Министерство общего и профессионального
образования Российской Федерации
Тверской государственный университет

Ю.С.ХОХЛОВ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие

Часть II

Тверь 1997

УДК 519.2

Хохлов Ю.С. Теория вероятностей и математическая статистика: Ч. II. Учебное пособие/ ТвГУ. — Тверь, 1997. — 75 с.

Пособие является второй частью курса лекций по теории вероятностей и математической статистике, в которой излагаются следующие разделы теории вероятностей: математическое ожидание случайной величины, числовые характеристики случайных величин, условные распределения и условные математические ожидания, многомерное нормальное распределение, закон больших чисел, центральная предельная теорема, цепи Маркова. Прилагаются задачи для самостоятельного решения.

Рекомендуется студентам математических специальностей, а также экономистам.

Библиогр. 10.

Рецензенты: кафедра математической статистики Московского госуниверситета им. М.В. Ломоносова;

доктор физико-математических наук В.В. Сенатов

Теория вероятностей

1.10 Математическое ожидание случайной величины

Как отмечалось ранее, наиболее полной характеристикой случайной величины ξ, которую можно получить из результатов эксперимента, является ее распределение P_ξ. Но сделать это реально довольно трудно, а часто и ненужно. Во многих задачах мы интересуемся только тем, около какой точки "в среднем " меняются значения случайной величины ξ и как далеко они (опять "в среднем") могут отклоняться от этого значения. Для этого в теории вероятностей используются понятия математического ожидания (как центра распределения) и дисперсии (для оценки отклонения от этого центра). Мы начнем изложение нашей теории с простейшего случая, когда ξ имеет дискретное распределение.

1.10.1 Математическое ожидание дискретной случайной величины

.

Прежде чем дать абстрактное определение, мы рассмотрим простой пример, который объяснит нам его разумность.

Пусть с.в. ξ принимает конечное множество значений $x_1, \ldots, x_n$ с вероятностями появления этих значений $p_k = P(\xi = x_k)$, $k = 1, \ldots, n$. Предположим, что мы провели N независимых измерений этой случайной величины ξ в одинаковых условиях и получили результаты $y_1, \ldots, y_N$. Пусть N_k есть число появлений значения x_k в этой серии измерений. Интуитивно мы представляем себе среднее в длинной серии измерений как среднее арифметическое. Тогда

$$\frac{y_1 + \ldots + y_N}{N} = x_1 \frac{N_k}{N} + \ldots + x_n \frac{N_n}{N} \approx x_1 p_1 + \ldots + x_n p_n \ ,$$

т.к. в длинной серии испытаний частоты "тяготеют" к вероятностям. Последнюю сумму естественно считать некоторым теоретическим средним или центром распределения случайной величины ξ.

Эти рассуждения приводят нас к следующему определению.

Определение 1 . **Математическим ожиданием (средним)** *дискретной случайной величины ξ называется число*

$$M\xi = \sum_n x_n p_n \ ,$$

где $\{x_1, x_2, \ldots\}$ – множество значений с.в. ξ, а $p_n = P(\xi = x_n)$, при условии, что последний ряд сходится абсолютно.

Иногда используют и другое обозначение для математического ожидания – $E\xi$. Абсолютная сходимость нужна для того, чтобы величина $M\xi$ не зависела от порядка нумерации значений с.в. ξ.

Посмотрим, какими свойствами обладают математические ожидания дискретных случайных величин. Для удобства сведем все эти свойства в следующее

Предложение 1 . *Математические ожидания дискретных случайных величин (если они существуют) обладают следующими свойствами.*

1. *Если $P(\xi = C) = 1$, то $M\xi = C$.*
2. *Если $C = const$, то $M(C\xi) = CM\xi$.*
3. *$M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$.*
4. *Если $\xi \geq 0$, то $M\xi \geq 0$.*
5. *Если $\xi_1 \geq \xi_2$, то $M\xi_1 \geq M\xi_2$.*
6. *Если ξ_1 и ξ_2 - независимы, то $M(\xi_1\xi_2) = M\xi_1 \cdot M\xi_2$.*

Доказательство. 1. Если $P(\xi = C) = 1$, то $M\xi = C \cdot 1 = C$.

2. Случайная величина $\eta = C\xi$ имеет множество значений

$$Y = \{Cx_1, Cx_2, \ldots\}$$

и

$$P(\eta = Cx_n) = p_n \ .$$

Тогда
$$M\eta = \sum_n (Cx_n)p_n = C\sum_n x_n p_n = CM\xi \ .$$

3. Пусть $X = \{x_1, \ldots, x_n, \ldots\}$ – множество значений с.в. ξ_1, $Y = \{y_1, \ldots, y_m, \ldots,\}$ - множество значений с.в. ξ_2. Тогда множеством значений с.в. $\eta = \xi_1 + \xi_2$ будет $Z = \{z_1, \ldots, z_k, \ldots\}$, где $z_k = x_n + y_m$ для некоторых x_n и y_m. Если $p_{nm} = P(\xi_1 = x_n, \xi_2 = y_m)$, то

$$q_k = P(\xi_1 + \xi_2 = z_k) = \sum_{x_n + y_m = z_k} p_{nm} \ .$$

Напомним, что

$$P(\xi_1 = x_n) = \sum_m p_{nm} \ , \ P(\xi_2 = y_m) = \sum_n p_{nm} \ .$$

Отсюда получаем

$$M(\xi_1 + \xi_2) = \sum_k z_k q_k = \sum_k z_k \sum_{x_n + y_m = z_k} p_{nm} =$$

$$= \sum_{n,m} (x_n + y_m) p_{nm} = \sum_n x_n (\sum_m p_{nm}) + \sum_m y_m (\sum_n p_{nm}) =$$
$$\sum_n x_n P(\xi_1 = x_n) + \sum_m y_m P(\xi_2 = y_m) = M\xi_1 + M\xi_2 \ .$$

4. $\xi \geq 0$ означает, что все значения $x_n \geq 0$. Тогда

$$M\xi = \sum_n x_n p_n \geq 0 \ ,$$

т.к. это сумма неотрицательных слагаемых.

5. Это свойство есть прямое следствие свойств 2,3 и 4. Действительно, $\eta = \xi_1 - \xi_2 \geq 0$. Тогда

$$M\eta = M(\xi_1 - \xi_2) = M\xi_1 + M(-\xi_2) = M\xi_1 - M\xi_2 \geq 0 \ .$$

6. Это свойство доказывается так же, как и свойство 3. Необходимо только отметить, что в случае независимых с.в.

$$p_{nm} = P(\xi_1 = x_n, \xi_2 = y_m) = P(\xi_1 = x_n) \cdot P(\xi_2 = y_m) \ .$$

1.10.2 Определение математического ожидания в общем случае

Теперь мы хотим определить математическое ожидание с.в. ξ, распределение которой не обязательно дискретно. Начнем со случая неотрицательных случайных величин. Наш план состоит в том, чтобы аппроксимировать такие случайные величины с помощью дискретных, для которых математическое ожидание уже определено, а математическое ожидание ξ положить равным пределу математических ожиданий приближающих ее дискретных с.в. Кстати, это очень полезная общая идея, состоящая в том, что некоторая характеристика сначала определяется для простых объектов, а затем для более сложных объектов она определяется с помощью аппроксимации их более простыми (смотри, например, определение площадей, исследование функций путем разложения их в ряды и т.п.).

Лемма 1 . *Пусть ξ есть произвольная неотрицательная случайная величина. Тогда существует последовательность $\{\xi_n\}$ дискретных случайных величин, таких, что*

1) $\xi_n(\omega) \geq 0$, $\forall n$,
2) $\xi_n(\omega) \leq \xi_{n+1}(\omega)$,
3) $\xi_n(\omega) \to \xi(\omega)$ равномерно по $\omega \in \Omega$ при $n \to \infty$.

Доказательство. Разобьем полуось $[0, \infty)$ на равные отрезки длины 2^{-n} и определим

$$\xi_n(\omega) = \frac{k}{2^n} \text{ , если } k2^{-n} \leq \xi(\omega) < (k+1)2^{-n}.$$

Тогда свойства 1 и 2 легко следуют из определения с.в. ξ_n, и

$$|\xi_n(\omega) - \xi(\omega)| \leq 2^{-n} \to 0, \; n \to \infty,$$

$\forall \omega \in \Omega$.

Лемма 2 . *Пусть ξ – неотрицательная случайная величина и $\{\xi_n\}$ и $\{\eta_n\}$ – две последовательности дискретных с.в., обладающих свойствами 1 - 3 из леммы 1. Тогда*

$$\lim_n M\xi_n = \lim_n M\eta_n \; .$$

Доказательство. Отметим, что для неотрицательных случайных величин мы допускаем $M\xi = \infty$.

В силу свойства 3 легко видеть, что существует последовательность $\{\varepsilon_n\}$ положительных чисел, такая, что

$$|\xi_n(\omega) - \eta_n(\omega)| \leq \varepsilon_n \to 0,\ n \to \infty\ .$$

Отсюда следует, что

$$\xi_n(\omega) - \varepsilon_n \leq \eta_n(\omega) \leq \xi_n(\omega) + \varepsilon_n$$

$\forall \omega \in \Omega$. Используя свойства математических ожиданий для дискретных случайных величин, получаем

$$M\xi_n - \varepsilon_n \leq M\eta_n \leq M\xi_n + \varepsilon_n\ .$$

Переходя к пределу при $n \to \infty$, получаем утверждение леммы 2.

Определение 2 . *Пусть ξ – неотрицательная случайная величина, $\{\xi_n\}$ – последовательность дискретных случайных величин, обладающих свойствами 1 - 3 из леммы 1. Математическим ожиданием с.в. ξ называется число*

$$M\xi = \lim_n M\xi_n\ .$$

Лемма 2 гарантирует, что $M\xi$ не зависит от выбора аппроксимирующей последовательности $\{\xi_n\}$.

Пусть теперь ξ – произвольная случайная величина. Определим

$$\xi^+(\omega) = \begin{cases} \xi(\omega) & , \quad \text{если } \xi(\omega) \geq 0\ , \\ 0 & , \quad \text{если } \xi(\omega) < 0\ , \end{cases}$$

$$\xi^-(\omega) = \begin{cases} 0 & , \quad \text{если } \xi(\omega) \geq 0\ , \\ -\xi(\omega) & , \quad \text{если } \xi(\omega) < 0\ . \end{cases}$$

Из определения ξ^+ и ξ^- легко следует, что

1) $\xi^+ \geq 0,\ \xi^- \geq 0$,

2) $\xi(\omega) = \xi^+(\omega) - \xi^-(\omega)$.

Определение 3 . **Математическим ожиданием** *произвольной случайной величины* ξ *называется число*

$$M\xi = M\xi^+ - M\xi^- \ ,$$

если хотя бы одно из чисел в правой части этого равенства конечно.

Случайная величина ξ *называется* **интегрируемой**, *если* $M\xi^+ < \infty$ *и* $M\xi^- < \infty$.

Всюду далее мы при вычислении математических ожиданий будем предполагать, что соответствующие с.в. ξ являются интегрируемыми.

1.10.3 Математическое ожидание как интеграл Римана-Стилтьеса

Выше мы уже отмечали, что единственной объективной характеристикой с.в. ξ, которую можно восстановить однозначно по результатам эксперимента (по - крайней мере, в принципе), является ее распределение. Отсюда следует, что все другие характеристики должны выписываться с помощью распределения. Ниже мы покажем, как это сделать для математического ожидания. Для этого нам потребуется одно понятие из математического анализа, а именно интеграл Римана – Стилтьеса.

Пусть $F(x)$ – неубывающая ограниченная функция, $f(x)$ – произвольная борелевская функция, которые заданы на интервале $[a, b)$. Рассмотрим разбиение $a = t_0 < t_1 < \ldots < t_k < t_{k+1} \ldots < t_n = b$ и выберем по точке $s_k \in [t_k, t_{k+1})$. Образуем интегральную сумму

$$S_n = \sum_{k=0}^{n-1} f(s_k)[F(t_{k+1}) - F(t_k)] \ .$$

Определение 4 . *Функция* $f(x)$ *называется* **интегрируемой в смысле Римана – Стилтьеса** *относительно функции* $F(x)$ *на* $[a, b)$, *если существует предел интегральных сумм* S_n *при* $\max\limits_k |t_{k+1} - t_k| \to 0$ *и этот предел не зависит от способа измельчения интервала* $[a, b)$ *и выбора точек* s_k.

Полученный предел называется **интегралом Римана – Стил - тьеса** *функции $f(x)$ относительно функции $F(x)$ и обозначается*

$$\int_a^b f(x)dF(x) \ .$$

Замечания. 1. Если $F(x) = x$, то мы получаем определение обычного интеграла Римана.

2. Если интервал бесконечный, то сначала интеграл определяют на конечном интервале, а затем переходят к пределу, когда один или оба конца интервала уходят в бесконечность.

3. Достаточным условием интегрируемости по Риману – Стилтьесу на конечном интервале является непрерывность функции $f(x)$ на замкнутом отрезке $[a, b]$.

4. Если существует $\rho_\xi(x) = F^{'}(x)$, то $\int\limits_a^b f(x)dF(x) = \int\limits_a^b f(x)\rho_\xi(x)dx$, где последний интеграл нужно понимать как интеграл Римана.

Применим эту конструкцию к нашей ситуации, когда мы вычисляем математическое ожидание. Пусть ξ – неотрицательная случайная величина. Напомним, что аппроксимирующая ее последовательность дискретных с.в. ξ_n строится по правилу:

$$\xi_n(\omega) = \frac{k}{2^n} \ , \text{ если } \frac{k}{2^n} \leq \xi(\omega) < \frac{k+1}{2^n} \ .$$

Тогда

$$M\xi = \lim_n M\xi_n = \lim_n \sum_{k=0}^{\infty} \frac{k}{2^n}[F_\xi(\frac{k+1}{2^n}) - F_\xi(\frac{k}{2^n})] \ .$$

Сравнивая последнее равенство с определением интеграла Римана – Стилтьеса для $f(x) = x$ и $F(x) = F_\xi(x)$, получаем

$$M\xi = \int\limits_0^\infty x dF_\xi(x) \ .$$

Разлагая ξ в разность положительной и отрицательной частей, аналогично получаем в общем случае

$$M\xi = \int\limits_{-\infty}^{\infty} x dF_\xi(x) \ . \tag{10.1}$$

Приведем без доказательства следующий полезный результат.

Предложение 2 . *Если с.в.* ξ *имеет функцию распределения* $F_\xi(x)$, $y = g(x)$ *– борелевская функция такая, что с.в.* $\eta = g(\xi)$ *– интегрируема, то*

$$M\eta = \int_{-\infty}^{\infty} g(x)dF_\xi(x) \ . \tag{10.2}$$

Если ξ – дискретная с.в. с множеством значений $X = \{x_1, x_2, \dots\}$ и $p_n = P(\xi = x_n)$, то

$$M\xi = \sum_n x_n p_n \ , \tag{10.3}$$

и

$$M\eta = \sum_n g(x_n)p_n \ . \tag{10.4}$$

Если ξ имеет плотность распределения $\rho_\xi(x)$, то

$$M\xi = \int_{-\infty}^{\infty} x\rho_\xi(x)dx \ . \tag{10.5}$$

и

$$M\eta = \int_{-\infty}^{\infty} g(x)\rho_\xi(x)dx \ . \tag{10.6}$$

1.10.4 Свойства математических ожиданий

Всюду в этом разделе предполагается, что все написанные математические ожидания конечны.

1. Если $P(\xi = C) = 1$, то $M\xi = C$.

2. $M(C\xi) = C \cdot M\xi$.

3. $M(\xi_1 + \xi_2) = M\xi_1 + M\xi_2$.

4. Если $\xi \geq 0$, то $M\xi \geq 0$, причем $M\xi = 0$ тогда и только тогда, когда $P(\xi = 0) = 1$.

Эти четыре свойства получаются предельным переходом из аналогичных свойств математических ожиданий дискретных случайных величин.

5. Если $\xi_1 \geq \xi_2$, то $M\xi_1 \geq M\xi_2$.

Это свойство есть прямое следствие свойств 2 - 4.

6. $|M\xi| \leq M|\xi|$.

Доказательство. По определению абсолютной величины вещественного числа имеем

$$-|\xi| \leq \xi \leq |\xi| \ .$$

Далее применяем свойство 5.

7. Если ξ – индикатор некоторого случайного события A, то $M\xi = P(A)$.

Доказательство. По определению индикатора события с.в. ξ принимает два значения: 1 – с вероятностью $P(A)$ и 0 – с вероятностью $1 - P(A)$. Тогда по определению математического ожидания для дискретной случайной величины имеем

$$M\xi = 1 \cdot P(A) + 0 \cdot (1 - P(A)) = P(A) \ .$$

8. Если ξ_1 и ξ_2 – независимы, то $M(\xi_1\xi_2) = M\xi_1 \cdot M\xi_2$.

