ITMO more than a UNIVERSITY

# Современная теория информации

Лекция 2. Выпуклые функции. Условная энтропия. Стационарные источники.

**Беляев Евгений Александрович**
eabelyaev@itmo.ru

# Содержание лекции

ИТМО

1. Выпуклые функции.
2. Условная энтропия.
3. Дискретные источники.
4. Дискретные стационарные процессы.

- Множество вещественных векторов $\boldsymbol{R} = \{\boldsymbol{x}\}$ выпукло, если $\forall \boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{R}$ и $\forall \alpha \in [0, 1]$, вектор $\boldsymbol{y} = \alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{x}'$ принадлежит $\boldsymbol{R}$.
- **Теорема.** Множество вероятностных векторов длины $M$ выпукло.
- **Доказательство.** Для множества $\boldsymbol{X} = \{1, 2, \ldots, M\}$, рассмотрим два распределения вероятностей $\boldsymbol{p} = (p_1, ..., p_M)$ и $\boldsymbol{p}' = (p'_1, ..., p'_M)$, и $\alpha \in [0, 1]$. Рассмотрим векторы вида

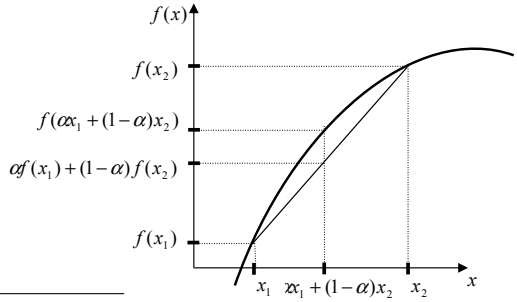
$$\boldsymbol{q} = \alpha \boldsymbol{p} + (1 - \alpha)\boldsymbol{p}'.$$

1. $q_i \geq 0$.
2. Сумма компонент вектора $\boldsymbol{q}$

$$\sum_{i=1}^{M} q_i = \alpha \sum_{i=1}^{M} p_i + (1 - \alpha) \sum_{i=1}^{M} p'_i = \alpha + 1 - \alpha = 1.$$

- Функция $f(\boldsymbol{x})$ выпуклая, если $\forall \boldsymbol{x}, \boldsymbol{x}' \in \boldsymbol{R}^1$ и $\forall \alpha \in [0,1]$ выполняется неравенство: $f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{x}') \geq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{x}')$
- Для случая функции одной переменной:

$$f(\alpha x_1 + (1-\alpha)x_2) \geq \alpha f(x_1) + (1-\alpha) f(x_2).$$




[1] $\boldsymbol{R}$ – выпуклая область

**Теорема.** Пусть $f(\boldsymbol{x})$ – выпуклая $\cap$ функция вектора $\boldsymbol{x}$, определённая на выпуклой области $\boldsymbol{R}$, и пусть константы $\alpha_1, ..., \alpha_M \in [0, 1]$ такие, что $\sum_{m=1}^{M} \alpha_m = 1$. Тогда $\forall \boldsymbol{x}_1, ..., \boldsymbol{x}_M \in \boldsymbol{R}$ справедливо неравенство:

$$f\left(\sum_{m=1}^{M} \alpha_m \boldsymbol{x}_m\right) \geq \sum_{m=1}^{M} \alpha_m f(\boldsymbol{x}_m).$$

Если $\alpha_m$ означает вероятность $\boldsymbol{x}_m$, то получим *неравенство Йенсена*:

$$f\left(\mathrm{E}\{\mathrm{x}\}\right) \geq \mathrm{E}\{f(\boldsymbol{x}))\}.$$

1. Сумма выпуклых функций выпукла.
2. Произведение выпуклой функции и положительной константы является выпуклой функцией.
3. Линейная комбинация выпуклых функций с неотрицательными коэффициентами – выпуклая функция.

