

## Конспект лекций по теории вероятностей

В данном конспекте в краткой форме излагается материал лекций по теории вероятностей, прочитанных на 2 потоке 3 курса Физического факультета МГУ в 2002 году. (Разбиение на лекции условное). Изложение, за исключением некоторых вопросов, соответствует учебнику Ю.П.Пытьева и И.А.Шимарева «Курс теории вероятностей и математической статистики для физиков», МГУ, 1983. По мере изложения лекций в 2003 году в конспекте предполагаются коррективы. Автор заранее приносит свои извинения за возможные опечатки.

Теория вероятностей есть математический анализ понятия случайного эксперимента. Событие и вероятность являются основными понятиями этой теории (аксиоматика Колмогорова, 1929, альтернативный, эмпирически-статистический подход — Р.Фишер, Р.Мизес.)

Лекция 1.

Предполагается, что результат или *исход* случайного эксперимента не может быть определен заранее. Однако сам эксперимент должен обладать свойством статистической устойчивости или устойчивости частот ( $\frac{N_A}{N}$ ).

Пусть  $\Omega = \{\omega\}$  — множество всех исходов и  $A$  есть некоторое событие, связанное с экспериментом. Естественно считать, что по исходу эксперимента можно сказать, осуществилось событие  $A$  или нет. Для наших дальнейших целей событие  $A$  можно отождествить с некоторой совокупностью исходов, то есть считать, что  $A$  есть подмножество элементов из  $\Omega$ . Сами элементы  $\omega$  мы будем называть элементарными событиями (предполагается, что их «нельзя» разбить на более мелкие).

Вероятность — числовая характеристика класса событий. Она имеет свойства, аналогичные относительной частоте события ( $\frac{N_A}{N}$ ), но не сводится к ней (не равна ей).

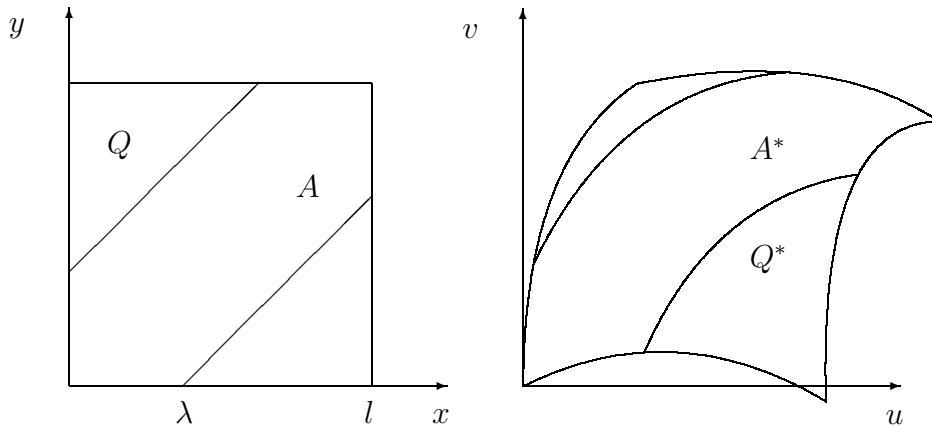
Примеры.

1. Бросание игральной кости — однородного куба. Выпадение граней  $\omega_1, \omega_2, \dots, \omega_6$  — элементарные события. Вероятность события  $\omega_i = \frac{1}{6}$ ,  $i = 1, 2, \dots, 6$ .  $A_1 = \{\omega_1, \omega_2, \omega_3\}$ ,  $A_2 = \{\omega_4, \omega_5\}$ ,  $A_3 = \{\omega_6\}$  — события, вероятности которых легко подсчитать:  $\frac{1}{2}$ ,  $\frac{1}{3}$ ,  $\frac{1}{6}$ .

2. Геометрические вероятности. Вероятность того, что точка, наудачу брошенная на отрезок  $[a, b]$  попадет в отрезок  $[\alpha, \beta]$ ,  $a \leq \alpha \leq \beta \leq b$ ,  $a < b$ , равна  $\frac{\beta - \alpha}{b - a} = \int_{\alpha}^{\beta} \frac{1}{b - a} dx = \int_{\alpha}^{\beta} p(x) dx$ , где  $p(x) = \frac{1}{b - a}$ .

Вероятность того, что точка, наудачу брошенная в квадрат  $Q$  размером  $l \times l$ , попадет в заданную область  $A$  с площадью  $S_A$  равна  $\frac{S_A}{l^2}$ , в частности, если область  $A$  определяется условием  $|\xi - \eta| < \lambda$ , где  $\xi$  и  $\eta$  — координаты точки, то  $P(A) = \frac{l^2 - (l - \lambda)^2}{l^2} = \int_A \frac{1}{l^2} dx dy = \int_{A^*} p(u, v) du dv$ . Здесь  $(x, y) \rightarrow (u, v)$  — невырожденное преобразование.  $p(u, v)$  — плотность вероятности.

В этих примерах элементарные события  $\omega$  — точки, множество  $\Omega$  — либо отрезок, либо квадрат  $Q$  (образ квадрата  $Q^*$ ).



**Алгебра событий.** (Алгебра множеств — это кольцо множеств с единицей).

Кольцо множеств — это класс множеств, замкнутый относительно операций  $A \cup B$  (объединение) и  $A \setminus B$  (вычитание).

Рассмотрим отношения между событиями, операции над событиями и некоторые специальные события<sup>1</sup>.

1.  $B \subset A$  —  $B$  влечет  $A$  ( $B$  — подмножество  $A$ ).
2.  $A = B$ , если  $B \subset A$  и  $A \subset B$ .
3.  $\bar{A}$  — противоположное событие ( $A$  не происходит).
4.  $\emptyset$  — невозможное и  $\Omega$  — достоверное события,  $\overline{\emptyset} = \Omega$ .
5.  $A \cap B$  — произведение (пересечение множеств) — события происходят одновременно.
6.  $A \cap B = \emptyset$  события несовместны (не могут произойти одновременно). Например,  $A \cap \bar{A} = \emptyset$ .
7.  $A \cup B$  объединение (происходит либо  $A$ , либо  $B$ ). Если они несовместны, то условимся писать  $A + B$  (вместо  $\bigcup_{i=1}^n$  писать  $\sum_{i=1}^n$ ).
8.  $A \setminus B = \{\omega \in \Omega : \omega \in A, \omega \notin B\}$  — вычитание.
9.  $A \Delta B = (A \setminus B) \cup (B \setminus A)$  — симметрическая разность.

Некоторые свойства операций над событиями.

- а)  $\overline{\bar{A}} = A$ ,  $\bar{\bar{A}} = \Omega \setminus A$ ,  $A \setminus B = A \cap \bar{B} = \bar{B} \setminus \bar{A}$ .
- б)  $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ,  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  — взаимная дистрибутивность.
- в)  $\overline{A \cup B} = \bar{A} \cap \bar{B}$ ,  $\overline{A \cap B} = \bar{A} \cup \bar{B}$  — принцип двойственности, верен также

<sup>1</sup>Их удобно иллюстрировать с помощью диаграмм Эйлера-Венна.

для любого числа элементов<sup>2</sup>.

$$г) A \subset B \iff \overline{B} \subset \overline{A}.$$

**Алгеброй событий** называется класс  $\mathcal{F}$  событий (система множеств), замкнутый относительно операций  $A \cup B$  и  $\overline{A}$ , то есть

- 1) из  $A, B \in \mathcal{F}$  следует  $A \cup B \in \mathcal{F}$ ,
- 2) из  $A \in \mathcal{F}$  следует  $\overline{A} \in \mathcal{F}$ .

Пример конечномерного множества  $\Omega$ . Пусть  $\Omega$  состоит из  $n$  несовместных событий (не считая  $\emptyset$  и  $\Omega$ ). Тогда  $\mathcal{F}$  — множество всех подмножеств  $\Omega$ , считая  $\emptyset$  и  $\Omega$  и всего их  $2^n$ .

**Классическая вероятность.**

Пусть дана полная группа  $\{A_i\}$ ,  $i = 1, 2, \dots, n$  попарно несовместных равновероятных событий. И пусть некоторое событие  $A$  имеет разложение  $A = A_{i_1} + \dots + A_{i_k}$ . В этом случае (по определению)  $P(A) = \frac{k}{n}$ . (Докажите корректность определения).

По иному: классическая вероятность равна отношению числа благоприятных (для данного события) исходов к числу всех возможных исходов эксперимента<sup>3</sup>.

**Свойства классической вероятности.**

1.  $0 \leq P(A) \leq 1$ .
2.  $P(\Omega) = 1, P(\emptyset) = 0$ .
3.  $P(A + B) = P(A) + P(B)$ . (См. парадокс Р.Мизеса о теннисисте<sup>4</sup>.)
4.  $P(\overline{A}) = 1 - P(A)$ .
5.  $A \subset B, P(A) \leq P(B), P(B \setminus A) = P(B) - P(A)$ . Следует из п. 3, если учесть  $B = A + B \setminus A$ .
6.  $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$  ( $A_1 \cup A_2 = A_1 + (A_2 \setminus (A_1 \cap A_2))$ ). Доказать самим формулу для  $n$  слагаемых:

$$P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{i < j}^n P(A_i \cap A_j) + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n).$$

Классическая вероятность применяется в основном в комбинаторных задачах, где подсчитывается число способов выбора (размещения)  $m$  объектов из  $n$  с или без учета порядка (различимых или неразличимых) и с возвращением (повторением) или без.

**Примеры.**

а. Размещение различных объектов без повторения (Аналог выборки с возвращением и с учетом порядка). Найти вероятность того, что хотя бы у одной пары в группе из  $n$  человек совпадают дни рождения.

Ответ:  $P(A) = 1 - \frac{r(r-1)\dots(r-n+1)}{r^n}$ . Если считать  $r = 365$ , то<sup>5</sup>

$n$	5	30	40	50	60
$P$	0.027	0.706	0.891	0.879	0.994

<sup>2</sup>Формулы де Моргана.

<sup>3</sup>Речь идет об определении числа исходов *до эксперимента*

<sup>4</sup>Вероятность выигрыша в Лондоне 0.6, в Москве 0.8. Эти события несовместны, однако их вероятности нельзя складывать, так как события принадлежат разным вероятностным пространствам.

<sup>5</sup>В числителе число  $A_r^n = \frac{r!}{(r-n)!}$  размещений  $n$  дней рождения по  $r$  дням года.

б. Выборка без возвращения и без учета порядка. Имеется  $n$  объектов, среди них  $k$  отмеченных. Выбирается (без возвращения)  $n_1$  объектов. Найти вероятность того, что среди них  $k_1$  отмеченных.

$$P(A) = \frac{C_k^{k_1} C_{n-k}^{n_1-k_1}}{C_n^{n_1}}.$$

(Это — гипергеометрическое распределение).

в. Еще три примера (из физики). Распределение  $r$  частиц по  $n$  ячейкам.

Статистика **Максвелла-Больцмана** — все частицы разные, запретов никаких нет. Число состояний  $n^r$ , вероятность каждого состояния  $p = n^{-r}$ .

Статистика **Бозе-Эйнштейна** — частицы тождественны (неразличимы), в каждой ячейке может быть сколько угодно частиц. Число состояний  $C_{n+r-1}^r$ , вероятность каждого  $p = \frac{1}{C_{n+r-1}^r}$ .

Статистика **Ферми-Дирака** — дополнительно к предыдущему действует принцип запрета: в каждой ячейке может быть только одна частица. Число состояний  $C_n^r$ , вероятность  $p = \frac{1}{C_n^r}$ .

### Аксиомы теории вероятностей.

Лекция 2.

Пусть  $\Omega$  — пространство элементарных событий,  $\mathcal{F}$  — алгебра событий (подмножеств  $\Omega$ ). Следующие условия являются аксиомами теории вероятностей:

1.  $\mathcal{F}$  является  $\sigma$ -алгеброй, т.е. алгеброй, замкнутой относительно операции счетного объединения событий<sup>6</sup>,  $A_n \in \mathcal{F}$ ,  $n = 1, 2, \dots, \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ .

2. На  $\sigma$ -алгебре  $\mathcal{F}$  определена функция  $P(\cdot)$ , принимающая числовые значения,  $P(A) \geq 0$ ,  $A \in \mathcal{F}$ , называемая вероятностью.

3.  $P(A + B) = P(A) + P(B)$  — аксиома сложения (верная также в случае конечного разложения,  $P(\sum_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ ).

4.  $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$  —  $\sigma$ -аддитивность.

5.  $P(\Omega) = 1$ .

Тройка  $(\Omega, \mathcal{F}, P(\cdot))$  называется вероятностным пространством (вероятностной моделью).

Аксиомы 1 – 5 непротиворечивы, поскольку существует пример — классическая вероятность, но не полны, поскольку не задана явно вероятность  $P(\cdot)$ .

Задание вероятности на  $\sigma$ -алгебре  $\mathcal{F}$  элементарно лишь для простейших случаев, например, в случае, когда  $\Omega$  конечно или счетно. Этот случай мы рассмотрим чуть позже.

Примерами  $\sigma$ -алгебр являются  $\mathcal{F}_* = \{\emptyset, \Omega\}$  и  $\mathcal{F}^*$  — множество всех подмножеств  $\Omega$ . Первая из них тривиальна, а вторая в общем случае слишком обширна, чтобы на ней можно непротиворечиво задать вероятность. На каких же  $\sigma$ -алгебрах можно задать вероятность и как это сделать? Решение этого вопроса лежит на следующем пути. Пусть  $\mathcal{A}$  — произвольная система множеств  $A_i$ ,  $\bigcup A_i = \Omega \in \mathcal{A}$ . Ясно, что любые  $\sigma$ -алгебры  $\mathcal{F}_1$  и  $\mathcal{F}_2$ , для которых  $\mathcal{A} \subset \mathcal{F}_1, \mathcal{F}_2$ ,

<sup>6</sup>Пример алгебры, не являющейся  $\sigma$ -алгеброй: конечное число полуинтервалов  $(a, b]$ ,  $0 < a < b \leq 1$  на  $(0, 1]$ .

удовлетворяют условию  $\mathcal{F}_* \subset \mathcal{F}_1, \mathcal{F}_2 \subset \mathcal{F}^*$ , причем  $\mathcal{F}_1 \cap \mathcal{F}_2$  — тоже  $\sigma$ -алгебра. Пусть  $\mathcal{F}_A = \bigcap_{\alpha: A \subset \mathcal{F}_\alpha} \mathcal{F}_\alpha$ , тогда  $\mathcal{F}_A$  — минимальная  $\sigma$ -алгебра, содержащая  $A$ .

Если, например, в качестве  $\mathcal{A}$  взять систему интервалов на прямой, то минимальная  $\sigma$ -алгебра, содержащая  $\mathcal{A}$ , даст  $\sigma$ -алгебру *борелевских множеств* на прямой. Оказывается, что к этой же  $\sigma$ -алгебре приводит система открытых множеств прямой или замкнутых множеств прямой и т.д. Сам процесс построения  $\sigma$ -алгебры называется борелевским замыканием класса  $\mathcal{A}$ . Этот класс достаточно обширный: он содержит все открытые и все замкнутые множества на прямой. Дальнейшая идея заключается в том, чтобы задать вероятность на более простом классе множеств  $\mathcal{A}$  (например, на полуалгебре<sup>7</sup>), а затем воспользоваться теоремой о единственном продолжении меры на минимальную  $\sigma$ -алгебру, содержащую  $\mathcal{A}$ . (Мера — функция множества — обобщение понятия длины; вероятность — мера, нормированная на единицу).

**Дискретное вероятностное пространство.**  $\Omega = \{\omega_i\}$  — конечно или счетно,  $\mathcal{F}$  — множество всех подмножеств  $\Omega$ ,  $P(\cdot)$  достаточно определить для каждого элементарного события  $P(\{\omega_i\}) = p_i$ , лишь бы  $\sum_{i=1}^{\infty} p_i = 1$ .

Тогда вероятность для любого события  $A \in \mathcal{F}$  равна  $P(A) = \sum_{\omega_i \in A} p_i$ .

Корректность следует из *Леммы о суммировании по блокам*:

Пусть ряд  $\sum_{i=1}^{\infty} c_i = S$  сходится абсолютно и пусть  $I = I_1 + I_2 + \dots$  — разбиение множества натуральных чисел  $I$ . Обозначим  $S_k = \sum_{i \in I_k} c_i$ . Тогда ряд  $\sum_{k=1}^{\infty} S_k$  сходится абсолютно и равен  $S$ .

*Доказательство.* Абсолютная сходимость следует из  $\sum_{k=1}^n |S_k| \leq \sum_{k=1}^n \sum_{i \in I_k} |c_i|$ .

Имеем  $\forall \varepsilon > 0 \exists N_1 = N_1(\varepsilon)$ , что при  $n \geq N_1$   $|\sum_{i=1}^n c_i - S| < \varepsilon/2$  и  $\sum_{i=n+1}^{\infty} |c_i| < \varepsilon/2$ .

Далее выберем  $N_2$  столь большим, чтобы в сумму  $\sum_{k=1}^{N_2} S_k = \sum_{k=1}^{N_2} \sum_{i \in I_k} c_i$  входили все числа  $c_i$  с номерами  $i \leq N_1$  и положим  $N_0 = \max(N_1, N_2)$ .

Тогда при  $n \geq N_0$  получаем

$$\left| \sum_{k=1}^n S_k - S \right| \leq \left| \sum_{i=1}^n c_i - S \right| + \left| \sum_{k=1}^n S_k - \sum_{i=1}^n c_i \right| \leq \left| \sum_{i=1}^n c_i - S \right| + \sum_{i=n+1}^{\infty} |c_i| \leq \varepsilon/2 + \varepsilon/2 = \varepsilon. \square$$

Обратное, вообще говоря, неверно, но если  $c_i \geq 0$ , то верно и обратное утверждение.

**Пример.** Бросание монеты до первого выпадения герба (герб — единица, аверс — ноль). Элементарные события  $\omega_1 = \{1\}$ ,  $\omega_2 = \{01\}$ , ...,  $\omega_\infty = \{00\dots\}$ .

<sup>7</sup>Полуалгебра — полукольцо с единицей; полукольцо — непустое множество, замкнутое относительно пересечения, и в котором каждая разность допускает конечное разложение

$$A \setminus B = \sum_{j=1}^n A_j$$

Вероятности:  $P(\omega_1) = p_1 = \frac{1}{2}$ ,  $P(\omega_2) = p_2 = \frac{1}{4}$ , ...,  $P(\omega_\infty) = p_\infty = 0$ ,  $\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1$ . Событие  $\omega_\infty$  возможное, но невероятное!

Аксиоматически определенная вероятность обладает всеми свойствами, которые мы отметили для классической вероятности, поскольку первые три свойства фактически повторяют аксиомы, а остальные свойства, как мы видели, выводятся из них.

1.  $0 \leq P(A) \leq 1$ .
2.  $P(\Omega) = 1$ ,  $P(\emptyset) = 0$ .
3.  $P(A + B) = P(A) + P(B)$ .
4.  $P(\overline{A}) = 1 - P(A)$ .
5.  $A \subset B \implies P(A) \leq P(B)$ ,  $P(B \setminus A) = P(B) - P(A)$ .
6.  $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$  (т.к.  $A_1 \cup A_2 = A_1 + (A_2 \setminus (A_1 \cap A_2))$ ).

Однако, есть еще одно свойство, которое вытекает из  $\sigma$ -аддитивности и называется непрерывностью вероятности, точнее, непрерывностью относительно предельного перехода.

Сначала определим понятие предела последовательности событий (множеств)  $\{A_i, i = 1, 2, \dots\}$ . Также, как для числовых последовательностей, можно это сделать через верхний и нижний пределы.