Доказательство будет разбито на несколько шагов.

a) Если ξ_1 и ξ_2 – дискретные случайные величины, то это свойство уже доказано ранее.

b) Пусть ξ_1 и ξ_2 – неотрицательные интегрируемые случайные величины. Построим две последовательности $\{\xi_1^{(n)}\}$ и $\{\xi_2^{(n)}\}$ дискретных случайных величин, обладающих свойствами из леммы 1. Из доказательства этой леммы мы знаем, что $\xi_1^{(n)}$ и $\xi_2^{(n)}$ можно построить как функции от ξ_1 и ξ_2. Отсюда следует, что они независимы. Предположим, что

$$|\xi_1^{(n)} - \xi_1| \leq \frac{1}{n} \ , \ |\xi_2^{(n)} - \xi_2| \leq \frac{1}{n} \ .$$

По определению математического ожидания неотрицательной случайной величины

$$M\xi_1 = \lim_n M\xi_1^{(n)} \ , \ M\xi_2 = \lim_n M\xi_2^{(n)} \ .$$

Используя все указанное выше, получаем

$$|M(\xi_1\xi_2) - M\xi_1 \cdot M\xi_2| =$$

$$= |M(\xi_1\xi_2) - M(\xi_1^{(n)}\xi_2) + M(\xi_1^{(n)}\xi_2)$$

$$-M(\xi_1^{(n)}\xi_2^{(n)}) + M(\xi_1^{(n)}\xi_2^{(n)}) - M\xi_1 \cdot M\xi_2|$$
$$\le M(|\xi_1 - \xi_1^{(n)}||\xi_2|) + M(|\xi_1^{(n)}||\xi_2 - \xi_2^{(n)}|)$$
$$+|M\xi_1^{(n)} \cdot M\xi_2^{(n)} - M(\xi_1) \cdot M(\xi_2)|$$
$$\le \frac{1}{n}M(|\xi_2|) + \frac{1}{n}M(|\xi_1|) + |M\xi_1^{(n)} \cdot M\xi_2^{(n)} - M(\xi_1) \cdot M(\xi_2)| \to 0 \ ,$$
при $n \to \infty$.

c) Произвольные случайные величины ξ_1 и ξ_2 представим в виде разностей $\xi_1 = \xi_1^+ - \xi_1^-$ и $\xi_2 = \xi_2^+ - \xi_2^-$ их положительных и отрицательных частей. В силу независимости ξ_1 и ξ_2 пары (ξ_1^+, ξ_1^-) и (ξ_2^+, ξ_2^-) также будут независимы. Используя пункт b), получаем

$$M(\xi_1\xi_2) = M[(\xi_1^+ - \xi_1^-)(\xi_2^+ - \xi_2^-)] =$$
$$= M(\xi_1^+\xi_2^+ - \xi_1^+\xi_2^- - \xi_1^-\xi_2^+ + \xi_1^-\xi_2^-) =$$
$$= M(\xi_1^+)M(\xi_2^+) - M(\xi_1^+)M(\xi_2^-) - M(\xi_1^-)M(\xi_2^+) + M(\xi_1^-)M(\xi_2^-) =$$
$$= M(\xi_1^+ - \xi_1^-)M(\xi_2^+ - \xi_2^-) = M\xi_1 \cdot M\xi_2 \ .$$

Далее мы докажем несколько неравенств, которые будут часто использоваться в нашем курсе.

9. **Неравенство Маркова**. Если $\xi \ge 0$, то $\forall \varepsilon > 0$

$$P(\xi \ge \varepsilon) \le \frac{M\xi}{\varepsilon} \ .$$

Доказательство. Рассмотрим случайное событие $A = (\xi \ge \varepsilon)$ и его индикатор I_A. Нетрудно проверить (задача!), что имеет место следующее неравенство:

$$\varepsilon \cdot I_A(\omega) \le \xi(\omega) \ .$$

Тогда, применяя свойства 3,5 и 7, получаем

$$M(\varepsilon I_A) = \varepsilon P(A) \le M\xi \ .$$

10. **Неравенство Коши - Буняковского**. Если ξ_1 и ξ_2 – случайные величины, то

$$(M(\xi_1\xi_2))^2 \le M(\xi_1^2) \cdot M(\xi_2^2) \ ,$$

причем знак равенства достигается тогда и только тогда, когда существует некоторая константа C, такая, что $\xi_1 = C\xi_2$ или $\xi_2 = C\xi_1$.

Доказательство. Возьмем $\lambda \in R^1$ и образуем случайную величину $\eta = (\xi_1 + \lambda\xi_2)^2$. Так как $\eta \geq 0$, то

$$M\eta = M(\xi_1^2) + 2M(\xi_1\xi_2)\lambda + M(\xi_2^2)\lambda^2 \geq 0 \ .$$

Мы имеем полином второго порядка от λ, который принимает только неотрицательные значения. В этом случае его дискриминант должен быть неположительным:

$$D = (M(\xi_1\xi_2))^2 - M(\xi_1^2)M(\xi_2^2) \leq 0 \ .$$

Если в последнем неравенстве мы имеем знак равенства, т. е. $D = 0$, то существует некоторое $\lambda_0 \in R^1$, такое, что

$$M\eta = M(\xi_1 + \lambda_0\xi_2)^2 = 0 \ .$$

Но тогда (смотри свойство 2) мы имеем

$$\xi_1 + \lambda_0\xi_2 = 0 \text{ или } \xi_1 = -\lambda_0\xi_2 \ .$$

11. **Неравенство Иенсена**. Пусть $y = g(x)$ – выпуклая вниз борелевская функция, ξ – некоторая случайная величина. Тогда

$$g(M\xi) \leq Mg(\xi) \ .$$

Доказательство. Из определения выпуклости следует, что $\forall x_0 \in R^1$ существует $\lambda(x_0) \in R^1$, такое, что $\forall x \in R^1$

$$g(x) - g(x_0) \geq \lambda(x_0)(x - x_0).$$

Мы построили так называемую опорную прямую. Подставим в это неравенство ξ и $M\xi$ вместо x и x_0:

$$g(\xi) - g(M\xi) \geq \lambda(M\xi)(\xi - M\xi) \ .$$

В силу свойства (5) получаем

$$Mg(\xi) - g(M\xi) \geq \lambda(M\xi)(M\xi - M\xi) = 0 \ .$$

12. **Неравенство Ляпунова**. Если ξ – некоторая случайная величина, $0 < s < t < \infty$, то

$$(M|\xi|^s)^{1/s} \leq (M|\xi|^t)^{1/t} \ .$$

Доказательство. Пусть $r = t/s > 1$, $g(x) = |x|^r$, $\eta = |\xi|^s$. Тогда $g(x)$ – выпуклая вниз функция. В силу неравенства Иенсена

$$g(M\eta) = |M|\xi|^s|^{t/s} \leq M(g(\eta)) = M(|\xi|^{s\cdot\frac{t}{s}}) = M(|\xi|^t) \ .$$

Отсюда получаем

$$(M|\xi|^s)^{1/s} \leq (M|\xi|^t)^{1/t} \ .$$

1.11 Числовые характеристики случайных величин и векторов

Как мы отмечали ранее, распределение случайной величины является ее полной характеристикой. Но эта характеристика является достаточно сложной и в реальных задачах редко бывает известна достоверно. Поэтому пытаются описать распределение с помощью конечного числа сравнительно простых числовых характеристик. Одной из них является математическое ожидание. В этом параграфе мы вводим некоторые новые характеристики.

1.11.1 Моменты

Определение 1 . Моментом порядка k относительно точки *a случайной величины ξ называется число*

$$M(\xi - a)^k \ .$$

Если число $a = 0$, то момент $\alpha_k = M(\xi^k)$ называется **начальным**.

Если число $a = M\xi$, то момент $\mu_k = M(\xi - M\xi)^k$ называется **центральным**.

Центральный момент второго порядка

$$\mu_2 = D(\xi) := M(\xi - M\xi)^2$$

называется **дисперсией**. Эта характеристика будет изучаться подробнее ниже.

Между центральными и начальными моментами существует простая связь:

$$\mu_k = \sum_{m=0}^{k} (-1)^m C_k^m \alpha_1^m \alpha_{k-m} \ .$$

Определяются также **абсолютные** моменты:

$$\beta_k = M(|\xi|^k) \ .$$

Предполагается, что задание достаточного числа моментов определяет распределение с нужной степенью точности. В связи с этой идеей важное значение имеет следующая

Проблема моментов. Пусть мы имеем некоторую последовательность чисел $\alpha_0, \alpha_1, \alpha_2, \ldots$.

1. При каких условиях последовательность $\{\alpha_k\}$ является последовательностью моментов какой-либо случайной величины?

2. При каких условиях последовательность моментов $\{\alpha_k\}$ однозначно определяет распределение?

На первый вопрос отвечает следующая

Теорема 1 . *Последовательность чисел $\{\alpha_k\}$ является последовательностью моментов некоторой случайной величины тогда и только тогда, когда $\alpha_0 = 1$ и для любого натурального $N \geq 1$ и любых комплекных чисел $c_0, \ldots, c_N$*

$$\sum_{m,n=0}^{N} \alpha_{m+n} c_m \overline{c_n} \geq 0 \ ,$$

т.е. матрица $(b_{mn} = \alpha_{m+n})$ является положительно определенной.

Второй вопрос является более сложным. Мы приводим только одно достаточное условие. Но прежде нам необходимо ввести одно вспомогательное понятие.

Определение 2 . **Производящей функцией моментов** *с.в. ξ (или ее распределения) называется функция*

$$\psi_\xi(t) = M(e^{t\xi}) \ .$$

Свойства производящих функций.

1. $\psi_\xi(0) = 1$.

2. Если с.в. ξ_1 и ξ_2 независимы, то

$$\psi_{\xi_1+\xi_2}(t) = \psi_{\xi_1}(t) \cdot \psi_{\xi_2}(t) \ .$$

3. Если для некоторого $t_0 > 0$ производящая функция моментов $\psi(t)$ существует для всех $t: \ |t| < t_0$, то $\forall k \geq 1$

$$\exists \frac{d^k}{dt^k}\psi(t)\Big|_{t=0} = M(\xi^k) \ .$$

Эти и другие свойства производящих функций моментов будут доказаны позднее.

Пример. Пусть $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ – число успехов в схеме Бернулли с параметрами n и p, где $\varepsilon_1, \ldots, \varepsilon_n$ – н.о.р. и имеют распределение Бернулли с параметром p. Нетрудно вычислить, что

$$\psi_{\varepsilon_k}(t) = e^{t\cdot 1}p + e^{t\cdot 0}(1-p) = 1 + p(e^t - 1) \ ,$$

$$\psi(t) = \psi_{S_n}(t) = \prod_{k=1}^{n} \psi_{\varepsilon_k}(t) = [1 + p(e^t - 1)]^n \ .$$

Дифференцируя, получаем

$$\frac{d}{dt}\psi(t) = n[1 + p(e^t - 1)]^{n-1}pe^t \ ,$$

$$\frac{d^2}{dt^2}\psi(t) = n(n-1)[1 + p(e^t - 1)]^{n-2}p^2e^{2t} + n[1 + p(e^t - 1)]^{n-1}pe^t \ .$$

Применяя свойство 3, получаем

$$M(S_n) = np, \ M(S_n^2) = n(n-1)p^2 + np \ .$$

Теорема 2 . *Пусть $\{\alpha_n\}$ – последовательность моментов с.в. ξ. Если существует $r > 0$, такое, что ряд*

$$\sum_{n=0}^{\infty} \frac{\alpha_n}{n!} r^n$$

сходится абсолютно (что эквивалентно условию $\psi_\xi(r) = M(e^{r\xi}) < \infty$, и в этом случае обе величины совпадают), то распределение с.в. ξ однозначно определяется своими моментами.

Смысл этого утверждения в том, что если нам удалось каким-либо образом вычислить моменты распределения, то мы можем утверждать, что это то или иное из известных нам распределений.

Пример. Пусть ξ имеет стандартное нормальное распределение. Тогда

$$\psi_\xi(r) = M(e^{r\xi}) = \int\limits_{-\infty}^{\infty} e^{rx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx = \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-r)^2}{2}}dxe^{\frac{r^2}{2}} = e^{\frac{r^2}{2}} < \infty\,.$$

Таким образом, нормальное распределение однозначно определяется своими моментами.

Пусть мы теперь имеем случайный вектор $\xi = (\xi_1, \ldots, \xi_d)$.

Определение 3 . **Моментом порядка** k *относительно точки* $a \in R^d$ *случайного вектора* ξ *называется число*

$$M[(\xi_1 - a_1)^{k_1} \cdot \ldots \cdot (\xi_d - a_d)^{k_d}]\ ,$$

где $k_1 + \ldots + k_d = k$.

Если $a_1 = \ldots = a_d = 0$*, то момент называется* **начальным**. *Если* $a_1 = M\xi_1, \ldots, a_d = M\xi_d$*, то момент называется* **центральным**.

В отличие от одномерного случая, теперь могут существовать несколько разных моментов одного порядка.

Для $k = 1$ начальные моменты совпадают с математическими ожиданиями координат: $M\xi_1, \ldots, M\xi_d$, а все центральные моменты равны нулю.

Для $k = 2$ имеем следующие начальные моменты:

$$M(\xi_1^2), \ldots, M(\xi_d^2),\ M(\xi_i\xi_j)\ ,\ i, j = \overline{1, d}.$$

Чаще используют центральные моменты:

$$\sigma_{11} = D(\xi_1) = M[(\xi_1 - M\xi_1)^2], \ldots, \sigma_{dd} = D(\xi_d) = M[(\xi_d - M\xi_d)^2]$$

– **дисперсии** и

$$\sigma_{ij} = cov(\xi_i, \xi_j) = M[(\xi_i - M\xi_i)(\xi_j - M\xi_j)]$$

– **ковариации** координат случайного вектора ξ.

Обычно все центральные моменты второго порядка собирают в одну матрицу

$$\Sigma = (\sigma_{ij}) \ ,$$

которая называется **матрицей ковариаций** или **ковариационной матрицей** случайного вектора ξ.

1.11.2 Дисперсия и ее свойства

Определение 4 . **Дисперсией** *случайной величины ξ называется число*

$$D(\xi) = M[(\xi - M\xi)^2] \ .$$

Если $F(x)$ – функция распределения с.в. ξ, то

$$D(\xi) = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 dF(x) \ .$$

Для дискретной случайной величины

$$D(\xi) = \sum_n (x_n - M\xi)^2 \cdot p_n \ .$$

В случае абсолютно непрерывного распределения с плотностью $\rho(x)$ имеем

$$D(\xi) = \int\limits_{-\infty}^{\infty} (x - M\xi)^2 \rho(x) dx \ .$$

Перечислим и докажем основные свойства дисперсии.

1. $D(\xi) \geq 0$, $D(\xi) = 0$ тогда и только тогда, когда $\xi = C$ п.н.

Доказательство. С.в. $(\xi - M\xi)^2 \geq 0$. Отсюда в силу свойства 4 для математических ожиданий случайных величин получаем нужное утверждение.

2. $D(\xi + C) = D(\xi)$.

3. $D(C\xi) = C^2 D(\xi)$.

Доказательство. По определению дисперсии

$$D(\xi + C) = M[((\xi + C) - M(\xi + C))^2] = M[(\xi + C - M\xi - C)^2] =$$

$$M[(\xi - M\xi)^2] = D(\xi) \ .$$

Свойство 3 доказывается аналогично.

4. $D(\xi) = M(\xi^2) - (M\xi)^2$.

Доказательство. Вновь по определению дисперсии и используя свойства математических ожиданий, получаем

$$D(\xi) = M[(\xi - M\xi)^2] = M[\xi^2 - 2\xi \cdot M\xi + (M\xi)^2] = M(\xi^2) - (M\xi)^2 \ .$$

Определение 5 . *С.в.* ξ_1 *и* ξ_2 *называются* **некоррелированными**, *если*

$$cov(\xi_1, \xi_2) = M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)] = 0 \ .$$

Лемма 1 . *Если с.в.* ξ_1 *и* ξ_2 *независимы, то они и некоррелированы.*

Доказательство. Если ξ_1 и ξ_2 независимы, то независимы и с.в. $\xi_1 - M\xi_1$ и $\xi_2 - M\xi_2$. Тогда

$$M[(\xi_1 - M\xi_1)(\xi_2 - M\xi_2)] = M(\xi_1 - M\xi_1) \cdot M(\xi_2 - M\xi_2) =$$

$$(M\xi_1 - M\xi_1) \cdot M(\xi_2 - M\xi_2) = 0 \ .$$

Приведем пример, который показывает, что обратное неверно.

Пусть случайная величина ξ имеет равномерное распределение на $[-1, 1]$. Положим $\xi_1 = \xi$, $\xi_2 = \xi^2$. В этом случае $\xi_2 = \xi_1^2$, т.е. эти величины функционально связаны и поэтому зависимы. Но

$$cov(\xi_1, \xi_2) = M[(\xi - M\xi)(\xi^2 - M(\xi^2))] = M(\xi^3) - M(\xi)M(\xi^2) = 0 \ ,$$

т.к. в силу симметрии распределения $M(\xi) = M(\xi^3) = 0$.

5. $D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2) + 2cov(\xi_1, \xi_2)$.

Если ξ_1 *и* ξ_2 *некоррелированы (в частности, независимы), то*

$$D(\xi_1 + \xi_2) = D(\xi_1) + D(\xi_2) \ .$$

Предлагается доказать это свойство самостоятельно.

Замечания. 1. $D(\xi)$ оценивает степень рассеивания с.в. ξ около своего "центра" $M\xi$.

2. Часто полезно (вспомним случай нормального распределения) перейти к так называемой нормированной случайной величине

$$\xi_0 = \frac{\xi - M\xi}{\sqrt{D(\xi)}} \ .$$

Используя свойства математических ожиданий и дисперсий, легко показать (задача!), что $M\xi_0 = 0,\ D(\xi_0) = 1$.

3. Некоррелированность является ослабленным вариантом понятия независимости. Геометрическая интерпретация некоррелированности будет дана ниже.

В заключение этого раздела выпишем значения математических ожиданий и дисперсий для стандартных распределений.