**Теорема**. Энтропия $H(\boldsymbol{p})$ ансамбля с распределением вероятностей $\boldsymbol{p}$ – выпуклая $\cap$ функция от $\boldsymbol{p}$.

**Доказательство**.

$$H(\boldsymbol{p}) = -\sum_{m=1}^{M} p_m \log p_m.$$

Рассмотрим слагаемые $f_m(\boldsymbol{p}) = -p_m \log p_m$.
Вторая производная $f_m''(\boldsymbol{p}) = -(\log e)/p_m$.
$f_m''(\boldsymbol{p}) < 0 \ \forall p_m \in (0, 1)$.

- $X = \{0, 1\}$. Пусть $p(1) = p$, $p(0) = 1 - p = q$.
- Энтропия двоичного ансамбля

$$H(X) = -p \log p - q \log q \triangleq \eta(p).$$

- Первая производная от $\eta(p)$.

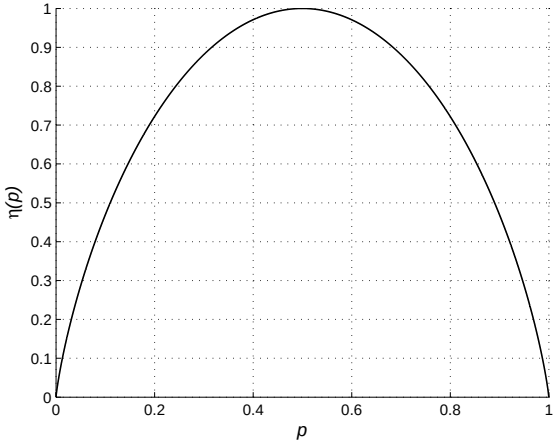
$$\eta'(p) = -\log p + \log(1 - p),$$

$\eta'(p)$=0, при $p = \frac{1}{2}$ – точка экстремума.

- Вторая производная от $\eta(p)$.

$$\eta''(p) = -\log e/p - \log e/(1 - p) < 0.$$

$$H(X) = \eta(p) = -p \log p - (1-p) \log(1-p)$$



- Обозначим $\tilde{\boldsymbol{p}} = ((p_1 + p_2)/2, (p_1 + p_2)/2, p_3, ..., p_M)$.
- Необходимо доказать, что

$$H(\tilde{\boldsymbol{p}}) \geq H(\boldsymbol{p}).$$

- Обозначим

$$\begin{aligned} \boldsymbol{p}' = \boldsymbol{p} &= (p_1, p_2, p_3, ..., p_M), \\ \boldsymbol{p}'' &= (p_2, p_1, p_3, ..., p_M). \end{aligned}$$

- Заметим, что: $H(\boldsymbol{p}') = H(\boldsymbol{p}'') = H(\boldsymbol{p})$.
- $\tilde{\boldsymbol{p}} = (\boldsymbol{p}' + \boldsymbol{p}'')/2$.
- Из выпуклости энтропии следует, что:

$$H(\tilde{\boldsymbol{p}}) = H\left(\frac{\boldsymbol{p}' + \boldsymbol{p}''}{2}\right) \geq \frac{1}{2}H(\boldsymbol{p}') + \frac{1}{2}H(\boldsymbol{p}'') = H(\boldsymbol{p}).$$

- Произведение ансамблей $XY = \{(x, y), p_{XY}(x, y)\}$.
- Условное распределение вероятностей:

$$p(x|y) = \begin{cases} \frac{p(x,y)}{p(y)}, & \text{if } p(y) \neq 0, \\ 0 & \text{иначе,} \end{cases} \qquad x \in X.$$

- Ансамбли $X$ и $Y$ – независимы, если

$$p(x, y) = p(x)p(y), \quad x \in X, \quad y \in Y.$$

- Условная собственная информация сообщения $x$ при фиксированном $y$

$$I(x|y) = -\log p(x|y),$$

- Условная энтропия $X$ при заданном $y \in Y$

$$H(X|y) = -\sum_{x \in X} p(x|y) \log p(x|y),$$

- Условная энтропия $X$ при фиксированном ансамбле $Y$

$$H(X|Y) = -\sum_{y \in Y} \left( p(y) \sum_{x \in X} p(x|y) \log p(x|y) \right) = \\ = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y).$$

