\mathsf{M}\left[X^2-2X\mathsf{M}X+(\mathsf{M}X)^2+Y^2-2Y\mathsf{M}Y+(\mathsf{M}Y)^2+2X(Y-\mathsf{M}Y)-2\mathsf{M}X(Y-\mathsf{M}Y)\right]=$$
$$=\mathsf{M}\left[(X-\mathsf{M}X)^2+(Y-\mathsf{M}Y)^2+2(X-\mathsf{M}X)(Y-\mathsf{M}Y)\right]=\mathsf{D}(X)+\mathsf{D}(Y)+2\operatorname{cov}(X,Y).$$

Заметим без доказательства, что существование математического ожидания $\mathsf{M}[(X-\mathsf{M}X)(Y-\mathsf{M}Y)]$ следует из конечности значений $\mathsf{D}(X)$ и $\mathsf{D}(Y)$.

§ 3. Числовые характеристики совместного распределения

3.1. Ковариация

Определение. *Пусть* (X,Y) *– дискретный или абсолютно непрерывный вектор. Величина* $\operatorname{cov}(X,Y)=\mathsf{M}[(X-\mathsf{M}X)(Y-\mathsf{M}Y)]$ *называется ковариацией величин* X *и* Y, *при условии, что все входящие в формулу математические ожидания существуют.*

Положив в определении ковариации $X=Y$, получим

$$\operatorname{cov}(X,X)=\mathsf{D}X\,.$$

Подобно тому, как это было сделано для дисперсии, легко проверить, что формулу ковариации можно записать в виде

$$\operatorname{cov}(X,Y)=\mathsf{M}(XY)-\mathsf{M}X\,\mathsf{M}Y\,.$$

Для дискретных величин ковариация вычисляется по формулам

$$\operatorname{cov}(X,Y)=\sum_i\sum_j \mathsf{M}\left[(x_i-\mathsf{M}X)(y_j-\mathsf{M}Y)\right]p_{ij}$$

или

$$\operatorname{cov}(X,Y)=\sum_i\sum_j x_i y_j p_{ij}-\mathsf{M}X\,\mathsf{M}Y\,.$$

Для абсолютно непрерывных случайных величин – по формулам

$$\operatorname{cov}(X,Y)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x-\mathsf{M}X)(y-\mathsf{M}Y)\,dxdy$$

или

$$\operatorname{cov}(X,Y)=\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}xy\,dxdy-\mathsf{M}X\,\mathsf{M}Y\,.$$

Определение. *Случайные величины X и Y, для которых верно равенство $\operatorname{cov}(X,Y)=0$ называются некоррелированными.*

Теорема. *Независимые случайные величины X и Y некоррелированы.*

Доказательство. Используя свойства математического ожидания, найдем

$$\operatorname{cov}(X,Y)=\mathsf{M}[(X-\mathsf{M}X)(Y-\mathsf{M}Y)]=\mathsf{M}(X-\mathsf{M}X)\mathsf{M}(Y-\mathsf{M}Y)=$$
$$=(\mathsf{M}X-\mathsf{M}X)(\mathsf{M}Y-\mathsf{M}Y)=0.$$

Замечание. Отметим без доказательства, что некоррелированные случайные величины могут быть зависимыми. Таким образом, из свойства 7) следует свойство 4), но не наоборот.

Теорема. *Для любых постоянных a_1, a_2, b_1, b_2,справедливо равенство*

$$\operatorname{cov}(a_1X+b_1,a_2Y+b_2)=a_1a_2\operatorname{cov}(X,Y). \qquad (12)$$

Доказательство. Имеем

$$\operatorname{cov}(a_1X+b_1,a_2Y+b_2)=\mathsf{M}[(a_1X+b_1-\mathsf{M}(a_1X+b_1))(a_2Y+b_2-\mathsf{M}(a_2Y+b_2))]=$$
$$=\mathsf{M}[(a_1X+b_1-a_1\mathsf{M}(X)+b_1)(a_2Y+b_2-a_2\mathsf{M}(Y)+b_2)]=a_1a_2\mathsf{M}[(X-\mathsf{M}X)(Y-\mathsf{M}Y)].$$

Замечание. Ковариация служит характеристикой взаимозависимости случайных величин. В некоторых случаях с ее помощью можно установить факт зависимости величин X и Y, даже не зная их совместного распределения. Действительно, если из каких-либо соображений известно, что $\operatorname{cov}(X,Y)\neq 0$, то величины X и Y заведомо зависимы.

3.2. Коэффициент корреляции

Существенным недостатком ковариации как меры взаимозависимости случайных величин является влияние на нее единиц измерения этих величин. Действительно, переход к другим единицам измерения равносилен умножению величины на некоторый положительный коэффициент. Но, как видно из (12), при умножении любой из величин на постоянное число, ковариация также умножается на это число. Данный недостаток устраняется, если снабдить ковариацию подходящим нормирующим множителем. Так возникает следующее понятие.

Определение. *Коэффициентом корреляции случайных величин* X *и* Y*, для которых* $0 < \mathsf{D}(X) < +\infty$, $0 < \mathsf{D}(Y) < +\infty$*, называется число*

$$r(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathsf{D}(X)\mathsf{D}(Y)}}.$$

Определение. *Говорят, что случайные величины* X *и* Y *положительно (отрицательно) коррелированы, если* $r(X,Y) > 0$ *(*$r(X,Y) < 0$*).*

Теорема. *Коэффициент корреляции обладает свойствами:*

1) *если случайные величины* X *и* Y *некоррелированы (и, тем более, независимы), то* $r(X,Y) = 0$;

2) $-1 \le r(X,Y) \le 1$;

3) $|r(X,Y)| = 1$ *тогда и только, когда случайные величины* X *и* Y *связаны линейной функциональной зависимостью* $Y = kX + b$*. При этом* $r(X,Y) = 1 \Leftrightarrow k > 0$ *и* $r(X,Y) = -1 \Leftrightarrow k < 0$.