1. Вырожденное распределение в точке a.

$$M\xi = a,\ D(\xi) = 0\ .$$

2. Распределение Бернулли с параметром p.

$$M\xi = p\ ,\ D(\xi) = p(1-p)\ .$$

3. Биномиальное распределение с параметрами n и p.

$$M\xi = np\ ,\ D(\xi) = np(1-p)\ .$$

4. Геометрическое распределение с параметром p.

$$M\xi = \frac{1-p}{p},\ D(\xi) = \frac{1-p}{p^2}\ .$$

5. Отрицательное биномиальное распределение с параметрами . r и p.

$$M\xi = \frac{1-p}{p}r + (r-1),\ D(\xi) = \frac{1-p}{p^2}r\ .$$

6. Распределение Пуассона с параметром λ.

$$M\xi = \lambda,\ D(\xi) = \lambda\ .$$

7. Равномерное распределение на отрезке $[a, b]$.

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

8. Показательное распределение с параметром λ.

$$M\xi = \frac{1}{\lambda}\ ,\ D(\xi) = \frac{1}{\lambda^2}\ .$$

9. Гамма-распределение с параметрами α и β.

$$M\xi = \frac{\alpha}{\beta}\ ,\ D(\xi) = \frac{\alpha}{\beta^2}\ .$$

10. Нормальное распределение с параметрами α и σ^2.

$$M\xi = a\ ,\ D(\xi) = \sigma^2\ .$$

Эти результаты будут получены на практических занятиях.

1.11.3 Коэффициент корреляции и его свойства

Выше мы уже отмечали, что некоррелированность, т.е. равенство нулю ковариации случайных величин, является ослабленным вариантом независимости. Интуитивно понятно, что если изменить единицу измерения и начало отсчета, то степень зависимости между случайными величинами не должна измениться. Таким образом, для описания силы связи между случайными величинами естественно использовать некоторый коэффициент, который был бы пропорционален ковариации, но не менялся бы при линейных преобразованиях.

Определение 6 . **Коэффициентом корреляции** *случайных величин* ξ_1 *и* ξ_2 *называется число*

$$\rho(\xi_1, \xi_2) = \frac{cov(\xi_1, \xi_2)}{\sqrt{D(\xi_1)D(\xi_2)}}\ .$$

Это определение имеет смысл, если случайные величины ξ_1 и ξ_2 невырождены. Основные свойства коэффициента корреляции собраны в следующей теореме.

Теорема 3 . *Пусть ξ_1 и ξ_2 – невырожденные случайные величины. Тогда справедливы следующие свойства:*

1. $|\rho(\xi_1, \xi_2)| \le 1$.

2. Если ξ_1 и ξ_2 некоррелированы, то $\rho(\xi_1, \xi_2) = 0$.

3. Пусть $\eta_1 = a_1\xi_1 + b_1$, $\eta_2 = a_2\xi_2 + b_2$, $a_1 \cdot a_2 \ne 0$. Тогда

$$\rho(\eta_1, \eta_2) = \rho(\xi_1, \xi_2) \cdot sign(a_1 a_2) \ .$$

4. Если $\rho(\xi_1, \xi_2) = 1$, то существуют $a > 0$, $b \in R^1$, такие, что $\xi_1 = a\xi_2 + b$.

Если $\rho(\xi_1, \xi_2) = -1$, то существуют $a < 0$, $b \in R^1$, такие, что $\xi_1 = a\xi_2 + b$.

Доказательство. Начнем с доказательства свойства 3. Из определения с.в. η_1 и η_2, используя свойства математических ожиданий и дисперсий, легко получаем

$$M\eta_1 = a_1 M\xi_1 + b_1 \ , \ M\eta_2 = a_2 M\xi_2 + b_2 \ , D(\eta_1) = a_1^2 D(\xi_1) \ ,$$

$$D(\eta_2) = a_2^2 D(\xi_2) \ , \ cov(\eta_1, \eta_2) = a_1 a_2 cov(\xi_1, \xi_2) \ .$$

Отсюда следует, что

$$\rho(\eta_1, \eta_2) := \frac{cov(\eta_1, \eta_2)}{\sqrt{D(\eta_1) D(\eta_2)}} =$$

$$\frac{a_1 a_2 cov(\xi_1, \xi_2)}{|a_1 a_2| \sqrt{D(\xi_1) D(\xi_2)}} = \rho(\xi_1, \xi_2) \cdot sign(a_1 a_2) \ .$$

Для доказательства свойства 1 введем с.в.

$$\eta_1 = \frac{\xi_1 - M\xi_1}{\sqrt{D(\xi_1)}} \ , \ \eta_2 = \frac{\xi_2 - M\xi_2}{\sqrt{D(\xi_2)}} \ .$$

В силу свойства 3 получаем

$$|\rho(\xi_1, \xi_2)|^2 = |\rho(\eta_1, \eta_2)|^2 = |M(\eta_1 \cdot \eta_2)|^2 \le$$

$$M(\eta_1^2) \cdot M(\eta_2^2) = D(\eta_1)D(\eta_2) = 1 \ .$$

На последнем шаге мы использовали неравенство Коши – Буняковского и замечание 2 о нормированной с.в.

Свойство 2 есть тривиальное следствие из определений некоррелированности и коэффициента корреляции.

Для доказательства свойства 4 вновь введем нормированные с.в. η_1 и η_2. В неравенстве Коши – Буняковского достигается знак равенства тогда и только тогда, когда существует такая константа c, что $\eta_1 = c\eta_2$. Возвращаясь к исходным величинам ξ_1 и ξ_2, мы получаем $\xi_1 = a\xi_2 + b$, где

$$a = c\sqrt{\frac{D(\xi_1)}{D(\xi_2}} \text{ и } b = M\xi_2 - c\sqrt{\frac{D(\xi_1)}{D(\xi_2}}M\xi_1 \ .$$

Используя свойство 3, легко показать, что $a > 0$ для $\rho = 1$ и $a < 0$ для $\rho = -1$.

Замечания. 1. Как мы увидим позднее, коэффициент корреляции является мерой **линейной** связи между случайными величинами.

2. На практических занятиях будет показано, что для с.в. ξ_1 и ξ_2, которые имеют двумерное нормальное распределение, параметр ρ равен коэффициенту корреляции для этих величин.

1.11.4 Характеристики расположения и формы распределения

В реальных задачах бывает полезно разметить прямую точками, которые делят ее на несколько отрезков, вероятности попадания в которые имеют определенные значения.

Определение 7 . **Квантилью порядка** *p, $0 < p < 1$, случайной величины ξ (или ее распределения) называется число $x_p \in R^1$:*

$$P(\xi \leq x_p) \geq p \ , \ P(\xi \geq x_p) \geq 1 - p \ .$$

Замечание. Если $P(\xi = x) = 0$ для всех x, то

$$P(\xi < x_p) = F_\xi(x_p) = p \ .$$

Примеры. 1. Величины $x_{1/4}, x_{1/2}, x_{3/4}$ называются **квартилями**. Число $x_{1/2}$ называется **медианой**. Оно определяет "центр" распределения. Разность $x_{3/4} - x_{1/4}$ называется **семиинтерквартильной широтой**. Она характеризует степень рассеивания распределения около центра.

2. Числа $x_{0,1}, x_{0,2}, \ldots, x_{0,9}$ называются **децилями**.

3. Наиболее часто используются на практике **процентили**, т.е. величины вида $x_{0,01}, \ldots, x_{0,99}$.

Часто хотелось бы знать, какие значения более вероятны, а какие менее. В связи с этим является полезным следующее понятие.

Определение 8 . Модой дискретного распределения $(X, \{p_n\})$ $(x_1 < x_2 < \ldots)$ *называется такое значение* x_n, *что* $p_n \geq p_{n-1}$, $p_n \geq p_{n+1}$.

Модой абсолютно непрерывного распределения *называется любая точка максимума плотности* $\rho_\xi(x)$.

Распределение называется **унимодальным**, *если оно имеет одну моду, и* **полимодальным** *– в противном случае.*

Примеры. 1. Нормальное распределение с параметрами a и σ^2 является унимодальным, мода которого совпадает с математическим ожиданием и равна a.

2. Биномиальное распределение с параметрами n и p является унимодальным, если число np – дробное, и **бимодальным**, если это число – целое (смотри выше).

Определение 9 . Коэффициентом асимметрии *называется число*

$$\frac{M(\xi - M\xi)^3}{\sigma^3} \ ,$$

где $\sigma^2 = D(\xi)$.

Коэффициентом эксцесса *называется число*

$$\frac{M(\xi - M\xi)^4}{\sigma^4} - 3 \ .$$

Эти характеристики применяются для грубого сравнения данного распределения с нормальным, для которого оба коэффициента равны нулю. Асимметрия измеряет величину "скошенности" распределения в ту или иную сторону. Эксцесс измеряет степень "островершинности" распределения. Если эксцесс положительный, то распределение является более плоским (размазанным) по сравнению с нормальным, в противном случае оно является более островершинным.

1.12 Гильбертово пространство случайных величин

В этом разделе мы выделим некоторое подмножество случайных величин и покажем, как в нем можно определить геометрическую структуру.

Рассмотрим множество L_2, состоящее из случайных величин ξ, таких, что $M(|\xi|^2) < \infty$. Используя неравенство Коши - Буняковского, нетрудно показать, что L_2 – это линейное пространство (задача!). Для каждой пары $\xi, \eta \in L_2$ определим число

$$(\xi, \eta) := M(\xi \cdot \eta) \ .$$

Будем говорить, что случайные величины ξ и η совпадают **почти наверное** (п.н.), если $P(\xi \neq \eta) = 0$. В дальнейшем мы не будем различать такие случайные величины и объединим их в один класс. Таким образом, мы рассматриваем L_2 как совокупность классов эквивалентности.

Задача. Введенный выше функционал обладает следующими свойствами:

a) $(\xi, \xi) \geq 0, \ (\xi, \xi) = 0 \Leftrightarrow \xi = 0$ п.н.,

b) $\forall \xi, \eta \in L_2, \ (\xi, \eta) = (\eta, \xi)$,

c) $\forall \xi_1, \xi_2, \eta \in L_2, \ c_1, \ c_2 \in R^1 \ (c_1\xi_1 + c_2\xi_2, \eta) = c_1(\xi_1, \eta) + c_2(\xi_2, \eta)$.

Функционал, который приписывает каждой паре элементов из линейного пространства некоторое число так, что выполнены свойства a) - c), называется **скалярным произведением**. Используя скалярное произведение, можно определить **норму** произвольного элемента $\xi \in L_2$:

$$\|\xi\| := \sqrt{(\xi, \xi)} = (M(|\xi|^2))^{1/2} .$$

Норма элемента ξ определяет его длину. Будем говорить, что последовательность с.в. $\{\xi_n\}$ **сходится в среднем квадратическом** (с.к.) к с.в. ξ, если

$$\|\xi_n - \xi\|^2 = M(|\xi_n - \xi|^2) \to 0, \ n \to \infty .$$

В этом случае мы будем писать $\xi_n \overset{\text{с.к.}}{\to} \xi$. Можно показать, что пространство L_2 полно относительно такой сходимости, т.е. всякая фундаментальная последовательность в L_2 имеет предел. Линейное пространство, в котором определено скалярное произведение и которое полно относительно порожденной этим скалярным произведением сходимости, называется **гильбертовым пространством**. Таким образом, мы показали, что L_2 – гильбертово пространство.

Выше мы отмечали, что норма вектора определяет его длину. Но в геометрии нам необходимо измерять не только длины, но и углы. Используя аналогию с понятием скалярного произведения в планиметрии, определим косинус угла α между элементами $\xi, \eta \in L_2$ по правилу

$$\cos\alpha = \frac{M(\xi\eta)}{\sqrt{M|\xi|^2 \cdot M|\eta|^2}} = \frac{(\xi, \eta)}{\|\xi\| \cdot \|\eta\|} .$$

Назовем ξ и η **ортогональными**, если $(\xi, \eta) = M(\xi\eta) = 0$.

Замечание. Пусть $M\xi = M\eta = 0$. Тогда имеет место следующее:

1) $M(\xi\eta) = cov(\xi, \eta) = 0$, т.е. понятие ортогональности совпадает с понятием некоррелированности;

2) $\cos\alpha = \rho(\xi, \eta)$, т.е. косинус угла между случайными величинами совпадает с коэффициентом корреляции.

Используя введенные выше понятия, мы решим очень важную с практической точки зрения задачу. Пусть $\mathcal{L} \subset L_2$ – некоторое линейное подпространство, которое замкнуто относительно сходимости в среднем квадратическом, а η – произвольный элемент из L_2.

Определение 1 . *Случайная величина $\hat{\eta} \in L_2$ называется* **наилучшим приближением с.в. η в пространстве $\mathcal{L}$**, *если*

1) $\hat{\eta} \in \mathcal{L}$,

2) $\|\eta - \hat{\eta}\|^2 = M|\eta - \hat{\eta}|^2 \le M|\eta - \xi|^2 = \|\eta - \xi\|^2$, $\forall \xi \in \mathcal{L}$.

Лемма о перпендикуляре. *С.в. $\hat{\eta}$ является наилучшим приближением с.в. η в линейном пространстве $\mathcal{L}$ тогда и только тогда, когда*

1) $\hat{\eta} \in \mathcal{L}$,

2) $(\eta - \hat{\eta}, \xi) = M[(\eta - \hat{\eta})\xi] = 0$, $\forall \xi \in \mathcal{L}$.

Доказательство. Пусть выполнено свойство 2 и $\xi \in \mathcal{L}$. Тогда

$$\|\eta - \xi\|^2 = M|\eta - \xi|^2 = M|(\eta - \hat{\eta}) + (\hat{\eta} - \xi)|^2 =$$

$$= M|\eta - \hat{\eta}|^2 + M|\hat{\eta} - \xi|^2 + M[(\eta - \hat{\eta})(\hat{\eta} - \xi)] =$$

$$= M|\eta - \hat{\eta}|^2 + M|\hat{\eta} - \xi|^2 + 0 \ge M|\eta - \hat{\eta}|^2 .$$

Здесь мы использовали тот факт, что $\hat{\eta} - \xi \in \mathcal{L}$.

Обратно, пусть мы знаем, что $\hat{\eta}$ – наилучшее приближение для η в $\mathcal{L}$. Возьмем произвольные $\xi \in \mathcal{L}$ и $\lambda \in R^1$. В силу оптимальности $\hat{\eta}$ получаем

$$M|\eta - \hat{\eta}|^2 \le M|\eta - \hat{\eta} + \lambda\xi|^2 =$$

$$M|\eta - \hat{\eta}|^2 + 2M[(\eta - \hat{\eta})\xi] \cdot \lambda + M|\xi|^2 \cdot \lambda^2 .$$

Последнее выражение есть полином второго порядка от λ, у которого коэффициент при λ^2 положительный и который принимает минимальное значение при $\lambda = 0$. Из курса математического анализа мы знаем, что в этом случае коэффициент при λ, т.е. $M[(\eta - \hat{\eta})\xi]$, равен нулю. А это и есть условие 2.

Дадим геометрическую интерпретацию полученному результату. Свойство 2 означает, что $\eta - \hat{\eta}$ является **перпендикуляром** к подпространству $\mathcal{L}$ (эквивалентно: $\hat{\eta}$ – **проекция** η на $\mathcal{L}$).

Иногда с.в. $\hat{\eta}$ называют **условным математическим ожиданием в широком смысле** с.в. η относительно подпространства $\mathcal{L}$. Смысл такого названия мы объясним позднее.

Рассмотрим несколько более общую ситуацию. Пусть $\mathcal{L} = \xi_0 + \mathcal{L}_0$, где $\xi_0 \in L_2$ – фиксированный элемент, а $\mathcal{L}_0$ – замкнутое линейное подпространство в L_2. Если $\xi_0 \neq 0$, то $\mathcal{L}$ не является линейным пространством и называется **гиперплоскостью**. Пусть далее, $\eta \in L_2$ – некоторая с.в. Определение наилучшего приближения $\hat{\eta}$ для с.в. η в гиперплоскости $\mathcal{L}$ дословно повторяет определение 1. Нетрудно доказать следующий результат.

Задача. $\hat{\eta}$ является наилучшим приближением для η в гиперплоскости $\mathcal{L}$, если

1) $\hat{\eta} \in \mathcal{L}$,

2) $(\eta - \hat{\eta}, \xi - \hat{\eta}) = M[(\eta - \hat{\eta}, \xi - \hat{\eta})] = 0 \ , \ \forall \xi \in \mathcal{L}$.

Применим эти результаты к следующей задаче. Пусть $\xi, \eta \in L_2$ – две случайные величины. Необходимо найти наилучшее линейное приближение с.в. η с помощью случайной величины ξ. В этом случае $\mathcal{L} = \{c_1 + c_2\xi \ , \ c_1, c_2 \in R^1\}$. Это двумерное линейное подпространство в L_2, базисом которого являются с.в. $\xi_1 = 1$ и $\xi_2 = \xi$. Наилучшее приближение должно иметь вид $\hat{\eta} = \hat{c}_1 + \hat{c}_2\xi$, т.е. мы должны найти $\hat{c}_1$ и $\hat{c}_2$. По лемме о перпендикуляре $(\eta - \hat{\eta}, \theta) = 0, \ \forall \theta \in \mathcal{L}$. В частности,

$$(\eta - \hat{\eta}, \xi_1) = M[(\eta - \hat{\eta}) \cdot 1] = M\eta - \hat{c}_1 - \hat{c}_2 M\xi = 0 \ ,$$

$$(\eta - \hat{\eta}, \xi_2) = M[(\eta - \hat{\eta}) \cdot \xi] = M(\eta\xi) - \hat{c}_1 M\xi - \hat{c}_2 M(\xi^2) = 0 \ .$$

Решая эти уравнения относительно $\hat{c}_1$ и $\hat{c}_2$, получим:

$$\hat{c}_2 = \frac{cov(\xi, \eta)}{D(\xi)} = \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \ ,$$

$$\hat{c}_1 = M\eta - \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot M\xi \ ,$$

где $\sigma_\eta^2 = D(\eta)$, $\sigma_\eta^2 = D(\eta)$. Таким образом, окончательно получаем

$$\hat{\eta} = M\eta + \rho(\xi, \eta) \cdot \frac{\sigma_\eta}{\sigma_\xi} \cdot (\xi - M\xi) \ .$$

1.13 Условные распределения и условные математические ожидания

Во многих прикладных задачах мы встречаемся с такой ситуацией, когда часть компонент случайного вектора фиксирована и необходимо найти распределение остальных компонент. Например, мы фиксируем цену некоторого товара и интересуемся распределением величины спроса на этот товар при заданной цене. Это приводит нас к понятию условного распределения. Общая теория условных распределений и условных математических ожиданий является одной из наиболее сложных тем в теории вероятностей. Поэтому мы ограничимся рассмотрением двух частных, но наиболее важных с практической точки зрения случаев.

