# Теория вероятностей и математическая статистика

Цель изложения



*Необходимо обеспечить переход от теории вероятностей к квантовой механике и распределениям вероятности. Производящие и характеристические функции (и еще факториальные моменты) нужны для введения в теорию операторов.*

*Все изложенные идеи - абстрактны. Ни один интеграл приведенный ниже считать не нужно. Весь этот материал-вводная часть к дискретной математике, которая следует из теории операторов в дискретном пространстве. Расчеты начинаются там, где вместо функций возникают вектора коэффициентов разложения и матрицы, а вместо уравнений в частных производных - полиномы и алгоритмы.*


- [1] Б.А. Севастьянов. Курс теории вероятностей и математической статистики.
- [2] А.Н. Колмогоров. Основные понятия теории вероятностей.
- [3] А.Н. Колмогоров, С.В. Фомин. Элементы теории функций и функционального анализа.

# Вероятностное пространство

Тройку $(\Omega, \mathcal{A}, \mathrm{P})$, где $\Omega$ — пространство элементарных событий, — σ-алгебра подмножеств $\Omega$, называемых событиями, P —
числовая функция, определенная на событиях и называемая вероятностью, будем называть вероятностным пространством, если выполнены следующие аксиомы:

1. $P(A) \geq 0$ для всех $A \in \mathcal{A}$ (неотрицательность P);
2. $P(\Omega) = 1$ (нормированность P);
3. $P(A + B) = P(A) + P(B)$, если $A \cap B = \emptyset$ (аддитивность P);
4. Если $A_n \downarrow \emptyset$, т. е. $A_1 \supseteq A_2 \supseteq \ldots$ и $\bigcap_{n=1}^{\infty} A_n = \emptyset$, то $\lim_{n\to\infty} P(A_n) = 0$ (непрерывность P).

Система аксиом 1◦, 2◦, 3◦, 4◦ определяет вероятностную меру на σ-алгебре пространства $\Omega$. Эта система аксиом предложена А. Н. Колмогоровым [2].

# Формула полной вероятности

Для любых событий $A$ и $B$:

$$P(AB) = P(A) + P(B) - P(A \cup B)\ .$$

Два события $A$ и $B$ *независимы* тогда и только тогда

$$P(AB) = P(A) + P(B)\ .$$

**Условная вероятность**. Пусть $P(B) > 0$. Условной вероятностью $P(A|B)$ события $A$ при условии, что произошло событие $B$ (или просто: при условии $B$), назовем отношение

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

Для условной вероятности $P(A|B)$ применяется также обозначение $P_B(A)$.

$$P(AB) = P(B)P(A|B)$$

Систему событий $A_1, A_2, \ldots, A_n$ будем называть конечным разбиением (в дальнейшем — просто разбиением), если они попарно несовместны и $A_1 + A_2 + \ldots + A_n = \Omega$ образуют полное пространство.

**Формула полной вероятности**. Если $A_1, \ldots, A_n$ — разбиение и все $P(A_k) > 0$, то для любого события $B$ имеет место формула

$$P(B)=\sum_{k=1}^{n}P(A_k)P(B|A_k)\ ,$$

называемая формулой полной вероятности.

**Формула Байеса**. Если выполнены условия (2): $A_1, ..., A_n$ — конечное разбиение и все $P(A_k)>0$ и $P(B)>0$, то имеют место формулы

$$P(A_k|B)=\frac{P(A_k)P(B|A_k)}{\sum\limits_{i=1}^{n}P(A_i)P(B|A_i)}\ ,$$

называемые формулами Байеса.

*из этих определений следует разложение в ряд по ортогональным функциям*

*Надо выстроить аналогию между условной вероятностью и линейным пространством функций, которая выражается через скалярное произведение и норму в гильбертовом пространстве.*

*Надо ввести аналогию между функцией Softmax из формулы Байеса. Откуда берется экспонента, из какого распределения плотности вероятности*

# Плотность вероятности

*в курсе теории функции вводится понятие $\sigma$-алгебра над обрезками абстрактного множества и алгебра (борелевская) над обрезками множества, к которому можно применить операцию сравнения (множество чисел). Теория вероятностей сопоставляет абстрактному множеству со своей нарезкой - вероятности. Множеству чисел, множеству векторов, множеству функций мы сопоставляем некоторую числовую меру - вероятность. Если множество упорядоченное, то ему можно сопоставить числовую функцию и говорить про функции распределения. Из двух положительных функций можно составить произвольную функцию. Из двух функций вещественных можно сделать комплексную. Из множества функций - вектор. Если функция распределения (вероятности) непрерывная и дифференцируемая, можно говорить про плотность (вероятности).*

Числовая функция $\xi=\xi(\omega)$ от элементарного события $\omega\in\Omega$ ($\xi:\Omega\mapsto\mathbb{R}$) называется *случайной величиной*, если для любого числа $x\in\mathbb{R}$

$$\{\xi \leqslant x\} = \{\omega : \xi(\omega) \leqslant x\} \in \mathcal{A}\,.$$

Функцию

$$F(x) = F_\xi(x) = P\{\xi \le x\},$$

определенную при всех $x \in \mathbb{R}$, назовем функцией распределения случайной величины $\xi$.

Для $x_1 < x_2$ можно записать событие $\{\xi \leqslant x_2\}$ как сумму интервалов (непересекающихся)

$$\{\xi \leqslant x_1\} + \{x_1 < \xi \leqslant x_2\}$$

откуда следует

$$P(\xi \leqslant x_2) = P(\xi \leqslant x_1) + P(x_1 < \xi \leqslant x_2)$$

$$P(x_1 < \xi \leqslant x_2) = F(x_2) - F(x_1)$$

$\sigma$-алгебра $\mathcal{B}$ числовых множеств, порожденная всевозможными интервалами вида $x_1 < x \leqslant x_2$, называется *борелевской*; множества B, входящие в $\mathcal{B}$, также называются *борелевскими*.