Доказательство. Свойство 1) очевидно.

Для доказательства 2) заметим, что в силу (12), не теряя общности, можно считать $\mathsf{M}X = \mathsf{M}Y = 0$. Тогда

$$r(X,Y) = \frac{\mathsf{M}(XY)}{\sqrt{\mathsf{M}(X^2)\mathsf{M}(Y^2)}}.$$

Рассмотрим неотрицательную функцию

$$g(t) = \mathsf{M}\left[(X + tY)^2\right] = \mathsf{M}(X^2) + 2t\mathsf{M}(XY) + t^2\mathsf{M}(Y^2) \ge 0.$$

Так как $g(t)$ – квадратный трехчлен, необходимым и достаточным условием ее неотрицательности будет неравенство $D = 4[\mathsf{M}(XY)]^2 - 4\mathsf{M}(X^2)\mathsf{M}(Y^2) \le 0$, которое эквивалентно следующему

$$|r(X,Y)| = \frac{|\mathsf{M}(XY)|}{\sqrt{\mathsf{M}(X^2)\mathsf{M}(Y^2)}} \le 1.$$

Свойство 3) докажем только в части достаточности. Пусть $Y = kX + b$, $k \ne 0$. Тогда, учитывая свойства дисперсии и равенство (12), найдем

$$r(X,Y) = \frac{\operatorname{cov}(X, kX + b)}{\sqrt{\mathsf{D}(X)\mathsf{D}(kX + b)}} = \frac{\operatorname{cov}(X, kX + b)}{\sqrt{\mathsf{D}(X)\mathsf{D}(kX + b)}} = \frac{k\operatorname{cov}(X, X)}{|k|\sqrt{\mathsf{D}(X)\mathsf{D}(X)}} = \operatorname{sgn}(k).$$

Замечание. Коэффициент корреляции достигает наибольшего по модулю значения $|r(X,Y)| = 1$, если между величинами X и Y существует наиболее сильный вид взаимосвязи – линейная функциональная зависимость $Y = kX + b$. При этом значению $r(X,Y) = 1$ соответствует положительный коэффициент k, так что росту (уменьшению) величины X будет соответствовать увеличение (уменьшение) величины Y. Если $r(X,Y)=-1$, то ситуация будет противоположной: чем больше X, тем меньше Y и наоборот. Аналогичные соотношения между X и Y сохраняются и при $|r(X,Y)|<1$. В этом случае зависимость между величинами может быть либо функциональной нелинейной, либо только статистической.

§ 4. Вектор математический ожиданий и матрица ковариаций

Многомерными аналогами математического ожидания и дисперсии являются вектор математический ожиданий и матрица ковариаций. Дадим соответствующие определения.

Определение. *Пусть* $\vec{X}=(X_1,X_2,...X_n)$ *– случайный вектор с известным многомерным распределением и* $\mathsf{M}X_i, i=1,...,n$, *– математические ожидания его компонент. Математическим ожиданием вектора* $\vec{X}$ *называется вектор-столбец* $\mathsf{M}\vec{X}=(\mathsf{M}X_1,\mathsf{M}X_2,...,\mathsf{M}X_n)^{\mathrm{T}}$ (*здесь верхний индекс* «Т» *– символ транспонирования*).

Свойства вектора $\mathsf{M}\vec{X}$ аналогичны свойствам обычного математического ожидания.

Определение. *Пусть* $\vec{X}=(X_1,X_2,...X_n)$ *– случайный вектор с известным многомерным распределением. Матрицей ковариаций вектора* $\vec{X}$ *называется матрица*

$$V(\vec{X})=\begin{pmatrix} \mathrm{cov}(X_1,X_1) & \mathrm{cov}(X_1,X_2) & \dots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \mathrm{cov}(X_2,X_2) & \dots & \mathrm{cov}(X_2,X_n) \\ \dots & \dots & \dots & \dots \\ \mathrm{cov}(X_n,X_1) & \mathrm{cov}(X_n,X_2) & \dots & \mathrm{cov}(X_n,X_n) \end{pmatrix}.$$

Матрица $V(\vec{X})$ – симметрическая, т.к. $\mathrm{cov}(X_i,X_j)=\mathrm{cov}(X_j,X_i)$ и на ее главной диагонали расположены дисперсии $\mathsf{D}X_i=\mathrm{cov}(X_i,X_i), i=1,...,n$. Если компоненты вектора $\vec{X}$ попарно некоррелированы (тем более, независимы), то матрица $V(\vec{X})$ будет диагональной.

Напомним, что квадратная матрица A называется *положительно (неотрицательно) определенной*, если для любого вектора $\vec{x}=(x_1,x_2,...,x_n)^{\mathrm{T}}\neq\vec{0}$ справедливо неравенство

$$\vec{x}^{\mathrm{T}}A\vec{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j>0 \quad (\vec{x}^{\mathrm{T}}A\vec{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}x_ix_j\geq 0).$$

Следующее свойство матрицы $V(\vec{X})$ обобщает на многомерный случай свойство дисперсии $\mathsf{D}X\geq 0$. Мы приведем его без доказательства.