Пусть мы имеем дискретный случайный вектор $\xi = (\xi_1, \ldots, \xi_m, \xi_{m+1}, \ldots, \xi_{m+n})$. Предположим, что мы фиксировали значения первых m координат. Тогда условное распределение последних n координат определяется по формуле

$$P(x_{m+1}, \ldots, x_{m+n} | x_1, \ldots, x_m) = \frac{P(\xi_1 = x_1, \ldots, \xi_m = x_m, \xi_{m+1} = x_{m+1}, \ldots, \xi_{m+n} = x_{m+n})}{P(\xi_1 = x_1, \ldots, \xi_m = x_m)}. \tag{13.1}$$

Можно проверить, что условные вероятности в (1) (как функции от $x_{m+1}, \ldots, x_{m+n}$) обладают всеми свойствами дискретного распределения. Это распределение называют **условным распределением** $\xi_{m+1}, \ldots, \xi_{m+n}$ при условии, что $\xi_1 = x_1, \ldots, \xi_m = x_m$. Ясно, что, меняя условия $x_1, \ldots, x_m$ мы получим разные условные распределения.

Пример. Дискретный случайный вектор (ξ_1, ξ_2) имеет следующее распределение:

$\xi_1 \backslash \xi_2$	- 1	0	1
0	0.1	0.3	0.1
1	0.2	0.2	0.1

Найти условное распределение ξ_2 при условии $\xi_1 = 1$. Вероятность того, что $\xi_1 = 1$, равна $0.2 + 0.2 + 0.1 = 0.5$. Тогда по формуле (1) находим, что условное распределение ξ_2 задается таблицей

- 1	0	1
0.4	0.4	0.2

Выпишем основные свойства условного распределения. Для упрощения обозначений ограничимся двумерным случаем.

1) $P(\xi_2 = y | \xi_1 = x) \geq 0$.
2) $\sum\limits_y P(\xi_2 = y | \xi_1 = x) = 1$.
3) $P(\xi_1 = x, \xi_2 = y) = P(\xi_1 = x) P(\xi_2 = y | \xi_1 = x)$.
4) $P(\xi_2 \in B) = \sum\limits_x P(\xi_2 \in B | \xi_1 = x) P(\xi_1 = x)$.

Как и в случае безусловных распределений, можно определить понятие математического ожидания. Пусть $\xi = (\xi_1, \dots, \xi_m, \xi_{m+1})$ – дискретный случайный вектор. **Условное математическое ожидание** с.в. ξ_{m+1} при условии, что $\xi_1 = x_1, \dots, \xi_m = x_m$, есть число

$$M(\xi_{m+1}|\xi_1 = x_1, \dots, \xi_m = x_m) =$$
$$\sum_y yP(\xi_{m+1} = y|\xi_1 = x_1, \dots, \xi_m = x_m) \ . \qquad (13.2)$$

Ясно, что при различных условиях мы будем получать разные условные математические ожидания, т.е. условное математическое ожидание ξ_{m+1} при условии $\xi_1 = x_1, \dots, \xi_m = x_m$ есть некоторая функция $g(x_1, \dots, x_m)$. Иногда нам будет удобно подставлять в эту функцию сами случайные величины $\xi_1, \dots, \xi_m$ и рассматривать условное математическое ожидание как случайную величину. В этом случае мы будем использовать обозначение $M(\xi_{m+1}|\xi_1, \dots, \xi_m)$. На самом деле последнее выражение просто равно $g(\xi_1, \dots, \xi_m)$. Функция $y = g(x_1, \dots, x_m)$ называется **функцией регрессии** случайной величины ξ_{m+1} на случайные величины $\xi_1, \dots, \xi_m$. Условное математическое ожидание, как функция от ξ_{m+1}, обладает всеми свойствами обычного математического ожидания. Но, кроме того, справедлив следующий аналог формулы полной вероятности:

$$M(\xi_{m+1}) = M[M(\xi_{m+1}|\xi_1, \dots, \xi_m)] \ . \qquad (13.3)$$

Этот результат называется **формулой полного математического ожидания**.

Задача. Доказать формулу (3).

В качестве задачи предлагается доказать следующие свойства условного математического ожидания:

1) $M(\varphi(\xi)|\xi) = \varphi(\xi)$.

2) $M(\varphi(\xi_1)\xi_2|\xi_1) = \varphi(\xi_1)M(\xi_2|\xi_1)$.

3) Если ξ_1 и ξ_2 – независимы, то для любой борелевской функции $y = \psi(x_1, x_2)$

$$M[\psi(\xi_1, \xi_2)|\xi_1 = x] = M[\psi(x, \xi_2)] \ .$$

В частности,
$$M(\xi_2|\xi_1) = M(\xi_2)\ .$$

Замечания. 1. Можно рассмотреть и более общую задачу: найти условное распределение случайного вектора (ξ_1, ξ_2) при некотором условии на обе координаты. Например, в качестве такого условия можно взять $\varphi(\xi_1, \xi_2) = y$ или $\varphi(\xi_1, \xi_2) > y$. Если распределение (ξ_1, ξ_2) задано в виде таблицы, то для того, чтобы найти условное распределение (ξ_1, ξ_2) при некотором условии, нужно выбрать те клетки таблицы, где это условие выполнено, и разделить соответствующие вероятности на сумму всех вероятностей в этих клетках, а в остальных клетках нужно поставить нули.

2. В прикладных задачах часто вначале заданы условное распределение и вероятности условий. Тогда, используя формулу полной вероятности, можно найти и безусловное распределение.

Пусть теперь случайный вектор $\xi = (\xi_1, \dots, \xi_m, \xi_{m+1}, \dots, \xi_{m+n}) = (\xi', \xi'')$ имеет абсолютно непрерывное распределение с плотностью $\rho_\xi(x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n})$. По аналогии с формулой (1) определим условную плотность подвектора $(\xi_{m+1}, \dots, \xi_{m+n})$ при условии $\xi_1 = x_1, \dots, \xi_m = x_m$ по формуле
$$\rho_{\xi''|\xi'}(x_{m+1}, \dots, x_{m+n}|x_1, \dots, x_m) = \frac{\rho_\xi(x_1, \dots, x_m, x_{m+1}, \dots, x_{m+n})}{\rho_{\xi'}(x_1, \dots, x_m)}\ . \tag{13.4}$$

При таком определении условной плотности естественно определить условное распределение по формуле

$$P_{\xi''|\xi'}(B|x_1, \dots, x_m) = \int\limits_B \rho_{\xi''|\xi'}(x_{m+1}, \dots, x_{m+n}|x_1, \dots, x_m) dx_{m+1} \dots dx_{m+n}\ , \tag{13.5}$$

где B – борелевское подмножество в R^n. В частности, мы можем определить условную функцию распределения
$$F_{\xi''|\xi'}(x_{m+1}, \dots, x_{m+n}|x_1, \dots, x_m) = P(\xi_{m+1} < x_{m+1}, \dots, \xi_{m+n} < x_{m+n}|x_1, \dots, x_m)\ . \tag{13.6}$$

Нетрудно проверить, что справедливы следующие свойства:

1) $\rho_{\xi''|\xi'}(x''|x') \geq 0$,

2) $\int\limits_{R^n} \rho_{\xi''|\xi'}(x''|x')dx'' = 1$,

3) $\rho_{\xi}(x', x'') = \rho_{\xi'}(x')\rho_{\xi''|\xi'}(x''|x')$,

4) $\rho_{\xi''}(x'') = \int\limits_{R^m} \rho_{\xi'}(x')\rho_{\xi''|\xi'}(x''|x')dx'$ – формула полной вероятности для плотностей.

Т.к. условная плотность $\rho_{\xi''|\xi'}(x''|x')$ зависит от x', то все вычисленные по ней характеристики будут функциями от x'. В частности, мы можем определить **условное математическое ожидание** ξ_{m+1} при условии $\xi' = x'$:

$$M[\xi_{m+1}|\xi' = x'] := \int\limits_{-\infty}^{\infty} y\rho_{\xi_{m+1}|\xi'}(y|x')dy = g(x') \ . \qquad (13.7)$$

Функция $y = g(x')$ называется **функцией регрессии** с.в. ξ_{m+1} на случайный вектор ξ'. Если в этом определении сделать условие случайным, т.е. рассмотреть $g(\xi')$, то мы получим новую случайную величину, которую вновь будем называть условным математическим ожиданием ξ_{m+1} относительно случайного вектора ξ' и использовать для него обозначение $M[\xi_{m+1}|\xi_1, \ldots, \xi_m]$. Оно обладает всеми обычными свойствами математического ожидания по аргументу ξ_{m+1}. Но, кроме того, справедливы следующие свойства, которые предлагается доказать самостоятельно.

Задача. 1) $M[\varphi(\xi')|\xi'] = \varphi(\xi')$.

2) $M[\xi_{m+1}\varphi(\xi')|\xi'] = \varphi(\xi')M[\xi_{m+1}|\xi']$.

3) $M(\xi_{m+1}) = M\{M[\xi_{m+1}|\xi']\}$ – формула полного математического ожидания.

4) Если ξ_{m+1} и ξ' – независимы, то

$$M[\psi(\xi', \xi_{m+1})|\xi' = x'] = M[\psi(x', \xi_{m+1})] \ ,$$

в частности,

$$M[\xi_{m+1}|\xi'] = M(\xi_{m+1}) \ .$$

Без доказательства рассмотрим несколько более общую ситуацию, которая иллюстрирует смысл условного математического ожидания и формулу полного математического ожидания. Для простоты рассмотрим только двумерный случай. Пусть (ξ_1, ξ_2) – двумерный случайный вектор и $\eta_1 = \psi_1(\xi_1, \xi_2)$ – некоторая борелевская функция от этого случайного вектора. Это новая случайная величина. Уравнение $y = \psi_1(x_1, x_2)$ выделяет некоторую линию уровня этой функции. Вся плоскость будет разбита на такие линии. Рассмотрим еще одну случайную величину $\eta_2 = \psi_2(\xi_1, \xi_2)$. В силу формулы полного математического ожидания мы имеем

$$M\eta_2 = M[\psi_2(\xi_1, \xi_2)] = M\{M[\psi_2(\xi_1, \xi_2)|\psi_1(\xi_1, \xi_2)]\} \ .$$

Смысл этого результата состоит в следующем. Сначала мы находим условное математическое ожидание вдоль каждой линии, а затем усредняем полученные результаты по всем линиям. Частный случай этого результата известен из курса математического анализа – это известная формула Фубини для вычисления кратных интегралов.

Используя сформулированные выше свойства, можно доказать следующие полезные для практических применений результаты.

Задача. Пусть (ξ_1, ξ_2) – случайный вектор. Тогда

1) $M(\xi_2) = M[M(\xi_2|\xi_1)]$;

2) $D(\xi_2) = M[D(\xi_2|\xi_1)] + D[M(\xi_2|\xi_1)]$.

Чаще всего этот результат применяется к случайным суммам, например в теории страхования. Пусть $X_1, X_2, \ldots$ – последовательность независимых и одинаково распределенных с.в., а N – независимая от них с.в., принимающая целые неотрицательные значения. Тогда величина $S_N = X_1 + \cdots + X_N$ $(S_0 = 0)$ называется случайной суммой. Используя предыдущую задачу, можно доказать следующее.

Задача. 1) $M(S_N) = M(X_1) \cdot M(N)$.

2) $D(S_N) = D(X_1) \cdot M(N) + (M(X_1))^2 \cdot D(N)$.

В заключение мы рассмотрим одно экстремальное свойство функции регрессии, которое будет использоваться в математической статистике. В качестве разминки предлагается решить следующую задачу.

Задача. Пусть ξ – с.в. с конечным вторым моментом. Тогда $\forall a \in R^1$

$$D(\xi) = M(\xi - M\xi)^2 \le M(\xi - a)^2 = f(a) \ ,$$

т.е. функция $f(a)$ принимает наименьшее значение при $a = M\xi$.

Аналогичный результат справедлив и для условных математических ожиданий.

Теорема 1 *(***экстремальное свойство функции регрессии***). Пусть ξ и η – две случайные величины, $y = f(x)$ – некоторая борелевская функция, причем $M(\eta^2) < \infty$ и $M([f(\xi)]^2) < \infty$. Если $y = g(x) = M[\eta|\xi = x]$ есть функция регрессии с.в. η на с.в. ξ, то*

$$M|\eta - g(\xi)|^2 \le M|\eta - f(\xi)|^2 \ .$$

Доказательство. В силу свойств условных математических ожиданий и сформулированной выше задачи при каждом x мы имеем

$$M[(\eta - g(\xi))^2|\xi = x] = M[(\eta - g(x))^2|\xi = x] \le$$
$$M[(\eta - f(x))^2|\xi = x] = M[(\eta - f(\xi))^2|\xi = x] \ .$$

Усредняя по условию, из формулы полного математического ожидания мы получаем нужный результат.

На самом деле мы доказали даже больше. А именно, что нужное неравенство справедливо для условных математических ожиданий. Объясним смысл полученного результата. По условию с.в. $\eta \in L_2$. Рассмотрим множество $\mathcal{L}$ с.в., которые можно представить в виде $f(\xi)$, причем $M(|f(\xi)|^2) < \infty$. Можно проверить, что $\mathcal{L}$ – замкнутое линейное подпространство в L_2. Теорема утверждает, что наилучшее приближение η в пространстве $\mathcal{L}$ (т.е. с помощью функций от ξ) дает нам условное математическое ожидание $g(\xi)$. Ранее мы получили аналогичный результат для линейных функций от ξ. Если рассмотреть произвольное линейное подпространство $\mathcal{L}$ в L_2, то естественно назвать наилучшее приближение для с.в. η в пространстве $\mathcal{L}$ **условным математическим ожиданием в широком смысле**.

Наиболее важным примером в статистике является многомерное нормальное распределение. Для простоты вычислений мы рассмотрим только двумерный случай.

Пример (двумерное нормальное распределение). Напомним, что случайный вектор $\xi = (\xi_1, \xi_2)$ имеет двумерное нормальное распределение, если он имеет плотность распределения следующего вида:

$$\begin{aligned}\rho_\xi(x_1, x_2) &= \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}}\exp\{-\frac{1}{2(1-\rho^2)}[\frac{(x_1-a_1)^2}{\sigma_1^2} - \\ &\quad -2\rho\frac{(x_1-a_1)}{\sigma_1}\cdot\frac{(x_2-a_2)}{\sigma_2} + \frac{(x_2-a_2)^2}{\sigma_2^2}]\}\ , (x_1, x_2) \in R^2\ .\end{aligned}$$

Выделяя полный квадрат по x_2 и интегрируя по этой переменной, мы получаем, что $\xi_1 \in N(a_1, \sigma_1^2)$. Аналогичный результат справедлив и для ξ_2. Отсюда мы получаем, что $a_1 = M\xi_1,\ a_2 = M\xi_2,\ \sigma_1^2 = D(\xi_1),\ \sigma_2^2 = D(\xi_2)$. Вычисляя соответствующий двумерный интеграл, можно показать, что $\rho = \rho(\xi_1, \xi_2)$. Вычислим теперь условное распределение ξ_2 при условии $\xi_1 = x_1$. По формуле (4) после стандартных вычислений получаем

$$\begin{aligned}\rho_{\xi_2|\xi_1}(x_2|x_1) &= \frac{\rho_{\xi_1}(x_1, x_2)}{\rho_{\xi_1}(x_1)} = \\ &= \frac{1}{\sqrt{2\pi\sigma_2^2(1-\rho^2)}}\cdot\exp\{-\frac{[x_2-(a_2+\rho\frac{\sigma_2}{\sigma_1}(x_1-a_1))]^2}{2\sigma_2^2(1-\rho^2)}\}\ .\end{aligned}$$

Таким образом, мы видим, что условное распределение ξ_2 при условии $\xi_1 = x_1$ является нормальным со средним

$$M(\xi_2|\xi_1 = x_1) = g(x_1) = a_2 + \rho\frac{\sigma_2}{\sigma_1}(x_1 - a_1)$$

и дисперсией

$$D(\xi_2|\xi_1 = x_1) = \sigma_2^2(1-\rho^2)\ .$$

Условное среднее является линейной функцией от x_1, а условная дисперсия не зависит от условия. Кроме того, в силу доказанной выше теоремы мы видим, что наилучшее приближение ξ_2 с помощью функций от ξ_1 является линейным в случае нормального распределения. Теперь мы можем записать с.в. ξ_2 в виде

$$\xi_2 = a_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(\xi_1 - a_1) + \varepsilon \ , \tag{13.8}$$

где

$$\varepsilon = \xi_2 - [a_2 + \rho \cdot \frac{\sigma_2}{\sigma_1}(\xi_1 - a_1)] \ .$$

Оба слагаемых в представлении (8) имеют нормальное распределение и некоррелированы (а значит, и независимы). Первое слагаемое описывает влияние на ξ_2 с.в. ξ_1, а второе – влияние каких-то других, неучтенных факторов.