1. $H(X|Y) \geq 0$
2. $H(X|Y) \leq H(X)$, равенство, если $X$ и $Y$ независимы.
3. $H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
4. $H(X|YZ) \leq H(X|Y)$ равенство, если $X$ и $Z$ условно независимы для всех $y \in Y$.

5.

$$\begin{aligned} H(X_1...X_n) = H(X_1) &+ H(X_2|X_1) + \\ &+ H(X_3|X_1X_2) + .... + \\ &+ H(X_n|X_1, ..., X_{n-1}). \end{aligned}$$

6. $H(X_1...X_n) \leq \sum_{i=1}^{n} H(X_i)$ равенство, если $X_1, \ldots, X_n$ являются совместно независимыми.

$$p(x) = \sum_{y \in Y} p(x|y)p(y)$$

$$\begin{aligned} H(X|Y) - H(X) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y) + \\ &+ \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x) = \end{aligned}$$

$$p(x) = \sum_{y \in Y} p(x|y)p(y)$$

$$\begin{aligned} H(X|Y) - H(X) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y) + \\ &\quad + \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x) = \\ &= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x)}{p(x|y)} \leq \end{aligned}$$

$$p(x) = \sum_{y \in Y} p(x|y)p(y)$$

$$\begin{aligned} H(X|Y) - H(X) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y) + \\ &\quad + \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x) = \\ &= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x)}{p(x|y)} \leq \\ &\leq \sum_{x \in X} \sum_{y \in Y} p(x,y) \left( \frac{p(x)}{p(x|y)} - 1 \right) \log e = \end{aligned}$$

$$p(x) = \sum_{y \in Y} p(x|y)p(y)$$

$$\begin{aligned} H(X|Y) - H(X) &= -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x|y) + \\ &\quad + \sum_{x \in X} \sum_{y \in Y} p(x,y) \log p(x) = \\ &= \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \frac{p(x)}{p(x|y)} \leq \\ &\leq \sum_{x \in X} \sum_{y \in Y} p(x,y) \left( \frac{p(x)}{p(x|y)} - 1 \right) \log e = \\ &= \left( \sum_{x \in X} \sum_{y \in Y} p(y)p(x) - \sum_{x \in X} \sum_{y \in Y} p(x,y) \right) \log e = 0. \end{aligned}$$

- Рассмотрим $X = \{x, p(x)\}$, $f(x)$, $Y = \{y = f(x), x \in X\}$.
- Нужно доказать, что
$$H(Y) \leq H(X).$$
- Используя свойство 3 условной энтропии:
$$H(XY) = \underbrace{H(X|Y)}_{\geq 0} + H(Y) = \underbrace{H(Y|X)}_{=0} + H(X).$$
- Поскольку $f(x)$ определена для каждого $x$, получим, что $H(Y|X) = 0$, $H(X|Y) \geq 0$.

- **Дискретный источник** это устройство, которое в каждый момент времени выбирает одно сообщение из дискретного множества.
- Если множество значений времени также дискретно, то источник называется **дискретным по времени**.
- Источник считается заданным, если известна его вероятностная модель. Другими словами, мы должны определить вероятностную модель **случайного процесса** генерирования случайных сообщений на выходе источника.

- Дискретный источник задан, если для $n = 1, 2, ...$ и $i = 0, 1, 2, ...$ известна вероятность

$$p(x_{i+1}, x_{i+2}, ..., x_{i+n})$$

случайной последовательности из $\{X_{i+1}X_{i+2}...X_{i+n}\}$, которая начинается с индекса $i + 1$ и имеет длину $n$, где $x_j \in X_j$, $j = i + 1, ..., i + n$.