Теорема. *Матрица ковариаций любого случайного вектора является неотрицательно определенной.*

Глава V. ОСНОВНЫЕ РАСПРЕДЕЛЕНИЯ И ИХ ЧИСЛОВЫЕ ХАРАКТЕРИСТИКИ

§ 1. Вырожденное распределение

Говорят, что случайная величина имеет вырожденное распределение, сосредоточенное в точке $C \in \mathbb{R}$, если

$$\mathsf{P}\{X = C\} = 1.$$

Вырожденное распределение задает неслучайные величины (константы). Для величины X такого типа математическое ожидание и дисперсия, соответственно, равны $\mathsf{M}X = C \cdot 1 = C$, $\mathsf{D}X = \mathsf{M}(X - \mathsf{M}X)^2 = \mathsf{M}(C - C)^2 = 0$.

§ 2. Распределение Бернулли

Распределение Бернулли – это распределение случайной величины X, которая равна 1 при наступлении «успеха» A в схеме Бернулли и 0 при наступлении «неудачи»:

$\mathsf{P}(A) = p, \mathsf{P}(\overline{A}) = q,$ где $p, q > 0,\ p + q = 1.$ Тогда

$$\mathsf{M}X = 1 \cdot p + 0 \cdot q = p,$$

$$\mathsf{D}X = (1-p)^2 \cdot p + (0-p)^2 \cdot q = q^2 p + p^2 q = pq(q+p) = pq.$$

Таким образом, для величины X, распределенной по закону Бернулли, имеем

$$\mathsf{M}X = p,\ \mathsf{D}X = pq.$$

§ 3. Биномиальное распределение

Биномиальное распределение – это распределение случайной величины X, равной числу k «успехов» A в серии из n испытаний по схеме Бернулли:

$$P_n(k) := \mathsf{P}(X = k) = C_n^k p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n,$$

где

$$p = \mathsf{P}(A),\ C_n^k = \frac{n!}{k!(n-k)!},\ \ n! = 1 \cdot 2 \cdots n,\ \ 0! = 1.$$

Величина X есть сумма n независимых, одинаково распределенных по закону Бернулли величин $\{X_i\}_{i=1}^n$. Поэтому

$$\mathsf{M}X = \mathsf{M}(X_1 + X_2 + \ldots + X_n) = \mathsf{M}(X_1) + \mathsf{M}(X_2) + \ldots + \mathsf{M}(X_n) = np,$$

$$\mathsf{D}X = \mathsf{D}(X_1 + X_2 + \ldots + X_n) = \mathsf{D}(X_1) + \mathsf{D}(X_2) + \ldots + \mathsf{D}(X_n) = npq.$$

§ 4. *Геометрическое распределение*

Геометрическое распределение – это распределение случайной величины X, равной числу k испытаний до первого «успеха» A в серии из n испытаний по схеме Бернулли:

$$\mathsf{P}(X = k) = pq^{k-1},\ k = 0,1,\ldots,$$

где

$$p = \mathsf{P}(A),\ q = 1 - p.$$

$$\mathsf{M}X = \sum_{k=1}^{\infty} kpq^{k-1} = p\sum_{k=1}^{\infty}\left(q^k\right)'_q = p\left(\sum_{k=1}^{\infty} q^k\right)'_q =$$

$$= p\left(\frac{q}{1-q}\right)'_q = p\frac{1}{(1-q)^2} = \frac{1}{p}.$$

Находим так называемый *второй факториальный момент*:

$$\mathsf{M}X(X-1) = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = pq\sum_{k=0}^{\infty}\left(q^k\right)''_{qq} = pq\left(\sum_{k=0}^{\infty} q^k\right)''_{qq} =$$

$$= pq\left(\frac{q}{1-q}\right)''_{qq} = pq\frac{2}{(1-q)^3} = \frac{2q}{p^2}.$$

Выразим теперь дисперсию через второй факториальный момент:

$$\mathsf{D}X = \mathsf{M}X(X-1) + \mathsf{M}X - (\mathsf{M}X)^2 = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q - 1 + p}{p^2} = \frac{q}{p^2}.$$

§ 5. *Гипергеометрическое распределение*

Определение. *Случайная величина X имеет гипергеометрическое распределение с параметрами* n_1, n_2 *и* $n \le n_1 + n_2$ $(n, n_1, n_2$ *— натуральные числа), если она принимает конечное множество натуральных значений* $\{k_1, k_1+1, \ldots, k_2\}$ *с вероятностями*

$$P\{X = k\} = \frac{C_{n_1}^{k} C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}},\ k_1 \le k \le k_2, \qquad (1)$$

причём $k_1 = \max\{0; n - n_2\}$ $k_2 = \min\{n_1; n\}$.

Гипергеометрическое распределение соответствует схеме выбора без возвращения (в отличие от биномиального распределения). Пусть урна содержит $n_1 + n_2$ шаров, из которых n_1 белых и n_2 чёрных. Наугад извлекают n шаров. Тогда величина (1) определяют вероятность того, что среди этих n шаров ровно k белых. На практике к гипергеометрическому распределению приводят задачи, где изделия из партии отбирают случайно (обеспечивая для каждого изделия равную возможность быть отобранным), но отобранные изделия не возвращают в партию. Такой отбор особенно важен в тех задачах, где проверка изделия связана с его разрушением (например, проверка изделия на срок службы).