1.14 Многомерное нормальное распределение

Нормальное распределение играет важную роль в теории вероятностей и математической статистике. В этом параграфе мы дадим определение и перечислим без доказательства основные свойства многомерного нормального распределения.

Определение 1 . *Случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет* **многомерное нормальное распределение**, *если он обладает плотностью распределения следующего вида:*

$$\rho_\xi(x_1, \ldots, x_n) = (2\pi)^{-\frac{n}{2}} det(A) \exp\{-\frac{1}{2}(A(x-m), x-m)\} =$$

$$(2\pi)^{-\frac{n}{2}} det(A) \exp\{-\frac{1}{2}\sum_{i,j=1}^{n} a_{ij}(x_i - m_i)(x_j - m_j)\} \ ,$$

где $m = (m_1, \ldots, m_n) \in R^n$, $A = (a_{ij})$ – симметричная положительно определенная $n \times n$-матрица.

Выясним смысл параметров многомерного нормального распределения. Пусть $\Sigma = (\sigma_{ij}) = A^{-1}$. Тогда $m_j = M\xi_j$,

$$\sigma_{ij} = cov(\xi_i, \xi_j) = M(\xi_i - m_i)(\xi_j - m_j) \ .$$

m называется **вектором средних**, а Σ – **матрицей ковариаций**. Если с.в. ξ имеет многомерное нормальное распределение с вектором средних m и матрицей ковариаций Σ, то мы будем использовать обозначение $\xi \in N(m, \Sigma)$.

Теорема 1 . *Если $C : R^n \to R^k$ – линейное отображение R^n на R^k, $\xi \in N(m, \Sigma)$, то $C\xi \in N(Cm, C\Sigma C^T)$.*

Разобьем случайный вектор ξ на два подвектора размерности n_1 и n_2: $n_1 + n_2 = n$, т.е. $\xi = (\xi', \xi'')$. Это порождает разложения вектора средних $m = (m', m'')$ и матрицы ковариаций

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} .$$

Теорема 2 . *Если $\xi = (\xi_1, \ldots, \xi_n) \in N(m, \Sigma)$, то*

1) $\xi' \in N(m', \Sigma_{11})$,

2) ξ' и ξ'' – независимы тогда и только тогда, когда

$$\Sigma_{12} = \Sigma_{21}^T = 0.$$

Теорема 3 . *Если $\xi = (\xi_1, \ldots, \xi_n) \in N(m, \Sigma)$, то условное распределение ξ'' при условии $\xi' = x'$ является многомерным нормальным распределением с вектором средних*

$$m'' + \Sigma_{21}\Sigma_{11}^{-1}(x' - m')$$

и матрицей ковариаций

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} .$$

1.15 Закон больших чисел

В самом начале нашего курса мы отметили следующий экспериментальный факт. В длинной серии независимых и одинаковых испытаний частота появления некоторого события, которая является случайной величиной, тяготеет к некоторому постоянному числу, которое мы назвали вероятностью этого события. Далее мы фиксировали в виде аксиом основные свойства вероятностей и в качестве следствия доказали более сложные свойства вероятностей. Но хорошая математическая теория должна не только аккумулировать в своих определениях те или иные экспериментальные факты, но и содержать некоторые теоремы, которые дают описание наиболее важных результатов, отмеченных первоначально эмпирически. Свойство устойчивости частот говорит об определенном типе сходимости случайных величин. Но это не есть сходимость в привычном для нас смысле, как это принято в математическом анализе. Если в некоторой серии испытаний событие A не произойдет ни разу, то его частота равна нулю. Наоборот, если оно будет происходить всегда, то его частота равна 1. Но таких серий будет немного. Эксперимент показывает, что *большинство* серий испытаний будут такими, что частота появления события A в каждой из них будет примерно одинаковой. Таким образом, наше утверждение об устойчивости частот само носит вероятностный характер. Эти предварительные рассуждения приводят нас к следующему определению.

Определение 1 . *Последовательность с.в.* $\{\xi_n\}$ **сходится по вероятности** *к с.в.* ξ, *если* $\forall \varepsilon > 0$

$$P(|\xi_n - \xi| > \varepsilon) \to 0 \ ,$$

или

$$P(|\xi_n - \xi| \leq \varepsilon) \to 1.$$

при $n \to \infty$.

Такая сходимость обозначается $\xi_n \xrightarrow{P} \xi$, $n \to \infty$.

В этом определении утверждается, что событие, состоящее в том, что ξ_n и ξ близки $(|\xi_n - \xi| \leq \varepsilon)$, имеет вероятность, близкую к 1, т.е. это *практически достоверное* событие.

Вернемся теперь к свойству устойчивости частот. Пусть ε_k равно 1, если в k - м испытании событие A произошло, и равно 0 – в противном случае. Тогда $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ есть число появлений события A в n испытаниях, а S_n/n есть относительная частота появления события A. Легко вычислить, что

$$M(\frac{S_n}{n}) = \frac{M\varepsilon_1 + \ldots + M\varepsilon_n}{n} = p \ (= P(A)) \ .$$

Тогда свойство устойчивости относительных частот означает, что

$$\frac{S_n}{n} - \frac{MS_n}{n} = \frac{S_n}{n} - p \to 0 \ , \ n \to \infty \ ,$$

в некотором смысле. Используя эти соображения, дадим следующее

Определение 2 . *Говорят, что к последовательности с.в.* $\{\xi_n\}$ *применим* **закон больших чисел** *(ЗБЧ), если*

$$\frac{\xi_1 + \ldots + \xi_n}{n} - \frac{M\xi_1 + \ldots + M\xi_n}{n} \xrightarrow{P} 0 \ , \ n \to \infty \ .$$

В определении дано только описание некоторого свойства. Ниже мы приведем условия, при которых это свойство имеет место. По определению нам нужно доказать сходимость по вероятности. Полезным техническим средством при этом является неравенство Чебышева

$$P(|\xi - M\xi| > \varepsilon) \leq \frac{D(\xi)}{\varepsilon^2} \ ,$$

которое было доказано ранее.

Лемма 1 . *Пусть с.в.* ξ_n *имеют конечные математические ожидания* $M\xi_n$ *и дисперсии* $D(\xi_n)$. *Если* $D(\xi_n) \to 0, \ n \to \infty$, *то* $\xi_n - M\xi_n \xrightarrow{P} 0$.

Доказательство следует непосредственно из определения и неравенства Чебышева.

Теорема 1 . *Пусть* $\{\xi_n\}$ *– последовательность случайных величин, таких, что*

1) $\{\xi_n\}$ *– независимы,*

2) существуют конечные $M\xi_n$ *и* $D(\xi_n)$, *3)* $n^{-2}\sum\limits_{k=1}^{n} D(\xi_k) \to 0\ ,\ n\to\infty.$

Тогда применим ЗБЧ, т.е.

$$\frac{\xi_1+\ldots+\xi_n}{n} - \frac{M\xi_1+\ldots+M\xi_n}{n} \xrightarrow{P} 0\ ,\ n\to\infty\ .$$

Доказательство. Положим, $S_n = \xi_1+\ldots+\xi_n$ и

$$\overline{S}_n = \frac{\xi_1+\ldots+\xi_n}{n}\ .$$

Тогда

$$M(\overline{S}_n) = \frac{M\xi_1+\ldots+M\xi_n}{n}\ ,\ D(\overline{S}_n) = \frac{1}{n^2}\sum_{k=1}^{n} D(\xi_k) \to 0\ .$$

Утверждение теоремы следует теперь из сформулированной выше леммы.

Замечание. Условие 1 нашей теоремы можно значительно ослабить, потребовав только попарной *некоррелированности*.

Теорема 2 . *Пусть $\{\xi_n\}$ – последовательность случайных величин:*

1) $\{\xi_n\}$ – независимы,

2) существуют конечные $M\xi_n$ и $D(\xi_n)$,

3) $\forall n\ D(\xi_n) \le C < \infty$.

Тогда применим ЗБЧ.

Доказательство. Достаточно проверить условие 3 теоремы 1.

$$n^{-2}\sum_{k=1}^{n} D(\xi_k) \le n^{-2}\cdot n\cdot C = C/n \to 0\ ,\ n\to\infty\ .$$

Теорема 3 . *Пусть $\{\xi_n\}$ – последовательность случайных величин:*

1) $\{\xi_n\}$ – независимы,

2) $\{\xi_n\}$ – одинаково распределены,

3) $\exists M\xi_n = a,\ D(\xi_n) = \sigma^2 < \infty$.

Тогда применим ЗБЧ, т.е.

$$\frac{\xi_1+\ldots+\xi_n}{n} \xrightarrow{P} a\ ,\ n\to\infty\ .$$

Это частный случай теоремы 2.

Пример. Для измерения некоторой постоянной величины произвели n независимых измерений в одинаковых условиях. Каждое измерение ξ_n можно представить в виде $\xi_n = \varepsilon_n + a$, где $\{\varepsilon_n\}$ – последовательность н.о.р.с.в.: $M(\varepsilon_n) = 0,\ D(\varepsilon_n) = \sigma^2$. Тогда, в силу ЗБЧ,

$$\frac{\xi_1 + \ldots + \xi_n}{n} = \frac{\varepsilon_n + \ldots + \varepsilon_n}{n} + a \xrightarrow{P} a\ ,\ n \to \infty\ ,$$

т.е. среднее арифметическое большого числа измерений сходится к истинному значению.

Замечание. А.Я. Хинчин получил более сильный результат: ЗБЧ справедлив для последовательности н.о.р.с.в. с конечным математическим ожиданием, т.е. условие конечности дисперсии является излишним.

В заключение приведем два исторически первых результата из числа ЗБЧ.

Теорема 4 *(**Я. Бернулли, 1713**). Пусть S_n/n – относительная частота появления события A в n независимых испытаниях, вероятность которого равна p. Тогда*

$$\frac{S_n}{n} \xrightarrow{P} p\ .$$

Эта теорема есть частный случай предыдущей теоремы. Действительно, выше мы показали, что $S_n = \varepsilon_1 + \cdots + \varepsilon_n$, где $\{\varepsilon_n\}$ – н.о.р. и $M\varepsilon_n) = p,\ D(\varepsilon_n) = p(1-p) \leq 1/4$.

В некоторых ситуациях вероятность появления события A меняется от испытания к испытанию. Тогда имеет место

Теорема 5 *(**Пуассон, 1837**). Пусть мы имеем последовательность независимых испытаний, где p_k – вероятность появления события A в k - м испытании. Если S_n – число появлений события A в n испытаниях, то*

$$\frac{S_n}{n} - \frac{p_1 + \ldots + p_n}{n} \xrightarrow{P} 0\ .$$

1.16 Центральная предельная теорема

1.16.1 Постановка задачи

Во многих теоретических и прикладных задачах теории вероятностей мы встречаемся с ситуацией, когда нам нужно найти распределение суммы независимых величин. Формально мы знаем ее решение. Нужно последовательно несколько раз применить формулу свертки. Но здесь мы встречаем два препятствия. Во-первых, для вычисления свертки нам нужно точно знать аналитический вид каждого слагаемого. Во-вторых, вычисление свертки довольно сложная операция, не всегда реализуемая в конкретной задаче. Во многих науках эти трудности преодолеваются путем построения тех или иных простых аппроксимаций для более сложных моделей. То же самое делают и в теории вероятностей. Можно даже сказать, что основная часть теории вероятностей посвящена описанию таких аппроксимаций. Одной из наиболее известных является так называемая *центральная предельная теорема* (ЦПТ), к изучению которой мы и приступаем.

Чтобы понять суть проблемы, вернемся к простой модели последовательности независимых испытаний, которую мы изучали ранее. Пусть мы имеем последовательность независимых испытаний с двумя исходами и постоянной вероятностью успеха p. Обозначим через ε_k случайную величину, равную 1, если в k-м испытании был успех, и 0 – в противном случае. Тогда число успехов в n испытаниях есть случайная величина $S_n = \varepsilon_1 + \ldots + \varepsilon_n$. Она имеет биномиальное распределение с параметрами n и p. При больших n расчеты по формулам биномиального распределения становятся довольно сложными. Поэтому применяют те или иные приближенные формулы. Одну из таких формул дает теорема Пуассона, другую – теорема Муавра – Лапласа. Рассмотрим нормированное число успехов

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} ,$$

где $np = M(S_n),\ np(1-p) = D(S_n)$.

Теорема Муавра – Лапласа. *При $n \to \infty$ для $x \in R^1$ имеет место сходимость*

$$P(S_n^* < x) = F_{S_n^*}(x) \to \Phi(x) \ ,$$

где

$$\Phi(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy$$

– функция распределения стандартного нормального закона.

Таким образом, эта теорема утверждает, что при большом n и фиксированном p распределение соответствующим образом нормированной суммы случайных величин аппроксимируется нормальным распределением. Нормальное распределение хорошо изучено, обладает целым рядом привлекательных свойств, для него составлены таблицы. Все это позволяет надеяться, что такая аппроксимация будет удобной в практических задачах. Хотелось бы получить подобный результат и в более общей ситуации. Для этого отметим существенные черты полученного результата. Мы имеем последовательность независимых с.в. $\{\xi_n\}$, из которых образована последовательность сумм $S_n = \xi_1 + \ldots + \xi_n, \ n \geq 1$. Если эти суммы соответствующим образом нормировать и центрировать, т. е. рассмотреть

$$S_n^* = \frac{S_n - A_n}{B_n} \ ,$$

где $A_n \in R^1, \ B_n > 0$, то функции распределения нормированных сумм хорошо аппроксимируются функцией распределения стандартного нормального закона, а именно

$$F_{S_n^*}(x) \to \Phi(x) \ , \ n \to \infty \ .$$

Это приводит нас к следующему определению.

Определение 1 . *Говорят, что к последовательности случайных величин $\{\xi_n\}$ применима* **центральная предельная теорема** *(ЦПТ), если для любого $n \geq$ существуют центрирующие и нормирующие* *константы* $A_n \in R^1$ *и* $B_n > 0$*, такие, что для*

$$S_n^* = \frac{\xi_1 + \ldots + \xi_n - A_n}{B_n} \qquad (16.1)$$

имеет место сходимость

$$P(S_n^* < x) = F_{S_n^*}(x) \to \Phi(x) \qquad (16.2)$$

при $n \to \infty$ *для любого* $x \in R^1$*.*

Замечания. 1. Обычно в реальных задачах $A_n = M(S_n)$, $B_n = D(S_n)$. При такой нормировке мы получаем $M(S_n^*) = 0$, $D(S_n^*) = 1$.

2. В этом определении сформулировано некоторое свойство последовательности с.в. Но нас, конечно, интересует, при каких условиях оно имеет место. Утверждения, которые обычно называют ЦПТ, как раз и содержат те условия, при которых это свойство выполнено.

В ЦПТ мы изучаем сходимость не самих случайных величин а соответствующих им функций распределения. Поэтому в дальнейшем будет полезно следующее определение.

Определение 2 . *Последовательность функций распределения* $\{F_n\}$ **слабо сходится** *к функции распределения* F_0*, если*

$$F_n(x) \to F_0(x)$$

при $n \to \infty$ *для всех точек* x*, где предельная ф.р.* F_0 *непрерывна. Для соответствующих с.в.* ξ_n *будем говорить, что они сходятся к с.в.* ξ_0 **по распределению***, и писать*

$$\xi_n \xrightarrow{d} \xi_0 \; .$$

Так как функция распределения нормального закона всюду непрерывна, то мы изучаем обычную сходимость функций распределений.

В ЦПТ рассматривается сходимость распределений для сумм. Даже если слагаемые независимы, то нужно применять сложную операцию – свертку распределений. Поэтому при доказательстве ЦПТ применяют особый метод – **метод интегральных преобразований**. Он оказывается полезным и в других задачах.

1.16.2 Характеристические и производящие функции

Характеристические функции в теории вероятностей применяли уже Лаплас и Коши. Но систематическое их использование началось только после того, как А.М. Ляпунов (ученик П.Л. Чебышева) применил их в 1901 году для доказательства очень сильного варианта ЦПТ. Сегодня метод характеристических функций один из основных инструментов теории вероятностей.

Определение 3 . **Характеристической функцией** *с.в. ξ (или соответствующей ф.р. F) называется комплекснозначная функция $\varphi_\xi(t)$, $t \in R^1$, определяемая по правилу*

$$\varphi_\xi(t) = M(e^{it\xi}) = \int\limits_{-\infty}^{\infty} e^{itx} dF(x) \ . \tag{16.3}$$

Для дискретной с.в. ξ мы имеем формулу

$$\varphi_\xi(t) = \sum_n e^{itx_n} \cdot p_n \ , \tag{16.4}$$

а для непрерывной с.в. ξ с плотностью распределения $\rho_\xi(x)$ –

$$\varphi_\xi(t) = \int\limits_{-\infty}^{\infty} e^{itx} \rho_\xi(x) dx \ . \tag{16.5}$$

Напомним, что формула (5) есть обычное преобразование Фурье для плотности $\rho_\xi(x)$.

Начнем с того, что перечислим некоторые простейшие свойства характеристических функций.

Теорема 1 . *1.* $\forall \in R^1$ $|\varphi_\xi(t)| \leq 1 = \varphi_\xi(0)$.

2. $\varphi_{-\xi}(t) = \varphi_\xi(-t) = \overline{\varphi_\xi(t)}$.