$A^* = \overline{\lim}_{n \rightarrow \infty} A_n = \limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$  — событие, заключающееся в том, что произошло бесконечно много событий из  $\{A_i\}$ .

$A_* = \underline{\lim}_{n \rightarrow \infty} A_n = \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k$  — событие, заключающееся в том, что произошли все события из  $\{A_i\}$  за исключением, быть может, конечного их числа<sup>8</sup>.

Очевидно, что  $\underline{\lim} A_n \subset \overline{\lim} A_n$ . Если  $\overline{\lim} A_n = \underline{\lim} A_n$ , то говорят, что последовательность событий имеет предел. Для монотонных последовательностей событий:  $A_1 \subset A_2 \dots \subset A_n \dots$  или  $B_1 \supset B_2 \dots \supset B_n \dots$  предел  $\lim \uparrow A_n$  ( $\lim \downarrow B_n$ ) всегда существует.

7.  $P(\lim A_n) = \lim P(A_n)$  — следствие 4 аксиомы, которую можно переформулировать как непрерывность вероятности относительно монотонных предельных переходов, в частности,  $\lim P(A_n) = 0$  если  $\lim \downarrow A_n = \emptyset$ . **7 Условная**

### вероятность.

Рассмотрим  $(\Omega, \mathcal{F}, P(\cdot))$  — вероятностное пространство. Пусть известно, что в ходе эксперимента произошло событие  $B$ ,  $P(B) > 0$ , в результате мы можем сузить множество исходов до  $\Omega_B = B$ , а вместо любого события  $A$  рассматривать  $A_B = A \cap B$ . Таким образом, мы переходим от  $(\Omega, \mathcal{F}, P(\cdot))$  к  $(\Omega_B, \mathcal{F}_B, P_B(\cdot))$ , где  $\Omega_B = \Omega \cap B$ ,  $\mathcal{F}_B = \{A \cap B, A \in \mathcal{F}\}$ , а для вероятности  $P(A)$  естественно использовать формулу

$$P_B(A_B) = P(A|B) = \frac{P(A \cap B)}{P(B)}. \quad (1)$$

<sup>8</sup>Полезное представление  $\chi_{A^*} = \overline{\lim} \chi_{A_n^*}$ ,  $\chi_{A_*} = \underline{\lim} \chi_{A_{**}}$ , где  $\chi_A$  — характеристическая функция множества  $A$ .

(Покажите, что  $\mathcal{F}_B$  —  $\sigma$ -алгебра, а  $P_B(\cdot)$  удовлетворяет аксиомам.)

Можно вернуться к старому вероятностному пространству  $(\Omega, \mathcal{F}, P(\cdot))$  и рассматривать вероятность  $P_B(\cdot)$  на  $\mathcal{F}$ . В этом случае ее называют условной вероятностью и обозначают  $P(\cdot|B)$ . Событие  $B$  можно рассматривать как параметр. Формула (1) легко интерпретируется в терминах классической вероятности, а в общем случае она является определением.

Свойства. (Докажите самостоятельно.)

$$P(\Omega|B) = 1, \quad P(A_1 + A_2|B) = P(A_1|B) + P(A_2|B),$$

$$P(\bar{A}|B) = 1 - P(A|B), \quad P\left(\sum_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B).$$

Эти свойства полезны для решения задач. **Пример:** Вероятность аварии ракеты 0.1, причем на старте 0.09. Какова вероятность аварии в случае успешного старта? Пусть событие  $A$  — авария,  $B$  — авария на старте,  $B \subset A$ ,  $\bar{A} \subset \bar{B}$ . Искомая вероятность  $P(A|\bar{B}) = 1 - P(\bar{A}|\bar{B}) = 1 - \frac{P(\bar{A} \cap \bar{B})}{P(\bar{B})} = 1 - \frac{P(\bar{A})}{P(\bar{B})} = 1 - \frac{1-0.1}{1-0.09} = \frac{1}{91}$ .

**Теорема умножения вероятностей.**

$$P(A \cap B) = P(A|B)P(B), \quad P(B) > 0.$$

Для трех событий

$$P(A \cap B \cap C) = P(A|B \cap C)P(B|C)P(C) = P(A \cap B|C)P(C), \text{ отсюда}$$

$$P(A \cap B|C) = P(A|B \cap C)P(B|C).$$

**Независимость.** События  $A$  и  $B$  независимы<sup>9</sup>, если  $P(A \cap B) = P(A) \cdot P(B)$ .

Если  $P(A) = 0$ , то  $A$  и  $B$  независимы.

Если  $P(A) > 0$ , то  $P(B|A) = P(B)$  означает «независимость»  $B$  от  $A$ .

Свойства.

1.  $A$  и  $\Omega$  независимы.
2.  $A$  и  $B$ , если  $P(A) = 0$ , независимы.
3. Если  $A$  и  $B$  независимы, то  $A$  и  $\bar{B}$ ,  $\bar{A}$  и  $B$ ,  $\bar{A}$  и  $\bar{B}$  — также независимы.
4. Если  $A$  и  $B_i$  попарно независимы  $i = 1, 2, \dots, n$ , то независимы  $A$  и  $\sum A_i$  (но не  $\bigcup A_i$ ).

Независимость в совокупности. События  $\{A_i, i = 1, 2, \dots, n\}$  независимы в совокупности, если для любых  $m \leq n$  и любых наборов различных индексов  $i_1, i_2, \dots, i_m$  имеет место равенство  $P\left(\bigcap_{k=1}^m A_{i_k}\right) = \prod_{k=1}^m P(A_{i_k})$ .

Задача (Пример Бернштейна). Пусть правильный тетраэдр раскрашен так, что на трех его гранях красный, синий и зеленый цвет соответственно, а на четвертой — все три цвета. Проверьте, что события «выпадение разных цветов» попарно независимы, но не независимы в совокупности.

**Формула полной вероятности. Формулы Байеса.**

Лекция 3.

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, P(\cdot))$ .

Пусть  $\{A_i, i = 1, 2, \dots\}$  — полная группа (необязательно конечная) попарно несовместных (необязательно равновероятных) событий,  $A_i \cap A_j = \emptyset, i \neq j$  и пусть  $B \in \mathcal{F}$  — некоторое событие,  $P(B) > 0$ . Тогда

$$B = B \cap \left(\sum_i A_i\right) \text{ и } P(B) = \sum_i P(B|A_i)P(A_i) \text{ — формула полной вероятности.}$$

Из  $P(B \cap A_k) = P(A_k|B)P(B) = P(B|A_k)P(A_k)$  следует

<sup>9</sup>стохастически независимы.

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_i P(B|A_i)P(A_i)} \text{ — формулы Байеса. } \square$$

**Пример построения независимых событий.**

Пусть  $(\Omega_1, \mathcal{F}_1, P_1(\cdot))$ ,  $(\Omega_2, \mathcal{F}_2, P_2(\cdot))$  — дискретные (для простоты) вероятностные пространства,  $\omega_i^1 \in \Omega_1$ ,  $\omega_j^2 \in \Omega_2$ ,  $P_1(\omega_i^1) = p_i^1$ ,  $P_2(\omega_j^2) = p_j^2$ . Рассмотрим множество упорядоченных пар  $\{\omega_i^1 \omega_j^2\}$ , обозначим его  $\Omega = \Omega_1 \times \Omega_2$  — прямое произведение. Множество всевозможных подмножеств  $\Omega$ , которое, очевидно,  $\sigma$ -алгебра, обозначим  $\mathcal{F}$ . Наконец, введем  $\Omega$ :  $P(\{\omega_i^1 \omega_j^2\}) = p_{ij} = P_1(\omega_i^1)P_2(\omega_j^2) = p_i^1 p_j^2$  (проверить корректность!). Полученное вероятностное пространство назовем прямым произведением:  $(\Omega, \mathcal{F}, P(\cdot)) = (\Omega_1, \mathcal{F}_1, P_1(\cdot)) \times (\Omega_2, \mathcal{F}_2, P_2(\cdot))$ .

Пусть в этой схеме  $A_1 \in \mathcal{F}_1$ ,  $A_2 \in \mathcal{F}_2$ ,

$$A = \sum_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in \Omega}} \{\omega_i^1 \omega_j^2\} \text{ и } B = \sum_{\substack{i:\omega_i^1 \in \Omega \\ j:\omega_j^2 \in A_2}} \{\omega_i^1 \omega_j^2\}, \quad A, B \in \mathcal{F}.$$

Найдем вероятность

$$P(A) = \sum_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in \Omega}} p_{ij} = \sum_{i:\omega_i^1 \in A_1} p_i^1 \sum_{j:\omega_j^2 \in \Omega} p_j^2 = \sum_{i:\omega_i^1 \in A_1} p_i^1.$$

Аналогично,  $P(B) = \sum_{j:\omega_j^2 \in A_2} p_j^2$ .

Наконец,  $P(A \cap B) = \sum_{\substack{i:\omega_i^1 \in A_1 \\ j:\omega_j^2 \in A_2}} p_{ij} = \sum_{i:\omega_i^1 \in A_1} p_i^1 \sum_{j:\omega_j^2 \in A_2} p_j^2 = P(A)P(B)$ ,

откуда следует их независимость.  $\square$

*Определение 1.* Пусть  $(\Omega, \mathcal{F}, P(\cdot))$  — дискретное вероятностное пространство. Последовательностью независимых испытаний называется вероятностное пространство  $(\Omega_n, \mathcal{F}_n, P_n(\cdot))$ , которое является прямым произведением  $n$  одинаковых пространств ( $n$ -й степенью):  $(\Omega, \mathcal{F}, P(\cdot))$ , т.е.  $(\Omega_n, \mathcal{F}_n, P_n(\cdot)) = \overset{n}{\times} (\Omega, \mathcal{F}, P(\cdot))$ . Подробнее:  $\Omega_n$  состоит из цепочек  $(\omega_{i_1} \omega_{i_2} \dots \omega_{i_n})$  длины  $n$  с необязательно различными индексами,  $\mathcal{F}_n$  — алгебра подмножеств  $\Omega_n$ ,  $P_n(\omega_{i_1} \omega_{i_2} \dots \omega_{i_n}) = p_{i_1} p_{i_2} \dots p_{i_n}$ .

*Определение 2.* Схемой Бернулли называется последовательность  $n$  независимых испытаний, полученная на основе вероятностного пространства, в котором содержится лишь два элементарных события (исхода):  $\omega_1$  — успех и  $\omega_2$  — неудача (1 и 0 соответственно). Обозначив  $P(\omega_1) = p$ ,  $P(\omega_2) = q = 1 - p$ , получим  $P_n(\omega_{i_1} \omega_{i_2} \dots \omega_{i_n}) = p^k q^{n-k}$ , где  $k$  — число успехов в серии из  $n$  испытаний Бернулли.

Какова вероятность при  $n$  испытаниях получить  $k$  успехов? Это  $p_n(k) = C_n^k p^k q^{n-k}$  — биномиальное распределение,  $\sum_{k=0}^n p_n(k) = 1$ .

На это распределение очень много задач и вы с ними познакомитесь на семинарских занятиях.

Отрицательное биномиальное распределение (распределение Паскаля):

Вероятность того, что для достижения  $n$  успехов в схеме Бернулли потре-

буется  $n + k$  испытаний равна  $p(n, n + k) = C_{n+k-1}^k p^n q^k$ .

$$\begin{aligned} \sum_{k=0}^{\infty} p(n, n + k) &= \sum_{k=0}^{\infty} \frac{(n + k - 1)(n + k - 2) \dots (n)}{k!} p^n q^k = \\ &= \sum_{k=0}^{\infty} \frac{(-n)(-n - 1) \dots (-n - k + 1)}{k!} p^n (-q)^k = p^n (1 - q)^{-n} = 1. \end{aligned}$$

Пусть  $\Omega = \{\omega_1 \omega_2 \dots \omega_r\}$ . Тогда

$$p_n(s_1, s_2, \dots, s_r) = \frac{n!}{s_1! s_2! \dots s_r!} p_1^{s_1} p_2^{s_2} \dots p_r^{s_r}, \quad \sum_i s_i = n$$

— полиномиальное распределение (общий член разложения полинома  $(p_1 + p_2 + \dots + p_r)^n$ ).

**Предельные теоремы** (для биномиального распределения).

**Теорема Пуассона.**

Рассмотрим последовательность биномиальных распределений. Пусть  $p = \frac{\lambda}{n}$  (т.е.  $\lambda = pn = \text{Const}$ ).

Тогда при  $n \rightarrow \infty$ :  $p_n(k) \Rightarrow p_{\infty}(k) = \frac{\lambda^k}{k!} e^{-\lambda}$ .

Доказательство.

$$\begin{aligned} p_n(k) &= C_n^k p^k q^{n-k} = \frac{n(n-1) \dots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^n = \\ &= \frac{1(1 - \frac{1}{n}) \dots (1 - \frac{k-1}{n}) \lambda^k}{(1 - \frac{\lambda}{n})^k k!} \left(1 - \frac{\lambda}{n}\right)^n \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}. \square \end{aligned}$$

Замечание. Теорема, очевидно, справедлива и в случае, когда величина  $pn$  не постоянна, но стремится к  $\lambda$ .

Теорему практически можно применять<sup>10</sup> при  $n \geq 100$ ,  $\lambda \leq 10$ .

**Теорема Муавра-Лапласа.** (В этом случае  $npq$  велико, так что  $p > 0$  и  $q > 0$  — фиксированы).

Локальная.

$$\lim_{n \rightarrow \infty} \frac{\sqrt{npq} p_n(m)}{\frac{1}{\sqrt{2\pi}} e^{-\frac{x_m^2}{2}}} = 1, \quad p_n(m) \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x_m^2}{2}}, \quad x_m = \frac{m - np}{\sqrt{npq}}.$$

Интегральная. Эта теорема будет получена как следствие из более общей, так называемой Центральной Предельной Теоремы.

$$P \left\{ a \leq \frac{m - np}{\sqrt{npq}} < b \right\} \approx \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

А локальную можно вывести из интегральной следующим образом.

<sup>10</sup> Аккуратную оценку погрешности можно посмотреть в книге *Боровков А.А.* Теория вероятностей. — М.: Наука, 1976.

Положим  $a = \frac{k-np}{\sqrt{npq}}$ ,  $b = \frac{k+1-np}{\sqrt{npq}}$ , так что  $b - a = \frac{1}{\sqrt{npq}}$ . Тогда, используя теорему о среднем, получим

$$p_n(k) = P\{m = k\} = P\left\{a \leq \frac{m - np}{\sqrt{npq}} < b\right\} \approx \frac{1}{\sqrt{2\pi}}(b - a)e^{-\frac{x^2}{2}} \Big|_{x \in (a,b)} \approx \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{npq}} e^{-\frac{x^2}{2}}. \square$$

*Замечание.* Если  $n^k p \rightarrow \lambda$ , где  $0 < k < 1$  то можно применять обе приближенные формулы.

**Пример.** Сколько раз нужно бросить монету, чтобы с вероятностью 0.99 частота появления герба отличалась от  $\frac{1}{2}$  не более, чем на 0.01?

Решение. Согласно интегральной теореме Муавра-Лапласа

$$P\left(\left|\frac{m}{n} - p\right| < 0.01\right) = P\left(\left|\frac{m}{n} - p\right| < \sqrt{\frac{pq}{n}}\varepsilon\right) = P\left\{\left|\frac{m - np}{\sqrt{npq}}\right| < \varepsilon\right\} = \approx \frac{1}{\sqrt{2\pi}} \int_{-\varepsilon}^{\varepsilon} e^{-\frac{x^2}{2}} dx = 1 - 2\Phi(-\varepsilon) = 0.99.$$

Здесь  $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$  — интеграл ошибок (табулирован).

$\varepsilon = \sqrt{\frac{n}{pq}} 0.01 = 2.58$  (из таблиц), откуда  $n = (129)^2 = 16641$ .  $\square$

### Цепи Маркова.

Лекция 4.

Рассмотрим последовательность испытаний, где элементарным событием является цепочка  $\omega_{i_1}\omega_{i_2}\dots\omega_{i_n}$ . Согласно теореме умножения вероятностей,  $P(\omega_{i_1}\omega_{i_2}\dots\omega_{i_n}) = P(\omega_{i_n} | \omega_{i_1}\dots\omega_{i_{n-1}})\dots p(\omega_{i_1})$ . Пусть последовательные испытания «минимально» зависимы:  $P(\omega_{i_s} | \omega_{i_1}\dots\omega_{i_{s-1}}) = P(\omega_{i_s} | \omega_{i_{s-1}})$  и не зависит от абсолютного номера испытания  $s$ , то есть  $P(\omega_j | \omega_i) = p_{ij}$ . Пусть известно начальное распределение вероятностей  $P(\omega_i) = a_i$ . Тогда

$$P(\omega_{i_1}\omega_{i_2}\dots\omega_{i_n}) = a_{i_1}p_{i_1i_2}\dots p_{i_{n-1}i_n}. \quad (2)$$

*Определение.* Пусть задано вероятностное пространство с конечным набором элементарных событий  $\{\omega_i\}$ ,  $i = 1, \dots, N$ . Конечной однородной цепью Маркова называется вероятностное пространство, в котором  $\Omega_n = \overset{n}{\times} \Omega$ ,  $\mathcal{F}_n = \overset{n}{\times} \mathcal{F}$ , а вероятность  $P_n(\omega_{i_1}\omega_{i_2}\dots\omega_{i_n})$  определяется формулой (2).

Для вычисления (2), в частности, должны быть заданы переходные вероятности  $p_{ij}$  или переходная матрица

$$\pi = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1N} \\ p_{21} & p_{22} & \dots & p_{2N} \\ \dots & \dots & \dots & \dots \\ p_{N1} & p_{N2} & \dots & p_{NN} \end{pmatrix}$$

Это так называемая стохастическая матрица ( $\sum_j p_{ij} = 1$ ). Отсюда следует корректность задания (2) — сумма по всем индексам равна 1. (Проверить!) Марковскую цепь можно интерпретировать как модель системы с  $N$  состояниями и вероятностями перехода  $p_{ij}$  из  $i$ -го в  $j$ -е состояние.

Примеры: случайное дискретное блуждание по прямой.

Частный случай — независимые испытания ( $p_{ij} = p_j$ ).

*Замечание.* Для практики удобен следующий критерий «Марковости».

Обозначим буквами П, Н и Б — три последовательных события в цепи Маркова: прошлое, настоящее и будущее. Тогда  $P(\text{ПБ}|\text{Н}) = P(\text{П}|\text{Н})P(\text{Б}|\text{Н})$  — при фиксированном настоящем прошлое и будущее независимы. Доказательство следует из равенств

$$\begin{aligned} P(\text{ПНБ}) &= P(\text{ПБ}|\text{Н})P(\text{Н}) = P(\text{Б}|\text{ПН})P(\text{П}|\text{Н})P(\text{Н}) = \\ &= P(\text{Б}|\text{Н})P(\text{П}|\text{Н})P(\text{Н}). \end{aligned}$$

(Докажите самостоятельно необходимость и достаточность!)

**Переход за  $n$  шагов.** Обозначим через  $p_{ij}^{(n)}$  вероятность перехода из  $i$ -го в  $j$ -е состояние за  $n$  шагов. Легко предположить (однородность!), что она не зависит от абсолютного номера шага  $s$  (речь идет о переходе из состояния  $i$  на  $s$ -м шаге в состояние  $j$  на  $s + n$ -м шаге). По формуле полной вероятности для  $0 < m < n$  имеем

$$p_{ij}^{(n)} = \sum_{k=1}^N p_{ik}^{(m)} p_{kj}^{(n-m)} \quad \text{или} \quad \pi^{(n)} = \pi^{(m)} \pi^{(n-m)}. \quad (3)$$

Отсюда следует, что  $\pi^{(n)} = \pi^n$  и  $\pi^n = \pi^m \pi^{n-m}$  (в данном случае это просто степень матрицы).