Математическое ожидание и дисперсия гипергеометрического распределения равны (доказательство не приводим):

$$\mathsf{M}X = \frac{nn_1}{n_1 + n_2}, \tag{2}$$

$$\mathsf{D}X = \frac{nn_1 n_2}{(n_1 + n_2)^2} \cdot \frac{n_1 + n_2 - n}{n_1 + n_2 - 1}. \tag{3}$$

Отметим, что если $n_1 + n_2$ очень велико по сравнению с n (практически, когда $n_1 + n_2 > 10n$), то не имеет существенного значения, возвращаются шары обратно или нет. Формула (1) при этом может быть приближённо заменена формулой биномиального распределения

$$\mathsf{P}\{\tilde{X} = k\} = C_n^k p^k (1-p)^{n-k} = C_n^k p^k q^{n-k}, k = 0,1,2,\ldots,n \tag{4}$$

где $p = n_1/(n_1 + n_2)$ – доля белых шаров среди их общего количества в урне. С этим выводом согласуется и тот факт, что, что при неограниченном росте $N := n_1 + n_2$ и постоянном p, матожидание (2) и дисперсия (3) стремятся, соответственно, к np и $np(1-p)$. Действительно,

$$\lim_{N\to\infty} \frac{nn_1}{n_1 + n_2} = \lim_{N\to\infty} \frac{nNp}{N} = np,$$

$$\lim_{N\to\infty} \frac{nn_1 n_2}{(n_1 + n_2)^2} \cdot \frac{n_1 + n_2 - n}{n_1 + n_2 - 1} =$$

$$= \lim_{N\to\infty} \frac{nNpN(1-p)}{N^2} \cdot \frac{N-n}{N-1} = \lim_{N\to\infty} np(1-p) \cdot \frac{1-n/N}{1-1/N} = np(1-p)$$

При этом предельные значения оказываются математическим ожиданием $np = \mathsf{M}\tilde{X}$ и дисперсией $np(1-p) = \mathsf{D}\tilde{X}$ биномиального распределения (4).

§ 6. Распределение Пуассона

Распределение Пуассона имеет вид

$$\mathsf{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \lambda > 0, \; k = 0,1,\ldots$$

$$\mathsf{M}X = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} =$$

$$= \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{s=0}^{\infty}\frac{\lambda^{s-1}}{(s-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

Теперь снова находим второй факториальный момент:

$$\mathsf{M}X(X-1) = \sum_{k=0}^{\infty}k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!} =$$

$$= \lambda^2 e^{-\lambda}e^{\lambda} = \lambda^2.$$

Отсюда получим:

$$\mathsf{M}X^2 = \mathsf{M}X(X-1) + \mathsf{M}X = \lambda^2 + \lambda,$$

и

$$\mathsf{D}X = \mathsf{M}X^2 - (\mathsf{M}X)^2 = \lambda.$$

§ 7. Равномерное распределение

Случайная величина X *равномерно распределена* на отрезке $[a,b]$, если ее плотность вероятности имеет вид

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b]; \\ 0, & x \in (-\infty, a) \cup (b, +\infty). \end{cases}$$

Так как функция плотности вероятности равномерно распределенной величины X симметрична относительно прямой $x = \dfrac{a+b}{2}$ имеем

$$\mathsf{M}X = \frac{a+b}{2}.$$

Чтобы вычислить DX найдем так называемый *второй момент*

$$\mathsf{M}X^2 = \int_{-\infty}^{+\infty} x^2 p(x)dx = \int_a^b x^2\frac{1}{b-a}dx = \frac{b^3-a^3}{3(b-a)} =$$

$$= \frac{b^2+ab+a^2}{3}.$$

Отсюда

$$\mathsf{D}X = \mathsf{M}X^2 - (\mathsf{M}X)^2 = \frac{(b-a)^2}{12}.$$

§ 8. Показательное распределение

Определение. *Говорят, что абсолютно непрерывная случайная величина X распределена по экспоненициальному (показательному) закону,* *если ее плотность вероятности имеет вид*

$$p(x) = \begin{cases} 0, & x \le 0; \\ \lambda e^{-\lambda x}, & x > 0. \end{cases} \qquad (5)$$

Интегрируя (5), легко найти функцию распределения $F(x)$ показательного распределения

$$F(x) = \begin{cases} 0, & x \le 0; \\ 1 - e^{-\lambda x}, & x > 0. \end{cases}$$

Математическое ожидание и дисперсия случайной величины X, распределённой по экспоненциальному закону, соответственно равны

$$\mathsf{M}X = \frac{1}{\lambda}, \qquad (6)$$

$$\mathsf{D}X = \frac{1}{\lambda^2}.$$

Выведем эти формулы. Интегрируя по частям, находим

$$\mathsf{M}X = \int_{-\infty}^{\infty} xp(x)dx = \lambda \int_{0}^{\infty} xe^{-\lambda x}dx = -\int_{0}^{\infty} xde^{-\lambda x} =$$

$$= -xe^{-\lambda x}\Big|_0^{\infty} + \int_{0}^{\infty} de^{-\lambda x} = 0 - \frac{1}{\lambda}\int_{0}^{\infty} de^{-\lambda x} =$$

$$= -\frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty} = -\frac{1}{\lambda}(0-1) = \frac{1}{\lambda}. \qquad (7)$$

Дисперсию вычислим по формуле

$$\mathsf{D}X = \mathsf{M}\left(X^2\right) - (\mathsf{M}X)^2.$$

Снова интегрируя по частям и учитывая найденное значение интеграла (7), вычислим второй момент

$$\mathsf{M}\left(X^2\right) = \int_{-\infty}^{\infty} x^2 p(x)dx = \lambda \int_{0}^{\infty} x^2 e^{-\lambda x}dx = -\int_{0}^{\infty} x^2 de^{-\lambda x} =$$

$$= -x^2 e^{-\lambda x}\Big|_0^{\infty} + 2\int_{0}^{\infty} xe^{-\lambda x}dx = 0 + \frac{2}{\lambda^2} = \frac{2}{\lambda^2}.$$

Отсюда

$$\mathsf{D}(X) = \mathsf{M}\left(X^2\right) - (\mathsf{M}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Экспоненциальное (показательное) распределение часто встречается в теории массового обслуживания (например, X – время ожидания при техническом обслуживании или X – продолжительность телефонных разговоров, ежедневно регистрируемых на телефонной станции) и в теории надёжности (например, X – срок службы радиоэлектронной аппаратуры).