3. $\varphi_\xi(t)$ *– равномерно непрерывная функция.*

4. Функция $\varphi_\xi(t)$ положительно определена, т.е. $\forall t_1, \ldots, t_n \in R^1$, $\forall c_1, \ldots, c_n \in C$

$$\sum_{k,m=1}^{n} \varphi_\xi(t_k - t_m) \cdot c_k \cdot \overline{c_m} \geq 0 \ .$$

5. Если $\eta = a\xi + b$, то $\varphi_\eta(t) = \varphi_\xi(at) \cdot e^{itb}$.

6. Если ξ_1 и ξ_2 независимы, то

$$\varphi_{\xi_1+\xi_2}(t) = \varphi_{\xi_1}(t) \cdot \varphi_{\xi_2}(t) \ .$$

7. Если $M(|\xi|^k) < \infty$ для некоторого натурального k, то существует k-я производная для функции $\varphi_\xi(t)$ и

$$\frac{d^k}{dt^k}\varphi_\xi(t)|_{t=0} = (i)^k \cdot M(\xi^k) \ .$$

Доказательство. 1. Для любого $t \in R^1$

$$\begin{aligned} |\varphi_\xi(t)| &= |M(e^{it\xi})| \le M|e^{it\xi}| = M(1) = 1 = \\ &= M(e^{i\cdot 0\cdot\xi}) = \varphi_\xi(0) \ . \end{aligned}$$

2. $\varphi_{-\xi}(t) = M(e^{it(-\xi)}) = M(e^{-it\xi}) = M(\overline{e^{it\xi}}) = \overline{M(e^{it\xi})} = \overline{\varphi_\xi(t)}$.

3. Пусть $t \in R^1, \ h > 0$. Тогда

$$\varphi_\xi(t+h) - \varphi_\xi(t) = M[e^{i(t+h)\xi} - e^{it\xi}] =$$

$$M[e^{it\xi}(e^{ih\xi} - 1)] = \int\limits_{-\infty}^{\infty} e^{itx}(e^{ihx} - 1)dF_\xi(x) \ .$$

Так как $F_\xi(x) \to 0$ при $x \to -\infty$ и $F_\xi(x) \to 1$ при $x \to \infty$, то для любого $\varepsilon > 0$ можно найти $A > 0$, такое, что

$$F_\xi(-A) < \varepsilon/6 \ , \ 1 - F_\xi(A) < \varepsilon/6 \ .$$

Тогда мы получаем

$$|\varphi_\xi(t+h) - \varphi_\xi(t)| \le \int\limits_{-\infty}^{\infty} |e^{ihx} - 1|dF_\xi(x) =$$

$$\int\limits_{|x|\le A} |e^{ihx} - 1|dF_\xi(x) + \int\limits_{|x|>A} |e^{ihx} - 1|dF_\xi(x) \le$$

$$\varepsilon/3 \cdot P(|\xi| \le A) + 2P(|\xi| > A) = \varepsilon/3 + 2\varepsilon/3 = \varepsilon \ ,$$

если $h < \delta$. На последнем шаге мы использовали свойство непрерывности в нуле функции e^{iy} и то, что hx можно сделать равномерно малым при $|x| \le A$ за счет выбора малого h.

4. Пусть $t_1,\dots,t_n\in R^1,\ c_1,\dots,c_n\in C$. Тогда

$$\sum_{k,m=1}^{n}\varphi_\xi(t_k-t_m)\cdot c_k\cdot\overline{c_m}=\sum_{k,m=1}^{n}M(e^{i(t_k-t_m)\xi})c_k\overline{c_m}=$$

$$M(\sum_{k=1}^{n}\sum_{m=1}^{n}e^{it_k\xi}c_k\overline{e^{it_m\xi}c_m})=M(\sum_{k=1}^{n}e^{it_k\xi}c_k)\overline{(\sum_{m=1}^{n}e^{it_k\xi}c_m)}=$$

$$M|\sum_{k=1}^{n}e^{it_k\xi}c_k|^2\geq 0\ .$$

5. Пусть $\eta=a\xi+b$, где $a,b\in R^1$. Тогда

$$\varphi_\eta(t)=M(e^{it\eta})=M(e^{i(at)\xi}e^{itb})=e^{itb}\varphi_\xi(at)\ .$$

6. Если ξ_1 и ξ_2 – независимы, то с.в. $\eta_1=e^{it\xi_1}$ и $\eta_2=e^{it\xi_2}$ также будут независимы. Отсюда мы получаем

$$\varphi_{\xi_1+\xi_2}(t)=M(e^{it(\xi_1+\xi_2)})=M(e^{it\xi_1}e^{it\xi_2})=$$

$$M(e^{it\xi_1})M(e^{it\xi_2})=\varphi_{\xi_1}(t)\varphi_{\xi_2}(t)\ .$$

Последние два свойства показывают, что характеристические функции для линейных преобразований и для сумм независимых случайных величин легко вычисляются.

7. Продифференцируем формально выражение для характеристической функции:

$$\begin{aligned}\frac{d^k}{dt^k}\varphi_\xi(t)&=\frac{d^k}{dt^k}(\int_{-\infty}^{\infty}e^{itx}dF_\xi(x))=\\&=\int_{-\infty}^{\infty}(ix)^ke^{itx}dF_\xi(x)\ .\end{aligned}$$

Последний интеграл сходится абсолютно, если $M(|\xi|^k)<\infty$, так как

$$|\int_{-\infty}^{\infty}(ix)^ke^{itx}dF_\xi(x)|\leq\int_{-\infty}^{\infty}|x|^kdF_\xi(x)=M(|\xi|^k)<\infty\ .$$

Известный результат из математического анализа гарантирует нам тогда возможность дифференцирования по параметру под знаком интеграла. При $t = 0$ получаем

$$\frac{d^k}{dt^k}\varphi_\xi(t)|_{t=0} = i^k \int\limits_{-\infty}^{\infty} x^k dF_\xi(x) = i^k M(\xi^k) \ . \tag{16.6}$$

В качестве следствия свойства 7 получаем следующее разложение характеристической функции:

$$\varphi_\xi(t) = 1 + M\xi \cdot \frac{it}{1!} + \ldots + M(\xi^k) \cdot \frac{(it)^k}{k!} + o(|t|^k) \tag{16.7}$$

при $t \to 0$.

Следующие свойства характеристических функций доказываются более громоздко, поэтому они будут приведены без доказательства. В силу своей важности они выделены в отдельные теоремы.

Теорема 2 (**единственности**). *Соответствие между функциями распределениями F и характеристическими функциями φ является взаимно однозначным. Более того, если x_1 и x_2 – точки непрерывности ф.р. F, то*

$$F(x_2) - F(x_1) = \lim_{A\to\infty} \frac{1}{2\pi} \int\limits_{-A}^{A} \frac{e^{-itx_2} - e^{-itx_1}}{-it}\varphi(t)dt \ . \tag{16.8}$$

В частности, если существует плотность $\rho(x)$, то

$$\rho(x) = \frac{1}{2\pi} \int\limits_{-\infty}^{\infty} e^{-itx}\varphi(t)dt \ . \tag{16.9}$$

Соотношения (8) и (9) называются **формулами обращения**. Эта теорема показывает, что по характеристической функции можно узнать, с каким распределением мы имеем дело.

Теорема 3 (**непрерывности**). *Пусть $F_0, F_1, F_2, \ldots$ – последовательность ф.р., а $\varphi_0, \varphi_1, \varphi_2, \ldots$ – соответствующая ей последовательность х.ф. Тогда следующие утверждения эквивалентны:*

1) F_n слабо сходится к F_0,

2) $\varphi_n(t) \to \varphi_0(t)$ при $n \to \infty$ для любого $t \in R^1$.

Эта теорема позволяет сводить задачу о слабой сходимости функций распределения к эквивалентной ей (а технически более простой) задаче о сходимости характеристических функций.

На практических занятиях будут вычислены характеристические функции некоторых стандартных распределений. Приведем без доказательства окончательные результаты.

Примеры. 1) Если $P(\xi = a) = 1$ для некоторого $a \in R^1$, то

$$\varphi_\xi(t) = e^{ita} \ .$$

2) Если ξ имеет распределение Бернулли с параметром p, т.е.

$$\xi = \begin{cases} 1 & , \quad \text{с вероятностью } p \ , \\ 0 & , \quad \text{с вероятностью } 1-p \ , \end{cases}$$

то

$$\varphi_\xi(t) = 1 + p(e^{it} - 1) \ .$$

3) Если ξ имеет биномиальное распределение с параметрами n и p, то

$$\varphi_\xi(t) = [1 + p(e^{it} - 1)]^n \ .$$

4) Если ξ имеет распределение Пуассона с параметром λ, то

$$\varphi_\xi(t) = e^{\lambda(e^{it} - 1)} \ .$$

5) Если ξ имеет геометрическое распределение с параметром p, то

$$\varphi_\xi(t) = \frac{p}{1 - (1-p)e^{it}} \ .$$

6) Если ξ имеет равномерное распределение на $[a, b]$, то

$$\varphi_\xi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)} \ .$$

В частности, для равномерного распределения на $[-a, a]$ имеем

$$\varphi_\xi(t) = \frac{\sin at}{at} \ .$$

7) Если ξ имеет показательное распределение с параметром λ, то

$$\varphi_\xi(t) = \frac{\lambda}{\lambda - it} \; .$$

8) Если ξ имеет нормальное распределение с параметрами a и σ^2, то

$$\varphi_\xi(t) = e^{ita - \frac{\sigma^2 t^2}{2}} \; .$$

В частности, для стандартного нормального распределения
($a = 0$, $\sigma^2 = 1$) имеем

$$\varphi_\xi(t) = e^{-\frac{1}{2}t^2} \; .$$

Для дискретных случайных величин, принимающих значения вида $a + h \cdot n$, $n \in N$, (в частности, целочисленных) чаще используют так называемые производящие функции.

Определение 4 . *Пусть с.в.* ξ *принимает значения* $0, 1, 2, \ldots$ *с вероятностями* $p_k = P(\xi = k)$. **Производящая функция** *с.в.* ξ *(дискретного распределения* $\{p_k\}$*) определяется по формуле*

$$P_\xi(z) = \sum_{k=0}^{\infty} p_k z^k = M(z^\xi) \; . \tag{16.10}$$

Здесь z – комплексное число. Ряд (10) сходится, по крайней мере, для $|z| \leq 1$. Производящие функции обладают свойствами, аналогичными свойствам характеристических функций, которые легко получить, используя следующее соответствие между этими функциями: если $z = e^{it}$, то

$$\varphi_\xi(t) = P_\xi(z) = P_\xi(e^{it}) \; .$$

Название "производящая функция" связано со следующим ее свойством:

$$p_k = P(\xi = k) = \frac{1}{k!} \frac{d^k}{dz^k} P_\xi(z)|_{z=0} \; . \tag{16.11}$$

Функция $P_\xi(z)$ **производит** вероятности p_k по формуле (11).

Если мы работаем с моментами, то более удобно использовать **производящую функцию моментов**:

$$M_\xi(t) = M(e^{t\xi}) \ . \tag{16.12}$$

Может случиться, что $M_\xi(t)$ существует только при $t = 0$. Если $M_\xi(t) < \infty$ для всех $|t| \le r$, где $r > 0$, то существуют все моменты, они однозначно определяют распределение, и имеет место следующий аналог формул (6) и (11):

$$\frac{d^k}{dt^k} M_\xi(t)|_{t=o} = M(\xi^k) \ . \tag{16.13}$$

1.16.3 Центральная предельная теорема

В этом разделе мы рассмотрим несколько вариантов ЦПТ. Доказательство будет приведено только в простейшей ситуации, когда мы имеем последовательность независимых одинаково распределенных величин. В более сложных случаях доказательство, по-существу, то же, но технически труднее.

Теорема 4 . *Пусть последовательность с.в.* $\{\xi_n\}$ *удовлятворяет условиям:*

1) $\{\xi_n\}$ *– независимы,*

2) $\{\xi_n\}$ *– одинаково распределены,*

3) существуют конечные $M(\xi_1) = a$ *и* $D(\xi_1) = \sigma^2$.

Тогда к этой последовательности применима ЦПТ. Более точно, если

$$S_n^* = \frac{\xi_1 + \ldots + \xi_n - na}{\sigma\sqrt{n}} \ ,$$

то

$$F_{S_n^*}(x) \to \Phi(x) \ , \ n \to \infty \ .$$

Доказательство. Для доказательства будем использовать характеристические функции. В силу теоремы непрерывности достаточно доказать, что

$$\varphi_{S_n^*}(t) \to e^{-\frac{1}{2}t^2} \ , \ n \to \infty \ .$$

Запишем S_n^* в виде

$$S_n^* = \sum_{k=1}^{n} \frac{\xi_k - a}{\sigma\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \xi_k^* \ .$$

По условию теоремы с.в. $\{\xi_k^*\}$ – н.о.р. и $M(\xi_k^*) = 0$, $D(\xi_k^*) = 1$. Обозначим через φ характеристическую функцию с.в. ξ_1^*. Тогда в силу формулы (7) имеем

$$\varphi(t) = 1 - \frac{1}{2}t^2 + o(t^2)$$

при $t \to 0$. Характеристическая функция с.в. S_n^* имеет вид (смотри свойства 5 и 6)

$$\varphi_{S_n^*}(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n \ .$$

Используя разложение для φ, получаем при любом фиксированном $t \in R^1$

$$\varphi_{S_n^*}(t) = \left[1 - \frac{1}{2}t^2 \cdot \frac{1}{n} + o(\frac{t^2}{n})\right]^n \to e^{-\frac{1}{2}t^2} \ .$$

Теорема доказана.

Из последней теоремы мы получаем как следствие интегральную теорему Муавра – Лапласа. Действительно, пусть $\varepsilon_k = 1$, если в k-м испытании был успех (это происходит с вероятностью p), и $\varepsilon_k = 0$ – в противном случае. Тогда $S_n = \varepsilon_1 + \ldots + \varepsilon_n$ есть число успехов в n независимых испытаниях. С.в. $\{\varepsilon_k\}$ – н.о.р., $M(\varepsilon_k) = p$, $D(\varepsilon_k) = p(1-p)$. Отсюда следует, что $M(S_n) = np$, $D(S_n) = np(1-p)$. Окончательно получаем следующее выражение для нормированной суммы:

$$S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}} \ .$$

Применяя только что доказанную теорему, получаем, что S_n^* имеет асимптотически стандартное нормальное распределение.

Выше мы доказали ЦПТ при довольно жестких условиях. Далее мы попытаемся без доказательства **объяснить**, в каких именно ситуациях следует ожидать нормальное предельное распределение для нормированных сумм случайных величин.

Пусть $\{\xi_k\}$ – последовательность независимых случайных величин, для которых существуют конечные $M(\xi_k) = a_k$ и $D(\xi_k) = \sigma_k^2$. Как и ранее, образуем суммы $S_n = \xi_1 + \ldots + \xi_n$. Тогда $M(S_n) = a_1 + \ldots + a_n = A_n$ и (в силу независимости) $D(S_n) = \sigma_1^2 + \ldots + \sigma_n^2 = B_n^2$. Определим нормированную сумму

$$S_n^* = \frac{S_n - A_n}{B_n} = \sum_{k=1}^{n} \xi_{kn} \ , \tag{16.14}$$

где $\xi_{kn} = (\xi_k - a_k)/B_n, \ k = \overline{1, n}$, есть независимые случайные величины, для которых $M(\xi_{kn}) = 0, \ D(\xi_{kn}) = \sigma_{kn}^2$ и $\sum\limits_{k=1}^{n} \sigma_{kn}^2 = 1$.

Теорема 5 . *Пусть для любого n мы имеем набор $\{\xi_{nk}\}$ независимых случайных величин, таких, что $M(\xi_{kn}) = 0$, $D(\xi_{kn}) = \sigma_{kn}^2$ и $\sum\limits_{k=1}^{n} \sigma_{kn}^2 = 1$. Если выполнено условие*

$$P(\max_{1 \le k \le n} |\xi_{kn}| > \varepsilon) \to 0, \ n \to \infty \ , \tag{16.15}$$

или, эквивалентно,

$$\sum_{k=1}^{n} P(|\xi_{kn}| > \varepsilon) \to 0, \ n \to \infty \ , \tag{16.16}$$

то применима ЦПТ, т.е.

$$P(\xi_{1n} + \ldots + \xi_{nn} < x) \to \Phi(x) \ , \ n \to \infty \ .$$

Это одна из наиболее общих форм ЦПТ. Смысл условия (15) состоит в том, что каждое из слагаемых ξ_{kn} суммы $S_n^* = \xi_{1n} + \ldots + \xi_{nn}$ вносит малый вклад по сравнению со всей суммой S_n^*. С математической точки зрения это наиболее ясная и простая форма ЦПТ. Но в реальных задачах проверить условие (15) бывает довольно трудно. Ниже мы приводим другие формулировки, которые содержат более легко проверяемые условия. Кстати, эти теоремы появились исторически раньше и помогли прояснить суть ЦПТ.

Теорема (Линдеберг, 1924). *Пусть введены те же обозначения, что и раньше. Если* $\{\xi_{kn}\}$ *– независимые случайные величины и выполнено* **условие Линдеберга**: $\forall\varepsilon > 0$

$$\sum_{k=1}^{n} \int\limits_{|x|>\varepsilon} x^2 dF_{kn}(x) \to 0 \ , \ n \to \infty \ . \tag{16.17}$$

где $F_{kn}(x) = P(\xi_{kn} < x)$, *то применима ЦПТ, т.е.*

$$P(\xi_{1n} + \ldots + \xi_{nn} < x) \to \Phi(x) \ , \ n \to \infty \ .$$

Условие (17) легко следует из условия (16) с помощью аналога неравенства Чебышева.