Отметим, что  $\pi^n$  — тоже стохастическая матрица (Доказать!).

**Эргодичность.** Рассмотрим абсолютную вероятность  $p_j^n = \sum_{i=1}^N a_i p_{ij}^{(n)}$  (нахождения системы в состоянии  $j$  на  $n$ -м шаге от начала) и предположим, что<sup>11</sup>  $\lim_{n \rightarrow \infty} p_{ij}^n = p_j$ .

Тогда  $\lim_{n \rightarrow \infty} p_j^n = \lim_{n \rightarrow \infty} \sum_{i=1}^N a_i p_{ij}^n = \sum_{i=1}^N a_i p_j = p_j$  — финальное распределение вероятностей. Из равенства  $\sum_{i=1}^N p_i^{n-1} p_{ij} = p_j^n$  ((3),  $m = n - 1$ ) при  $n \rightarrow \infty$  следует

$$\sum_{i=1}^N p_i p_{ij} = p_j. \quad (4)$$

Это значит, что финальное распределение стационарно и является решением системы алгебраических уравнений<sup>12</sup> с дополнительными условиями  $p_j \geq 0$  и  $\sum_{i=1}^N p_i = 1$ .

<sup>11</sup> Далее опустим скобки у верхнего индекса.

<sup>12</sup> Левый собственный вектор матрицы  $\pi$ , отвечающий собственному значению, равному 1.

Если при этом  $p_j > 0$ ,  $j = 1, \dots, N$ , то говорят, что Марковская цепь (система) эргодична. В этом случае безусловные (абсолютные) вероятности  $p_j^{(n)}$  стремятся к  $p_j$  независимо от начального распределения, а система (4) при указанных условиях имеет единственное решение.

Полезна следующая теорема для конечных марковских цепей:

**Теорема Маркова.** Пусть существуют целые  $\nu$  и  $j_0$ , такие, что  $\min_i p_{ij_0}^\nu = \delta > 0$ . Тогда существуют числа  $p_j$ ,  $j = 1, 2, \dots, N$ , такие, что  $\lim_{n \rightarrow \infty} p_{ij}^n = p_j$ ,  $j = 1, 2, \dots, N$  независимо от  $i$ .

Доказательство. Обозначим  $M_j^r = \max_i p_{ij}^r$ ,  $m_j^r = \min_i p_{ij}^r$ ,  $m_j^r \leq M_j^r$ .

Далее  $m_j^{r+1} = \min_i \sum_k p_{ik} p_{kj}^r \geq \min_i \sum_k p_{ik} \min_k p_{kj}^r = m_j^r$ , аналогично  $M_j^{r+1} \leq M_j^r$ .

Таким образом,

$$m_j^1 \leq m_j^2 \leq \dots \leq m_j^r \leq \dots \leq M_j^r \leq \dots \leq M_j^2 \leq M_j^1.$$

Из монотонности и ограниченности следует, что существуют пределы  $\lim_{n \rightarrow \infty} M_j^n = M_j$  и  $\lim_{n \rightarrow \infty} m_j^n = m_j$ .

Фиксируем две строки матрицы  $\pi^\nu \alpha$  и  $\beta$  (т.е. рассмотрим  $p_{\alpha k}^\nu$  и  $p_{\beta k}^\nu$ ),  $k = 1, 2, \dots, N$ .

Обозначим  $\sum^+ = \sum_{k: p_{\alpha k}^\nu \geq p_{\beta k}^\nu}$  и  $\sum^- = \sum_{k: p_{\alpha k}^\nu < p_{\beta k}^\nu}$ .

Тогда

$$\begin{aligned} \sum^+ p_{\alpha k}^\nu + \sum^- p_{\alpha k}^\nu &= 1, \quad \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) + \sum^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) = 0, \\ \sum^+ p_{\alpha k}^\nu - \sum^+ p_{\beta k}^\nu &= 1 - \sum^- p_{\alpha k}^\nu - \sum^+ p_{\beta k}^\nu \leq 1 - \delta. \end{aligned}$$

Далее

$$\begin{aligned} M_j^{\nu+n} - m_j^{\nu+n} &= \max_\alpha \sum_k p_{\alpha k}^\nu p_{kj}^n - \min_\beta \sum_k p_{\beta k}^\nu p_{kj}^n = \\ &= \max_{\alpha, \beta} \sum_k (p_{\alpha k}^\nu p_{kj}^n - p_{\beta k}^\nu p_{kj}^n) = \\ &= \max_{\alpha, \beta} \left\{ \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) p_{kj}^n + \sum^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) p_{kj}^n \right\} \leq \\ &\leq \max_{\alpha, \beta} \left\{ \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) M_j^n - \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) m_j^n \right\} = \\ &= \max_{\alpha, \beta} \left\{ \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) M_j^n + \sum^- (p_{\alpha k}^\nu - p_{\beta k}^\nu) m_j^n \right\} = \\ &= \max_{\alpha, \beta} \sum^+ (p_{\alpha k}^\nu - p_{\beta k}^\nu) (M_j^n - m_j^n) \leq \\ &\leq (1 - \delta) (M_j^n - m_j^n). \end{aligned}$$

Переходя к пределу при  $n \rightarrow \infty$ , получим

$$M_j - m_j \leq (1 - \delta) (M_j - m_j),$$

что возможно лишь при  $M_j = m_j$ .  $\square$

**Теорема** (без доказательства). Пусть  $A = \{\omega_i\}$  — подмножество состояний эргодической системы,  $T_A$  — время нахождения системы в  $A$ ,  $T$  — общее время функционирования системы. Тогда

$$\lim_{T \rightarrow \infty} \frac{T_A}{T} = \sum_{\omega_i \in A} p_i.$$

(Среднее по времени совпадает со средним по реализациям.)

### Случайная величина.

Лекция 5.

Пусть задано вероятностное пространство  $(\Omega, \mathcal{F}, P(\cdot))$ .

1. Случайной величиной называется вещественная функция  $\xi(\omega)$  (отображение  $\Omega \rightarrow \mathcal{R}_1$ ), такая, что для любого вещественного  $x$  множество  $\{\omega \in \Omega : \xi(\omega) < x\} \in \mathcal{F}$ , т.е. *событие*.

2. Вероятность  $P(\{\xi(\omega) < x\}) = F_\xi(x)$  называется функцией распределения случайной величины  $\xi$ .

Свойства функции распределения.

1. Для  $x_1 < x_2$ :  $\{\xi < x_2\} = \{\xi < x_1\} + \{x_1 \leq \xi < x_2\} \Rightarrow P(\xi < x_2) = P(\xi < x_1) + P(x_1 \leq \xi < x_2) \Rightarrow F(x_1) \leq F(x_2)$ ,  $P(x_1 \leq \xi < x_2) = F(x_2) - F(x_1)$ .

2.  $0 \leq F(x) \leq 1$ ,  $x \in \mathcal{R}_1$ .

3.  $F(x)$  непрерывна слева,  $\lim_{x_k \uparrow x} F(x_k) = F(x)$ . Пусть  $x_1 < x_2 < \dots < x_n < \dots < x$ . Доказательство следует из того, что  $\{\xi < x\} = \bigcup_k \{\xi < x_k\}$  и  $F(x) =$

$\lim_{x_k \uparrow x} F(x_k)$  в силу непрерывности вероятности для монотонных последовательностей событий.

4.  $P\{\xi \leq x\} = \lim_{x_k \downarrow x} F(x_k) = F(x+0)$ . В этом случае для  $x_1 > x_2 > \dots > x_n > \dots > x$  имеем  $\{\xi \leq x\} = \bigcap_k \{\xi < x_k\}$  и далее используем непрерывность вероятности для монотонных последовательностей событий.

5. Выпишем соотношения для  $x_1 \leq x_2$ :

$$P\{x_1 \leq \xi < x_2\} = F(x_2) - F(x_1),$$

$P\{x_1 \leq \xi \leq x_2\} = F(x_2+0) - F(x_1)$ , в частности, при  $x_1 = x_2 = x$ :  $P\{\xi = x\} = F(x+0) - F(x)$ ,

$$P\{x_1 < \xi \leq x_2\} = F(x_2+0) - F(x_1+0),$$

$$P\{x_1 < \xi < x_2\} = F(x_2) - F(x_1+0).$$

6.  $F(-\infty) = \lim_{n \rightarrow \infty} F(-n) = 0$ ,  $F(+\infty) = \lim_{n \rightarrow \infty} F(+n) = 1$ .

Существование пределов следует из монотонности и ограниченности. Равенства из соотношения  $1 = P\{-\infty < \xi < \infty\} = F(+\infty) - F(-\infty)$ .

Отметим еще, что функция распределения имеет не более, чем счетное число скачков. (Число  $N(\frac{1}{k})$  скачков размером не более  $\frac{1}{k}$  ограничено величиной  $k$ ).

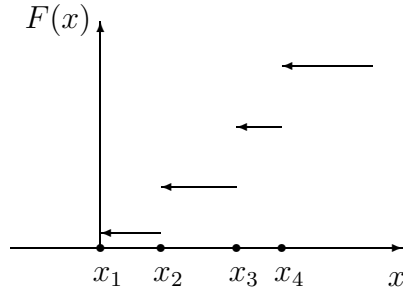
Функция, удовлетворяющая условиям 1 – 6 является функцией распределения (некоторой случайной величины). Однако, из равенства  $F_\xi(x) = F_\eta(x)$  не следует совпадения  $\xi$  и  $\eta$ ; они могут различаться с вероятностью единица.

Пример:  $\eta = -\xi$  для симметричного распределения.

### I. Дискретные случайные величины.

Множество значений конечно или счетно.  $P\{\xi = x_k\} = p_k, \sum p_k = 1$ . Для любого подмножества  $A$  значений  $\{x_k\}$  имеем  $p\{\xi \in A\} = \sum_{k: x_k \in A} p_k$ . Корректность следует из Леммы о суммировании по блокам.

Функция распределения  $F(x) = \sum_{k: x < x_k} p_k$  полностью определяет распределение: значения  $x_k$  — точки разрыва, вероятности  $p_k$  — величины скачков.



Функция распределения

**II. Случайная величина  $\xi$  называется непрерывной** (абсолютно непрерывной), если  $F_\xi(x) = \int_{-\infty}^x p_\xi(x) dx$ , где  $p_\xi(x)$  — неотрицательная кусочно-непрерывная функция, которая называется плотностью вероятности (плотностью распределения вероятности) случайной величины  $\xi$ ,  $\int_{-\infty}^{\infty} p_\xi(x) dx = 1$ .

В точках непрерывности  $p_\xi(x) = \frac{dF_\xi(x)}{dx}$ .

$$P\{x_1 \leq \xi < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p_\xi(x) dx.$$

*Замечания.*

1. Интеграл понимается в смысле Лебега (в простейших случаях он совпадает с Римановским).

2. Для любого борелевского (измеримого) множества  $A$   $P(A) = \int_A dF(x)$

3. Если  $p(x)$  непрерывна на  $[x, x + \Delta x]$ , то по теореме о среднем  $P\{x \leq \xi < x + \Delta x\} = p(x)\Delta x + o(\Delta x)$ .

4.  $P\{\xi = x\} = F(x+0) - F(x) = 0$  для непрерывных случайных величин, поэтому для них неравенства  $\leq, \geq$  можно заменить на  $<, >$  и наоборот.

**III. Сингулярные случайные величины.** Для них множество точек роста функции распределения имеет (Лебегову) меру нуль. Пример — Канторовская лестница<sup>13</sup>. Общая длина отрезков стационарности стремится к  $\sum_{n=1}^{\infty} \frac{2^{n-1}}{3^n} = 1$ .

Показано, что случайные величины исчерпываются этими тремя типами, но существуют смешанные типы:  $F(x) = \alpha F_{\xi_1}(x) + (1 - \alpha) F_{\xi_2}(x)$ ,  $0 < \alpha < 1$ .

**Многомерные случайные величины.** Пусть задано  $(\Omega, \mathcal{F}, P(\cdot))$ . Векторной случайной величиной называется векторная функция  $\xi(\omega) = (\xi_1(\omega), \xi_2(\omega), \dots, \xi_n(\omega))$ , такая, что для любых  $x_1, x_2, \dots, x_n$ :  $\{\omega : \xi_1(\omega) < x_1, \xi_2(\omega) < x_2, \dots, \xi_n(\omega) < x_n\} \in$

<sup>13</sup>См. Колмогоров А.Н. и Фомин С.В. Элементы теории функций и функционального анализа. М., Наука, 1976, стр. 341.

$\mathcal{F}$ .  $P\{\xi_1 < x_1, \xi_2 < x_2, \dots, \xi_n < x_n\} = F(x_1, x_2, \dots, x_n)$  —  $n$ -мерная функция распределения.

Далее рассмотрим случай  $n = 2$ .

$$F(x, y) = P\{\xi < x, \eta < y\}.$$

Свойства.

1.  $F(x, y)$  не убывает по  $x$  и по  $y$ .
2.  $F(x, y)$  непрерывна слева по каждому аргументу.
3.  $F(+\infty, +\infty) = 1$ ,  $F(-\infty, y) = F(x, -\infty) = 0$ .
4.  $P\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\} = F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1)$ .
5.  $F_\xi(x) = F(x, +\infty)$ ,  $F_\eta(y) = F(+\infty, y)$  — *маргинальные распределения*.

$$\begin{aligned} \{\xi < x\} &= \sum_{k=-\infty}^{\infty} \{\xi < x, k \leq \eta < k+1\} \Rightarrow \\ P\{\xi < x\} &= \sum_{k=-\infty}^{\infty} [F(x, k+1) - F(x, k)] = F(x, \infty). \end{aligned}$$

Для абсолютно непрерывных случайных величин  $F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(x, y) dx dy$  и далее:

1.  $p(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$  в точках непрерывности.
2.  $P(\xi \in D) = \int_D p(x) dx$  ( $\xi, x \in \mathcal{R}_n$ ,  $D \subset \mathcal{R}_n$ ).
3.  $P\{x \leq \xi < x + \Delta x, y \leq \eta < y + \Delta y\} = p(x, y) \Delta x \Delta y + o(\Delta x \Delta y)$  в окрестности точки непрерывности.
4.  $p_\xi(x) = \int_{-\infty}^{\infty} p(x, y) dy$ ,  $p_\eta(y) = \int_{-\infty}^{\infty} p(x, y) dx$ .

**Условные распределения.**

Пусть  $B$  — событие,  $P\{B\} \neq 0$ . Тогда  $P\{\xi < x | B\} = F_\xi(x | B) = \int_{-\infty}^x p_\xi(x | B) dx$ .

Пусть задана  $p(x, y)$  — плотность распределения случайной величины  $(\xi, \eta)$ . Рассмотрим  $F(\xi < x | B)$ , где  $B = \{y \leq \eta < y + \Delta y\}$ :

$$\begin{aligned} F(\xi < x, B) &= \frac{P\{\xi < x, y \leq \eta < y + \Delta y\}}{P\{y \leq \eta < y + \Delta y\}} = \frac{\int_{-\infty}^x dx \int_y^{y+\Delta y} p(x, y) dy}{\int_{-\infty}^{\infty} dx \int_y^{y+\Delta y} p(x, y) dy} = \\ &= \frac{\int_{-\infty}^x p(x, y) dx \Delta y + o(\Delta y)}{p_\eta(y) \Delta y + o(\Delta y)}. \end{aligned}$$

Разделив на  $\Delta y$  и переходя к пределу при  $\Delta y \rightarrow 0$ , получаем

$$F_{\xi|\eta}(x|y) = \frac{\int_{-\infty}^x p(x, y) dx}{p_\eta(y)} \quad \text{и} \quad p_{\xi|\eta}(x|y) = \frac{p(x, y)}{p_\eta(y)}.$$

Наконец, на основе этой формулы можно записать

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p_{\xi|\eta}(x|y)p_{\eta}(y)dy.$$

(аналог формулы полной вероятности). Аналогичные формулы можно написать и для дискретного случая.

**Назависимость случайных величин.**

Лекция 6.

Случайные величины  $\xi_1, \dots, \xi_n$  называются независимыми в совокупности (попарно), если  $\forall x_1, \dots, x_n$  события  $\{\xi_1 < x_1\}, \dots, \{\xi_n < x_n\}$  независимы в совокупности (попарно). Пусть теперь  $n = 2$ .

$$F(x, y) = P\{\xi < x, \eta < y\} = P\{\xi < x\} P\{\eta < y\} = F_{\xi}(x) F_{\eta}(y). \quad (5)$$

$$\begin{aligned} P\{x_1 \leq \xi < x_2, y_1 \leq \eta < y_2\} &= \\ &= F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = \\ &= F_{\xi}(x_2)F_{\eta}(y_2) - F_{\xi}(x_1)F_{\eta}(y_2) - F_{\xi}(x_2)F_{\eta}(y_1) + F_{\xi}(x_1)F_{\eta}(y_1) = \\ &= (F_{\xi}(x_2) - F_{\xi}(x_1)) (F_{\eta}(y_2) - F_{\eta}(y_1)) = \\ &= P\{x_1 \leq \xi < x_2\} P\{y_1 \leq \eta < y_2\}. \end{aligned} \quad (6)$$

Формулы (5) и (6) эквивалентны ( $x_1, y_1 \rightarrow \infty$ ). Для дискретных случайных величин  $P(\xi = x_k, \eta = y_j) = P(\xi = x_k)P(\eta = y_j)$ . Впрочем индексы можно опустить.

$p_{\xi\eta}(x, y) = p_{\xi}(x)p_{\eta}(y)$  в точках непрерывности этих функций.

Вернемся к (5). Верно ли обратное?

Теорема. Пусть  $F_{\xi\eta}(x, y) = F_1(x) \cdot F_2(y)$ , причем  $F_1(= \infty) = 1$ . Тогда  $F_1(x) = F_{\xi}(x)$ ,  $F_2(y) = F_{\eta}(y)$ .

Доказательство следует из свойств двумерной функции распределения.

Случайные величины  $\xi, \eta$  называются эквивалентными, если  $P(\xi \neq \eta) = 0$  (совпадают почти всюду).

**Функции от случайных величин.**

Пусть  $\eta = f(\xi)$ , где  $f(\cdot)$  — измеримая функция,  $\mathcal{R}_n \mapsto \mathcal{R}_m$ . Обычно требуется найти распределение  $\eta$ , если известно распределение  $\xi$ .

Пример. Пусть  $y = f(x)$  — невырожденное преобразование (биекция). Пусть дана плотность  $p_{\xi}(x)$ . Найти плотность  $p_{\eta}(y)$ , если  $\eta = f(\xi)$ .

Для произвольной области имеем:

$$\begin{aligned} \int_D p_{\eta}(y)dy &= P(\eta \in D) = P(f(\xi) \in D) = P(\xi \in f^{-1}(D)) = \\ &= \int_{f^{-1}(D)} p_{\xi}(x)dx = \int_D p_{\xi}(f^{-1}(y)) \left| \frac{\partial(x)}{\partial(y)} \right| dy. \end{aligned}$$

Отсюда в силу произвольности  $D$  получаем, что

$$p_{\eta}(y) = p_{\xi}(f^{-1}(y)) \left| \frac{\partial(x)}{\partial(y)} \right| dy.$$

Пример. Задана плотность  $p_{\xi}(x_1, x_2)$ . Найти распределение  $\xi_1 + \xi_2$ .