$\sigma$-алгебра борелевских множеств $\mathcal{B}$ содержит всевозможные интервалы вида

$$x_1 \le x \le x_2, \quad x_1 < x < x_2, \quad x_1 < x \le x_2, \quad x_1 \le x < x_2$$

с конечными и бесконечными концами, их конечные и счетные суммы, все открытые и замкнутые множества.

Для каждого $B \in \mathcal{B}$ определена вероятность $P\{\xi \in B\}$, которую мы будем обозначать $P_\xi(B)$.

Функция $P_\xi(B)$, определенная для всех $B \in \mathcal{B}$, называется распределением вероятностей случайной величины $\xi$.

- [3] Колмогоров, Фомин. *Гл.VI Пар. 6 Интеграл Стилтьеса.*
- [1] Севастьянов. § 27. Случайные величины и их распределения

В теории функции [3] вводится понятие меры Лебега и меры Стилтьеса на подмножествах отрезков на интервале [a,b]. Функция распределения монотонна и не убывает. Интеграл

строится как разность двух функций.
Интегралу Лебега по мере образованной функцией распределения $F(x)$

$$\int_a^b f(x) d\mu_F$$

сопоставляется интеграл в форме Лебега-Стилтьеса

$$\int_a^b f(x)\, dF(x) = \int_a^b f(x) F'(x)\, dx$$

Функция F'(x) будет называться *плотностью вероятности*.

Вместе с тем вводится понятие *линейный непрерывный функционал* и доказывается теорема Ф.Рисса. Всякий линейный непрерывный функционал на $S$ в пространстве $C[a, b]$ представим в виде

$$S(f) = \int_a^b f(x)\, dF(x)$$

Мы будем говорить, что функция $p(x) = p_\xi(x)$ есть плотность распределения случайной величины $\xi$, если для любых $x_1 < x_2$

$$P\{a < \xi < b\} = \int\limits_{x_1}^{x_2} p_\xi(x) dx\ .$$

Из определения следует

1. $F'_\xi(x) = p_\xi(x)$;
2. $F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(u)\, du$;
3. $F_\xi(x_2) - F_\xi(x_1) = \int_{x_1}^{x_2} p_\xi(u)\, du$ для $x_1 < x_2$.

*очень важный пункт 2. Я запишу в другом виде*

$$F_\xi(x) = P(\xi < x) = \int\limits_{-\infty}^{x} p_\xi(u)\, du = \int\limits_{-\infty}^{x} dF_\xi(u)$$

Если распределение имеет плотность $p_\xi(x)$, то мы будем говорить, что случайная величина $\xi$ имеет *абсолютно непрерывное распределение*. Через плотность $p_\xi(x)$ можно выразить любую вероятность $P\{\xi \in B\}$, если мы умеем вычислять интеграл по области $B$ в следующей форме:

$$P\{\xi \in B\} = \int_B p_\xi(x)dx \ .$$

Для множеств $B$, равных сумме интервалов, интеграл вычисляется обычным способом. Для того чтобы равенство имело смысл при любом борелевском множестве $B$, нам нужно обобщить понятие интеграла, перейдя от интеграла Римана к интегралу Лебега.

Плотность распределения $p(x)$ обладает следующими двумя свойствами:

$$p(x) \geq 0, \qquad \int\limits_{-\infty}^{\infty} p(x)dx = 1 \ .$$

# Математическое ожидание случайной величины

Математическое ожидание $\mathsf{E}\xi$ случайной величины $\xi = \xi(\omega)$, заданной на вероятностном пространстве $(\Omega, \mathcal{A}, \mathrm{P})$, определяется последовательно
сначала для целых случайных величин, затем для неотрицательных случайных величин и, наконец, в общем случае.

Ввести понятие функции можно через определение индикатора события A

$$I_A(\omega) = \begin{cases} 1, & \text{если} \quad \omega \in A \\ 0, & \text{если} \quad \omega \notin A \end{cases}$$

Случайную величину можно выразить с помощью индикаторов конечного разбиения на множестве … через сумму – разложение по индикаторам.

$$\xi(\omega) = \sum_{i=1}^{N} x_i I_{A_i}(\omega)$$

*Индикаторная функция тесно связана с понятием финитной функции и обобщенной функции. Так вводится понятие $\delta$-функции [3].*

Функцией распределения случайной величины $\xi$ будем называть вероятность $\mathbf{P}(\xi \leqslant x_i)$ того что величина примет значение не более $x_i$.

Математическое ожидание случайной величины при разложении по полному множеству независимых событий.

$$\mathsf{E}\xi = \sum_{i=1}^{N} x_i \mathrm{P}(A_i)$$

В теории вероятностей можно определить разложение по положительно определенным функциям.

В общем случае случайная величина однозначно представима в виде $\xi = \xi^+ - \xi^-$

$$\mathsf{E}\xi = \mathsf{E}\xi^+ - \mathsf{E}\xi^-$$

Можно определить мат.ожидание от комплексной величины $\zeta = \xi + i\eta$

$$\mathsf{E}\zeta = \mathsf{E}\xi + i\mathsf{E}\eta$$

Предел $\lim\limits_{n\to\infty} \to \mathsf{E}\xi_n$ обозначим как интеграл Лебега

$$\int_\Omega \xi(\omega) dP(\omega)$$

Функция распределения $F(x) = P(\xi < x)$. Функция F монотонно не убывает, непрерывная и удовлетворяет условиям $F(-\infty) = 0$, $F(+\infty) = 1$.