Теорема (Ляпунов, 1900). *Пусть* $\{\xi_k\}$ *– последовательность независимых случайных величин, для которых существуют конечные* $M(\xi_k) = a_k$, $D(\xi_k) = \sigma_k^2$ *и* $\beta_k = M(|\xi_k - a_k|^3)$. *Обозначим* $S_n = \xi_1 + \ldots + \xi_n$, $A_n = M(S_n) = a_1 + \ldots + a_n$, $B_n^2 = D(S_n) = \sigma_1^2 + \ldots + \sigma_n^2$ *и* $C_n = \beta_1 + \ldots + \beta_n$. *Если выполнено* **условие Ляпунова**

$$\frac{C_n}{B_n^{3/2}} \to 0 \ , \ n \to \infty \ , \tag{16.18}$$

то применима ЦПТ, т.е. для $S_n^* = (S_n - A_n)/B_n$

$$P(S_n^* < x) \to \Phi(x) \ .$$

Преимущество последней теоремы состоит в том, что в условие (18) входят только моменты отдельных слагаемых, вычислять которые в практических задачах обычно довольно легко.

Во всех сформулированных выше теоремах присутствует условие независимости случайных величин. На самом деле это условие можно значительно ослабить и потребовать, чтобы слагаемые в сумме были зависимы в определенном смысле слабо.

Наряду с ЗБЧ и теоремой Пуассона о сходимости к пуассоновскому распределению в схеме Бернулли ЦПТ есть наиболее часто применяемый результат в практических задачах. Мы проиллюстрируем, как это делается, на примере одной задачи из теории страхования.

Пример. Некоторая страховая компания продала 100 страховых полисов, каждый из которых представляет собой договор страхования от несчастного случая и смерти. При наступлении несчастного случая лицу, указанному в договоре, выплачивается сумма 100 000 рублей, а в случае смерти застрахованного выплачивается сумма 10 000 000 рублей. Какова должна быть стоимость такого полиса, чтобы вероятность безубыточной деятельности компании была не менее 0.95? Собранный статистический материал показывает, что для той группы лиц, которой предлагается данная страховка, вероятность несчастного случая равна 0.001, а вероятность смерти равна 0.0001.

Решение. В данном случае выплаты X_k по k-му полису представляют собой с.в. со следующим распределением:

$$X_k = \begin{cases} 10^7 & , \quad \textit{с вероятностью } 10^{-4} \ , \\ 10^5 & , \quad \textit{с вероятностью } 10^{-3} \ , \\ \quad 0 & , \quad \textit{с вероятностью } 0.9989 \ . \end{cases}$$

Если пренебречь случаями каких-либо катастроф, эпидемий и т.п., то можно предположить, что $\{X_k\}$ – н.о.р.с.в. Легко вычислить, что $a = M(X_k) = 1010$, а $\sigma^2 = D(X_k) \approx 10^{10}$. Пусть стоимость страховки есть величина b, которую необходимо определить. Тогда общая собранная с застрахованных сумма равна $nb = 100b$, а суммарные выплаты по проданным полисам равны $S_n = X_1 + \ldots + X_n$. Компания понесет убытки, если $(S_n > nb)$. Вероятность этого события должна быть не более 0.05. Вычислим эту вероятность в нашей задаче:

$$P(S_n > nb) = P\left(\frac{S_n - na}{\sigma\sqrt{n}} > \frac{n(b-a)}{\sigma\sqrt{n}}\right) \approx 1 - \Phi\left(\left(\frac{b-a}{\sigma}\right)\sqrt{n}\right) \ .$$

Приравнивая

$$1 - \Phi\left(\left(\frac{b-a}{\sigma}\right)\sqrt{n}\right) = 0.05 \ ,$$

находим по таблицам нормального распределения, что

$$\frac{(b-a)}{\sigma}\sqrt{n} = 1.64 \ .$$

Отсюда получаем

$$b = a + 1.64\frac{\sigma}{\sqrt{n}} = 1010 + 1.64\frac{10^5}{10} = 17410 \ .$$

1.17 Цепи Маркова

До сих пор мы изучали последовательности независимых случайных величин. Но поведение реальных систем в каждый момент времени зависит, вообще говоря, от предыстории процесса, в простейшем случае – от состояния системы в данный момент времени. Это приводит нас к понятию цепи Маркова. Впервые модели такого типа начал изучать известный российский математик академик Андрей Андреевич Марков, в честь которого они и получили свое название.

1.17.1 Определение цепи Маркова

Определение цепи Маркова уже было дано при изучении последовательности независимых испытаний. Для удобства мы повторим его, но уже в терминах случайных величин.

Определение 1 . Цепью Маркова *называется последовательность с.в.* $\{X_n, n \geq 0\}$*, принимающих значения в конечном или счетном множестве* $X = \{x_1, x_2, \ldots\}$*, такая, что* $\forall n \geq 1$*,* $\forall x_{i_0}, \ldots, x_{i_n} \in X$

$$P(X_n = x_{i_n} | X_0 = x_{i_0}, \ldots, X_{n-1} = x_{i_{n-1}}) =$$
$$P(X_n = x_n | X_{n-1} = x_{i_{n-1}}) \ . \qquad (17.1)$$

Множество X *называется* **пространством состояний** *цепи Маркова, а вероятности*

$$P_{ij}^{(n)} := P(X_n = x_j | X_{n-1} = x_i) \qquad (17.2)$$

называются **вероятностями перехода** *на n-м шаге.*

Набор вероятностей

$$P_i^{(0)} = P(X_0 = x_i) \ , \ x_i \in X \ , \qquad (17.3)$$

определяет **начальное распределение** *цепи Маркова.*

Во многих задачах необходимо знать также **распределение цепи на n-м шаге**, т.е. вероятности

$$P_i^{(n)} = P(X_n = x_i) \ , \ x_i \in X \ , \qquad (17.4)$$

и **вероятности перехода за m шагов**

$$P_{ij}^{(n)}(m) = P(X_{n-1+m} = x_j | X_{n-1} = x_i) \ . \qquad (17.5)$$

Для упрощения обозначений мы будем писать как и раньше $P_{ij}^{(n)}(1) = P_{ij}^{(n)}$.

Легко показать (задача!), что вероятности перехода обладают следующими свойствами: для любых натуральных m, n, i, j

1) $P_{ij}^{(n)}(m) \geq 0$,

2) $\sum\limits_j P_{ij}^{(n)}(m) = 1$.

Используя теорему умножения и формулу полной вероятности, нетрудно доказать (задача!) следующие соотношения: $\forall n \geq 1, \ \forall x_{i_0}, \ldots, x_{i_n} \in X$

$$P(X_0 = x_{i_0}, X_1 = x_{i_1}, \ldots, X_n = x_{i_n}) = P_{i_0}^{(0)} \cdot P_{i_0 i_1}^{(1)} \cdot \ldots \cdot P_{i_{n-1} i_n}^{(n)} \quad (17.6)$$

и

$$P_j^{(n)} = \sum_i P_i^{(0)} P_{ij}^{(1)}(n) \ . \qquad (17.7)$$

Таким образом, зная начальное распределение и матрицы переходных вероятностей, можно вычислить вероятности любых событий, связанных с поведением цепи Маркова. На самом деле достаточно знать матрицы перехода за один шаг. Действительно, имеет место следующее соотношение (задача!): $\forall n, m_1, m_2, i, j \geq 1$

$$P_{ij}^{(n)}(m_1 + m_2) = \sum_k P_{ik}^{(n)}(m_1) \cdot P_{kj}^{(n+m_1)}(m_2) \ , \qquad (17.8)$$

откуда следует

$$P_{ij}^{(n)}(m) = \sum_{k_1} \cdots \sum_{k_{m-1}} P_{ik_1}^{(n)} \cdots P_{k_{m-1}j}^{(n+m-1)} \ . \qquad (17.9)$$

Цепь Маркова называется **однородной** (по времени), если вероятности перехода одни и те же на каждом шаге, т.е. $\forall n \geq 1 \ P_{ij}^{(n)} = P_{ij}$. Всюду далее мы будем рассматривать только такие цепи Маркова. В этом случае формулы (6), (7) и (8) можно переписать соответственно в следующем виде:

$$P(X_0 = x_{i_0}, X_1 = x_{i_1}, \ldots, X_n = x_{i_n}) = P_{i_0}^{(0)} \cdot P_{i_0 i_1} \cdot \ldots \cdot P_{i_{n-1} i_n} \ ; \qquad (17.10)$$

$$P_j^{(n)} = \sum_i P_i^{(0)} P_{ij}(n) \ ; \qquad (17.11)$$

$$P_{ij}(m_1 + m_2) = \sum_k P_{ik}(m_1) \cdot P_{kj}(m_2) \ . \qquad (17.12)$$

Последнее соотношение называется **уравнением Маркова**.

Используя матричные обозначения

$$P^{(n)} = (P_1^{(n)}, P_2^{(n)}, \ldots), \ P(n) = (P_{ij}(n)), \ P := P(1),$$

можно переписать соотношения (11) и (12) в следующем виде:

$$P^{(n)} = P^{(0)} \cdot P(n) \ , \qquad (17.13)$$

$$P(m+n) = P(m) \cdot P(n) \ . \qquad (17.14)$$

Из последнего соотношения получаем, что

$$P(n) = P^n \ . \qquad (17.15)$$

Отсюда следует, что для полного описания однородной цепи Маркова достаточно знать начальное распределение и матрицу перехода за один шаг.

Для анализа динамики цепи Маркова полезно использовать граф, построенный по матрице P переходных вероятностей за один шаг. Его вершинами являются элементы пространства состояний X. Если $P_{ij} > 0$, то проводим **ориентированное** ребро, ведущее из

i в j. Примеры применения этого понятия для анализа поведения цепи Маркова будут приведены ниже.

В заключение этого раздела рассмотрим классическую **задачу о разорении игрока**, которая впервые была сформулирована еще Гюйгенсом в 1656 году и которой интересовались многие известные математики в прошлом. Два игрока играют в некоторую игру, в которой первый игрок выигрывает с вероятностью p и проигрывает с вероятностью $q = 1 - p,\ 0 < p < 1$. Тот, кто выиграл, получает от проигравшего одну единицу его капитала. Игроки вступают в игру, имея на руках суммарный капитал величины a: z – у первого игрока и $a - z$ – у второго. Игра заканчивается, когда один из игроков полностью разорится. Нетрудно показать (задача!), что игра закончится с вероятностью 1 за конечное (но случайное) время. Необходимо вычислить вероятности разорения для каждого из игроков. Многие экономические задачи можно свести к задаче подобного типа.

Обозначим через q_z вероятность разорения первого игрока, если он начинает игру, имея на руках капитал $z,\ 0 < z < a$. По смыслу задачи естественно доопределить $q_0 = 1$ и $q_a = 0$. Это типичная задача о цепи Маркова. Выделим одного из игроков, например первого. Он может находиться в одном из состояний $z,\ 0 \le z \le a$. После каждой игры он переходит либо в состояние $z - 1$ с вероятностью q, либо в состояние $z + 1$ с вероятностью p, если $0 < z < a$. Попав в состояние 0 или a, он навсегда в нем остается. Необходимо найти вероятность попадания в состояние 0, если мы начинаем движение из состояния z.

Т.к. на каждом шаге мы имеем ту же задачу (однородная цепь Маркова!), но с разными z, то в силу формулы полной вероятности мы имеем следующее соотношение:

$$q_z = pq_{z+1} + qq_{z-1}\ ,\ \ 0 < z < a - 1\ . \tag{17.16}$$

Кроме того, в граничных точках мы имеем

$$q_0 = 1\ ,\ \ q_a = 0\ . \tag{17.17}$$

Соотношение (16) есть однородное разностное уравнение второго порядка. При $p \neq q$ оно имеет два независимых частных решения: $q_z \equiv 1$ и $q_z = (q/p)^z$ (проверить!). Тогда общее решение имеет вид

$$q_z = A + B(q/p)^z \ .$$

Используя граничные условия (17), окончательно получим

$$q_z = \frac{(q/p)^a - (q/p)^z}{(q/p)^a - 1} \ . \tag{17.18}$$

Для $p = q$ уравнение (16) имеет независимые частные решения $q_z \equiv 1$ и $q_z = z$ (проверить!). Это приводит нас к решению

$$q_z = 1 - \frac{z}{a} \ . \tag{17.19}$$

1.17.2 Классификация состояний

Далее в этом параграфе мы будем рассматривать только конечные однородные цепи Маркова, т.е. такие, у которых множество $X = \{x_1, x_2, \dots, x_r\}$ конечно.

Вероятностные свойства систем проявляются асимптотически, когда мы наблюдаем за ними достаточно долго и следим за тем, как часто происходят те или иные события. В силу этого нас будут интересовать асимптотические свойства цепей Маркова. С этой точки зрения разные состояния цепи обладают неодинаковыми свойствами. Поэтому мы начнем изложение нашей теории с классификации состояний. Для этой цели оказывается очень полезным введенное нами ранее понятие графа цепи Маркова, т.к. оно позволяет представить наглядно определяемые понятия.

Будем говорить, что состояние j **достижимо** из состояния i, если существует $n \geq 1$, такое, что $P_{ij}(n) > 0$. Состояния i и j называются **сообщающимися**, если i достижимо из j, а j достижимо из i. Состояние i называется **несущественным**, если можно найти состояние j и число $n_0 \geq 1$, такие, что $P_{ij}(n_0) > 0$, но $P_{ji}(n) = 0, \ \forall n \geq 1$. В противном случае состояние называется **существенным**.

Все состояния объединяются в некоторое число **замкнутых** классов. Внутри одного класса все состояния сообщаются, но это не так для состояний из разных классов. Доказательство этого проводится дословно так же, как и при определении классов смежности относительно линейного подпространства. Цепь Маркова называется **неразложимой**, если она имеет один замкнутый класс существенных состояний, и **разложимой** – в противном случае. Состояние i называется **поглощающим**, если $P_{ii} = 1$. Попав в такое состояние, мы остаемся в нем навсегда. Поглощающее состояние образует отдельный замкнутый класс.

Несущественное состояние обладает тем свойством, что с вероятностью 1 мы когда-то в очередной раз выйдем из этого состояния и больше в него не вернемся (задача!). Поэтому оно и является **несущественным** с точки зрения асимптотических свойств цепи.

Попав в один из замкнутых существенных классов, мы далее продолжаем движение только в нем. Это означает, что цепь Маркова распадается на некоторое число в определенном смысле "независимых" цепей. Поэтому достаточно изучить только цепи Маркова с одним замкнутым классом.

Периодом состояния i называется наибольший общий делитель d_i чисел n, для которых $P_{ii}(n) > 0$. Можно показать, что все состояния внутри одного замкнутого класса имеют один и тот же период d, который называют периодом данного класса. Класс называют **апериодическим**, если для него $d = 1$. **Периодический** класс разбивается на $d \geq 2$ подклассов, которые можно упорядочить так, что за один шаг возможны переходы только из одного подкласса в соседний.

1.17.3 Предельные теоремы для цепей Маркова

Большинство реальных систем ведут себя следующим образом. Некоторое время в системе происходят так называемые переходные явления, но через достаточно большой промежуток времени система выходит на стационарный режим и ее характеристики стабилизируются. Если динамика такой системы исследуется в рамках теории цепей Маркова, то нам необходимо ответить на следующие вопросы: что такое стационарный режим для цепи Маркова и что означает выход цепи в пределе на стационарный режим?

Определение 2 . *Распределение вероятностей* $Q = (q_1, q_2, \ldots, q_r)$ *на пространстве состояний* X *называется* **стационарным** *для однородной цепи Маркова с матрицей* $P = (P_{ij})$ *вероятностей перехода за один шаг, если имеет место соотношение*

$$Q = Q \cdot P \ , \tag{17.20}$$

или, более подробно,

$$q_j = \sum_i q_i P_{ij} \ . \tag{17.21}$$

Если взять такое распределение в качестве начального, т.е. $P^{(0)} = Q$, то в силу формул (13) и (15) имеем

$$P^{(n)} = P^{(0)} P(n) = Q P^n = Q \ .$$

Распределение цепи Маркова на n-м шаге будет совпадать с ее начальным распределением, т.е. не будет меняться со временем. Но именно это и означает стационарность системы.

Определение 3 . *Цепь Маркова называется* **эргодической**, *если для любого начального распределения* $P^{(0)}$ *существуют пределы*

$$\pi_j = \lim_{n\to\infty} P_j^{(n)} \tag{17.22}$$

и предельное распределение $\pi = (\pi_1, \ldots, \pi_r)$ *не зависит от начального распределения. В частности, для любых* i, j *существуют* **финальные вероятности**

$$\pi_j = \lim_{n\to\infty} P_{ij}(n) \ . \tag{17.23}$$

В последнем случае мы начинаем движение из фиксированного состояния x_i, т.е. $P(X_0 = x_i) = P_i^{(0)} = 1$.

Напомним, что мы рассматриваем **конечные** цепи Маркова. Тогда существование пределов (22) и (23) гарантирует, что финальные вероятности образуют распределение вероятностей π на X, т.е.

$$\pi_j \geq 0, \ \sum_j \pi_j = 1 \ . \tag{17.24}$$

Когда число состояний бесконечно, условие (24) необходимо явно вписать в определение эргодичности.

Нетрудно доказать (задача!), что финальные вероятности, если они существуют, образуют стационарное распределение, т.е.

$$\pi = \pi \cdot P \ . \tag{17.25}$$

Из определения эргодичности цепи Маркова следует, что в этом случае существует только одно стационарное распределение, называемое **эргодическим**, и оно совпадает с набором π финальных вероятностей (почему?). В общем случае это не так.