Пусть

$$\begin{aligned} \eta_1 &= \xi_1 + \xi_2, & y_1 &= x_1 + x_2, & x_1 &= y_2 - y_1, & \left| \frac{\partial(x)}{\partial(y)} \right| &= 1. \\ \eta_2 &= \xi_2, & y_2 &= x_2, & x_2 &= y_2, \end{aligned}$$

Тогда  $p_\eta(y) = p_\xi(y_1 - y_2, y_2)$  и

$$p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_\xi(y_1 - y_2, y_2) dy_2.$$

Если  $\xi$  и  $\eta$  независимы, то  $p_{\eta_1}(y_1) = \int_{-\infty}^{\infty} p_{\xi_1}(y_1 - y_2) p_{\xi_2}(y_2) dy_2$  (свертка!).

**Независимость функций от независимых случайных величин.** Рассмотрим дискретный случай. Пусть  $\xi$  и  $\eta$  — независимые случайные величины, тогда  $f_1(\xi)$  и  $f_2(\eta)$  (где  $f_1$  и  $f_2$  — измеримые функции) также независимые случайные величины. Пусть  $\xi$  принимает значения  $x_i$ , а  $\eta$  — значения  $y_j$ . Для любых  $s$  и  $t$  имеем:

$$\begin{aligned} P\{f_1(\xi) = s, f_2(\eta) = t\} &= \sum_{\substack{i: f_1(x_i)=s \\ j: f_2(y_j)=t}} P\{\xi = x_i, \eta = y_j\} = \\ &= \sum_{i: f_1(x_i)=s} P\{\xi = x_i\} \sum_{j: f_2(y_j)=t} P\{\eta = y_j\} = P\{f_1(\xi) = s\} \cdot P\{f_2(\eta) = t\}. \end{aligned}$$

**Числовые характеристики случайных величин. Моменты.**

Начальный момент порядка  $k$ :

$$M\xi^k = \int_{-\infty}^{\infty} x^k p_\xi(x) dx, \text{ если } \int_{-\infty}^{\infty} |x|^k p_\xi(x) dx < +\infty \text{ или}$$

$$M\xi^k = \sum_{i=1}^{\infty} x_i^k p_i, \text{ если } \sum_{i=1}^{\infty} |x_i|^k p_i < +\infty.$$

Математическое ожидание или среднее значение:  $M\xi = \int_{-\infty}^{\infty} x dF_\xi(x)$ .

Теорема. Пусть случайная величина  $\eta = f(\xi)$ . Тогда<sup>14</sup>  $M\eta = \int_{-\infty}^{\infty} f(x) dF_\xi(x)$ .

Доказательство (для дискретных случайных величин)

$$M\eta = \sum_{k=1}^{\infty} y_k P(\eta = y_k) = \sum_k y_k \sum_{i: f(x_i)=y_k} p_i = \sum_i f(x_i) p_i.$$

Следствие:  $M(\sum C_k \xi_k) = \sum C_k M\xi_k$ .

Центральный момент  $k$ -го порядка  $M(\xi - M\xi)^k = \int_{-\infty}^{\infty} (x - M\xi)^k dF_\xi(x)$ .

Центральный момент второго порядка  $M(\xi - M\xi)^2 = \int_{-\infty}^{\infty} (x - M\xi)^2 dF_\xi(x) = D\xi$

носит название дисперсии,  $\sqrt{D\xi}$  — стандартное отклонение.

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

**Свойства математического ожидания и дисперсии.**

1.  $M(\xi + \eta) = M\xi + M\eta =$

$$= \int_{-\infty}^{\infty} z p_{\xi+\eta}(z) dz = \int_{-\infty}^{\infty} (x + y) \int_{-\infty}^{\infty} p(x, y) dx dy = \iint_{-\infty}^{\infty} xp(x, y) dx dy +$$

<sup>14</sup>если момент существует

$$+ \iint_{-\infty}^{\infty} yp(x, y) dx dy = \int_{-\infty}^{\infty} xp_{\xi}(x) dx + \int_{-\infty}^{\infty} yp_{\eta}(y) dy.$$

(Аналогично для дискретных случайных величин.)

2.  $M(C_1\xi + C_2\eta) = C_1M\xi + C_2M\eta$  и  $MA\xi = AM\xi$  ( $A$  — матрица,  $\xi$  — случайный вектор).

3. Если  $\xi$  и  $\eta$  независимы, то  $M(\xi\eta) = M\xi M\eta$ . (При  $M\xi < \infty$ ,  $M\eta < \infty$ . Заметим, что  $M(\xi\eta)$  в этом случае существует.)

Обратное неверно! Пусть, например,  $\xi$  и  $\eta$  — независимые случайные величины и  $M\eta = M\xi = 0$ . Тогда  $M(\xi\eta)\xi = M\xi^2\eta = 0$ , но отсюда не следует, что  $\xi\eta$  и  $\eta$  независимы. В самом деле,

$\eta \backslash \xi$	-1	2
-1	4/9	2/9
2	2/9	1/9

но  $P\{\xi\eta = 1, \xi = -1\} = \frac{4}{9} \neq P\{\xi\eta = 1\}P\{\xi = -1\} = \frac{8}{27}$ .

### Примеры моментов

Пусть  $\xi_i$  — число успехов в  $i$ -м испытании Бернулли.  $M\xi_i = 1 \cdot p + 0 \cdot q = p$ ,  $M\xi_i^2 = 1^2 \cdot p + 0^2 \cdot q = p$ ,  $D\xi_i = p - p^2 - pq$ .

Распределение Пуассона

$M\xi = \sum_{k=0}^{\infty} k \frac{\lambda^k}{n!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{n!} e^{-\lambda} = \lambda e^{-\lambda} \frac{d}{d\lambda} e^{\lambda} = \lambda$ . Аналогичным приемом

получаем  $M\xi^2 = \lambda + \lambda^2$  и  $D\xi = \lambda$ .

Нормальное распределение  $\mathcal{N}(\mu, \sigma^2)$ :

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

$$\begin{aligned} M\xi &= \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} dx = \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (y + \mu) \exp\left\{-\frac{1}{2\sigma^2}y^2\right\} dy = \mu. \end{aligned}$$

Аналогично

$$D\xi = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\} dx = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2}} dt = \sigma^2.$$

### 4. Неравенства.

а/  $\xi \geq \eta \Rightarrow M\xi \geq M\eta$  (Следствие 2.)

б/ Неравенство Коши-Буняковского

$$M(\xi - \lambda\eta)^2 = M\xi^2 - 2\lambda M\xi\eta + \lambda^2 M\eta^2 \geq 0 \Rightarrow (M\xi\eta)^2 \leq M\xi^2 M\eta^2.$$

Если равенство, то  $\exists \lambda$ ,  $M(\xi - \lambda\eta)^2 = 0$ .

Лекция 7.

в/ Неравенство Маркова.

$$P\{|\xi| > \varepsilon\} \leq \frac{M|\xi|^k}{\varepsilon^k}, \quad k = 1, 2, \dots$$

Пусть

$$\eta = \begin{cases} 0, & |\xi| \leq \varepsilon, \\ \varepsilon, & |\xi| > \varepsilon. \end{cases}$$

Тогда  $|\eta|^k \leq |\xi|^k \Rightarrow \varepsilon^k P\{|\xi| > \varepsilon\} \leq M|\xi|^k$ .

При  $k = 2$  — неравенство Чебышёва<sup>15</sup>.

5.  $DC = 0$ . Обратно: если  $D\xi = 0$ , то  $P\{\xi = \text{const}\} = 1$ . Доказательство.  $\forall \varepsilon > 0 \quad P\{|\xi| > \varepsilon\} = 0$ . Рассмотрим последовательность  $A_k = \{|\xi - M\xi| > \frac{1}{k}\}$ .  $A_k \uparrow A = \{|\xi - M\xi| > 0\} \Rightarrow P\{|\xi - M\xi| > 0\} = \lim_{k \rightarrow \infty} P\{|\xi - M\xi| > \frac{1}{k}\} = 0$ .

Возвращаясь к неравенству Коши: если равенство, то  $\exists \lambda$ , такое, что  $\xi - \lambda\eta = \text{const}$  с вероятностью 1 (почти наверное).

6.  $DC\xi = C^2 D\xi$ .

$$\begin{aligned} 7. \quad D \sum_{i=1}^n \xi_i &= M \left[ \sum_{i=1}^n (\xi_i - M\xi_i) \right]^2 = M \sum_{i=1}^n (\xi_i - M\xi_i)^2 + \sum_{i,j=1, i \neq j}^n M(\xi_i - M\xi_i)(\xi_j - M\xi_j) \\ &= \sum_{i=1}^n D\xi_i + \sum_{i,j=1, i \neq j}^n \text{cov} \xi_i \xi_j \end{aligned}$$

( $= \sum_{i=1}^n D\xi_i$  если  $\xi_i, \xi_j$  попарно независимы — формула Бьенемэ.)

Моменты векторных случайных величин.

$$M\xi = (M\xi_1, M\xi_2, \dots, M\xi_n), \quad MA\xi = AM\xi,$$

$$D\xi = (D\xi_1, D\xi_2, \dots, D\xi_n)$$

$\|\text{cov} \xi_i \xi_j\|$  — ковариационная матрица. Введем  $r_{ij} = \frac{\text{cov} \xi_i \xi_j}{\sqrt{D\xi_i D\xi_j}}$ ,  $|r_{ij}| \leq 1$ .

Каков смысл  $r_{ij}$ ? Пусть  $|r_{ij}| = 1$  и  $\varepsilon = \pm 1$ , тогда  $M \left( \frac{\xi_i}{\sqrt{D\xi_i}} + \varepsilon \frac{\xi_j}{\sqrt{D\xi_j}} \right)^2 = 2(1 + \varepsilon r_{ij}) = 0$  и величины  $\xi_i$  и  $\xi_j$  линейно зависимы с вероятностью 1.

### Условное математическое ожидание

$$M(\xi|\eta = y) = \int_{-\infty}^{\infty} x p_{\xi|\eta}(x|y) dx.$$

Это случайная величина, функция от  $\eta$ . Она обладает всеми свойствами математического ожидания с вероятностью 1.

Если учесть, что  $p_{\xi|\eta}(x|y) = \frac{p_{\xi\eta}(x,y)}{p_{\eta}(y)}$ , то

$$\begin{aligned} M[M(\xi|\eta = y)] &= \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{\infty} x p_{\xi|\eta}(x|y) dx \right] p_{\eta}(y) dy = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x p_{\xi\eta}(x, y) dx dy = M\xi. \end{aligned}$$

<sup>15</sup>Пафнутий Львович Чебышёв 1821–1894 — русский математик, создатель Петербургской научной школы.

**Задача о наилучшем среднеквадратичном приближении.**

Пусть  $\xi$  и  $\eta$  — случайные величины, наблюдаем  $\xi$ , требуется оценить  $\eta$ , т.е. найти такую функцию  $f(\cdot)$ , что  $M(\eta - f(\xi))^2 \sim \min_{f(\cdot)}$

**Теорема.**  $f(\xi) = M(\eta|\xi)$  (п.н.).

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

$$\begin{aligned} &= M[M((\eta - M(\eta|\xi) + M(\eta|\xi) - f(\xi))^2|\xi)] = \\ &= M\{M((\eta - M(\eta|\xi))^2|\xi) + 2M((\eta - M(\eta|\xi))(M(\eta|\xi) - f(\xi))|\xi) + \\ &+ (M(\eta|\xi) - f(\xi))^2|\xi)\} = M(\eta - M(\eta|\xi))^2 + M(M(\eta|\xi) - f(\xi))^2 \geq \\ &\geq M(\eta - M(\eta|\xi))^2, \text{ т.к. } M((\eta - M(\eta|\xi))(M(\eta|\xi) - f(\xi))|\xi) = 0, \end{aligned}$$

причем равенство выполняется, только если  $f(\xi) = M(\eta|\xi)$  (п.н.).  $\square$

Другие задачи наилучшего приближения.

1) Приближение постоянной.  $M(\xi - C)^2 \sim \min_C \Rightarrow M\xi^2 - 2CM\xi + C^2 \sim \min_C \Rightarrow C = M\xi$  и  $\min M(\xi - C)^2 = D\xi$ .

2) Приближение  $\eta$  по наблюдениям  $\xi$  в классе линейных функций.

$$M(\eta - a\xi - b)^2 = M(\eta - a\xi)^2 - 2bM(\eta - a\xi) + b^2 \sim \min_{a,b}. \quad (*)$$

Дифференцируя (\*) по  $(b)$ , находим  $b = M\eta - aM\xi$ , после чего (\*) принимает вид  $M(\eta - M\eta - a(\xi - M\xi))^2 = D\eta - 2a\text{cov}\xi\eta + a^2D\xi$ , дифференцируя по  $(a)$ , получаем  $-2\text{cov}\xi\eta + 2aD\xi = 0 \Rightarrow a = \frac{\text{cov}\xi\eta}{D\xi}$  и  $a\xi + b = a(\xi - M\xi) + M\eta = \frac{\text{cov}\xi\eta}{\text{cov}\xi\xi}(\xi - M\xi) + M\eta$ .

Среднеквадратичная ошибка при этом равна  $M(\eta - a\xi - b)^2 =$

$$= M(\eta - M\eta - a(\xi - M\xi))^2 = D\eta - 2a\text{cov}\xi\eta + a^2D\xi = D\eta - \frac{(\text{cov}\xi\eta)^2}{D\xi}. \square$$

**Задача наилучшего линейного приближения в векторном случае.**

Пусть  $\xi$  — измеренный вектор,  $M\xi = 0$ ,  $M\eta = 0$ . Нужно найти матрицу  $A$ , такую, что  $M\|\eta - A\xi\|^2 \sim \min_A$ .

$$\begin{aligned} Y(A) &= M\|\eta - A\xi\|^2 = M \sum_i (\eta - A\xi)_i^2 = \\ &= M \text{tr}(\eta - A\xi)(\eta - A\xi)^* = \text{tr} \Sigma_{\eta\eta} - \text{tr} A \Sigma_{\xi\eta} - \text{tr} \Sigma_{\eta\xi} A^* + \text{tr} A \Sigma_{\xi\xi} A^*. \end{aligned}$$

Здесь  $\Sigma_{\eta\eta} = M\eta\eta^*$ ,  $\Sigma_{\eta\xi} = M\eta\xi^*$ ,  $\Sigma_{\xi\xi} = M\xi\xi^*$ , а знак  $(*)$  означает сопряжение (транспонирование).

Варьируя  $Y$  по  $A$ , получим  $\delta Y = Y(A + \delta A) - Y(A) =$

$$= -\text{tr} \delta A \Sigma_{\xi\eta} - \text{tr} \Sigma_{\eta\xi} \delta A^* + \text{tr} \delta A \Sigma_{\xi\xi} A^* + \text{tr} A \Sigma_{\xi\xi} \delta A^* = 0.$$

Отсюда, учитывая, что  $\text{tr} Q^* = \text{tr} Q$ , имеем  $\text{tr}(\Sigma_{\eta\xi} - A \Sigma_{\xi\xi}) \delta A = 0$ , откуда в силу произвольности  $\delta A$  следует  $\Sigma_{\eta\xi} = A \Sigma_{\xi\xi}$ , таким образом,  $A = \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1}$  и

$$\hat{\eta} = \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \xi.$$

При этом с.к. ошибка (погрешность) линейного приближения равна

$$M\|\eta - \hat{\eta}\|^2 = \text{tr}(\Sigma_{\eta\eta} - \Sigma_{\eta\xi} \Sigma_{\xi\xi}^{-1} \Sigma_{\xi\eta})$$

и  $M\|\eta - M\eta\|^2 = \text{tr } \Sigma_{\eta\eta}$  — априорная погрешность.  $\square$

**Последовательности случайных величин. Сходимость. Виды сходимости.** Лекция 8.

**Закон больших чисел в форме Чебышёва.**

Пусть  $\xi_i$  попарно независимы, причем  $D\xi_n < C$  (ограничены в совокупности).

Тогда

$$\eta_n = \frac{1}{n} \sum_{i=1}^n (\xi_i - M\xi_i) \xrightarrow[n \rightarrow \infty]{P} 0.$$

Доказательство.  $D\eta_n = \frac{\sum D\xi_i}{n^2} < \frac{C}{n}$ .

Из неравенства Чебышёва следует

$$P\{|\eta_n| > \varepsilon\} \leq \frac{D\eta_n}{\varepsilon^2} < \frac{C}{n\varepsilon^2} \xrightarrow[n \rightarrow \infty]{} 0. \square$$

**Следствия.**

1. Пусть  $M\xi_k = \mu$ . Тогда  $\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{P} \mu$ .

2. Пусть  $\xi_i$  — число успехов при одном испытании в схеме Бернулли. Тогда  $M\xi_i = p$ ,  $D\xi_i = pq$  и

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{P} p.$$

3. Теорема Маркова. Снимем условие независимости, но потребуем, чтобы  $D\frac{1}{n} \sum_{i=1}^n \xi_i \rightarrow 0$ . Тогда из неравенства Чебышёва получаем непосредственно

$$\frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow[n \rightarrow \infty]{P} \frac{1}{n} \sum_{i=1}^n M\xi_i = \mu.$$

**Характеристические функции.**

Определение. Характеристической функцией называется

$$f_\xi(t) = Me^{i\xi t},$$

причем этот момент существует всегда, так как  $|e^{i\xi t}| = 1$ .

**Примеры.**

Распределение Пуассона:

$$f_\xi(t) = Me^{i\xi t} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{e^{ikt} \lambda^k}{k!} = e^{\lambda(e^{it}-1)}.$$

Нормальное распределение:

$$f_\xi(t) = Me^{i\xi t} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixt} e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}.$$

Из очевидных свойств отметим следующие:  $f(0) = 1$ ,  $|f(t)| \leq 1$ , четность и вещественность  $f(t)$  эквивалентны и в этом случае  $f(t) = M \cos(i\xi t)$ . Следующие свойства сформулированы в виде теорем.

**Теорема 1.**  $f_{\sigma\xi+\mu}(t) = e^{i\mu t} M e^{i\xi\sigma t}$ .

**Теорема 2.** Пусть  $\xi_i$ ,  $i = 1, 2, \dots$  независимы в совокупности. Тогда  $f_{\xi_1+\xi_2+\dots+\xi_n}(t) = f_{\xi_1}(t)f_{\xi_2}(t)\dots f_{\xi_n}(t)$ .

**Теорема 3.** Пусть существует момент  $k$ -го порядка. Тогда существует  $k$ -я производная характеристической функции  $f_{\xi}^{(k)}(t)$ , она равномерно непрерывна на  $-\infty, \infty$  и имеет место равенство

$$f_{\xi}^{(k)}(0) = i^k M \xi^k. \quad (7)$$

Первое и третье свойства очевидны, второе доказывается цепочкой равенств (неравенств):

$$|f_{\xi}^{(k)}(t+h) - f_{\xi}^{(k)}(t)| \leq \int_{-A}^A |e^{ixh} - 1| p(x) dx + 2 \int_{|x|>A} p(x) dx. \quad (8)$$

Выберем  $A > 0$  так, чтобы второе слагаемое было менее  $\varepsilon/2$  (следствие сходимости). Заметим, что

$$|e^{ixh} - 1| = \left| \int_0^{xh} e^{i\alpha} d\alpha \right| \leq |xh| \leq A|h|.$$

Выбрав теперь  $|h| \leq \frac{\varepsilon}{2A}$ , мы получаем  $|f_{\xi}^{(k)}(t+h) - f_{\xi}^{(k)}(t)| \leq \varepsilon$  независимо от  $t$ .

**Теорема 4.** *Без доказательства.*

Если х.ф.  $f_{\xi}(t)$  абсолютно интегрируема на  $R_1$ , то случайная величина  $\xi$  непрерывна, а ее плотность вероятности равна

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} f_{\xi}(t) dt \quad (9)$$

и равномерно непрерывна на  $R_1$ .

(По существу — это теорема об обратном Фурье-преобразовании.)