§ 9. Нормальное распределение

Говорят, что случайная величина X *нормально* распределена (обозначение: $X \sim N_{a,\sigma}$), если ее плотность вероятности равна

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, a \in \mathbb{R},\ \sigma > 0. \tag{8}$$

Нормальное распределение имеет фундаментальное значение для естествознания, техники, социальных и экономических наук (см. § 4 гл. VI).

Вычислим $\mathsf{M}X$ и $\mathsf{D}X$. Сначала найдем $\mathsf{M}X_0$ и $\mathsf{D}X_0$ величины X_0, имеющей так называемое *стандартное нормальное* распределение, $X_0 \sim N_{0,1}$:

$$\mathsf{M}X_0 = \int_{-\infty}^{+\infty} xp(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{x^2}{2}} dx = 0.$$

Последний интеграл равен нулю т.к. он сходится и подынтегральная функция является нечетной. Далее, интегрируя по частям, находим

$$\mathsf{M}X_0^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^2 e^{-\frac{x^2}{2}} dx =$$

$$-\frac{2}{\sqrt{2\pi}} \int_0^{+\infty} xde^{-\frac{x^2}{2}} = -\frac{2x}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \Big|_0^{+\infty} + \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} e^{-\frac{x^2}{2}} dx =$$

$$= 0 + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\left(\frac{x}{\sqrt{2}}\right)^2} d\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-t^2} dt = 1.$$

Здесь мы использовали известную из курса высшей математики формулу $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$. Таким образом,

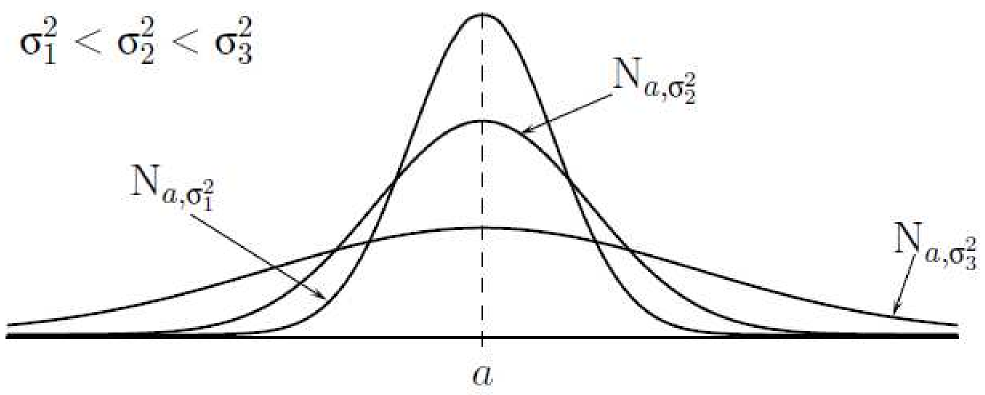
$$\mathsf{D}X_0 = \mathsf{M}X_0^2 - (\mathsf{M}X_0)^2 = 1 - 0 = 1.$$

Пусть теперь $X \sim N_{a,\sigma}$. Тогда $X_0 = \dfrac{X-a}{\sigma} \sim N_{0,1}$. Поскольку $\mathsf{M}X_0 = 0$ и $\mathsf{D}X_0 = 1$, из свойств матожидания и дисперсии получаем

$$\mathsf{M}X = \mathsf{M}(\sigma X_0 + a) = \sigma \mathsf{M}X_0 + a = 0 + a = a,$$

$$\mathsf{D}X = \mathsf{D}(\sigma X_0 + a) = \sigma^2 \mathsf{D}X_0 = \sigma^2.$$

Мы получили, что параметры a и σ^2 нормального распределения являются, соответственно, его математическим ожиданием и дисперсией.



На рисунке показаны графики плотности вероятности нормального распределения при различных значениях параметра σ^2 (дисперсии). Можно заметить, что меньшей дисперсии соответствует более узкий и высокий «пик» графика. Это связано с тем, что в точке a функция (8) имеет максимум, равный $1/\sqrt{2\pi}\sigma$, а площадь под графиком любой плотности равна 1. Изменение параметра a приводит к сдвигу графика вдоль оси абсцисс, не изменяя его форму.

Функция распределения для $X_0 \sim N_{0,1}$ имеет вид

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt,$$

в общем случае функция распределения $F(x;a,\sigma)$ величины $X \sim N_{a,\sigma}$ вычисляется по формуле

$$F(x;a,\sigma) = \Phi\left(\frac{x-a}{\sigma}\right).$$

Для функции $\Phi(x)$ составлены подробные таблицы. Отметим, что в силу равенства $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$, функция $\Phi(x)$ и введенный нами ранее (см. п. 7.2 гл. I) интеграл вероятностей $\Phi_0(x)$ связаны соотношением $\Phi(x) = \frac{1}{2} + \Phi_0(x)$. Вычислим вероятность того, что нормально распределенная величина X отклонится от своего среднего значения a более, чем на $k\sigma, k > 0$. Имеем

$$\mathsf{P}(|X-a| > k\sigma) = 1 - \mathsf{P}(a - k\sigma \le X \le a + k\sigma) = 1 - \Phi(k) + \Phi(-k) = 1 - 2\Phi_0(k).$$

Здесь мы учли нечетность функции $\Phi_0(x)$. Найденная вероятность не зависит от σ и быстро убывает с ростом k. Например, при $k = 3$ указанная вероятность приблизительно равна 0,0026998. С этим фактом связано так называемое *«правило трех сигма»*: *событие, состоящее в том, что нормально распределенная величина X отклонится от своего среднего значения $\mathsf{M}X = a$ более, чем на 3σ, является практически невозможным.* Это правило широко используется в различных прикладных областях, связанных с применением теории вероятностей.