Пример. Пространство состояний X состоит из четырех элементов, которые мы обозначим их номерами 1,2,3 и 4. $P_{12} = P_{21} = P_{34} = P_{43} = 1/2$. Все остальные вероятности перехода равны нулю. Положим $Q_1 = (1/4, 1/4, 1/4, 1/4)$ и $Q_1 = (1/6, 1/6, 1/3, 1/3)$. Непосредственная проверка показывает, что Q_1 и Q_2 – стационарные распределения для нашей цепи Маркова. Построив граф этой цепи, мы видим, что она является разложимой. Это и есть причина неединственности стационарного распределения.

Задача. Пусть мы имеем разложимую цепь Маркова и $Q_1, \ldots,$ Q_k – стационарные распределения для каждой компоненты, т.е. стационарное распределение Q_l сосредоточено на l-й компоненте. Доказать, что распределение $Q = \alpha_1 Q_1 + \ldots \alpha_r Q_r$, $\alpha \geq 0$, $\sum\limits_l \alpha_l = 1$, является стационарным распределением для всей цепи Маркова. Показать, что любое стационарное распределение представимо в таком виде.

Таким образом, одной из причин неэргодичности цепи Маркова является ее разложимость. Другой причиной является периодичность класса состояний. Пусть мы имеем один замкнутый класс с периодом d. Разобьем его на d подклассов, как это описано выше. Тогда $P_{ij}(nd) > 0$, если i и j лежат в одном подклассе, и $P_{ij}(nd) = 0$, если i и j лежат в разных подклассах. Отсюда видно, что, двигаясь по подпоследовательности $\{nd\}$, мы получим разные пределы для переходных вероятностей в зависимости от начального состояния. Для конечных цепей Маркова этим и исчерпывается список причин, приводящих к неэргодичности.

Теорема 1 . *Пусть мы имеем конечную однородную цепь Маркова, все состояния которой сообщаются. Тогда эта цепь Маркова является эргодической, т.е. имеют место соотношения (22). Более того, если число $\lambda > 0$ таково, что $\forall i$ и некоторого j_0*

$$P_{ij_0} \geq \lambda > 0 \ , \tag{17.26}$$

то для любых двух начальных распределений $Q_1 = (q_1, \ldots, q_r)$ и $Q_2 = (p_1, \ldots, p_r)$ соответствующие распределения $Q_1^{(n)} = (q_1^{(n)}, \ldots, q_r^{(n)})$ и $Q_2^{(n)} = (p_1^{(n)}, \ldots, p_r^{(n)})$ на n-м шаге удовлетворяют соотношению

$$\sum_{j=1}^{r} |q_j^{(n)} - p_j^{(n)}| \leq 2(1-\lambda)^n \ ,$$

т.е. эти вероятности сближаются при $n \to \infty$. Эргодическое распределение π является единственным решением уравнения (20), удовлетворяющим условиям (24).

Отметим, что рассматриваемая цепь Маркова является неразложимой и апериодической без несущественных состояний. Можно показать, что для любой цепи с такими свойствами все состояния сообщаются (задача!). Такая цепь Маркова называется **регулярной**.

Доказательство. Во-первых, заметим, что в нашем случае условие (26) всегда выполнено, т.к. $P_{ij} > 0 \ \forall i, j$ и этих вероятностей конечное число.

Начнем с доказательства неравенства (27), которое называется **неравенством Бернштейна**. Т.к. $Q_1^{(n)}$ и $Q_2^{(n)}$ являются распределениями вероятностей, то

$$\sum_j q_j^{(n)} = \sum_j p_j^{(n)} = 1 \ .$$

Отсюда следует, что

$$0 = \sum_j (q_j^{(n)} - p_j^{(n)}) = \sum_j{}^{+} (q_j^{(n)} - p_j^{(n)}) + \sum_j{}^{-} (q_j^{(n)} - p_j^{(n)}) \ .$$

Σ^+ содержит положительные слагаемые, а Σ^- – неположительные. Из последнего соотношения следует, что

$$\sum_j{}^{+} (q_j^{(n)} - p_j^{(n)}) = -\sum_j{}^{-} (q_j^{(n)} - p_j^{(n)})$$

и

$$L_n := \sum_j |q_j^{(n)} - p_j^{(n)}| = 2\sum_j{}^{+} (q_j^{(n)} - p_j^{(n)}) \ .$$

На $(n-1)$-м шаге мы имеем распределения $Q_1^{(n-1)}$ и $Q_2^{(n-1)}$. Тогда на следующем шаге получаем

$$q_j^{(n)} = \sum_i q_i^{(n-1)} P_{ij} \ , \ p_j^{(n)} = \sum_i p_i^{(n-1)} P_{ij}$$

и

$$L_n = 2\sum_j{}^{+} \sum_i (q_i^{(n-1)} - p_i^{(n-1)}) \cdot P_{ij} \ .$$

Предположим, что $q_{j_0}^{(n)} - p_{j_0}^{(n)} \leq 0$ (если имеет место противоположное неравенство, то доказательство проводится аналогично). Тогда в $\sum\limits_j{}^{+}$ все $j \neq j_0$ и $\sum\limits_j{}^{+} P_{ij} \leq (1-\lambda)$ для любого i. Оставляя в сумме по i только положительные слагаемые, получаем

$$L_n \leq 2\sum_j{}^{+} \sum_i{}^{+} (q_i^{(n-1)} - p_i^{(n-1)}) \cdot P_{ij} = 2\sum_i{}^{+} (q_i^{(n-1)} - p_i^{(n-1)}) \sum_j{}^{+} P_{ij}$$

$$\leq 2(1-\lambda) \sum_i{}^{+} (q_i^{(n-1)} - p_i^{(n-1)}) = (1-\lambda) L_{n-1} \ .$$

Повторив это рассуждение несколько раз, приходим к неравенству

$$\begin{aligned} L_n &\le (1-\lambda)^n L_0 = (1-\lambda)^n \sum_j |q_j^{(0)} - p_j^{(0)}| \\ &\le (1-\lambda)^n \left(\sum_j q_j^{(0)} + \sum_j p_j^{(0)} \right) = 2(1-\lambda)^n \ . \end{aligned}$$

Это доказательство было предложено еще А.А. Марковым и усовершенствовано С.Н. Бернштейном.

Покажем теперь, что финальные вероятности π_j существуют. Пусть мы начинаем движение из состояния i, т.е. $P_i^{(0)} = 1$. Тогда $P_j^{(n)} = P_{ij}(n)$. Положим $m_j(n) = \min_i P_{ij}(n)$. Последовательность $\{m_j(n)\}$ монотонно не убывает по n. Действительно

$$m_j(n) = \min_i \sum_k P_{ik} P_{kj}(n-1) \ge \min_i \sum_k P_{ik} m_j(n-1) = m_j(n-1) \ .$$

В силу монотонности существует

$$\pi_j = \lim_{n\to\infty} m_j(n) \ .$$

Обозначим $M_j(n) = \max_i P_{ij}$. По определению величин $m_j(n)$ и $M_j(n)$ имеем

$$m_j(n) \le P_{ij}(n) \le M_j(n) \ .$$

Из неравенства Бернштейна можно вывести (задача!), что

$$M_j(n) - m_j(n) \le 2(1-\lambda)^n \to 0 \ , \ n \to \infty \ .$$

Отсюда следует, что существует

$$\lim_{n\to\infty} P_{ij}(n) = \pi_j \ .$$

Утверждения о стационарности финального распределения и единственности стационарного распределения сформулированы выше в виде задачи.

Сформулируем без доказательства аналог закона больших чисел для цепей Маркова.

Теорема 2 . *Пусть* $(X_n,\ n \geq 1)$ *– эргодическая цепь Маркова с финальным распределением* π. *Если* f *– вещественная функция на пространстве состояний* X, *то*

$$\frac{f(X_0) + \ldots f(X_{n-1})}{n} \xrightarrow{P} \sum_j f(x_j)\pi_j \ .$$

В частности, если $f(x_j) = 1$ *для* $j = i_0$ *и* $f(x_j) = 0$ *для* $j \neq i_0$, *то*

$$\frac{\nu_{i_0}(n)}{n} \xrightarrow{P} \pi_{i_0}\ ,\ n \to \infty\ ,$$

где $\nu_{i_0}(n)$ *– число посещений состояния* i_0 *за первые* n *шагов.*

Замечание. Обозначим через τ_j момент первого возвращения в состояние j (после выхода из него же). Более тонкий анализ позволяет доказать, что $\pi_j = 1/M(\tau_j)$.

Всюду в этом параграфе мы рассматривали только конечные цепи Маркова с дискретным временем. Свойства цепей Маркова со счетным множеством состояний, а также с непрерывным временем будут рассматриваться позднее в курсе теории случайных процессов.

1.17.4 Примеры

Приведем несколько примеров, которые описывают ситуации, близкие к реальным.

Пример 1. Пусть мы имеем схему Бернулли с вероятностью успеха p в одном испытании. Это простейшая однородная цепь Маркова с пространством состояний $X = \{0, 1\}$ и матрицей перехода за один шаг

$$P = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}\ .$$

Пример 2. В предыдущем примере нас может интересовать появление различных сочетаний символов на последовательных шагах. Например, для двух последовательных шагов мы имеем четыре различные комбинации: $x_1 = (00)$, $x_2 = (01)$, $x_3 = (10)$,

$x_4 = (11)$. В этом случае мы имеем дело с однородной цепью Маркова с матрицей перехода

$$P = \begin{pmatrix} 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \\ 1-p & p & 0 & 0 \\ 0 & 0 & 1-p & p \end{pmatrix}$$

Пример 3. Компания по страхованию автомобилей провела обследование среди потенциальных клиентов и получила следующие результаты. Из водителей, попавших в аварию в течение года, 20% попали в аварию и в следующем году. Из тех, кто не попал в аварию в течение года, только 10% попали в аварию в следующем. Данную ситуацию можно описать с помощью однородной цепи Маркова с двумя состояниями: $x_1 = A$, $x_2 = \overline{A}$, где A обозначает событие, состоящее в том, что водитель попал в аварию, $\overline{A}$ – противоположное событие. По имеющимся данным получаем следующую матрицу перехода:

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.1 & 0.9 \end{pmatrix}$$

Предположим, что в этом году среди застрахованных водителей 8% попали в аварию. Какова вероятность того, что случайно выбранный водитель попадет в аварию в следующем году?

Мы начинаем с начального распределения $P^{(0)} = (0.08, 0.92)$. Тогда

$$P^{(1)} = P^{(0)} \cdot P = (0.108, 0.892) \ ,$$

т.е. искомая вероятность равна 0.108.

Найдем стационарное распределение в этой задаче. Стационарное распределение $\pi = (\pi_1, \pi_2)$ удовлетворяет уравнению

$$\pi P = \pi$$

и условию нормировки

$$\pi_1 + \pi_2 = 1 \ .$$

Это приводит к следующей системе уравнений:

$$\begin{cases} 0.2\pi_1 + 0.1\pi_2 & = & \pi_1 \ , \\ 0.8\pi_1 + 0.9\pi_2 & = & \pi_2 \ , \\ \quad \pi_1 + \pi_2 & = & 1 \ . \end{cases}$$

Она имеет единственное решение $\pi_1 = 1/9, \ \pi_2 = 8/9$.

Пример 4. В стране Оз никогда не бывает подряд двух ясных дней. Если сегодня ясно, то завтра будет плохая погода (снег или дождь с равной вероятностью). Если сегодня снег (или дождь), то погода на следующий день не изменится с вероятностью $1/2$. Если она все-таки изменится, то лишь в половине случаев будет ясно. Построить модель изменения погоды в виде цепи Маркова с тремя состояниями, вычислить матрицу перехода и найти стационарное распределение.

Пример 5. Обучение в некотором колледже продолжается 4 года. Многолетняя статистика показывает, что студент каждого курса выбывает из колледжа с вероятностью p, остается на том же курсе с вероятностью q и переходит на следующий курс с вероятностью r. Образуем цепь Маркова, введя состояния: x_0 – выбыл, x_k – студент k-го курса, $k = \overline{1,4}$, x_5 – окончил колледж. Выписать матрицу переходных вероятностей.

Следующий пример имеет иллюстративный характер.

Пример 6. Рассмотрим цепь Маркова с семью состояниями, которая имеет следующую матрицу переходных вероятностей

$$P = \begin{pmatrix} 0 & 0.2 & 0.8 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0.3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 0 & 0.3 & .07 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 & 0 & 0 \end{pmatrix}$$

Нарисовать граф этой цепи и провести классификацию состояний.

1.18 Задачи для самостоятельного решения

1. Дискретная случайная величина ξ имеет распределение

x_n	-2	-1	0	1	2
p_n	0.2	0.1	0.2	0.4	0.1

С.в. $\eta = \xi^2$. Вычислить $M\xi, M\eta, D(\xi), D(\eta)$.

2. ξ_1 и ξ_2 – независимы, $M(\xi_1) = 2$, $M(\xi_2) = 3$, $D(\xi_1) = 4$, $D(\xi_2) = 9$. $\eta = 2\xi_1 + 4\xi_2 + 5$. Вычислить $M(\eta)$ и $D(\eta)$.

3. Вычислить математические ожидания и дисперсии для всех стандартных распределений.

4. С.в. ξ имеет $M(\xi) = a$ и $D(\xi) = \sigma^2 > 0$. Положим $\xi_0 = (\xi - a)/\sigma$. Доказать, что $M(\xi_0) = 0$, $D(\xi_0) = 1$.

5. С.в. ξ имеет нормальное распределение с параметрами $a = 2$, $\sigma^2 = 9$. Вычислить: $P(\xi \geq 5)$, $P(\xi < 8)$, $P(-7 < \xi < 11)$.

6. С.в. ξ имеет плотность распределения

$$\rho_\xi(x) = \begin{cases} 2x & , \quad x \in [0,1] \;, \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Вычислить $M(\xi^2)$ и $D(\xi^2)$.

7. С.в. ξ имеет показательное распределение с параметром $\lambda = 2$. Найти медиану и квантиль порядка 0.95 с.в. ξ.

8. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет распределение

$\xi_2 \backslash \xi_1$	-1	0	1	2
-1	0.05	0.15	0.05	0.05
0	0.1	0.1	0.1	0.1
1	0.05	0.1	0.15	0

Найти: 1) условное распределение вектора ξ при условии $\xi_1\xi_2 = 0$,

2) условное распределение ξ_1 при условии $\xi_2 = -1$,

3) условное математическое ожидание ξ_1 при условии $\xi_2 = -1$,

4) $\rho(\xi_1, \xi_2)$.

9. Случайный вектор $\xi = (\xi_1, \xi_2)$ имеет плотность распределения

$$\rho_\xi(x_1, x_2) = \begin{cases} C(2x_1 + 3x_2) & , \quad x_1, x_2 \in [0,1] \ , \\ 0 & , \quad \text{в пр. сл.} \end{cases}$$

Найти: 1) константу C,

2) условную плотность ξ_1 при условии $\xi_2 = 1/2$,

3) условное математическое ожидание ξ_1 при условии $\xi_2 = 1/2$.

4) $\rho(\xi_1, \xi_2)$.

10. Вычислить характеристические и производящие функции для стандартных распределений.

11. Доказать, что сумма двух независимых с.в. ξ_1 и ξ_2, имеющих распределения Пуассона с параметрами λ_1 и λ_2, также имеет распределение Пуассона, но с параметром $\lambda_1 + \lambda_2$.

12. С.в. ξ имеет геометрическое распределение с параметром p. Положим $\eta_p = p \cdot \xi$. Найти предельное распределение для η_p при $p \to 0$.

13. Используя характеристическую функцию, вычислить математическое ожидание и дисперсию распределения Пуассона с параметром λ.

14. Случайные величины ξ_1, ξ_2, ... – н.о.р., $M(\xi_1) = 1$, $D(\xi_1) = 4$. Положим $S_n = \xi_1 + \ldots \xi_n$. При каком минимальном n $P(S_n > 120) \geq 0.95$?

15. Ресторан под открытым небом работает только в погожие дни. Если в данной местности сегодня идет дождь, то с вероятностью 0.4 он будет и завтра, если же сегодня дождя нет, то завтра он будет с вероятностью 0.06. Описать эту ситуацию с помощью однородной цепи Маркова и найти матрицу перехода за один шаг. Найти стационарное распределение. Сегодня ресторан был закрыт. Найти вероятность того, что он будет закрыт еще два дня.

16. Однородная цепь Маркова имеет следующую матрицу перехода за один шаг

$$P = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 2/3 & 1/3 & 0 \end{pmatrix}$$

Нарисовать соответствующий граф и провести классификацию состояний.

Список литературы

a) Основной

1. Севастьянов Б.А. Курс теории вероятностей и математической статистики. - М.: Наука, 1982.

2. Хохлов Ю.С. Теория вероятностей и математическая статистика. Ч. I: Учебное пособие / ТвГУ. – Тверь, 1997.

3. Чистяков В.П. Курс теории вероятностей. - М.: Наука, 1982.

4. Ширяев А.Н. Вероятность. - М.: Наука, 1980.

5. Агапов Г.И. Задачник по теории вероятностей: Учебное пособие для студентов втузов. - М.: Высш. шк., 1986.

6. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей. - М.: Наука, 1989.

b) Дополнительный

7. Колмогоров А.Н. Основные понятия теории вероятностей. - М.: Наука, 1974.

8. Феллер В. Введение в теорию вероятностей: В 2 т.- М.: Мир, 1984.

9. Хургин Я.И. Да, нет или может быть. - М.: Наука, 1977.

10. Хургин Я.И. Как объять необъятное. - М.: Знание, 1979.