**Теорема 5. Теорема непрерывности для характеристических функций.** *Без доказательства*

Пусть последовательность  $\{f_{\xi_n}(t)\}$  х.ф. случайных величин сходится при  $n \rightarrow \infty$  к х.ф.  $f_{\xi}(t)$  случайной величины  $\xi$  равномерно по  $t$  в каждом конечном интервале  $|t| \leq T$ . Тогда при  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} F_{\xi_n}(x) = F_{\xi}(x) \quad (10)$$

в точках непрерывности  $F_{\xi}(x)$ , а если последняя непрерывна, то равномерно по  $x \in R_1$ .

**Следствие.** Если у двух случайных величин совпадают х.ф., то совпадают и законы распределения (функции распределения).

**Центральная предельная теорема (ЦПТ).**

Пусть  $\{\xi_n\}$  — последовательность одинаково распределенных, независимых в совокупности случайных величин,  $M\xi_n = \mu$ ,  $D\xi_n = \sigma^2$ . Тогда

$$\eta_n = \frac{\sum_{i=1}^n (\xi_i - \mu)}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow \infty]{d} \eta \sim \mathcal{N}(0, 1), \quad (11)$$

т.е.,  $P\{\eta_n < x\} \approx \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp(-\frac{s^2}{2}) ds$ .

**Доказательство.**

Пусть  $\varphi(t)$  — х.ф. случайной величины  $\xi_i - \mu$ . Тогда

$$\begin{aligned} f_{\eta_n}(t) &= \left[ \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n = \left[ \varphi(0) + \frac{t}{\sigma\sqrt{n}}\varphi'(0) + \frac{t^2}{2\sigma^2 n}\varphi''(0) + o\left(\frac{1}{n}\right) \right]^n = \\ &= \left[ 1 + 0 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right) \right]^n \xrightarrow[n \rightarrow \infty]{} e^{-\frac{t^2}{2}} \end{aligned}$$

— х.ф. нормального распределения. Далее воспользуемся теоремой непрерывности для характеристических функций.  $\square$

**Следствие. Теорема Муавра-Лапласа.** Рассмотрим схему Бернулли, где  $m$  — число успехов и  $n \rightarrow \infty$  при фиксированных  $0 < p < 1$  и  $0 < q < 1$ . Тогда справедливо

$$P\left\{a \leq \frac{m - np}{\sqrt{npq}} < b\right\} \approx \Phi(b) - \Phi(a).$$

**Центральные предельные теоремы для неодинаково распределенных случайных величин.** Обзор, без доказательств.

Пусть случайные величины  $\{\xi_k\}$  независимы, существуют моменты 1 и 2 порядка  $M\xi_k = \mu_k$  и  $D\xi_k = \sigma_k^2$

**Теорема А.М.Ляпунова.** Пусть для некоторого  $\delta > 0$  существуют  $c_k^{2+\delta} = M|\xi_k - \mu_k|^{2+\delta}$ . Обозначим  $C_n^{2+\delta} = \sum_{k=1}^n c_k^{2+\delta}$ ,  $B_n^2 = \sum_{k=1}^n \sigma_k^2$ . Если  $\frac{C_n^{2+\delta}}{B_n^{2+\delta}} \xrightarrow[n \rightarrow \infty]{} 0$  (условие Ляпунова), то  $P\{\eta_n < x\} \xrightarrow[n \rightarrow \infty]{} \Phi(x)$ .

*Замечание 1.* Условие Ляпунова не является необходимым, однако его практически легче проверять и доказывать теорему, особенно при<sup>16</sup>  $\delta = 1$ .

*Замечание 2.* Еще более слабым является условие Линдеберга, которым можно заменить условие Ляпунова: для всякого  $\tau > 0$

$$\frac{1}{B_n^2} \sum_{k=1}^n \int_{|x - \mu_k| \geq \tau B_n} (x - \mu_k)^2 dF_k(x) \xrightarrow[n \rightarrow \infty]{} 0,$$

но и оно не является необходимым, т.е. существуют распределения, которые ему не удовлетворяют, но сходимость к нормальному закону имеет место. Однако,

<sup>16</sup>См. доказательство в учебнике Пытьева, Шишмарева, стр. 128.

если добавить требование пренебрежимой (асимптотической) малости величин

$$\zeta_k = \frac{\xi_k - M\xi_k}{B_n},$$

$$\max_{1 \leq k \leq n} P\{|\zeta_k| > \varepsilon\} \xrightarrow{n \rightarrow \infty} 0,$$

то условие Линдеберга становится необходимым.

### Перечислим виды сходимости случайных величин.

1. Среднеквадратичная  $\lim_{n \rightarrow \infty} M|\xi_n - \xi|^2 = 0$  или  $\xi_n \xrightarrow[n \rightarrow \infty]{\text{с.к.}} \xi$  или  $\text{l.i.m.}_{n \rightarrow \infty} \xi_n = \xi$ .
2. С вероятностью 1, почти наверное, по модулю  $P \pmod{P}$ <sup>17</sup>:  $P\{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega)\} = 1$  или  $\xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi$ .
3. По вероятности  $\forall \varepsilon > 0: \lim_{n \rightarrow \infty} P\{|\xi - \xi_n| > \varepsilon\} = 0$  или  $\xi_n \xrightarrow[n \rightarrow \infty]{P} \xi$ .
4. По распределению («слабая»)  $F_{\xi_n}(x) \rightarrow F_\xi(x)$  или  $\xi_n \xrightarrow[n \rightarrow \infty]{d} \xi$ .

### Отношения между видами сходимости.

$$\begin{array}{ccc} \xi_n \xrightarrow[n \rightarrow \infty]{\text{с.к.}} \xi & \Downarrow & \xi_n \xrightarrow[n \rightarrow \infty]{P} \xi \implies \xi_n \xrightarrow[n \rightarrow \infty]{d} \xi. \\ & \nearrow & \\ \xi_n \xrightarrow[n \rightarrow \infty]{\text{п.н.}} \xi & & \end{array}$$

Других отношений без дополнительных условий не существует.

Напомним, что речь всюду идет о случайных величинах  $\xi_n(\omega)$ , зависящих от элементарного исхода  $\omega \in \Omega$ . При каждом элементарном исходе мы получаем числовую последовательность, называемую выборочной.

Если последовательность  $\{\xi_n\}$  сходится в с.к., по вероятности или по распределению, то может так случиться, что соответствующие выборочные последовательности не будут сходиться ни при каком  $\omega \in \Omega$ .

Однако, если  $\xi_n \xrightarrow{P} \xi$ , то существует подпоследовательность  $\{\xi_n\}$  последовательности  $\{\xi_n\}$ , сходящаяся к  $\xi$  с вероятностью единица.

Связь между сходимостью по вероятности и с вероятностью единица. Покажем, что из сходимости с вероятностью единица следует сходимост по вероятности. Действительно, пусть  $A_k = \{|\xi_k - \xi| > \varepsilon\}$ ,  $\varepsilon > 0$ . Сходимость с вероятностью единица эквивалентна равенству  $P(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k) = 0$ , т.е.  $\lim_{n \rightarrow \infty} P(\bigcup_{k \geq n} A_k) = 0$ . Но  $P(\bigcup_{k \geq n} A_k) \geq P(A_n)$ , следовательно,  $\lim_{n \rightarrow \infty} P(A_n) = 0$ .  $\square$

<sup>17</sup>Заметим, что множество  $A$  тех  $\omega \in \Omega$ , для которых существует  $\lim_{n \rightarrow \infty} \xi_n$  действительно является событием, так как его можно представить в виде  $A = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m_1, m_2 > n} \{\omega : |\xi_{m_1} - \xi_{m_2}| < \frac{1}{k}\}$ .

$A$  — множество  $\omega$ , удовлетворяющее условиям: для  $\forall k = 1, 2, \dots \exists n \geq 1$ , такое, что  $\forall m_1, m_2 > n: |\xi_{m_1}(\omega) - \xi_{m_2}(\omega)| < \frac{1}{k}$ .

В каждой точке  $\omega \in A$  последовательность  $\{\xi_n\}$  фундаментальна и, следовательно, сходится. Измеримость  $A$  следует из его представления, а в случае сходимости  $\{\xi_n\}$  почти наверное  $P(A) = 1$ .

Однако, если  $\xi = \text{const}$  с вероятностью единица, то  $\xi_n \xrightarrow{P} \xi \Leftrightarrow \xi_n \xrightarrow{d} \xi$  и, в частности,  $\xi_n - \xi \xrightarrow{P} 0 \Leftrightarrow \xi_n - \xi \xrightarrow{d} 0$

Поэтому  $\xi_n \xrightarrow{d} \xi$ ,  $\xi_n - \xi \xrightarrow{d} 0$  — разные сходимости.

Если последовательность  $\{\xi_n\}$  с.к. сходится к  $\xi$  «достаточно быстро», так что  $\sum_{k=1}^{\infty} M|\xi_k - \xi|^2 < \infty$ , то она сходится к  $\xi$  и почти наверное:  $\xi_n \xrightarrow{\text{с.к.}} \xi$ ,  $\sum_{k=1}^{\infty} M|\xi_k - \xi|^2 < \infty \Rightarrow \xi_n \xrightarrow{\text{п.н.}} \xi$ .

Кроме того,  $\xi_n \xrightarrow{\text{с.к.}} \xi \Leftrightarrow \xi_n \xrightarrow{P} \xi$  и  $M\xi_n^2 \rightarrow M\xi^2$  (Лозэв).

Закон больших чисел может пониматься и в смысле сходимости почти наверное («Усиленный ЗБЧ»). Различные его варианты изложены в изящных теоремах, однако он мало интересен с прикладной точки зрения.

В ЦПТ мы показали сходимость  $\eta_n$  лишь по распределению. К сожалению, никакая другая сходимость не имеет места: последовательность  $\eta_n$  не сходится ни к какой случайной величине по вероятности, а следовательно и с вероятностью единица и в среднем квадратичном. Можно показать, что сходимость  $\eta_n \xrightarrow{P} \eta$ , а следовательно, и сходимости  $\eta_n \xrightarrow{\text{с.к.}} \eta$  и  $\eta_n \xrightarrow{\text{п.н.}} \eta$  не имеют места.

## СЛУЧАЙНЫЕ ПРОЦЕССЫ.

Лекция 9.

Пусть задано вероятностное пространство  $(\Omega, \mathcal{F}, P(\cdot))$  и некоторое множество  $T$  значений  $t$  (непрер. или дискр.). Случайной функцией называется  $\xi(t) = \xi(\omega, t)$ , такая, что  $\forall t_0 \in T$   $\xi(t_0)$  — сл. величина. Если зафиксировать  $t = t_0$ , то получим *сечение* сл. функции (или процесса, если  $t$  интерпретировать как время). Если зафиксировать  $\omega = \omega_0$ , то полученная  $\xi(\omega_0, t)$  есть *выборочная функция* (реализация) сл. процесса. Распределение вероятностей в каждом сечении  $t$  можно задать функцией  $F(x, t) = P\{\xi(t) < x\}$ , но она не дает никакого представления о связи сл. величин, характеризующих разные сечения. Более полным является задание семейства функций совместного распределения  $F(x_1, t_1, \dots, x_n, t_n)$  для  $n$  сечений,  $n = 1, 2, \dots$ . Однако при этом должно быть выполнено условие *согласованности*, касающееся редукции (уменьшению) числа аргументов и их перестановок.

При каком  $n$  функции совместного распределения достаточно для описания процесса? Если рассматривать дискретные значения параметра  $t$ , то для схемы независимых испытаний  $n = 1$ , для марковских цепей  $n = 2$ . Обычно этим ( $n = 2$ ) ограничиваются в случае непрерывного времени для так называемых процессов с независимыми приращениями.

Рассмотрим в качестве примера такие два процесса 2 порядка.

**Процесс Пуассона** (одномерный случай).

*Определение.* Сл. функция  $\eta(t)$   $0 \leq t < \infty$ , называется процессом Пуассона или пуассоновским потоком событий, если

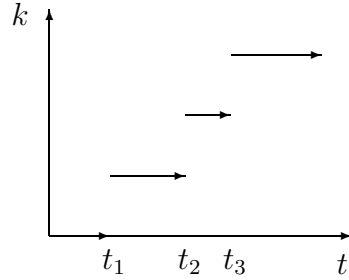
1. для любых  $0 \leq t_1 < \dots < t_n$  сл. величины  $\eta_i - \eta_{i-1}$ ,  $i = 1, 2, \dots, n$ , независимы в совокупности (процесс с независимыми приращениями),

2. сл. величина  $\eta(t) - \eta(s)$ ,  $0 \leq s < t$ , имеет распределение Пуассона с параметром  $\lambda(t - s)$ :

$$P\{\eta(t) - \eta(s) = k\} = \frac{(\lambda(t - s))^k}{k!} e^{-\lambda(t - s)}.$$

3. Если  $\eta(0) = 0$ , то говорят, что процесс *начинается в нуле*.

Процесс регулярный: траектория непрерывна справа и имеет предел слева.



Выборочная функция Пуассоновского процесса

**Теорема.** Пусть  $\eta(t)$  — процесс с независимыми приращениями и пусть выполняются условия (при  $t \rightarrow 0$ ):

а)  $P\{\eta(t) - \eta(0) = 1\} = \lambda t + o(\lambda t)$ ,

б)  $P\{\eta(t) - \eta(0) > 1\} = o(\lambda t)$ .

Тогда  $\eta(t)$  — процесс Пуассона.

**Доказательство.** В силу независимости приращений достаточно доказать, что выполняется пункт 2 в определении.

Разобьем интервал  $t = n\Delta$  на  $n$  (одинаковых) подынтервалов. Обозначим через  $\Delta_i$   $i$ -ый подынтервал,  $i = 1, 2, \dots, n$ . При ( $n > k$ ) событие  $\{\eta(t) = k\}$  представим в виде  $A + B$ , где  $A$  — событие, в котором в каждый подынтервал попадает не более 1 точки,  $B$  — событие, в котором по крайней мере в один подынтервал попадает более одной точки. Тогда  $P(A) = C_n^k p^k (1-p)^{n-k}$ , где  $p = \lambda\Delta + o(\lambda\Delta)$ . Поскольку  $pn = \lambda n\Delta + n o(\frac{\lambda t}{n}) \xrightarrow{n \rightarrow \infty} \lambda t$ , то применяя теорему Пуассона, получаем  $P(A) \xrightarrow{n \rightarrow \infty} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ . Аналогично,

$$P(B) < \sum_{i: \eta(\Delta_i) > 1} P\{\eta(\Delta_i) > 1\} < n o\left(\frac{\lambda t}{n}\right) \xrightarrow{n \rightarrow \infty} 0. \square$$

**Задача.** Найти распределение сл. величины — времени ожидания  $\tau$  первого события в пуассоновском потоке событий.

Из условия  $P\{\tau \geq t\} = P\{\eta(t) - \eta(0) = 0\} = e^{-\lambda t}$  получаем  $F_\tau(t) = P\{\tau < t\} = 1 - e^{-\lambda t}$ ,  $t \geq 0$ .

Плотность вероятности  $p_\tau(t) = \lambda e^{-\lambda t}$ ,  $t \geq 0$ .

Среднее время ожидания  $M\tau = \int_0^\infty t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}$ .

Заметим, что из свойств сл. процесса Пуассона следует, что время ожидания не зависит от момента начала ожидания, а лишь от величины интервала ожидания.

**Радиоактивный распад.** Вероятность распада для каждого атома  $P\{\eta(t) > 0\} = P\{\tau < t\} = 1 - e^{-\lambda t}$ . При  $t = t_0$ ,  $1 - e^{-\lambda t_0} = \frac{1}{2}$ , половина всего вещества распадается, так что  $e^{-\lambda t_0} = \frac{1}{2}$ ,  $\lambda t_0 = \ln 2$ ,  $t_0 = \frac{\ln 2}{\lambda}$  — время полураспада.

**Винеровский сл. процесс.** Случайный процесс  $\xi(t)$  называется Винеровским, если

1. для  $0 \leq t_1 < t_2 < \dots < t_n$   $\eta_i = \xi(t_i) - \xi(t_{i-1})$  — независимы в совокупности,
2.  $\xi(t) - \xi(s) \sim \mathcal{N}(0, t - s)$ ,  $0 < s < t$ ,

3.  $\xi(0) = 0$  (начинается в нуле).

Таким образом, вектор  $\eta = (\eta_1, \eta_2, \dots, \eta_n)$  распределен с плотностью

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_1 - t_{i-1})}} \exp \left\{ -\frac{x_i^2}{2(t_1 - t_{i-1})} \right\}.$$

Если перейти к  $\xi_i = \sum_{k=1}^i \eta_k$  с помощью невырожденного преобразования с матрицей  $A$  вида

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}, \quad A^{-1} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

то получим для вектора  $\xi$  плотность

$$p_\xi(x, t) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi(t_1 - t_{i-1})}} \exp \left\{ -\frac{(x_i - x_{i-1})^2}{2(t_1 - t_{i-1})} \right\}.$$

**Пример. Броуновское движение** (одномерная задача).

Пусть  $\xi(t)$  — координата броуновской частицы на прямой,  $\xi(0) = 0$ .

$$\xi(t+s) = [\xi(t+s) - \xi(s)] + [\xi(s) - \xi(0)], \quad t, s > 0.$$

Из однородности следует, что слагаемые в правой части независимы и поэтому  $D\xi(t+s) = D\xi(t) + D\xi(s)$ , откуда  $D\xi(t) = \sigma^2 t$ , где  $\sigma^2$  — некоторая константа.

Рассмотрим дискретный аналог — случайное блуждание по одномерной сетке в дискретном времени:  $\xi_n = \sum_{i=1}^n \xi_i$ ,  $P\{\xi_i = \pm \Delta x\} = \frac{1}{2}$ ,  $t = n\Delta t$ .

$M\xi_n = \sum M\xi_i = 0$ ,  $D\xi_n = \sum D\xi_i = n(\Delta x)^2$ , приравнявая  $\sigma^2 n\Delta t$ , получим, что  $\frac{(\Delta x)^2}{\Delta t} = \sigma^2 = \text{const}$ . Переходя к пределу при  $\Delta x \rightarrow 0$ ,  $\Delta t \rightarrow 0$ ,  $\frac{(\Delta x)^2}{\Delta t} = \sigma^2$ , по ЦПТ получаем

$$\frac{\xi_n}{\sigma\sqrt{t}} \xrightarrow[n \rightarrow \infty]{d} \frac{\xi(t)}{\sigma\sqrt{t}} \sim \mathcal{N}(0, 1).$$

Отсюда  $\xi(t) \sim \mathcal{N}(0, \sigma^2 t)$  или  $\xi(t) - \xi(s) \sim \mathcal{N}(0, \sigma^2(t-s))$ .

В общем случае (если нет симметрии)

$\xi(t) - \xi(s) \sim \mathcal{N}(\vartheta(t-s), \sigma^2(t-s))$ , где  $\vartheta$  — коэффициент сноса.

**Характеризация винеровского процесса.**

**Теорема.** Пусть сл. процесс удовлетворяет условиям:

- 1) для любых непересекающихся промежутков приращения независимы;
- 2) для любого промежутка вероятность зависит лишь от длины промежутка (однородность);

3) для функции распределения приращения  $\xi(t_0+t) - \xi(t_0)$ ,  $F(x, t)$  существует плотность  $p(x, t)$ , имеющая непрерывные и ограниченные по  $x$ ,  $-\infty < x < \infty$ , производные до 3-го порядка  $\forall t > 0$ ;

4) для малых  $\Delta t$  малые приращения более вероятны, чем большие, а именно

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} xp(x, \Delta t) dx = A; \quad (12)$$

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} x^2 p(x, \Delta t) dx = B > 0; \quad (13)$$

$$\lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \int_{-\infty}^{\infty} |x|^3 p(x, \Delta t) dx = 0; \quad (14)$$

Тогда  $\xi(t)$  — (обобщенный) винеровский процесс.