§ 10. Многомерное нормальное распределение

Определение. *Говорят, что случайный вектор*

$$\vec{X} = (X_1,\dots,X_n)^{\mathrm{T}} : \Omega \to \mathbb{R}^n$$

имеет многомерное нормальное распределение, если существует вектор $\vec{\mu} \in \mathbb{R}^n$ *и положительно определённая симметрическая матрица* Σ *размера* $n \times n$, *такие что плотность вероятности вектора* $\vec{X}$ *имеет вид:*

$$p(\vec{x}) = \frac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{x}-\vec{\mu})}, \; \vec{x} = (x_1, x_2, \dots, x_n), \qquad (9)$$

где $|\Sigma|$ – определитель матрицы Σ, а Σ^{-1} – матрица, обратная к Σ.

В случае $n=1$, многомерное нормальное распределение сводится к нормальному распределению с плотностью

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

Если случайный вектор $\vec{X}$ имеет многомерное нормальное распределение, то пишут $\vec{X} \sim N(\vec{\mu}, \Sigma)$. Многомерное нормальное распределение имеет многочисленные приложения в математической статистике, регрессионном анализе и других дисциплинах, использующих стохастические (вероятностные) методы.

Выражение в показателе экспоненты является квадратичной формой от переменных $(x_i - \mu_i)$. Если обозначить элементы матрицы Σ^{-1} через s_{ij}, то можно записать

$$(\vec{x}-\vec{\mu})^{\mathrm{T}}\Sigma^{-1}(\vec{x}-\vec{\mu}) = \sum_{i=1}^{n}\sum_{j=1}^{n} s_{ij}(x_i-\mu_i)(x_j-\mu_j). \qquad (10)$$

Рассмотрим важный частный случай. Пусть матрица Σ – диагональная, на главной диагонали которой расположены элементы $\sigma_1^2,\dots,\sigma_n^2$ (в этом случае будем писать $\Sigma = \mathrm{diag}\{\sigma_1^2,\dots,\sigma_n^2\}$). Обратная матрица примет вид $\Sigma^{-1} = \mathrm{diag}\{\sigma_1^{-2},\dots,\sigma_n^{-2}\}$, кроме того, в силу (10)

$$(\vec{x}-\mu)^{\mathrm{T}}\Sigma^{-1}(\vec{x}-\mu) = \sum_{i=1}^{n}\sigma_i^{-2}(x_i-\mu_i)^2$$

и $|\Sigma| = \sigma_1^2 \cdots \sigma_n^2$. Подставляя эти выражения в (8), получим

$$p(x_1,x_2,\dots,x_n) = \frac{1}{(2\pi)^{n/2}\sigma_1\cdots\sigma_n}\exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\sigma_i^{-2}(x_i-\mu_i)^2\right\} =$$

$$= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x_1-\mu_1)^2}{2\sigma_1^2}\right\}\cdots\frac{1}{\sqrt{2\pi}\sigma_n}\exp\left\{-\frac{(x_n-\mu_n)^2}{2\sigma_n^2}\right\}. \qquad (11)$$

Мы видим, что совместная плотность разлагается в произведение одномерных нормальных плотностей вида $\frac{1}{\sqrt{2\pi}\sigma_i}\exp\{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\}$. Убедимся, что эти функции совпадают с маргинальными плотностями $p_i(x_i)$ величин X_i. Напомним, что для получения частной плотности $p_i(x_i)$ необходимо проинтегрировать совместную плотность (11) по всем «лишним» переменным $x_j,\ j=1,\ldots,n,\ j\neq i$. Учитывая, что при любых μ_j и σ_j

$$\frac{1}{\sqrt{2\pi}\sigma_j}\int_{-\infty}^{+\infty}\exp\left\{-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}\right\}dx_j=1,$$

получим

$$p_i(x_i)=\int_{-\infty}^{+\infty}\ldots\int_{-\infty}^{+\infty}p(x_1,x_2,\ldots,x_n)dx_1\ldots dx_{i-1}dx_{i+1}\ldots dx_n=$$

$$=\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left\{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right\},$$

что и требовалось. Таким образом, при указанных условиях, совместная плотность нормального вектора представима в виде произведения маргинальных плотностей. Это означает, что компоненты вектора независимы. Мы доказали, что если матрица Σ является диагональной $\Sigma=\mathrm{diag}\{\sigma_1^2,\ldots,\sigma_n^2\}$, то компоненты $X_1,\ldots,X_n$ нормального вектора $\vec{X}\sim N(\vec{\mu},\Sigma)$ независимы и $X_i\sim N(\mu_i,\sigma_i^2)$. Следующую теорему приведем без доказательства.

Т е о р е м а. *Математическое ожидание вектора* $\vec{X}\sim N(\vec{\mu},\Sigma)$ *с плотностью* (9) *равно векторному параметру* $\vec{\mu}$, *а матрица ковариаций совпадает с матрицей* Σ

$$\mathsf{M}\vec{X}=\vec{\mu},\quad V\left(\vec{X}\right)=\Sigma.$$

Глава VI. ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 1. Различные типы сходимости последовательностей случайных величин

О п р е д е л е н и е. *Пусть случайные величины* $X,X_1,X_2,\ldots,X_n,\ldots$ *определены на одном и том же вероятностном пространстве* $(\Omega,\mathscr{A},\mathsf{P})$. *Говорят, что последовательность* $\{X_n\}_{n=1}^{\infty}$ *сходится к случайной величине* X *почти наверное, если событие*

$$A=\{\omega\in\Omega:\lim_{n\to\infty}X_n(\omega)=X(\omega)\}$$

имеет вероятность, равную единице: $\mathsf{P}(A)=1$.