*Математическое ожидание* случайной величины (в форме интеграла Лебега-Стилтьеса):

$$\mathsf{E}\xi = \int\limits_{-\infty}^{\infty} x\, dF(x)\ .$$

*Дисперсия*:

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

Если случайная величина $\xi$ имеет плотность $p_\xi(x)$ то

$$\mathsf{E}\xi = \int\limits_{-\infty}^{\infty} x p_\xi(x) dx\ .$$

Если $\xi$ имеет плотность $p_\xi(x)$, функция $g(x)$ непрерывна, то

$$\mathsf{E}g(\xi) = \int\limits_{-\infty}^{\infty} g(x)p_\xi(x)dx\ .$$

# Формула композиции

Пусть $\xi$ и $\eta$ – независимые случайные величины, $p_\xi$ и $p_\eta$ – их плотности вероятности. Плотность их совместного распределения равна

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

Функция распределения суммы равна следующему интегралу

$$F_{\xi+\eta}(z) = P(\xi+\eta \leqslant z) = \int\limits_{x+y\leqslant z} p_\xi(x)p_\eta(y)\ dx\ dy$$

Интеграл можно выразить, как повторный (для непрерывных плотностей — это факт из анализа, — следствие теоремы Фубини, доказываемой в теории интеграла Лебега), поэтому

$$F_{\xi+\eta}(z) = \int\limits_{-\infty}^{+\infty} p_\xi(x)\ dx \int\limits_{-\infty}^{z-x} p_\eta(y)\ dy = \int\limits_{-\infty}^{+\infty} F_\eta(z-x)p_\xi(x)\ dx$$

$$= \int\limits_{-\infty}^{+\infty} p_\xi(x) \int\limits_{-\infty}^{z} p_\eta(y-x)\ dx\ dy = \int\limits_{-\infty}^{z} dy \int\limits_{-\infty}^{+\infty} p_\eta(y-x)p_\xi(x)\ dx$$

Формулы

$$F_{\xi+\eta}(z) = \int\limits_{-\infty}^{+\infty} F_\eta(z-x)p_\xi(x)\ dx \quad \text{– композиция}$$

$$p_{\xi+\eta}(y) = \int\limits_{-\infty}^{+\infty} p_\eta(y-x)p_\xi(x)\ dx \quad \text{– свертка}$$

**Компози́ция (суперпози́ция) фу́нкций** — это применение одной функции к результату другой. Суперпозиция это операция, которую мы хотим использовать для генерализации выражений связанных с разложениями.

Пусть даны две функции $f : X \to Y$ и $g : f[X] \to Z$, где $f[X] \subseteq Y$ — образ (отображение) множества $X$. Тогда их композицией называется функция $g \circ f : X \to Z$, определённая равенством: $(g \circ f)(x) = g(f(x)), x \in X$.
Это утверждение такое, что мы применяем некоторое обобщение операции сложения, когда работаем с более сложными объектами. В простейшем случае это может быть свертка с весовыми коэффициентами, в более сложном случае это более сложная функция, которую можно представить как *композицию* или *суперпозицию функций*.

Свойства композиции:

1. Композиция ассоциативна: $(h \circ g) \circ f = h \circ (g \circ f)$
2. Если $f(x) = I_X$ - тождество — тождественное отображение на $X$, то есть $\forall x \in X : f(x) = I_X(x)$, то $g \circ f = g$.
3. Если $g = I_Y$ — тождественное отображение на $Y$, то есть $\forall y \in Y : g(y) = I_Y(y) = y$, то $g \circ f = f$
4. Композиция отображений $f : X \to X, g : X \to X$, в общем случае не коммутативна, $f \circ g \neq g \circ f$. Для некоторого класса функций коммутативность существует.

- Свойства 2 и 3 выражаются через «identity» - инвариант множества, который обозначается как $I$.
- $\circ$ – операция разложения, которая соответствует функции от функции $f(g(x))$.

*Композиции функций мы хотим сопоставить ассоциативную алгебру. см. понятие линейность. Операцию умножения на скаляр. Композиция - линейна.*

# Евклидово пространство случайных величин

см. [1] Севастьянов § 16.
Векторное $n$-мерное пространство квадратно- суммируемых функций на интервале $L^2([0,1]^n)$ с определенными свойствами:

1. мера Лебега, заданной плотностью вероятности на интервале $\rho(x) > 0$.
2. Скалярное произведение векторов $(\xi, \eta) = \sum_\omega \xi(\omega)\eta(\omega)\rho(\omega) = \mathsf{E}\xi\eta$
3. Норма $\|\xi\| = \sqrt{(\xi, \xi)}$
4. Дистанция $d(\xi, \eta) = \sqrt{\mathsf{E}(\xi - \eta)^2} = \|\xi - \eta\|$

Механическое истолкование $\mathsf{E}(\xi)$ и $\mathsf{D}(\xi)$

$$\mathsf{E}\xi = \sum_{k=0}^{n} x_k p_k\ , \quad \mathsf{D}\xi = \sum_{k=0}^{n} (x_k - \mathsf{E}\xi)^2 p_k\ , \quad \mathrm{Cov}(\xi,\eta) = \mathsf{E}\{(\xi - \mathsf{E}\xi)(\eta - \mathsf{E}\eta)\}$$

-- центр масс и момент инерции (разброс масс) относительно центра. В трехмерном пространстве момент инерции — тензорная величина, представленная матрицей 3х3, собственным вектором и собственными значениями. Ковариация имеет смысл коэффициента корреляции — скалярного произведения.

*дан пример в дискретном пространстве. Тоже самое можно определить для непрерывных вероятностей.*

# Гильбертово пространство случайных величин

[4] Бахтин В. И., Пиндрик О. И. Геометрические методы в статистике

В математических классах вводится аксиоматика из теории функции [3] - пространство со счетно-аддитивной мерой. Теория вероятностей представлена, как вероятностное пространство со своей мерой. В курсе теории вероятностей показано, как вводится понятие меры и интеграла Лебега. Аналогичная связь (двойственность представления) существует между разделами математической статистики и теорией Гильбертова пространства.