**Доказательство.** Рассмотрим два смежных промежутка  $(0, t)$  и  $(t, t + \Delta t)$ , а также его сумму (объединение). Пользуясь формулой для плотности суммы независимых сл. величин, получаем

$$p(x, t + \Delta t) = \int_{-\infty}^{\infty} p(x - s, t) p(s, \Delta t) ds. \quad (15)$$

Разлагая сомножитель  $p(x - s, t)$  в ряд Тейлора по степеням  $s$ , получим

$$p(x - s, t) = p(x, t) - \frac{\partial p(x, t)}{\partial x} s + \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2} s^2 - \frac{1}{6} \frac{\partial^3 p(\tilde{x}, t)}{\partial x^3} s^3,$$

подставляя это в (15), получаем

$$p(x, t + \Delta t) = p(x, t) - \frac{\partial p(x, t)}{\partial x} \int_{-\infty}^{\infty} sp(s, \Delta t) ds + \\ + \frac{1}{2} \frac{\partial^2 p(x, t)}{\partial x^2} \int_{-\infty}^{\infty} s^2 p(s, \Delta t) ds - \frac{1}{6} \int_{-\infty}^{\infty} \left[ \frac{\partial^3 p(\tilde{x}, t)}{\partial x^3} \right] s^3 p(s, \Delta t) ds.$$

Разделим на  $\Delta t$  и перейдем к пределу при  $\Delta t \rightarrow 0$ , учитывая свойства (12, 13, 14):

$$\frac{\partial p(x, t)}{\partial t} = -A \frac{\partial p(x, t)}{\partial x} + \frac{B}{2} \frac{\partial^2 p(x, t)}{\partial x^2}, \quad \int_{-\infty}^{\infty} \bar{p}(x, t) dx = 1, \quad p(x, t) \geq 0.$$

Подстановкой  $y = x - At$ ,  $\tau = Bt$ ,  $p(x, t) = \bar{p}(y, \tau)$  приводим к

$$\frac{\partial \bar{p}}{\partial \tau} = \frac{1}{2} \frac{\partial^2 \bar{p}}{\partial y^2}, \quad -\infty < y < \infty$$

— уравнению теплопроводности с решением

$$\bar{p}(y, \tau) = \frac{1}{\sqrt{2\pi\tau}} e^{-\frac{y^2}{2\tau}}, \quad \int_{-\infty}^{\infty} \bar{p} dy = 1.$$

В прежних переменных это выглядит так:

$$p(x, t) = \frac{1}{\sqrt{2\pi Bt}} e^{-\frac{(x-At)^2}{2Bt}}. \square$$

**Задача.** Найти распределение  $\tau_x$  — времени первого достижения броуновской частицей точки  $x$  ( $x > 0$ ).

$$P\{\xi(t) > x | \tau_x < t\} = \frac{1}{2} = \frac{P\{\xi(t) > x\}}{P\{\tau_x < t\}}.$$

Отсюда  $P\{\tau_x < t\} = 2P\{\frac{\xi(t)}{\sqrt{t}} > \frac{x}{\sqrt{t}}\} = \sqrt{\frac{2}{\pi}} \int_{\frac{x}{\sqrt{t}}}^{\infty} e^{-\frac{\alpha^2}{2}} d\alpha$  и  $p_{\tau_x} = \frac{1}{\sqrt{2\pi}} \frac{x}{t^{3/2}} e^{-\frac{x^2}{2t}}$ .

Аналогично для  $\xi_t = \max_{0 \leq s \leq t} \xi(s)$  — максимальной координаты броуновской частицы имеем:  $P\{\max_{0 \leq s \leq t} \xi(s) > x\} = P\{\tau_x < t\}$  и  $p_{\xi_t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$ ,  $x \geq 0$ .

Интересно, что  $\forall x: P\{\tau_x < \infty\} = 1$  и  $\forall t: P\{\max_{0 \leq s \leq t} \xi(s) > 0\} = 1$ .

**Теория второго порядка (корреляционная теория).**

Лекция 10.

Далее будут встречаться комплексные случайные процессы  $\xi(t) = \eta(t) + i\zeta(t)$ , где  $\eta(t)$  и  $\zeta(t)$  — действительные случайные процессы.

Корреляционной функцией случайного процесса  $\xi(t)$  называется

$$K(t, s) = M(\xi(t) - M\xi(t))(\overline{\xi(s) - M\xi(s)}).$$

*Примеры.* (1) Корреляционная функция винеровского случайного процесса,  $\xi(t) - \xi(s) \sim \mathcal{N}(0, t - s)$ ,  $s < t$ . При  $0 < s < t$

$$M\xi(t)\xi(s) = M[(\xi(t) - \xi(s)) + (\xi(s) - \xi(0))](\xi(s) - \xi(0)) = M(\xi(s) - \xi(0))^2 = s.$$

При  $0 < t < s$  аналогично получаем  $M\xi(t)\xi(s) = t$ . Следовательно,  $M\xi(t)\xi(s) = \min(t, s)$ .

(2). Корреляционная функция пуассоновского процесса,

$$P(\xi(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad M\xi(t) = \lambda t.$$

Для  $0 < s < t$ :

$$\begin{aligned} K(t, s) &= M(\xi(t) - \lambda t)(\xi(s) - \lambda s) = M[(\xi(t) - \xi(s) - \lambda(t - s)) + \\ &+ (\xi(s) - \lambda s)](\xi(s) - \lambda s) = M(\xi(s) - \lambda s)^2 = \lambda s. \end{aligned}$$

Поэтому и в этом случае  $K(t, s) = \lambda \min(t, s)$ .

В обоих примерах использована независимость приращений.

**Определение.** Случайный процесс  $\xi(t)$  называется непрерывным в среднем квадратичном в точке  $t$ , если

$$M|\xi(t+h) - \xi(t)|^2 \rightarrow 0 \text{ при } h \rightarrow 0,$$

или, иначе говоря, если

$$\xi(t) = \text{l.i.m.}_{h \rightarrow 0} \xi(t+h).$$

Речь идет о случайных процессах с конечными моментами второго порядка;  $|M\xi| \leq M|\xi| \leq [M|\xi|^2]^{1/2} < \infty$ , в этом случае и  $M|\xi| < \infty$ .

Напомним, что последовательность  $\{\xi_k\}$  сходится в среднем квадратичном к  $\xi$ , если  $M|\xi_k - \xi|^2 \rightarrow 0$ ,  $k \rightarrow \infty$ ;  $\xi = \text{l.i.m.}_{k \rightarrow \infty} \xi_k$ .

Свойства с.к. сходимости.

Критерий 1. (фундаментальность) Для сходимости в с.к. последовательности  $\{\xi_k\}$  необходимо и достаточно, чтобы  $M|\xi_k - \xi_n|^2 \rightarrow 0$  при  $k, n \rightarrow \infty$ .<sup>18</sup>

Критерий 2. Последовательность  $\{\xi_k\}$  сходится в среднем квадратичном, если и только если числовая последовательность  $M\xi_k \bar{\xi}_n \rightarrow M\xi \bar{\xi}$  при  $k, n \rightarrow \infty$  независимо.

Доказательство.  $M(\xi_k - \xi_n)(\bar{\xi}_k - \bar{\xi}_n) = M\xi_k \bar{\xi}_k + M\xi_n \bar{\xi}_n - M\xi_k \bar{\xi}_n - M\xi_n \bar{\xi}_k \rightarrow 0$ , если  $M\xi_k \bar{\xi}_n$  сходится.

Наоборот,  $M\xi_k \bar{\xi}_n - M\xi \bar{\xi} = M(\xi_k - \xi)(\bar{\xi}_n - \bar{\xi}) + M\xi(\bar{\xi}_n - \bar{\xi}) + M(\xi_k - \xi)\bar{\xi} \rightarrow 0$ , если  $\xi_k \xrightarrow{\text{с.к.}} \xi$ .  $\square$

Другие свойства: если  $\xi = \text{l.i.m.}_{k \rightarrow \infty} \xi_k$ , то

(1)  $M\xi = \lim_{k \rightarrow \infty} M\xi_k$ . Действительно,  $|M\xi - M\xi_k| \leq M|\xi - \xi_k| \leq (M|\xi - \xi_k|^2)^{1/2} \rightarrow 0$ ,  $k \rightarrow \infty$ .

(2)  $M|\xi|^2 = \lim_{k \rightarrow \infty} M|\xi_k|^2$ . Доказывается на основе критерия 2 или непосредственно:  $M|\xi - \xi_k|^2 = M|\xi|^2 - M\xi \bar{\xi}_k - M\xi_k \bar{\xi} + M|\xi_k|^2 \geq M|\xi|^2 - 2(M|\xi|^2 M|\xi_k|^2)^{1/2} + M|\xi_k|^2 = (\sqrt{M|\xi|^2} - \sqrt{M|\xi_k|^2})^2$ .  $\square$

(3) Если, кроме того,  $\eta = \text{l.i.m.}_{k \rightarrow \infty} \eta_k$ , то  $M\xi \bar{\eta} = \lim_{k \rightarrow \infty} M\xi_k \bar{\eta}_k$ . Действительно,  $|M(\xi_n \bar{\eta}_n - \xi \bar{\eta})| \leq M|\xi_n \bar{\eta}_n - \xi \bar{\eta}| \leq M|\xi_n \bar{\eta}_n - \xi_n \bar{\eta} + \xi_n \bar{\eta} - \xi \bar{\eta}| \leq M|\xi_n \bar{\eta}_n - \xi_n \bar{\eta}| + M|\xi_n \bar{\eta} - \xi \bar{\eta}| \leq \sqrt{M|\xi_n|^2 M|\eta_n - \eta|^2} + \sqrt{M|\xi_n - \xi|^2 M|\eta|^2} \xrightarrow{n \rightarrow \infty} 0$ .  $\square$

Вернемся к сл. процессам. Всюду далее будем считать  $M\xi(t) = 0$ .

**Теорема.**

1. Случайный процесс  $\xi(t)$  с.к. непрерывен в точке  $t_0$  тогда и только тогда, когда  $K(t, s)$  непрерывна в точке  $(t_0, t_0)$ .

2. Если  $K(t, s)$  непрерывна на диагонали  $t = s$ , то  $K(t, s)$  непрерывна всюду.

*Доказательство.* 1. Достаточность. Пусть  $K(t, s)$  непрерывна в точке  $(t_0, t_0)$ . Тогда

$$M|\xi(t_0+h) - \xi(t_0)|^2 = K(t_0+h, t_0+h) - K(t_0+h, t_0) - K(t_0, t_0+h) + K(t_0, t_0) \rightarrow 0$$

при  $h \rightarrow 0$ .

<sup>18</sup>(Если бы  $M$  символизировало римановское интегрирование, не было бы полноты).

Необходимость. Пусть  $\xi(t)$  с.к. непрерывна при  $t = t_0$ , тогда

$$K(t_0 + h, t_0 + k) - K(t_0, t_0) = \mathbb{M}[(\xi(t_0 + h) - \xi(t_0))(\overline{\xi(t_0 + k) - \xi(t_0)}) + (\xi(t_0 + h) - \overline{\xi(t_0)})\xi(t_0) + \xi(t_0)(\overline{\xi(t_0 + k) - \xi(t_0)})].$$

Далее для каждого из слагаемых правой части воспользуемся свойством (3). Например, для первого:

$$\begin{aligned} & |\mathbb{M}(\xi(t_0 + h) - \xi(t_0))(\overline{\xi(t_0 + k) - \xi(t_0)})| \leq \\ & \leq \sqrt{\mathbb{M}|\xi(t_0 + h) - \xi(t_0)|^2 \mathbb{M}|\xi(t_0 + k) - \xi(t_0)|^2} \xrightarrow{h \rightarrow 0, k \rightarrow 0} 0. \end{aligned}$$

2. Если  $K(t, s)$  непрерывна на диагонали, например, в точках  $(t, t)$ ,  $(s, s)$ , то  $\xi(t+h) \xrightarrow{\text{с.к.}} \xi(t)$  и  $\xi(s+k) \xrightarrow{\text{с.к.}} \xi(s)$  при  $h \rightarrow 0$ . Следовательно,  $\mathbb{M}\xi(t+h)\xi(s+k) \rightarrow \mathbb{M}\xi(t)\xi(s)$ , т.е.  $K(t+h, s+k) \rightarrow K(t, s)$ .  $\square$

*Следствие.* Пуассоновский и винеровский случайные процессы непрерывны в среднем квадратичном. Однако, реализации пуассоновского процесса разрывные (ступенчатые) функции.

**Определение.** Случайная функция  $\xi(t)$  называется с.к. дифференцируемой в точке  $t$ , если существует с.к. предел  $\frac{\xi(t+h) - \xi(t)}{h}$  при  $h \rightarrow 0$ . Этот предел называется с.к. производной  $\xi'(t)$ :

$$\xi'(t) = \text{l.i.m.}_{h \rightarrow 0} \frac{\xi(t+h) - \xi(t)}{h}, \text{ или } \mathbb{M} \left| \frac{\xi(t+h) - \xi(t)}{h} - \xi'(t) \right|^2 \rightarrow 0, h \rightarrow 0.$$

Согласно критерию 2,  $\xi(t)$  с.к. дифференцируема в точке  $t$ , если и только если

$$Q = \frac{1}{hk} \mathbb{M}(\xi(t+h) - \xi(t))(\xi(t+k) - \xi(t)) \text{ сходится при } h, k \rightarrow 0 :$$

$$Q = \frac{K(t+h, t+k) - K(t, t+k) - K(t+h, t) + K(t, t)}{hk} \xrightarrow{h \rightarrow 0, k \rightarrow 0} \left. \frac{\partial^2 K(t, s)}{\partial t \partial s} \right|_{t=s}.$$

Это разностное отношение для второй производной сходится к правой части, если, например,  $\frac{\partial^2 K(t, s)}{\partial t \partial s}$  непрерывна.

В этом случае  $\frac{\partial^2 K(t, s)}{\partial t \partial s} = \mathbb{M}\xi'(t)\xi'(s)$  - корреляционная функция с.к. производной  $\xi(t)$ . Это утверждение — следствие свойства 2: если  $\frac{\xi(t+h) - \xi(t)}{h} \xrightarrow{\text{с.к.}}_{h \rightarrow 0} \xi'(t)$ ,  $\frac{\xi(s+k) - \xi(s)}{k} \xrightarrow{\text{с.к.}}_{k \rightarrow 0} \xi'(s)$ , то

$$\mathbb{M} \frac{\xi(t+h) - \xi(t)}{h} \frac{\xi(s+k) - \xi(s)}{k} \xrightarrow{h \rightarrow 0, k \rightarrow 0} \mathbb{M}\xi'(t)\xi'(s) = \frac{\partial^2 K(t, s)}{\partial t \partial s},$$

если эта производная непрерывна.

**Определение.** Случайная функция  $\xi(t)$  с.к. интегрируема по Риману на  $[a, b]$ , если последовательность римановских интегральных сумм  $\sum_{[a, b]} \xi(t_k) \delta t_k$  с.к. сходится при  $\max_k \delta t_k \rightarrow 0$ .

**Теорема.** Случайная функция  $\xi(t)$  с.к. интегрируема на  $[a, b]$  по Риману, если  $K(t, s)$  интегрируема по Риману на  $[a, b] \times [a, b]$ . Действительно, пусть

$$S = \sum_{[a,b]} \xi(t_k) \delta t_k, \quad S' = \sum_{[a,b]} \xi(t'_k) \delta t'_k$$

Тогда

$$MSS' = \sum_{[a,b] \times [a,b]} K(t_i, t'_k) \delta t_i \delta t'_k$$

и факт существования предела справа эквивалентен с.к. интегрируемости  $\xi(t)$  (с.к. сходимости S).  $\square$

## ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

### Распределение ортогональных проекций нормального вектора.

Лекция 11.

Рассмотрим  $n$ -мерное евклидово пространство,  $x = (x_1, \dots, x_n)$ ,  $(x, y) = \sum_{k=1}^n x_k y_k$ . Пусть  $\xi = (\xi_1, \dots, \xi_n)$  — случайный вектор в этом пространстве, причем  $\xi \sim \mathcal{N}(0, \sigma^2 I)$ .

Напомним, что в общем случае, когда  $\xi \sim \mathcal{N}(0, \Sigma)$ ,

$$p_\xi(x) = \frac{(\det \Sigma^{-1})^{1/2}}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} (x, \Sigma^{-1} x) \right\}.$$

Если  $\Sigma = \sigma^2 I$ , то координаты  $\xi_i$  независимы и

$$p_{\xi_i}(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} x_i^2 \right\}.$$

Пусть  $\Pi$  — ортогональный проектор на  $\mathcal{L} \subset \mathcal{R}_n = \mathcal{L} \oplus \mathcal{L}^\perp$ .

Это значит, что если  $x = x_1 + x_2$ ,  $x_1 \in \mathcal{L}_1$ ,  $x_2 \in \mathcal{L}_2$ , то  $\Pi x = x_1$ .

Оператор  $\Pi$  линейный, самосопряженный и идемпотентный ( $\Pi^2 = \Pi$ ). Обрато, эти свойства полностью определяют ортогональный проектор.

Обозначим через  $U$  оператор ортогонального преобразования (оператор перехода от одного ортонормированного базиса к другому),  $\tilde{e}_i = U e_i$ . Он обладает свойствами:  $U^* = U^{-1}$  и  $\|Ux\| = \|x\|$  ( $\|U^*x\| = \|x\|$ ),  $\det \hat{U} = \pm 1$ . (Здесь  $\hat{U}$  — матрица оператора  $U$ ).

**Теорема.** Распределение вектора  $\eta = U\xi$  совпадает с распределением вектора  $\xi$  (т.е.  $\eta_i \sim \mathcal{N}(0, \sigma^2)$  и независимы).

**Доказательство.**

$$\begin{aligned} p_\eta(y) &= p_\xi(U^{-1}y) |\det \hat{U}^{-1}| = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \|U^{-1}y\|^2 \right\} = \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{\|y\|^2}{2\sigma^2} \right\}. \quad \square \end{aligned}$$

**Определение 1.** Распределением  $\chi_m^2$  или Пирсона (E.S.Pearson) с  $m$  степенями свободы называется распределение сл. величины, равной сумме квадратов  $\sum_{k=1}^m \xi_k^2$  независимых сл. величин  $\xi_k \sim \mathcal{N}(0, 1)$ .

Свойства<sup>19</sup>. Плотность:

$$p_{\chi_m^2}(x) = \frac{1}{2^{m/2} \Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}.$$

Моменты:  $M\chi_m^2 = m$ ,  $D\chi_m^2 = 2m$ .

**Следствие 1.**  $\|\Pi_m \xi\|^2 = \sigma^2 \chi_m^2$ .

**Доказательство.** Пусть  $\Pi_m x \in L_m \subset \mathcal{R}_n$ . Рассмотрим новый базис  $\{\tilde{e}_i\}$ ,  $i = 1, \dots, n$ , удовлетворяющий условиям  $\tilde{e}_i \in L_m$ ,  $i = 1, \dots, m$ ,  $\tilde{e}_i \in L_m^\perp$ ,  $i = m+1, \dots, n$  и пусть  $U$  — ортогональный оператор перехода от старого базиса  $\{e_i\}$  к новому  $\{\tilde{e}_i\}$ ,  $\tilde{e}_i = Ue_i$ ,  $i = 1, \dots, n$ . Тогда  $\|\Pi_m \xi\|^2 = \sum_{i=1}^m (\xi, \tilde{e}_i)^2 = \sigma^2 \chi_m^2$ , т.к. по теореме  $(\xi, \tilde{e}_i) \sim \mathcal{N}(0, \sigma^2)$ .  $\square$

**Определение 2.** Пусть  $\chi_k^2$  и  $\chi_m^2$  независимы. Тогда сл. величина  $F_{k,m} = \frac{\frac{1}{k} \chi_k^2}{\frac{1}{m} \chi_m^2}$  контролируется распределением Фишера (R.A.Fischer).