Иначе говоря, последовательность сходится почти наверное, если событие B, состоящее из тех $\omega\in\Omega$, для которых сходимости нет,

$$\lim_{n\to\infty} X_n(\omega)\neq X(\omega)$$

имеет вероятность ноль: $\mathsf{P}(B)=0$. Если последовательность $\{X_n\}_{n=1}^{\infty}$ сходится к случайной величине X почти наверное, принято писать

$$\lim_{n\to\infty} X_n = X \text{ п.н. или } X_n\to X \text{ п.н при } n\to\infty.$$

В математическом анализе сходимости «почти наверное» соответствует понятие сходимости «почти всюду».

Если потребовать, чтобы вероятность множества тех элементарных исходов $\omega\in\Omega$, для которых $X_n(\omega)$ не попадает в ε-окрестность числа $X(\omega)$, стремилась к нулю с ростом n, то мы придем к определению сходимости «по вероятности». Отметим, что в математическом анализе такой вид сходимости называется сходимостью «по мере».

Определение. *Говорят, что последовательность случайных величин $\{X_n\}_{n=1}^{\infty}$ сходится по вероятности к случайной величине X, если для любого $\varepsilon>0$ выполняется равенство*

$$\lim_{n\to\infty}\mathsf{P}(|X_n-X|\geq\varepsilon)=0,$$

или, что то же самое,

$$\lim_{n\to\infty}\mathsf{P}(|X_n-X|<\varepsilon)=1.$$

Тот факт, что $\{X_n\}_{n=1}^{\infty}$ сходится к случайной величине X по вероятности принято обозначать

$$X_n\xrightarrow{p}X \textit{ при } n\to\infty \textit{ или } p\lim_{n\to\infty}X_n=X,$$

(здесь p от английского probability – вероятность).

Определение. *Говорят, что последовательность случайных величин $X_1,X_2,\ldots,X_n,\ldots$ с функциями распределения $F_{X_1}(x),F_{X_2}(x),\ldots,$ $F_{X_n}(x),\ldots$ сходится слабо (или по распределению) к случайной величине X с функцией распределения $F_X(x)$, если для любой точки x, в которой функция $F_X(x)$ непрерывна, имеет место равенство*

$$\lim_{n\to\infty}F_{X_n}(x)=F_X(x).$$

Сходимость по распределению обозначается как

$$X_n\xrightarrow{d}X \text{ при } n\to\infty \text{ или } d\lim_{n\to\infty}X_n=X.$$

(здесь d от английского distribution – распределение).

Замечание. Как всякая монотонная функция, любая функция распределения непрерывна всюду, за исключением, быть может, некоторого конечного или счетного множества. Поэтому множество точек непрерывности функции $F_X(x)$ заведомо непусто.

Замечание. Поскольку функция распределения не задает однозначно случайную величину, правильнее говорить не о сходимости по распределению последовательноти $\{X_n\}_{n=1}^{\infty}$ к случайной величине $X(\omega)$, а лишь о сходимости соответствующих распределений.

Приведем без доказательства два факта относительно взаимосвязи рассмотренных видов сходимости.

Теорема. *Из сходимости почти наверное следует сходимость по вероятности, а из сходимости по вероятности – сходимость по распределению.*

Теорема. *Если последовательность сходится по вероятности, то из нее можно выделить подпоследовательность, сходящуюся почти наверное.*

§ 2. Закон больших чисел в форме Чебышева

Сначала сформулируем и докажем важное неравенство, принадлежащее Чебышеву (Пафнутий Львович Чебышёв – выдающийся русский математик, 1821-1894).

Теорема (неравенство Чебышева). *Пусть случайная величина* $X:\Omega\to\mathbb{R}$ *определена на вероятностном пространстве* $(\Omega,\mathscr{A},\mathsf{P})$, *и её математическое ожидание* $\mathsf{M}X$ *и дисперсия* $\mathsf{D}X$ *конечны. Тогда*

$$\mathsf{P}(|X-\mathsf{M}X|\ge a)\le\frac{\mathsf{D}X}{a^2},$$

где $a>0$.

Доказательство. Докажем теорему в предположении, что X – абсолютно непрерывна и ее плотность вероятности есть $p(x)$. Пусть сначала $X\ge 0$. Тогда для любого $a>0$ верно неравенство

$$\mathsf{P}(X\ge a)\le\frac{\mathsf{M}X}{a}. \tag{1}$$

Действительно,

$$\mathsf{M}X=\int_0^{+\infty}xp(x)dx\ge\int_a^{+\infty}xp(x)dx\ge a\int_a^{+\infty}p(x)dx=a\mathsf{P}(X\ge a),$$

что эквивалентно (1). Пусть теперь X может принимать значения любого знака. Тогда, применяя неравенство (1) к неорицательной случайной величине $(X-\mathsf{M}X)^2$, получим

$$\mathsf{P}(|X-\mathsf{M}X|\ge a)=\mathsf{P}((X-\mathsf{M}X)^2\ge a^2)\le\frac{\mathsf{M}[(X-\mathsf{M}X)^2]}{a^2}=\frac{\mathsf{D}X}{a^2},$$

что и требовалось.

Для дискретной случайной величины доказательство проводится аналогично, только интеграл заменяется на сумму (конечную или бесконечную).