*Вероятностное пространство можно образовать от любого множества с разбиением на непересекающиеся подмножества. Множеству можно сопоставить набор базисных функций. Гильбертово пространство можно образовать от набора ортогональных базисных функций.*

Например, условные математические ожидания, линейные регрессии, частные и множественные корреляции проще всего определять с помощью ортогональных проекций в гильбертовом пространстве случайных величин.

Вероятностным пространством называется тройка $(\Omega, \mathcal{A}, \mathbf{P})$, в которой
$\Omega$ — некоторое множество (элементы которого называются элементарными событиями),
$\mathcal{A}$ — некоторая σ-алгебра подмножеств $\Omega$, а $\mathbf{P}$ — вероятностная мера на этой σ-алгебре. (то есть такая неотрицательная мера, что $\mathbf{P}(\Omega) = 1$).

Математическим ожиданием $\mathsf{E}$ и дисперсией $\mathsf{D}$ скалярной случайной величины $\xi$ называются соответственно интегралы

$$\mathsf{E}\xi = \int_\Omega \xi dP \quad \text{и} \quad \mathsf{D}\xi = \mathsf{E}\{(\xi - \mathsf{E}\xi)^2\} = \int_\Omega (\xi - \mathsf{E}\xi)^2 dP$$

Если случайные величины $\xi$ и $\eta$ имеют конечные дисперсии, то их ковариацией называется число

$$\mathrm{Cov}\{\xi,\eta\} = \mathsf{E}\{(\xi - \mathsf{E}\xi)(\eta - \mathsf{E}\eta)\} = \mathsf{E}\xi\eta - \mathsf{E}\xi \cdot \mathsf{E}\eta\ ,$$

а корреляцией

$$\mathrm{Corr}\{\xi,\eta\} = \frac{\mathrm{Cov}\{\xi,\eta\}}{\sqrt{\mathsf{D}\xi \cdot \mathsf{D}\eta}}\ .$$

Обозначим через $H$ множество всех случайных величин $\xi$, для которых $\mathsf{E}\{\xi^2\} < \infty$.

Определим скалярное произведение любых двух случайных величин $\xi, \eta \in H$ формулой

$$(\xi,\eta) = \mathsf{E}\xi\eta = \int_\Omega \xi\eta dP$$

Скалярное произведение порождает норму

$$\|\xi\| = \sqrt{(\xi,\xi)}, \text{или} \|\xi\| = \int_\Omega \xi^2 dP\ .$$

По определению, H является гильбертовым пространством, оно совпадает с $L^2(\Omega, \mathcal{A}, \mathbf{P})$ - линейное квадратично суммируемое пространство.

Для элементов этого гильбертова пространства математическое ожидание, дисперсия, ковариация и корреляция представляются через скалярное произведение в виде:

$$\mathsf{E}\xi = (\xi, 1)\ , \quad \mathsf{D}\xi = \|\xi - \mathsf{E}\xi\|^2\ ,$$

$$\mathrm{Cov}\{\xi,\eta\} = (\xi - \mathsf{E}\xi, \eta - \mathsf{E}\eta)\ ,$$

$$\mathrm{Corr}\{\xi,\eta\} = \frac{(\xi - \mathsf{E}\xi, \eta - \mathsf{E}\eta)}{\|\xi - \mathsf{E}\xi\| \cdot \|\eta - \mathsf{E}\eta\|}\ .$$

Корреляция может интерпретироваться как косинус угла между векторами $\xi - \mathsf{E}\xi \in H$ и $\eta - \mathsf{E}\eta \in H$. Если случайные величины не коррелируют то центрированные величины $\xi - \mathsf{E}\xi$ и

$\eta - \mathsf{E}\eta$ ортогональны.

# Разложение в ряд по ортогональным функциям

В случае сепарабельных гильбертовых пространств полная ортонормированная система $\{e_k\}$ является базисом. Для каждого элемента $f$ пространства $H$ имеет место разложение по ортонормированному базису $\{e_k\}$:

$$f = \sum_{k=1}^{\infty} \alpha_k e_k = \sum_{k=1}^{\infty} (f, e_k) e_k \ .$$

Коэффициенты разложения в ряд $\alpha_k = (f, e_k)$ называют коэффициентами Фурье. Норма определенная через скалярное произведение будет давать равенство

$$\|f\|^2 = \sum_{k=1}^{\infty} |(f, e_k)|^2 \ .$$

Любые два гильбертовы пространства, имеющие одинаковую размерность, изоморфны. В частности, любые два бесконечномерные сепарабельные гильбертовы пространства изоморфны друг другу и пространству квадратично-суммируемых последовательностей $\ell^2$.

Квантовая механика первоначально была разработана в виде двух эквивалентных теорий: матричной механики Гейзенберга, использующей пространство $\ell^2$, и волновой механики Шрёдингера, использующей изоморфное ему гильбертово пространство $L^2$ [^3]

# Пространство суммируемых функций

[3]: Колмогоров, Фомин. *Гл.VII Пространства суммируемых функций. Пар. 2 Пространство* $L_2$

*Определение #1*. Функция $f$ называется *функцией с интегрируемым квадратом* на X, если интеграл

$$\int f^2(x)\, d\mu$$

существует (конечен). Совокупность всех таких функций обозначим $L_2(X, \mu)$.