Свойства. Плотность:

$$p_{F_{k,m}}(x) = \frac{k^{\frac{k}{2}} m^{\frac{m}{2}} x^{\frac{k}{2}-1} (kx+m)^{-\frac{k+m}{2}}}{\Gamma(\frac{k}{2}) \Gamma(\frac{m}{2})} \Gamma(\frac{k+m}{2}), \quad x > 0.$$

Моменты:  $MF_{m,k} = \frac{m}{m-2}$ ,  $DF_{m,k} = \frac{2m^2}{(m-2)^2(m-4)} (1 + \frac{m-2}{k})$ .

**Следствие 2.** Пусть  $\mathcal{R}_n = L_k \oplus L_m \oplus L_s$ ,  $k+m+s = n$ ,  $\Pi_k$  — ортогональный проектор на  $L_k$ ,  $\Pi_m$  — ортогональный проектор на  $L_m$ . Тогда  $F_{k,m} = \frac{\frac{1}{k} \|\Pi_k \xi\|^2}{\frac{1}{m} \|\Pi_m \xi\|^2}$ .

**Доказательство.** Пусть новый базис  $\{\tilde{e}_i\}$ ,  $i = 1, \dots, n$  таков, что  $\tilde{e}_i \in L_k$ , если  $i = 1, \dots, k$ ,  $\tilde{e}_i \in L_m$ , если  $i = k+1, \dots, k+m$ ,  $\tilde{e}_i \in L_s$ , если  $i = k+m+1, \dots, n$ . Тогда  $\|\Pi_k \xi\|^2 = \sigma^2 \chi_k^2$ ,  $\|\Pi_m \xi\|^2 = \sigma^2 \chi_m^2$  и независимы.  $\square$

**Определение 3.** Пусть  $\xi \sim \mathcal{N}(0, 1)$  и  $\chi_m^2$  независимы. Тогда сл. величина  $t_m = \frac{\xi}{\sqrt{\frac{\chi_m^2}{m}}}$  контролируется распределением Стьюдента (В.С.Госсета) с  $m$  степенями свободы.

Свойства. Плотность:

$$p_{t_m}(x) = \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \frac{1}{\sqrt{\pi m}} \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \quad x > 0.$$

Моменты:  $Mt_m = 0$ ,  $Dt_m = \frac{m}{m-2}$ .

**Следствие 3.** Обозначим  $e = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$ ,  $L_1$  — одномерное подпространство, определенное вектором  $e$ ,  $\Pi_1$  — ортогональный проектор на  $L_1$ ,  $\Pi_1 \xi = (e, \xi)e = (\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i)e = (\frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i, \dots, \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i)$ .

Пусть  $\xi \sim \mathcal{N}(0, \sigma^2 I)$ , тогда

$$\frac{(\xi, e)}{[\frac{1}{n-1} \|(I - \Pi_1)\xi\|^2]^{1/2}} = t_{n-1} \quad (16)$$

<sup>19</sup>Вывод соответствующих формул см. в учебнике Пытьева, Шишмарева, стр. 97 – 102.

— распределение Стьюдента с  $n - 1$  степенью свободы.

**Доказательство.** Перейдем от базиса  $\{e_i\}$  к базису  $\{e, \tilde{e}_1, \dots, \tilde{e}_{n-1}\}$ , так что  $L = \mathcal{L}(e)$ ,  $L^\perp = \mathcal{L}(\tilde{e}_1, \dots, \tilde{e}_{n-1})$ . Тогда координаты в этом новом базисе независимы и распределены как  $\mathcal{N}(0, \sigma^2)$ , а  $\|(I - \Pi_1)\xi\|^2 = \sigma^2 \chi_{n-1}^2$  и по определению 3 получаем (16).

Перепишем (16) еще раз

$$\begin{aligned} \frac{(\xi, e)}{\left[\frac{1}{n-1}\|(I - \Pi_1)\xi\|^2\right]^{1/2}} &= \frac{\frac{1}{\sqrt{n-1}} \sum_{i=1}^n \xi_i}{\left[\frac{1}{n-1} \sum_{k=1}^n \left(\xi_k - \frac{1}{n} \sum_{i=1}^n \xi_i\right)^2\right]^{1/2}} = \\ &= \frac{\frac{1}{n} \sum_{i=1}^n \xi_i}{\left[\frac{1}{n(n-1)} \sum_{k=1}^n \left(\xi_k - \frac{1}{n} \sum_{i=1}^n \xi_i\right)^2\right]^{1/2}}. \square \end{aligned}$$

Если  $\xi \sim \mathcal{N}(\bar{\mu}, \sigma^2 I)$ , где  $\bar{\mu} = (\mu, \dots, \mu)$ , то последнюю формулу можно заменить на

$$t_{n-1} = \frac{\frac{1}{n} \sum_{i=1}^n (\xi_i - \mu)}{\left[\frac{1}{n(n-1)} \sum_{k=1}^n \left(\xi_k - \frac{1}{n} \sum_{i=1}^n \xi_i\right)^2\right]^{1/2}} = \frac{\hat{\mu} - \mu}{\sqrt{\frac{1}{n} \hat{\sigma}^2}}, \quad (17)$$

где  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \xi_i$ ,  $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n \left(\xi_k - \frac{1}{n} \sum_{i=1}^n \xi_i\right)^2 = \frac{1}{n-1} \|(I - \Pi_1)\xi\|^2$ .

### Интервальные оценки нормального распределения.

Пусть  $\{\xi_i\}$ ,  $i = 1.2. \dots$  — последовательность независимых нормально распределенных  $\mathcal{N}(\mu, \sigma^2)$  сл. величин (независимых измерений). Требуется оценить значения неизвестных параметров  $\mu$  и  $\sigma^2$ . Рассмотрим четыре случая.

1. Оценивание  $\mu$  при известном  $\sigma^2$ .

Очевидно,  $\frac{\sum_{i=1}^n (\xi_i - \mu)}{\sqrt{n\sigma^2}} \sim \mathcal{N}(0, 1)$ .

Тогда

$$P\left(\left|\frac{\sum_{i=1}^n (\xi_i - \mu)}{\sqrt{n\sigma^2}}\right| < \varepsilon\right) = \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon),$$

или, преобразуя неравенство, получаем

$$P\left(\hat{\mu} - \varepsilon \sqrt{\frac{\sigma^2}{n}} < \mu < \hat{\mu} + \varepsilon \sqrt{\frac{\sigma^2}{n}}\right) = \alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon),$$

где  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \xi_i$  — середина интервала, шириной  $2\varepsilon \sqrt{\frac{\sigma^2}{n}}$ , которому с вероятностью  $\alpha(\varepsilon) = 1 - 2\Phi(-\varepsilon)$  принадлежит неизвестный параметр  $\mu$ .

Упомянутый интервал называется *оценивающим множеством* параметра  $\mu$ . Отметим, что чем точнее мы хотим оценить параметр, тем меньше вероятность (определенность) этой оценки  $\alpha(\varepsilon)$ .

2. Оценивание  $\sigma^2$  при известном  $\mu$ . Из определения 1 следует, что  $\frac{\sum_{i=1}^n (\xi_i - \mu)^2}{\sigma^2} \sim \chi_n^2$ , поэтому

$$P\left(\varepsilon_1 < \frac{\sum_{i=1}^n (\xi_i - \mu)^2}{\sigma^2} < \varepsilon_2\right) = \alpha(\varepsilon_1, \varepsilon_2),$$

или

$$P\left(\frac{\sum_{i=1}^n (\xi_i - \mu)^2}{\varepsilon_2} < \sigma^2 < \frac{\sum_{i=1}^n (\xi_i - \mu)^2}{\varepsilon_1}\right) = \alpha(\varepsilon_1, \varepsilon_2),$$

причем обычно  $\varepsilon_1$  и  $\varepsilon_2$  выбирают так, чтобы  $P(\chi_n^2 < \varepsilon_1) = P(\chi_n^2 > \varepsilon_2)$ . Интервал, которому удовлетворяет  $\sigma^2$  с вероятностью  $\alpha$ , называется интервальной оценкой  $\sigma^2$ .

3. Оценивание  $\mu$  при неизвестном  $\sigma^2$ .

Воспользуемся формулой (17)

$$P(|t_{n-1}| < \varepsilon) = P\left(\left|\frac{\hat{\mu} - \mu}{\sqrt{\frac{1}{n}\hat{\sigma}^2}}\right| < \varepsilon\right) = \alpha(\varepsilon)$$

и получаем выражение, аналогичное пункту 1:

$$P\left(\hat{\mu} - \varepsilon\sqrt{\frac{\hat{\sigma}^2}{n}} < \mu < \hat{\mu} + \varepsilon\sqrt{\frac{\hat{\sigma}^2}{n}}\right) = \alpha_{n-1}(\varepsilon),$$

но с тем отличием, что вместо  $\sigma^2$  стоит  $\hat{\sigma}^2$  и  $\alpha_{n-1}(\varepsilon)$  соответствует распределению Стьюдента с  $n - 1$  степенями свободы.

4. Оценивание  $\sigma^2$  при неизвестном  $\mu$ .

Здесь по аналогии с пунктом 2 получаем

$$P\left(\frac{\sum_{i=1}^n (\xi_i - \hat{\mu})^2}{\varepsilon_2} < \sigma^2 < \frac{\sum_{i=1}^n (\xi_i - \hat{\mu})^2}{\varepsilon_1}\right) = \alpha_{n-1}(\varepsilon_1, \varepsilon_2),$$

где  $\alpha_{n-1}(\varepsilon_1, \varepsilon_2)$  вычисляется по распределению  $\chi_{n-1}^2$  с  $n - 1$  степенями свободы.

### Точечные оценки.

Лекция 12.

Пусть  $\{\xi_i\}$ ,  $i = 1, 2, \dots, n$  — независимая выборка из распределения  $P(x, \vartheta)$ , где  $\vartheta$  — неизвестный параметр. Нас интересует оценка величины  $\tau(\vartheta)$  (здесь  $\tau(\cdot)$  — известная функция), роль которой играет некоторая *статистика*  $t(\xi)$ .

Терминология:  $X$  — выборочное пространство,  $n$  — объем выборки, всякая измеримая функция от выборки называется статистикой, следовательно по определению любая точечная оценка — статистика.

Желательные свойства оценок:

1. Несмещенность  $Mt(\xi) = \tau(\vartheta)$ . (Гарантирует от накопления систематических ошибок).

2. Состоятельность  $t_n(\xi) \xrightarrow[n \rightarrow \infty]{P} \tau(\vartheta)$ .

3. Минимальность дисперсии (если оценка несмещенная) — качество оценки при фиксированном объеме выборки

Примеры.

Несмещенность  $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \xi_i$  очевидна. Состоятельность  $\hat{\mu}$  — утверждение З.Б.Ч.

Если  $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$ , то  $\hat{\sigma}^2 = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \hat{\mu})^2$  — несмещенная и состоятельная оценка. Действительно, при этом  $\hat{\sigma}^2 = \frac{1}{n-1} \sigma^2 \chi_{n-1}^2$ ,  $M\hat{\sigma}^2 = \sigma^2$ , а  $D\hat{\sigma}^2 = \frac{\sigma^4}{(n-1)^2} D\chi_{n-1}^2 = \frac{\sigma^4}{(n-1)^2} 2(n-1)$  и по неравенству Чебышёва  $P\{|\hat{\sigma}^2 - \sigma^2| > \varepsilon\} < \frac{2\sigma^4}{\varepsilon^2(n-1)} \xrightarrow[n \rightarrow \infty]{} 0$ . Если  $\xi$  не является нормальной, то несмещенность оценки сохраняется:

$$\begin{aligned} M \sum (\xi_i - \hat{\mu})^2 &= M \sum [(\xi_i - \mu)^2 - 2(\xi_i - \mu)(\hat{\mu} - \mu) + (\hat{\mu} - \mu)^2] = \\ &= M \left[ \sum (\xi_i - \mu)^2 - 2(\hat{\mu} - \mu) \sum (\xi_i - \mu) + \sum (\hat{\mu} - \mu)^2 \right] = \\ &= M \left[ \sum (\xi_i - \mu)^2 - 2 \frac{1}{n} \sum (\xi_i - \mu) \sum (\xi_i - \mu) + \frac{n}{n^2} \sum (\xi_i - \mu) \sum (\xi_i - \mu) \right] = \\ &= n\sigma^2 - 2\sigma^2 + \sigma^2 = (n-1)\sigma^2, \end{aligned}$$

а состоятельность — нет.

Минимальность дисперсии — желательное свойство, однако заметим, что смещение может уменьшить ср. кв. уклонение. Например, задача

$$M(k \sum (\xi_i - \hat{\mu})^2 - \sigma^2)^2 \sim \min_k$$

имеет решение для  $\xi_i \sim \mathcal{N}(\mu, \sigma^2)$  при  $k = \frac{1}{n+1}$ . (Доказать).

Рассмотрим специально несмещенные оценки минимальной дисперсии (НОМД).

**Лемма.** Если существует НОМД, то она единственна (с вер. 1).

Доказательство. Пусть  $t_1(\xi)$  и  $t_2(\xi)$  — НОМД, т.е.  $Mt_i(\xi) = \tau(\vartheta)$ ,  $Dt_i(\xi) = \delta$ .

Рассмотрим  $t_3 = \frac{1}{2}(t_1 + t_2)$ . Тогда  $Dt_3 =$

$$= \frac{1}{4}(Dt_1 + 2\text{cov}t_1t_2 + Dt_2) \leq \frac{1}{4}(Dt_1 + 2\sqrt{Dt_1Dt_2} + Dt_2) = \frac{1}{4}(\sqrt{\delta} + \sqrt{\delta})^2 = \delta,$$

но  $Dt_3 \geq \delta$ , т.е. возможно лишь равенство, откуда следует, что  $t_1(\xi) - \tau(\vartheta) = k(\vartheta)(t_1(\xi) - \tau(\vartheta))$  с вер. 1 ( $k^2 = 1$ ), и далее, из  $\text{cov}t_1t_2 = \delta$  получаем  $k = 1$ .  $\square$

*Определение.* Назовем статистику  $t(\xi)$ , удовлетворяющую условию  $M(t(\xi))^2 < +\infty$ , гильбертовой.

**Теорема.** Пусть  $\xi = (\xi_1, \dots, \xi_n)$  — выборка из распределения  $P(x, \vartheta)$ . Для того, чтобы гильбертова статистика  $t(\xi)$  была НОМД, необходимо и достаточно, чтобы для всякой центрированной гильбертовой статистики  $\eta = s(\xi)$  (такой, что  $M\eta = 0$ ), выполнялось  $Mt(\xi)\eta = 0$ .

**Доказательство.** Пусть  $t(\xi)$  — гильбертова несмещенная оценка  $\tau(\vartheta)$ . Тогда  $t(\xi) + \lambda\eta$  — тоже гильбертова несмещенная оценка  $\tau(\vartheta)$  для всех  $\lambda$ . Обозначим

$$\varphi_\lambda = \mathbf{M}(t(\xi) - \tau(\vartheta) + \lambda\eta)^2.$$

$$\min_{\lambda} \varphi_\lambda = \mathbf{M}(t - \tau)^2 - \frac{[\mathbf{M}t\eta]^2}{\mathbf{M}\eta^2} \quad \left( \text{при } \lambda = \lambda_* = -\frac{\mathbf{M}t\eta}{\mathbf{M}\eta^2} \right).$$

(Необходимость). Если последнее слагаемое не равно нулю, то существует несмещенная статистика с дисперсией, меньшей, чем у  $t(\xi)$ .

(Достаточность). Если последнее слагаемое при любых  $\eta$ ,  $\mathbf{M}\eta = 0$  равно нулю, то статистика  $t(\xi)$  имеет наименьшую дисперсию среди всех гильбертовых несмещенных оценок типа  $t(\xi) + \lambda\eta$ , а следовательно, всех гильбертовых несмещенных оценок, поскольку любую гильбертову несмещенную оценку можно представить в таком виде.  $\square$

Иногда качество оценки можно охарактеризовать, зная минимально возможное значение ее дисперсии (неравенство Рао-Крамера).

*Определение.* Функцией правдоподобия для некоторого распределения  $P(x, \vartheta)$  называется  $L(x, \vartheta) = f(x_1, \vartheta)f(x_2, \vartheta) \dots f(x_n, \vartheta)$ , где  $f(x_i, \vartheta)$  — либо плотность распределения  $p_\xi(x, \vartheta)$  сл. величины  $\xi$ , либо  $P_\vartheta\{\xi = x\}$ <sup>20</sup>.

**Теорема Рао-Крамера.** Пусть  $L(x, \vartheta)$  — функция правдоподобия,  $\vartheta \in R_1$  и выполнены условия:

1.  $t(\xi)$  — несмещенная оценка  $\tau(\vartheta)$ .
2. Функции  $L(x, \vartheta)$  и  $\tau(\vartheta)$  дифференцируемы по  $\vartheta$ .
3. Множество тех  $x$ , для которых  $L(x, \vartheta) > 0$  не зависит от  $\vartheta$  и

$$\frac{d}{d\vartheta} \int L(x, \vartheta) dx = \int \frac{\partial}{\partial \vartheta} L(x, \vartheta) dx$$

и

$$\frac{d}{d\vartheta} \int t(x)L(x, \vartheta) dx = \int t(x) \frac{\partial}{\partial \vartheta} L(x, \vartheta) dx.$$

Тогда

$$\mathbf{D}t(\xi) \geq \frac{|\tau'(\vartheta)|^2}{\mathbf{M} \left[ \frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} \right]^2}, \quad (18)$$

причем знак равенства имеет место тогда и только тогда, когда

$$\frac{\partial \ln L(x, \vartheta)}{\partial \vartheta} = a(\vartheta)[t(\xi) - \tau(\vartheta)] \quad (19)$$

с вероятностью единица для некоторого  $a(\vartheta)$ .

**Доказательство.** Дифференцируя тождества  $\int L(x, \vartheta) dx = 1$  и  $\int t(x)L(x, \vartheta) dx = \tau(\vartheta)$ , получим

$$\int \frac{d \ln L(x, \vartheta)}{d\vartheta} L(x, \vartheta) dx = 0, \quad \int t(x) \frac{d \ln L(x, \vartheta)}{d\vartheta} L(x, \vartheta) dx = \tau'(\vartheta)$$

<sup>20</sup> Обычно функция правдоподобия рассматривается как функция от  $\vartheta$ , а значения  $x_1, x_2, \dots, x_n$  (выборка) — параметры.

или

$$\int [t(x) - \tau(\vartheta)] \frac{d \ln L(x, \vartheta)}{d\vartheta} L(x, \vartheta) dx = \tau'(\vartheta)$$

и отсюда по неравенству Коши-Буняковского получаем (18).  $\square$

Если в условиях теоремы Рао-Крамера имеет место равенство, то справедливо (19) с вероятностью единица. В этом случае оценка называется эффективной и ее дисперсия равна  $Dt(\xi) = \frac{|\tau'(\vartheta)|}{|a'(\vartheta)|}$ .