- Обычно рассматривается случай, когда $X_j = X$ для всех $j$.

Пусть $\boldsymbol{x} = (x_1, x_2, \ldots, x_n, \ldots)$ – дискретный случайный процесс.
Рассмотрим случайный вектор $\boldsymbol{x}_{j+1}^{j+n} = (x_{j+1}, x_{j+2}, \ldots, x_{j+n})$.

- **Стационарность** процесса означает, что для любого $n$ и $j$, $p(\boldsymbol{x}_{j+1}^{j+n})$ не зависит от сдвига во времени $j$, т.е., $p(\boldsymbol{x}_{j+1}^{j+n}) = p(\boldsymbol{x}_1^n)$.
- Если

$$p(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} p(x_i),$$

то источник называют **дискретным источником без памяти**.

Дискретный случайный процесс называется **цепью Маркова порядка** $s$ если для любого $n$ и $\boldsymbol{x} = (x_1, \ldots, x_n) \in X^n$ выполняется следующее равенство:

$$p(\boldsymbol{x}) = p(x_1, \ldots, x_s)p(x_{s+1}|x_1, \ldots, x_s)\times$$

$$\times p(x_{s+2}|x_2, \ldots, x_{s+1}) \ldots p(x_n|x_{n-s}, \ldots, x_{n-1})$$

Другими словами, для цепи Маркова справедливо равенство:

$$p(x_n|x_1, \ldots, x_{n-1}) = p(x_n|x_{n-s}, \ldots, x_{n-1}),$$

условная вероятность текущего значения зависит от $s$ предыдущих значений и не зависит от остальных.

Марковская цепь порядка $s = 1$ с состояниями $X = \{0, 1, \ldots, M-1\}$ определяется начальным распределением $\{p(x_1), x_1 \in X\}$ и условными вероятностями

$$\pi_{ij} = P(x_t = j | x_{t-1} = i),\ i, j = 0, 1, \ldots, M-1$$

**Матрица переходных вероятностей:**

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} & \ldots & \pi_{0,M-1} \\ \pi_{10} & \pi_{11} & \ldots & \pi_{1,M-1} \\ \ldots & \ldots & \ldots & \ldots \\ \pi_{M-1,0} & \pi_{M-1,1} & \ldots & \pi_{M-1,M-1} \end{pmatrix}.$$

Обозначим через $\boldsymbol{p}_t = (p_t(0), \ldots, p_t(M-1)$ стохастический вектор, компоненты которого – вероятности состояний цепи Маркова в момент времени $t$, где $p_t(i)$, $i = 0, 1, \ldots, M-1$ – вероятность состояния $i$ в момент времени $t$.

Из формулы полной вероятности следует:

$$p_{t+1}(i) = \sum_{j=0}^{M-1} p_t(j)\pi_{ji}.$$

В матричном виде

$$\boldsymbol{p}_{t+1} = \boldsymbol{p}_t \Pi$$

Для произвольного числа шагов $n$

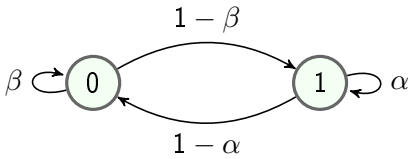
$$\boldsymbol{p}_{t+n} = \boldsymbol{p}_t \Pi^n. \tag{1}$$

Из формулы (1) следует, что распределение вероятностей в момент времени $t$ зависит от величины $t$ и от начального распределения $\boldsymbol{p}_1$. Отсюда следует, что в общем случае рассматриваемый случайный процесс нестационарен.