*Определение #2*. Евклидовым пространством $L_2$ называется линейное пространство, состоящее из классов эквивалентных между собой функций с интегрируемым квадратом, в котором скалярное произведение определяется формулой

$$(f, g) = \int f(x)g(x)d\mu$$

Норма и расстояние определяются через скалярное произведение

$$\|f\| = \sqrt{(f, f)}\,, \quad \rho(f, g) = \|f - g\|\,.$$

Согласно теореме Рисса всякий линейный функционал в гильбертовом пространстве $H$ записывается в виде скалярного произведения $F(h) = (h, a)$, где a - фиксированный вектор из $H$. Поэтому всякий линейный функционал в $L_2$ имеет вид

$$F(f) = \int f(x)g(x)\,d\mu\,,$$

где g - фиксированная функция с интегрируемым квадратом.

Комплексная функция $f$, определенная на некотором пространстве $X$ с заданной на нем мерой $\mu$, называется функцией с интегрируемым квадратом, если интеграл

$$\int_X |f(x)|^2\,d\mu$$

существует(конечен).

Определив обычное сложение функций, умножение на скаляр и введя скалярное произведение по формуле

$$(f, g) = \int_X f(x)\overline{g(x)}d\mu$$

определяем комплексное пространство $L_2$

# Ортогональные системы функций в $L_2$

Если в $L_2$ выбрана некоторая полная ортогональная система $\{\psi_n\}$, то каждый элемент $f \in L_2$ можно представить как сумму ряда

$$f = \sum_{n=1}^{\infty} c_n \psi_n$$

-- ряда Фурье функции $f$ по ортогональной системе $\{\psi_n\}$. При этом коэффициенты $c_n$ определяются формулами

$$c_n = \frac{1}{\|\psi_n\|^2} \int f(x)\psi_n(x)\, d\mu\ , \quad \|\psi_n\|^2 = \int \psi_n^2(x)\, d\mu\ .$$

Далее следует рассмотреть ортогональные системы полиномов... одна из таких систем базисные полиномы Бернштейна рождаются из биномиального распределения.

# Биномиальное распределение

**Схема Бернулли**. Частный случай независимых испытаний.

Пусть некоторый исход $A$, который мы будем называть успехом, может произойти при каждом испытании с одной и той же вероятностью $p$; противоположный исход $\bar{A}$ (неуспех) может произойти при каждом испытании с дополнительной вероятностью $q = 1 - p$. В элементарном событии $\omega = (\omega_1, ..., \omega_n)$ имеем $\omega_i = 1$, если при i-м испытании произошел успех, и $\omega_i = 0$ в противоположном случае.

Обозначим $Bk = \{\omega : \omega_1 + ... + \omega_n = k\}$ событие, состоящее в том, что при $n$ независимых испытаниях в схеме Бернулли произошло ровно $k$ успехов. Поскольку при $\omega \in B_k : p(\omega) = p^k q^{n-k}$, то

$$P(B_k) = p^k q^{n-k} \times (\text{число элементарных событий } \omega \in B_k)$$

$$P(B_k) = C_n^k p^k q^{n-k}, \quad k = 0, 1, ..., n\ .$$

**Теорема Пуассона**. Если $n \to \infty$ и $p \to 0$ так, что $np \to a$, то для любого фиксированного $m = 0, 1, ...$

$$P\{\mu = m\} = C_n^m p^m q^{n-m} \to \frac{a^m}{m!} e^{-a} .$$

Биномиальное распределение числа успехов $\mu$ при $n$ независимых испытаниях в схеме Бернулли с вероятностью успеха $p$ в каждом испытании
задается вероятностями
$P\{\mu = m\} = C_n^m p^m q^{n-m}, m = 0, 1, ..., n, q = 1 - p$ .
Биномиальное распределение случайной величины $\mu$ имеет $\mathsf{E}\mu = np$
и $\mathsf{D}\mu = npq$

**Локальная предельная теорема Муавра–Лапласа**. Если в схеме Бернулли $\sigma = \sqrt{npq} \to \infty$, то для любого $C > 0$ равномерно по всем $|x| \leq C$ вида

$$x = \frac{m - np}{\sigma} \ ,$$

где $m$ — целые неотрицательные числа,

$$P\{\frac{\mu - np}{\sqrt{npq}} = x\} \to \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2}$$

**Интегральная предельная теорема Муавра–Лапласа**. При $\sigma = \sqrt{npq} \to \infty$ равномерно по $-\infty \leqslant a < b \leqslant \infty$

$$P\{a \leqslant \frac{\mu - np}{\sqrt{npq}} \leqslant b\} \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

Приближение, получаемое с помощью теорем Муавра–Лапласа, называется *нормальным распределением*.

# Производящие функции последовательности

*{смысл этого раздела в том чтобы через вероятности ввести S-пространство Лапласа}*
Сложно найти эту связь, но изначально под производящей функцией понимали функцию представимую в форме интеграла Лапласа-Стилтьеса

$$f(x) = \int_0^\infty e^{-xt} \, d\mu(t)$$

В непрерывном распределении, если функцию $\mu$ возможно представить в виде интеграла

$$\mu(t) = \int\limits_0^t \phi(u)\, du \ ,$$

тогда интеграл примет вид

$$f(x) = \int\limits_0^\infty e^{-xt} \phi(t)\, d(t)$$

Или в дискретном виде (дискретное преобразование Лапласа), если функция $\mu$ является ступенчатой функцией со ступеньками в точках $\lambda_k$.

Обозначим прямое преобразование Лапласа функции $f(t)$ (переменной $t$) как функцию новой переменной $s$ как

$$\mathcal{L}\left\{f(t)\right\}(s) = \int\limits_0^\infty e^{-st} f(t)\, dt \ .$$

где s — комплексный параметр частотной области
$s = \sigma + i\omega$ с действительными числами $\sigma$ и $\omega$.