К таким оценкам, например, приводят распределения, плотности которых можно представить в виде

$$f(x, \vartheta) = \exp\{a(\vartheta)b(x) + c(\vartheta) + d(x)\}, \quad x \in R_1, \quad \vartheta \in R_1.$$

(Они называются экспоненциальными семействами.) Тогда

$$L(x, \vartheta) = \exp\{a(\vartheta) \sum b(x_i) + nc(\vartheta) + \sum d(x_i)\}$$

и

$$\frac{\partial \ln(\xi, \vartheta)}{\partial \vartheta} = a'(\vartheta)n \left\{ \frac{1}{n} \sum b(x_i) + \frac{c'(\vartheta)}{a'(\vartheta)} \right\}.$$

Пусть условия теоремы Рао-Крамера выполнены, тогда  $t(\xi) = \frac{1}{n} \sum b(x_i)$  есть эффективная оценка  $\tau(\vartheta) = -\frac{c'(\vartheta)}{a'(\vartheta)}$  с дисперсией

$$\left| \frac{\tau'(\vartheta)}{na'(\vartheta)} \right|.$$

Экспоненциальному семейству принадлежат многие важные для практики распределения: нормальное, Пуассона, Бернулли (биномиальное), гамма-распределение и другие.

К сожалению, класс эффективных оценок весьма узок: если такая оценка существует для функции  $\tau(\vartheta)$ , то она не существует ни для какой функции, отличной от  $c_1\tau(\vartheta) + c_2$ .

**Оценки максимального правдоподобия.** Значение параметра  $\vartheta = \hat{\vartheta}$ , при котором функция правдоподобия имеет максимум, называется оценкой максимального правдоподобия. Она, к сожалению, не связана с каким-либо принципом оптимальности и не обладает, например, свойством несмещенности, однако легко находится и имеет хорошие асимптотические свойства (состоятельность).

### Теорема Гаусса-Маркова.

Лекция 13.

Предположим, что наблюдению доступны лишь линейные комбинации неизвестных величин (наблюдения *косвенные*)

$$\xi_i = \sum_{j=1}^k a_{ij}\alpha_j + \nu_i, \quad i = 1, 2, \dots, n. \quad (20)$$

Пусть  $\nu_i$  — независимые сл. величины,  $M\nu_i = 0$ ,  $D\nu = \sigma^2$ ,  $i = 1, 2, \dots, n$ .

Требуется оценить  $\alpha_j$ ,  $j = 1, 2, \dots, k$ , точнее, найти линейные несмещенные оценки  $\hat{\alpha}_j$  с минимальной дисперсией.

Запишем (20) в виде

$$\xi = \sum_{j=1}^k a_j \alpha_j + \nu, \quad (21)$$

где  $a_j = (a_{1j}, a_{2j}, \dots, a_{nj})^*$ ,  $\xi, \nu, a_j \in \mathcal{R}_n$ ,  $n \geq k$ , векторы-столбцы  $a_j$  линейно независимы, или в виде

$$\xi = A\alpha + \nu, \quad \alpha \in \mathcal{R}_k, \quad (22)$$

причем  $M\nu = 0$ ,  $M\nu\nu^* = \sigma^2 I$ .

1. (Линейность) Будем искать оценку  $\hat{\alpha}_j$  в виде  $\hat{\alpha}_j = \sum_{i=1}^n b_{ji} \xi_i = (b_j, \xi)$ ,  $b_j = (b_{j1}, b_{j2}, \dots, b_{jn})^*$ .

2. Требование несмещенности дает:

$$M\hat{\alpha}_j = \sum_{i=1}^n b_{ji} \sum_{s=1}^k a_{is} \alpha_s = \sum_{s=1}^k \left( \sum_{i=1}^n b_{ji} a_{is} \right) \alpha_s = \alpha_j. \quad (23)$$

Отсюда  $(b_j, a_s) = \delta_{js}$ ,  $j, s = 1, 2, \dots, k$ .

3. Вычислим дисперсию  $D\hat{\alpha}_j = D \sum_{i=1}^n b_{ji} \xi_i = \sigma^2 \sum_{i=1}^n b_{ji}^2 = \sigma^2 \|b_j\|^2$ .

Требование минимальности дисперсии приводит к следующей задаче на условный экстремум:

Найти  $\min \|b_j\|^2$  при условии  $(b_j, a_s) = \delta_{js}$ ,  $j, s = 1, 2, \dots, k$ . Воспользуемся методом множителей Лагранжа<sup>21</sup>: введем функцию Лагранжа

$$\mathcal{L} = \|b_j\|^2 - 2 \sum_{s=1}^k \lambda_{js} (b_j, a_s) \quad (24)$$

и, дифференцируя по  $b_{ji}$ , получаем  $2b_{ji} - 2 \sum_{s=1}^k \lambda_{js} a_{si}$  или  $b_j = \sum_{s=1}^k \lambda_{js} a_s$ . Испол-

зуем условие несмещенности:  $(b_j, a_s) = \sum_{p=1}^k \lambda_{jp} (a_s, a_p) = \delta_{jp}$ .

Отсюда  $\lambda_{js} = (a_j, a_s)^-$  и окончательно<sup>22</sup>

$$\hat{\alpha}_j = \sum_{s=1}^k (a_j, a_s)^- (a_s, \xi).$$

Поскольку в векторно-матричной форме  $(a_j, a_s) = (A^* A)_{js}$ , то

$$\hat{\alpha} = (A^* A)^{-1} A^* \xi. \quad (25)$$

<sup>21</sup>Для нахождения минимума  $\varphi(x)$  при условиях  $g_i(x) = 0$ ,  $i = 1, 2, \dots, m$ , нужно, чтобы градиент  $\text{grad } \varphi(x)$  был ортогонален всем поверхностям  $g_i(x) = 0$ ,  $i = 1, 2, \dots, m$ , т.е. градиент  $\text{grad } \varphi(x)$  может быть разложен по векторам  $\text{grad } g_i(x)$ ,  $i = 1, 2, \dots, m$ :  $\text{grad}(\varphi(x) - \sum_{i=1}^m \lambda_i g_i(x)) = 0$  при некоторых  $\lambda_i$ . Выражение в скобках — так называемая функция Лагранжа.

<sup>22</sup>Знак  $-$  говорит о том, что берется соответствующий элемент матрицы, обратной к матрице  $\|(a_j, a_s)\|$ .

Найдем матрицу ковариаций  $\widehat{\alpha}$ . Поскольку

$$\begin{aligned}\widehat{\alpha} - \alpha &= (A^*A)^{-1}A^*\xi - \alpha = (A^*A)^{-1}A^*(A\alpha + \nu) - \alpha = \\ &= (A^*A)^{-1}A^*A\alpha + (A^*A)^{-1}A^*\nu - \alpha = (A^*A)^{-1}A^*\nu,\end{aligned}$$

то

$$\mathbf{M}(\widehat{\alpha} - \alpha)(\widehat{\alpha} - \alpha)^* = \mathbf{M}(A^*A)^{-1}A^*\nu\nu^*A(A^*A)^{-1} = \sigma^2(A^*A)^{-1}. \quad (26)$$

Рассмотрим *метод наименьших квадратов*. Пусть  $\widetilde{\alpha}$  выбираются из условия<sup>23</sup>

$$\sum_{i=1}^n (\xi_i - \sum_{j=1}^k a_{ij}\alpha_j)^2 \sim \min_{\alpha_j}.$$

Дифференцируя по  $\alpha_s$ , получим

$$2 \sum_{i=1}^n \sum_{i=1}^n (\xi_i - \sum_{j=1}^k a_{ij}\widetilde{\alpha}_j) a_{is} = 0 \Rightarrow \sum_{i=1}^n \sum_{j=1}^k a_{ij} a_{is} \widetilde{\alpha}_j = \sum_{i=1}^n a_{is} \xi_i. \quad (27)$$

Отсюда получаем

$$\widetilde{\alpha} = (A^*A)^{-1}A^*\xi,$$

т.е. ту же оценку, что и  $\widehat{\alpha}$ .

Таким образом, справедлива **Теорема Гаусса-Маркова**:

Пусть  $\xi$  измеряется по схеме (20). Тогда ЛНОМД дается формулой (25), а матрица ковариаций — формулой (26).  $\square$

Как оценить  $\sigma^2$ ?

Заметим, что из (27) следует  $(\xi - A\widehat{\alpha})a_i = 0$ ,  $i = 1, 2, \dots, k$ , т.е.

$$(I - A(A^*A)^{-1}A^*)\xi \perp \mathcal{L}(a_1, \dots, a_k) = (I - \Pi_a)\xi.$$

Таким образом,  $\Pi_a = A(A^*A)^{-1}A^*$  — ортогональный проектор на  $\mathcal{L}(a_1, \dots, a_k)$  (это можно проверить непосредственно).

Пусть  $k < n$ . Обозначим

$$\begin{aligned}s^2 &= \|\xi - A\alpha\|^2 = \|\nu\|^2, \\ s_1^2 &= \|\xi - \Pi_a\xi\|^2 = \|\xi - \Pi_a(A\alpha + \nu)\|^2 = \|(I - \Pi_a)\nu\|^2, \\ s_2^2 &= \|\Pi_a\xi - A\alpha\|^2 = \|\Pi_a(\xi - A\alpha)\|^2 = \|\Pi_a\nu\|^2.\end{aligned}$$

Далее,  $\mathbf{M}s^2 = \text{tr } \sigma^2 I = n\sigma^2$ ,  $\mathbf{M}s_1^2 = \sigma^2 \text{tr}(I - \Pi_a) = \sigma^2(n - k)$ .

Отсюда  $\widehat{\sigma}^2 = \frac{1}{n-k}s_1^2 = \frac{1}{n-k}\|\xi - \Pi_a\xi\|^2$  — несмещенная оценка  $\sigma^2$ .

**Доверительные множества в нормальной регрессии.**

Доверительные множества — аналог интервальных оценок.

---

<sup>23</sup>Здесь не делается никаких предположений о  $\xi_i$ ,  $i = 1, 2, \dots, n$ .

Пусть  $\nu \sim \mathcal{N}(0, \sigma^2 I)$ . Тогда  $s_2^2 = \|\Pi_a \xi - A\alpha\|^2 = \|\Pi_a \nu\|^2 = \sigma^2 \chi_k^2$ ,  $s_1^2 = \|\xi - \Pi_a \xi\|^2 = \|(I - \Pi_a)\nu\|^2 = \sigma^2 \chi_{n-k}^2$  и независимы, поэтому

$$F_{k,n-k} = \frac{\frac{1}{k} \chi_k^2}{\frac{1}{n-k} \chi_{n-k}^2}.$$

Пусть  $P\{F_{k,n-k} \leq \varepsilon\} = \gamma_F(\varepsilon)$ , тогда с вероятностью  $\gamma_F(\varepsilon)$

$$\|A(\alpha - \hat{\alpha})\|^2 = (A^* A(\alpha - \hat{\alpha}), (\alpha - \hat{\alpha})) \leq \varepsilon \frac{k}{n-k} \|(I - \Pi_a)\xi\|^2. \quad (28)$$

Левая часть неравенства (28) представляет собой квадратичную форму относительно координат  $\alpha$  с матрицей  $A^* A > 0$ , поэтому (28) определяет в координатах  $\alpha_j$  эллипсоид (доверительный эллипсоид Хотеллинга).

Если нам нужно оценить одну координату  $\alpha_j$ , то вспомним, что ее дисперсия равна  $\sigma^2(a_j, a_j)^-$ , поэтому  $\frac{\alpha_j - \hat{\alpha}_j}{\sqrt{\sigma^2(a_j, a_j)^-}} \sim \mathcal{N}(0, 1)$ , а

$$\frac{\alpha_j - \hat{\alpha}_j}{\sqrt{(a_j, a_j)^- \frac{1}{n-k} \|(I - \Pi_a)\xi\|^2}} = t_{n-k},$$

и если  $P\{|t_{n-k}| < \varepsilon\} = \gamma_t(\varepsilon)$ , то с вероятностью  $\gamma_t(\varepsilon)$  неравенство

$$|\alpha_j - \hat{\alpha}_j| \leq \varepsilon \sqrt{\frac{(a_j, a_j)^- \|(I - \Pi_a)\xi\|^2}{n-k}}$$

дает интервальную оценку  $\alpha_j$ .

### Задачи редукции измерений.

Лекция 14.

1° Постановка задачи несмещенной редукции измерений.

Для схемы измерений  $\xi = Af + \nu$ ,  $M\nu = 0$ ,  $M\nu\nu^* = \sigma^2 I$  ставится **задача несмещенной редукции**:

$$\inf\{M\|R\xi - f\|^2 \mid R, RA = I\} = \inf\{\sigma^2 \text{tr} RR^* \mid R, RA = I\} = h_0.$$

Решаем уравнение  $RA = I$ :  $R = R_0 + Y$ , где  $R_0 = (A^* A)^{-1} A^*$ , а  $Y$  — решение уравнения  $YA = 0 \Leftrightarrow Y\Pi_a = 0 \Leftrightarrow Y = Z(I - \Pi_a)$ ,  $\forall Z$ .

Т.о., общее решение  $R = (A^* A)^{-1} A^* Z(I - \Pi_a)$ .

В этом случае  $\sigma^2 \text{tr} RR^* = \text{tr}(A^* A)^{-1} + \text{tr} Z(I - \Pi_a)Z^*$  и  $\inf$  достигается на  $R = R_0 = (A^* A)^{-1} A^*$  и равен  $h_0$ . Очевидно, этот результат совпадает с результатом, полученным в теореме Гаусса-Маркова.

В этом случае  $R\xi = f + R\nu$ , где  $R\nu$  — шум, суммарная энергия которого равна  $h_0$ .

2° Задача редукции с ограничением на уровень шума. Часто шум, полученный при решении задачи несмещенной редукции, неприемлемо велик. Вспомним, что ошибка складывается из двух:

$$R\xi = f + (RA - I)f + R\nu.$$

Введем расстояние между матрицами<sup>24</sup> (операторами)  $A$  и  $B$ :  $\rho^2(A-B) = \text{tr}(A-B)(A-B)^*$ .

Поставим задачу **синтеза прибора с ограничением на уровень шума**:

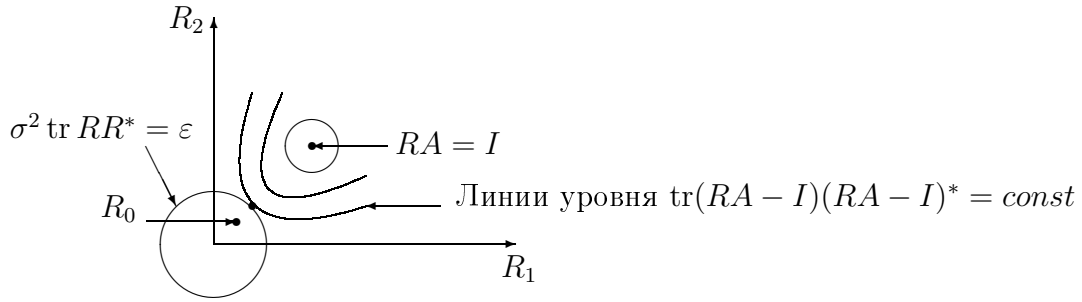
$$\inf\{\text{tr}(RA - I)(RA - I)^* \mid R, M \mid \|R\nu\|^2 \leq \varepsilon\}. \quad (29)$$

Заметим, что  $M\|R\nu\|^2 = \sigma^2 \text{tr} RR^* = \sigma^2 \sum_{i,j} r_{ij}^2$ .

Далее рассмотрим два случая.

1.  $h_0 < \varepsilon$ . В этом случае условие в задаче (29) выполняется и  $R_0 = (A^*A)^{-1}A^*$  — есть решение, так как любое  $R = (A^*A)^{-1}A^*Z(I - \Pi_a)$  минимизирует  $\|RA - I\|_2^2$ .

2. Введем систему координат в пространстве матричных элементов (изобразим лишь два!):



Очевидно, решение есть точка касания, где  $\sigma^2 \text{tr} RR^* = \varepsilon$  (равно!). Тогда решаем задачу методом множителей (одного!) Лагранжа. Функция Лагранжа

$$L(R) = \text{tr}(RA - I)(RA - I)^* + \omega \sigma^2 \text{tr} RR^*.$$

$$\nabla_R L = 2(RA - I)A^* + 2\omega \sigma^2 R = 0.$$

$R(AA^* + \omega \sigma^2 I) = A^*$ ,  $R = R(\omega) = A^*(AA^* + \omega \sigma^2 I)^{-1} = (A^*A + \omega \sigma^2 I)^{-1}A^*$ . Обозначим  $h = \sigma^2 \text{tr} R(\omega)R^*(\omega) = \sigma^2 \text{tr}(A^*A + \omega \sigma^2 I)^{-1}A^*A(A^*A + \omega \sigma^2 I)^{-1}$  и<sup>25</sup>  $g(\omega) = \text{tr}(RA - I)(RA - I)^* = \omega^2 \sigma^4 \text{tr}(A^*A + \omega \sigma^2 I)^{-2}$ .

Пусть далее  $\{e_i\}$  — ортонормированный базис из собственных векторов оператора  $A^*A$ :  $A^*Ae_i = \lambda_i e_i$ ,  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0$ <sup>26</sup>. Тогда

$$h = \sigma^2 \sum_{i=1}^m \frac{\lambda_i}{(\lambda_i + \omega \sigma^2)^2}, \quad g = \sigma^4 \sum_{i=1}^m \frac{\omega^2}{(\lambda_i + \omega \sigma^2)^2}.$$

Вычислим производные:  $\frac{dh}{d\omega} = -2\sigma^4 \sum_{i=1}^m \frac{\lambda_i}{(\lambda_i + \omega \sigma^2)^3} < 0$ ,  $0 < \omega < \infty$ ,  $h \xrightarrow{\omega \rightarrow 0} h_0 = \sigma^2 \sum_{i=1}^m \frac{1}{\lambda_i} = \sigma^2 \text{tr}(A^*A)^{-1}$ ,  $h \xrightarrow{\omega \rightarrow 0} 0$ . Поэтому уравнение  $h(\omega) = \varepsilon$  при  $\varepsilon < \varepsilon_0 = h_0$  имеет единственное решение.

<sup>24</sup>Можно также ввести скалярное произведение  $(AB)_2 = \text{tr} AB^*$  и норму  $\|A\|_2 = \{\text{tr} AA^*\}^{1/2}$ , которая называется нормой Гильберта-Шмидта.

<sup>25</sup>Здесь используется равенство  $I - RA = I - (A^*A + \omega \sigma^2 I)^{-1}A^*A = \omega \sigma^2 (A^*A + \omega \sigma^2 I)^{-1}$ .

<sup>26</sup> $\text{rang } A = m \leq n$ .

Кроме того,  $\frac{dg}{d\omega} = \sigma^4 \sum_{i=1}^m \left[ \frac{2\omega}{(\lambda_i + \omega\sigma^2)^2} - \frac{2\omega^2\sigma^2}{(\lambda_i + \omega\sigma^2)^3} \right] = 2\omega\sigma^4 \sum_{i=1}^m \frac{\lambda_i}{(\lambda_i + \omega\sigma^2)^3}$  и поэтому имеет место дифференциальный закон сохранения:

$$\omega \frac{dh}{d\omega} + \frac{dg}{d\omega} = 0. \quad (30)$$

Итак, общее решение задачи (29) имеет вид:

$$R = \begin{cases} R(\omega) = (A^*A + \omega\sigma^2 I)^{-1}, & 0 < \varepsilon < \varepsilon_0 = h_0 = \sigma^2 \operatorname{tr}(A^*A)^{-1}, \\ 0 & \varepsilon = 0, \\ R_0 = (A^*A)^{-1}A^*, & \varepsilon \geq \varepsilon_0 = h_0, \end{cases} \quad (31)$$

при этом выполняется (30).

Зависимость  $g$  от  $\varepsilon$  носит название оперативной характеристики. При этом характеристика, график которой лежит ниже, соответствует равномерно лучшему прибору.