Однако, если существует стохастический вектор $\boldsymbol{p}$, такой что

$$\boldsymbol{p} = \boldsymbol{p}\Pi, \tag{2}$$

то выбрав $\boldsymbol{p}_1 = \boldsymbol{p}$ мы получим стационарный процесс. Вектор $\boldsymbol{p}$, удовлетворяющий (2) называется **стационарным распределением вероятностей** для марковской цепи с матрицей переходных вероятностей $\Pi$.




$$\Pi = \begin{bmatrix} \beta & 1-\beta \\ 1-\alpha & \alpha \end{bmatrix}$$

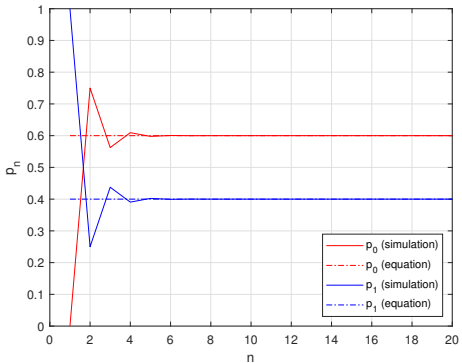
$$\mathbf{p} = \mathbf{p}\Pi,$$

где $\mathbf{p} = \{p_0, p_1\}$. Поэтому

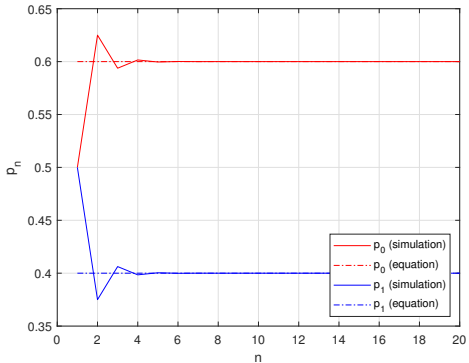
$$\begin{cases} p_0 = p_0 \cdot \beta + p_1 \cdot (1-\alpha) \\ p_0 + p_1 = 1, \end{cases}$$

откуда $p_0 = \frac{(1-\alpha)}{2-\beta-\alpha}$, $p_1 = \frac{(1-\beta)}{2-\beta-\alpha}$.

$$\boldsymbol{p}_n = \boldsymbol{p}_1 \Pi^n$$
$$\lim_{n \to \infty} \boldsymbol{p}_n = \boldsymbol{p}.$$




(a) $\boldsymbol{p}_1 = \{0, 1\}, \boldsymbol{p} = \{0.6, 0.4\}$




(b) $\boldsymbol{p}_1 = \{0.5, 0.5\}, \boldsymbol{p} = \{0.6, 0.4\}$

В первой урне находятся $N_1 = 3$ черных шара, во второй урне находятся $N_2 = 3$ белых шара. Число черных шаров в первой урне определяет состояние системы. На каждом шагу случайно выбирается по одному шару из каждой урны, и эти выбранные шары меняются местами. Построить цепь Маркова, которая соответствует данной задаче. Вычислить вероятность того, что если на шаге $t$ в первой урне нет черных шаров, то на шаге $t + 3$ в ней будет 3 черных шара. Вычислить стационарные вероятности каждого из состояний.

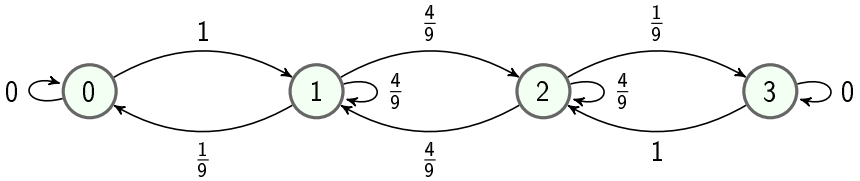



Рис.: Цепь Маркова для числа черных шаров в первой урне

Матрица переходов:

$$\mathbf{\Pi} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{9} & \frac{4}{9} & \frac{4}{9} & 0 \\ 0 & \frac{4}{9} & \frac{4}{9} & \frac{1}{9} \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Для поиска стационарного распределения нужно решить следующую систему линейных уравнений:

$$\boldsymbol{\pi} = (\pi_0 \pi_1 \pi_2 \pi_3) = (\pi_0 \pi_1 \pi_2 \pi_3)\mathbf{\Pi}. \tag{3}$$

Выражение (3) можно записать как $\boldsymbol{\pi}\mathbf{\Pi} - \boldsymbol{\pi} = \boldsymbol{\pi}(\mathbf{\Pi} - \boldsymbol{I}) = 0$, то есть мы получили однородное уравнение.