Преобразование Лапласа конечной борелевской меры $\mu$ можно определить с помощью интеграла Лебега:

$$\mathcal{L}\{\mu\}(s) = \int_{[0,\infty)} e^{-st}\, d\mu(t) \ .$$

Важный особый случай — когда μ — вероятностная мера, например, дельта-функция Дирака. В операторном исчислении меры преобразование Лапласа рассматривается так, как если бы мера давала функцию плотности вероятности $f$. Важно отметить что, в случае когда масса сосредоточена в точке 0 при использовании $\delta$- функции, функция полностью входит под интеграл.

В теории вероятностей преобразование Лапласа определяется как математическое ожидание. Если $\xi$ — случайная величина с функцией плотности вероятности $f$, то преобразование Лапласа $f$ задается математическим ожиданием

$\mathcal{L}\{f\}(s) = \mathsf{E}\left[e^{-s\xi}\right]$, где
$\mathsf{E}[r]$ - математическое ожидание случайной величины $r$.

Замена $s$ на $-t$ дает производящую функцию функцию моментов $\xi$.
$\mathsf{M}_\xi(t) = \mathcal{L}\{f_\xi\}(-t)$.

$Z$-преобразование — это преобразование Лапласа решётчатой функции, производимое с помощью замены переменных:
$z \equiv e^{sT}$, где $T = 1/f_s$ — период дискретизации, а $f_s$ — частота дискретизации сигнала.

Пусть $\Delta_T(t) \stackrel{\text{def}}{=} \sum_{n=0}^{\infty} \delta(t - nT)$ будет последовательность выделяющая дискретные значения в определенные моменты времени (решетчатая функция Дирака) и

$$\begin{aligned} x_q(t) &\stackrel{\text{def}}{=} x(t)\Delta_T(t) = x(t)\sum_{n=0}^{\infty} \delta(t - nT) \\ &= \sum_{n=0}^{\infty} x(nT)\delta(t - nT) = \sum_{n=0}^{\infty} x[n]\delta(t - nT) \end{aligned}$$

получаем дискретное представление непрерывной функции $x(t)$.
$x[n] \stackrel{\text{def}}{=} x(nT)$.

Преобразование Лапласа от дискретного сигнала $x_q(t)$, выделенного решетчатой функцией запишется:

$$\begin{aligned} X_q(s) &= \int_{0^-}^{\infty} x_q(t)e^{-st}\,dt \\ &= \int_{0^-}^{\infty} \sum_{n=0}^{\infty} x[n]\delta(t - nT)e^{-st}\,dt \\ &= \sum_{n=0}^{\infty} x[n] \int_{0^-}^{\infty} \delta(t - nT)e^{-st}\,dt \\ &= \sum_{n=0}^{\infty} x[n]e^{-nsT}\,. \end{aligned}$$

Определение Z-преобразования от дискретной функции x[n] с учетом подстановки $z \to e^{sT}$:

$$X(z) = \sum_{n=0}^{\infty} x[n]z^{-n}$$

Преобразование Лапласа образует линейное пространство операторов.

$$\mathcal{L}\{\alpha f(t) + \beta g(t)\} = \alpha\mathcal{L}\{f(t)\} + \beta\mathcal{L}\{g(t)\}$$

В пространстве линейных операторов можно определить вторую операцию - композиция функций (суперпозиция).
Свертке оригиналов соответствует произведение изображений

$$\mathcal{L}\{f(t) \circ g(t)\} = F(s)G(s).$$

Для аналитического решения линейных систем дифференциальных уравнений используется операционное исчисление, которое сопоставляет интегралам и производным рациональные функции (т.н. передаточные функции).

Дискретное пространство получается при использовании дискретного преобразования Лапласа. Производящие функции позволяют напрямую получить переход в дискретное Z-пространство. Производящие функции (формальные степенные ряды) и выражения с использованием рациональных функций дают простой алгебраический путь в дискретное Z-пространство.

[1] Б.А. Севастьянов Курс теории вероятностей и математической статистики. Гл.8 Пар. 32

*Определение*. Дискретную случайную величину $\xi$ принимающую только целые не отрицательные значения будем называть целочисленной случайной величиной. Закон распределения целочисленной случайной величины определяется элементарными вероятностями $p_k = P(\xi = k), k = 0, 1 \ldots$ для которых

$$\sum_{k=0}^{\infty} p_k = 1 \ .$$

Распределение случайной величины удобно изучать с помощью производящей функции последовательности $\{p_k\}$, как суммы ряда

$$f_\xi(s) = \sum_{k=0}^{\infty} p_k s^k \ ,$$

которая сходится при $|s| \leqslant 1$. s определено на множестве $\mathbb{C}$ или $\mathbb{R}$.

*Определение*. Пусть $a_0, a_1, a_2, \ldots$ — произвольная (бесконечная) последовательность чисел. Производящей функцией для этой последовательности будем называть выражение вида

$$a_0 + a_1 s + a_2 s^2 + \ldots$$

или

$$\sum_{k=0}^{\infty} a_k s^k \ .$$

Если все члены последовательности, начиная с некоторого $n$, равны нулю, то производящая функция является производящим многочленом.

*Замечение*. Употребляя слово «функция», мы вовсе не имеем в виду, что написанное выражение действительно является функцией. Так, не следует думать, будто мы можем сказать, чему равно «значение $A(s_0)$ производящей функции A в точке $s_0$». Переменная s является формальной, и сумма ряда $a_0 + a_1 s_0 + a_2 s_0^2 + ...$ смысла не имеет.

*Следует понимать производящую функцию, как разложение по степеням (оператора) формальной переменной s. Операции с производящими функциями такие же как с полиномами, мы работаем с коэффициентами разложения. Т.е. функция вычисления значения ряда не используется.*

Производящая функция определяется через математическое ожидание

$$f_\xi(s) = \mathsf{E} s^\xi$$

Производящей функцией любой числовой последовательности является сумма степенного ряда.

$$f(x) = \sum_{n=0}^{+\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

-- разложение в ряд Тейлора позволяет говорить о явных формулах для вычисления коэффициентов разложения для аналитических функций распределения. Однако, в контексте производящих функций мы можем говорить о разложении в начальной или конечной точке интервала.