Теорема (закон больших чисел в форме Чебышева). *Для любой последовательности* $X_1,X_2,\ldots$ *попарно некоррелированных и одинаково распределенных случайных величин с математическим ожиданием* $\mathsf{M}X_n=m$ *и* *дисперсией* $\mathsf{D}X_n = \sigma^2$, $n = 1,2,\ldots$, *имеет место сходимость*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{p}} m.$$

Доказательство. Обозначим $S_n = X_1 + \ldots + X_n$. Из линейности математического ожидания получим

$$\mathsf{M}\left(\frac{S_n}{n}\right) = \frac{\mathsf{M}X_1 + \ldots + \mathsf{M}X_n}{n} = \frac{nm}{n} = m.$$

Зафиксируем произвольное $\varepsilon > 0$ и воспользуемся неравенством Чебышёва

$$\mathsf{P}\left(\left|\frac{S_n}{n} - \mathsf{M}\left(\frac{S_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{\mathsf{D}\left(\frac{S_n}{n}\right)}{\varepsilon^2} = \frac{\mathsf{D}S_n}{n^2\varepsilon^2} = \frac{\mathsf{D}X_1 + \ldots + \mathsf{D}X_n}{n^2\varepsilon^2} =$$

$$= \frac{n\sigma^2}{n^2\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to 0$$

при $n \to \infty$, так как $\mathsf{D}X_i = \sigma^2 < \infty, i = 1,2,\ldots$ Дисперсия суммы оказалась равной сумме дисперсий в силу попарной некоррелированности слагаемых. В силу произвольности ε сходимость по вероятности, а с ней и теорема, доказана.

Замечание. Закон больших чисел (ЗБЧ) утверждает, что среднее арифметическое большого числа случайных слагаемых "стабилизируется" с ростом этого числа. Как бы сильно каждая случайная величина ни отклонялась от своего среднего значения, при суммировании эти отклонения "взаимно гасятся", так что среднее арифметическое приближается к постоянной величине.

Замечание. Требование конечности второго момента (или дисперсии) связано со способом доказательства, утверждение (с незначительными изменениями) останется верным, если требовать существования только первого момента.

§ 3. Закон больших чисел в форме Бернулли

В качестве следствия из ЗБЧ Чебышёва получим закон больших чисел Бернулли.

Теорема (закон больших чисел в форме Бернулли). *Пусть событие* A *может произойти в любом из* n *независимых испытаний с одной и той же вероятностью* p, *и пусть* $\nu_n(A)$ – *число появлений события* A *в* n *испытаниях. Тогда*

$$\frac{\nu_n(A)}{n} \xrightarrow{p} p \text{ при } n \to \infty.$$

При этом для любого $\varepsilon > 0$ *выполняется неравенство*

$$\left(\left|\frac{\nu_n(A)}{n} - p\right| \geq \varepsilon\right) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Доказательство. Величина $\nu_n(A)$ равна сумме независимых, одинаково распределенных случайных величин, имеющих распределение Бернулли с параметром $p = \mathsf{P}(A)$: $\nu_n(A) = X_1 + \ldots + X_n$, где $X_i = 1$, если A произошло в i-м испытании и $X_i = 0$, если A не произошло в i-м испытании и $\mathsf{M}X_i = \mathsf{P}(A) = p$, $\mathsf{D}X_i = p(1-p)$. Осталось воспользоваться ЗБЧ в форме Чебышёва и полученным в ходе его доказательства неравенством

$$\left(\left|\frac{X_1 + \ldots + X_n}{n} - m\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Приведем без доказательства еще один (принадлежащий А. Н. Колмогорову) вариант закона больших чисел

Теорема (усиленный закон больших чисел). *При указанных выше условиях имеет место равенство*

$$\frac{X_1 + \ldots + X_n}{n} \to m \text{ п.н.}$$

§ 4. Центральная предельная теорема

Приведем без доказательства следующее утверждение, являющееся одним из вариантов т.н. центральной предельной теоремы.

Теорема. *Пусть $X_1, X_2, \ldots$ – независимые и одинаково распределенные случайные величины с матожиданием m дисперсией $0 < \sigma^2 < \infty$ и пусть, как и ранее, $S_n = X_1 + \ldots + X_n$. Тогда имеет место сходимость по распределению*

$$\frac{S_n - mn}{\sigma\sqrt{n}} \xrightarrow{\mathsf{d}} N_{0,1}$$

последовательности центрированных и нормированных сумм случайных величин $\frac{S_n - mn}{\sigma\sqrt{n}}$ к стандартному нормальному распределению. В частности, для любых вещественных $x < y$ при $n \to \infty$ верно предельное соотношение

$$\left(x \leq \frac{S_n - mn}{\sigma\sqrt{n}} \leq y\right) \to \frac{1}{\sqrt{2\pi}} \int_x^y e^{-\frac{t^2}{2}}\, dt.$$

Замечание. С качественной точки зрения, содержание ЦПТ состоит в следующем: *сумма большого числа независимых случайных величин, максимальная из которых мала по сравнению со всей суммой, имеет распределение, близкое к нормальному*. Многие важные для практики величины представляют собой именно такого вида суммы. Поэтому центральная предельная теорема объясняет то исключительное значение, которое имеет нормальное распределение для теории вероятностей и ее приложений.

ЛИТЕРАТУРА

1. Гнеденко Б.В. Курс теории вероятностей. М.: Эдиториал УРСС, 2005. – 448 с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. М.: Высш. шк., 2003. – 479 с.
3. Чистяков В.П. Курс теории вероятностей. М.: Наука, 1982. – 224 с.

СОДЕРЖАНИЕ

Новиков Владимир Васильевич

ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

Учебное пособие

Ответственный за выпуск В.В. Новиков
Оригинал-макет В.В. Новиков

Подписано в печать 25.05.2016г.
Формат 60x84 1/16. Бумага офсетная. Гарнитура Times New Roman. Печать RISO.
Объем 3,69 усл.печ.л. Тираж 50 экз. Заказ № 272

413100, Россия, Саратовская область, г. Энгельс, пл. Свободы, 17
ЭТИ (филиал) СГТУ имени Гагарина Ю.А.