Рассмотрим матрицу $\boldsymbol{A} = \boldsymbol{\Pi} - \boldsymbol{I}$.

1. Во-первых, сумма строк матрицы $\boldsymbol{\Pi}$ равна единице, поэтому сумма строк матрицы $\boldsymbol{A}$ равна нулю.
2. Во-вторых, это означает, что сумма столбцов матрицы $\boldsymbol{A}$ тоже равна нулю, т.е. присутствует линейная зависимость строк. Поэтому ранг $rank(\boldsymbol{A}) < 4$, т.е. система имеет бесконечное число решений.
3. Единственность решения можно получить, принимая во внимание, что вероятность пребывания в одном из состояний равна единие, т.е. $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$.

В нашем случае (3) можно записать в виде следующей системы:

$$\begin{cases} \pi_0 = \frac{1}{9}\pi_1 \\ \pi_1 = \pi_0 + \frac{4}{9}\pi_1 + \frac{4}{9}\pi_2 \\ \pi_2 = \frac{4}{9}\pi_1 + \frac{4}{9}\pi_2 + \pi_3 \\ \pi_3 = \frac{1}{9}\pi_2, \end{cases}$$

решение которой выглядит как $\boldsymbol{\pi} = c(1, 9, 9, 1)$, где $c$ – константа. Принимая во внимание, что $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$, получим $c = \frac{1}{20}$ или $\pi_0 = \pi_3 = \frac{1}{20}$, $\pi_1 = \pi_2 = \frac{9}{20}$.

Каждое состояние цепи Маркова с памятью $s$ это буква английского языка или пробел[2]:

$s = 0$ OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVANAH BRL

$s = 1$ ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE.

$s = 2$ IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE.

[2] Shannon, C.E. ,"A Mathematical Theory of Communication", Bell System Technical Journal, 1948.

Каждое состояние цепи Маркова с памятью $s$ это слово английского языка[3]:

$s = 0$ REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

$s = 1$ THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED.

[3] Shannon, C.E. ,"A Mathematical Theory of Communication", Bell System Technical Journal, 1948.

Энтропия символа $x_t$ (из ансамбля $X_t = X$) сгенерированного в момент времени $t$ не зависит от $t$ $H(X_t) = H(X)$ и называется **одномерной энтропией источника** (процесса). Обозначим её как $H_1(X)$.

$H_1(X)$ **не учитывает зависимость между символами**, порождёнными источником.

Рассмотрим $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ из $X_1X_2 \ldots X_n = X^n$.
Энтропия $H(X_1X_2 \ldots X_n) = H(X^n)$ называется $n$-**мерной энтропией** процесса.
Энтропия на символ для последовательности длины $n$ определяется как:

$$H_n(X) = \frac{H(X^n)}{n},$$

Другой способ:

$$H(X_n|X_1, \ldots, X_{n-1}) = H(X|X^{n-1}).$$

Энтропия на сообщение:

$$\lim_{n\to\infty} H_n(X) \text{ и } \lim_{n\to\infty} H(X|X^n)$$

Theorem

*Для дискретного стационарного процесса (источника)*

A. $H(X|X^n)$ *не возрастает с увеличением* $n$;

B. $H_n(X)$ *не возрастает с увеличением* $n$;

C. $H_n(X) \geq H(X|X^{n-1})$;

D. $\lim\limits_{n\to\infty} H_n(X) = \lim\limits_{n\to\infty} H(X|X^n)$.

ITMO more than a UNIVERSITY

# Спасибо за внимание!