Можно установить соответствие между функцией распределения и числовой последовательностью

$$p_n = \frac{1}{n!} f^{(n)}(0) \ , \quad n = 0, 1, 2, \ldots ,$$

**Мультипликативное свойство** производящих функций. Если имеются несколько независимых целочисленных случайных величин, то производящая функция композиции является произведением производящих функций.

## Арифметические действия с формальными степенными рядами

Можно образовать формальный степенной ряд. В отличие от обычных рядов сходимость формальных рядов не требуется, и значение функции (суммы ряда) не вычисляется.

> The name "generating function" is due to Laplace. Yet, without giving it a name, Euler used the device of generating functions long before Laplace [..].

Понятие производящей функции числовой последовательности ввел Лаплас, с тех пор этот инструмент используется в теории вероятностей. Мы хотим немного расширить инструментарий, чтобы провести параллель между преобразованием Лапласа и операторным методом решения систему уравнений. Аналогичный результат и действия с операторами получается путем Z-преобразования из S-пространства.

Можно рассмотреть, какие операции над производящими функциями бывают: Сложение, умножение и суперпозиция функций. Из производящих функций можно составить рациональные дроби (формальные) и исследовать свойства рациональных производящих функций. Центральной теоремой является утверждение, что произведение рациональных производящих функций - рационально.

Выводы полученные для рациональных производящих функций применимы в теории операторов.

**Суммой** двух производящих функций

$$A(s) = a_0 + a_1 s + a_2 s^2 + \dots \quad B(s) = b_0 + b_1 s + b_2 s^2 + \dots$$

называется производящая функция

$$A(s) + B(s) = (a_0 + b_0) + (a_1 + b_1)s + (a_2 + b_2)s^2 + \dots$$

**Произведением** производящих функций A и B называется производящая функция

$$A(s)B(s) = a_0 b_0 + (a_0 b_1 + a_1 b_0)s + (a_0 b_2 + a_1 b_1 + a_2 b_0)s^2 + \dots$$

Операции сложения и умножения производящих функций коммутативны и ассоциативны. Операции сложения и умножения рациональных функций также будут коммутативны.

Пусть $A(s) = a_0 + a_1 s + a_2 s^2 + \ldots$ и $B(t) = b_0 + b_1 t + b_2 t_2 + b_3 t_3 + \ldots$ — две производящие функции, причем $B(0) = b0 = 0$.
Подстановкой производящей функции B в производящую функцию A называется производящая функция

$$A(B(t)) = a0 + a_1 b_1 t + (a_1 b_2 + a_2 b_1^2) t^2 + (a_1 b_3 + 2 a_2 b_1 b_2 + a_3 b_1^3) t^3 + \ldots$$

**Суперпозиция** функций (или сложная функция, или композиция функций, англ. function composition) — это функция, полученная из некоторого множества функций путем подстановки одной функции в другую.

Из производящих функций вводится понятие *оператор задержки* и задержка сопоставляется с производной.

*Теорема (о существовании и единственности обратной подстановки)* Пусть функция $B(t) = b_1 t + b_2 t^2 + b_3 t^3 + \ldots$ такова, что $B(0) = b_0 = 0$, а $b_1 \neq 0$. Тогда существуют такие функции
$A(s) = a_1 s + a_2 s^2 + a_3 s^3 + \ldots, A(0) = 0,$
и
$C(u) = c_1 u + c_2 u^2 + c_3 u^3 + \ldots, C(0) = 0$, что
$A(B(t)) = t$ и $B(C(u)) = u$. Функции $A$ и $C$ единственны. Функция A называется левой обратной, а функция C — правой
обратной к функции B.

*Мне иногда кажется, что я занимаюсь археологией, добываю артефакты из забытых теорий. Понятие производящей функции ввел Лаплас. По сути все эти выводы можно изложить в рамках единой теории операторов. Теория развитая вокруг производящих функций воспроизводится в виде схемы курса в теории операторов и теории линейных стационарных систем (LTI). При анализе физических процессов (электрических цепей) используется т.н операторный метод, который производным и интегралам сопоставляет полиномы и рациональные функции полиномов.*

*Теорема (обращение ряда)* Пусть $A(s) = a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \ldots$ — формальный степенной ряд, причем $A(0) \neq 0$. Тогда существует единственный формальный степенной ряд $B(s) = b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \ldots$, такой что $A(s)B(s) = 1$.

*Доказательство теорем в курсе Ландо [^3]. Доказательства не всегда даются четко.*

В итоге мы рассматриваем три операции: сложение, умножение и суперпозиция функций. Для каждой такой операции мы определяем понятие обратная операция и обратная функция. Для каждой операции мы определяем понятие единица или ноль - выделенный элемент множества

функций - инвариант. Таким образом мы рассматриваем множество производящих функций, как математический класс - кольцо рациональных производящих функций со своей ассоциативной алгеброй.
С алгебраической точки зрения множество формальных степенных рядов (с коэффициентами в поле комплексных, вещественных или рациональных чисел) образует (бесконечномерное) векторное пространство над этим полем.

Мы вводим понятие интегрирование и дифференцирования формального степенного ряда:

$$A'(s) = a_1 + 2a_2 s + 3a_3 s^2 + ... + n a_n s^{n-1} + ... ,$$

$$\int A(s) = a_0 s + a_1 \frac{s^2}{2} + a_2 \frac{s^3}{3} + ... + a_n \frac{s^{n+1}}{n+1} + ...$$

Операция дифференцирования обратна операици интегрирования.

Интегрирование соответсвует понятию интеграл с переменным верхним пределом.

$$\int A(s) = \int_0^s A(\xi) d\xi$$

## Рекуррентные соотношения и рациональные производящие функции

[3] С.А. Ландо. Лекции о производящих функциях, 2007.

Теория рациональных производящих функций совпадает, по существу, с теорией решений обыкновенных дифференциальных уравнений с постоянными коэффициентами.

Пример 1. В качестве примера рассмотрим геометрическую прогрессию $G(s) = 1 + s + s^2 + s^3 + ...$. Домножив на s получаем формулу

$sG(s) = s + s^2 + s^3 + s^4 + ... = G(s) - 1$, откуда

$$G(s) = \frac{1}{1-s}$$

Пример 2. Последовательность Фибоначчи определяется начальными членами ряда $f_0 = f_1 = 1$ и рекуррентным соотношением
$f_{n+2} = f_{n+1} + f_n$

дает формулу производящей функции последовательности
$F(s) = 1 + s + 2s^2 + 3s^3 + 5s^4 + \dots.$
Умножив левую и правую часть равенства на $s + s^2$

$$(s + s^2)F(s) = \dots = F(s) - 1$$

Откуда получается

$$F(s) = \frac{1}{1 - s - s^2}$$

Полученную формулу можно понимать как композицию двух производящих функций

*Теорема (рациональные производящие функции заданные линейной рекурсией)* Пусть последовательность $\{a_k\}$ задается линейным
рекуррентным соотношением порядка $n$ с постоянными $c_1, \dots, c_n$:

$$a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \dots + c_k a_n$$

и числа $a_0, a_1, \dots, a_{k-1}$ заданы. Тогда производящая функция $A(s) = a_0 + a_1 s + a_2 s^2 + \dots$ рациональна,
$A(s) = \frac{P(s)}{Q(s)}$, причем степень многочлена $Q(s)$ равна $k$, а степень многочлена $P(s)$ не превосходит $k - 1$.

Доказательство:
Умножив производящую функцию $A(s)$ на $c_1 s + c_2 s^2 + \dots + c_k s^k$, получаем

$$\begin{aligned} (c_1 s + \dots + c_k s^k)A(s) = c_1 a_0 s + c_1 a_1 s^2 + c_1 a_2 s^3 + \dots + c_1 a_{k-1} s^k + \dots \\ + c_2 a_0 s^2 + c_2 a_1 s^3 + \dots + c_2 a_{k-2} s^k + \dots \\ + c_3 a_0 s^3 + \dots + c_3 a_{k-3} s^k + \dots \\ \dots \\ + c_k a_0 s^k + \dots \quad & = -P(s) + A(s) \end{aligned}$$

где P- некоторый многочлен, степень которого не превосходит k-1. Коэффициенты при степенях $s^{n+k}$ равны правой части реккурентного соотнощения.

Полином $Q(s)$ имеет вид:

$$Q(s) = 1 - c_1 s - c_2 s^2 - \dots - c_k s^k$$

для P(s) также можно составить выражение.

Операция деления для бесконечных степенных рядов не определена. Но можно говорить о делении полиномов данной сетпени с остатком.

*Теорема 2 (обратная)*. Если производящая функция $A(s) = a_0 + a_1 s + +a_2 s^2 + \dots$ рациональна, $A(s) = \frac{P(s)}{Q(s)}$, где многочлены $P(s)$ и $Q(s)$ взаимно просты, то начиная с некоторого номера $n$ последовательность
$a_0, a_1, a_2, \dots$ задается линейным рекуррентным соотношением $a_{n+k} = c_1 a_{n+k-1} + c_2 a_{n+k-2} + \dots + c_k a_n$,
где $k$—степень многочлена $Q(s)$, а $c_1, c_2, \dots, c_k$ - некоторые константы.

*Далее см. раздел математического аланиза, Разложение рациональных функций с комплекснмыи и действительными коэффициентами.*

## Производящие функции моментов

Пусть есть случайная величина $\xi$ с распределением $\mathsf{P}_\xi$. Тогда её производящей функцией моментов называется функция, имеющая вид:
$\mathsf{M}_\xi(t) = \mathsf{E}\left[e^{t\xi}\right]$.

Пользуясь формулами для вычисления математического ожидания, определение производящей функции моментов можно переписать в виде:
$\mathsf{M}_\xi(t) = \int\limits_{-\infty}^{\infty} e^{tx}\, \mathsf{P}_\xi(dx)$ ,
то есть производящая функция моментов — это двустороннее преобразование Лапласа плотности распределения случайной величины (с точностью до отражения).

## Факториальные моменты

Для вывода интерполяционных формул понадобится понятие момента $\mathsf{E}\xi^r$ и факториального момента $\mathsf{E}\xi^{[r]}$, где

$$\xi^{[r]} = \xi(\xi-1)\cdots(\xi-r+1)\ , \quad \xi^{[0]} = 1\ .$$

Факториальные моменты можно выразить через моменты и наоборот. Например, $\mathsf{E}\xi = \mathsf{E}\xi^{[2]} + \mathsf{E}\xi$.

Факториальные моменты вычисляются через производные производящих функций в точке $s = 1$.

$$\mathsf{E}\xi^{[r]} = f_\xi^{(r)}(1)$$

# Характеристические функции последовательности

Характеристической функцией случайной величины $\xi$ мы будем называть функцию $f_\xi(t)$ от действительного аргумента $t$, равную

$$f_\xi(t) = \mathsf{E}e^{it\xi} \ .$$

Раскрывая в $e^{i\phi}$ по формуле Эйлера
$e^{i\phi} = \cos\phi + i\sin\phi$,
мы имеем $f_\xi(t) = \mathsf{E}\cos\xi t + i\mathsf{E}\sin\xi t$.

*{дописать}*