bilet

- Понятие энтропии случайной величины. Основные свойства энтропии (верхняя и нижняя оценки на величину энтропии дискретной СВ)
- Понятие Hq(U) и связь H(U), U~p и Hq(U)
- Условная энтропия и дивергенция Кульбака-Лейбнера. Свойства условной энтропии
- Совместная энтропия. Энтропия системы независимых СВ. Свойства совместной энтропии.
- Понятие взаимной информации.
- Базовые свойства взаимной информации.
- Выпуклость дивергенций Кульбака-Лейбнера
- Закон больших чисел
- AEP теорема
- Определение типичного множества. Основные свойства типичных множеств
- Мощность типичного множества
- Теорема о вероятности типичного множества
- Теорема Шеннона о кодировании источника
- Понятие высоковероятного множества. Связь типичного множества и высоковероятного множества.
- Понятие префиксного и однозначно-декодируемого кода. Связь между этими понятиями
- Кодирование источника с диадическим распределением
- Свойства диадического распределения
- Коды Шеннона
- Неравенство Крафта
- Коды Хаффмана. Оптимальность кодов Хаффмана
- Понятие кодовой схемы. Достижимая скорость передачи кодовой схемы.
- Пропускная способность канала. Примеры

1

1. Понятие энтропии случайной величины. Основные свойства энтропии (верхняя и нижняя оценки на величину энтропии дискретной СВ)

Функция - ”surprise”

Пусть задана некоторая дискретная случайная величина U, которая принимает значения из множества $\mathcal{U}$ с вероятностями $P_U(u) = P(U = u) = p(u)$.

Рассмотрим функцию:

$$s(u) = \log_2 \frac{1}{p(u)}$$

Данная функция тем больше, чем ”неожиданнее” ее аргумент: $s(u) \to \infty$ при $p(u) \to 0$. ”Чем реже u, тем больший ”сюрприз” его наблюдать”.

Энтропия

Определение

Пусть U — дискретная случайная величина, которая принимает значения из множества $\mathcal{U}$. Тогда энтропия — это ожидаемый ”сюрприз” случайной величины:

$$H(U) = \mathbb{E}[s(u)] = \sum_{u \in \mathcal{U}} p(u) \log_2 \frac{1}{p(u)}$$

Энтропия представляет собой ожидаемую величину неожиданности, которую имеет распределение. Интуитивно понятно, что чем больше ожидаемая ”неожиданность” или энтропия распределения, тем труднее его представить и тем труднее предсказать поведение источника.

Свойства энтропии - неравенство Йенсена

Теорема

Если $f(x)$ — выпуклая фукнция, $q_i \geq 0$, $\sum_{i=1}^{n} q_i = 1$, $x_i \in D_f$ то

$$f\left(\sum_{i=1}^{n} q_i x_i\right) \leq \sum_{i=1}^{n} q_i f(x_i).$$

Важные частные случаи:

- Если Q — выпуклая, то $\mathbb{E}Q(X) \geq Q(\mathbb{E}X)$ $(\mathbb{E}e^X \geq e^{\mathbb{E}X})$
- Если Q — вогнутая, то $\mathbb{E}Q(X) \leq Q(\mathbb{E}X)$ $(\mathbb{E}\log X \leq \log(\mathbb{E}X))$

Равенство достигается для детерминированных СВ.

Ограничения на энтропию - верхняя оценка

Теорема

Если $\mathcal{U} = \{1, 2, ..., m\}$, то $H(U) \leq \log m$, причем равенство достигается тогда и только тогда, когда $p(u) = 1/m$ для всех $u \in \mathcal{U}$.

Доказательство:

$H(U) = \mathbb{E}\left(\log \frac{1}{p(u)}\right) \geq \log\left(\mathbb{E}\left(\frac{1}{p(u)}\right)\right) = \log \sum_{\mathcal{U}} p(u) \frac{1}{p(u)} = \log m$. Равенство достигается, когда $\frac{1}{p(u)}$ постоянна, т. е. $p(u) = 1/m$ для всех u.

Нижняя оценка

Теорема

$H(U) \geq 0$, причем равенство достигается для детерминированной СВ U.

Доказательство:

$$H(U) = \mathbb{E}\left(\log \frac{1}{p(u)}\right) \geq 0,$$

так как

$$\log \frac{1}{p(u)} \geq 0.$$

Равенство достигается в том случае, когда $\log \frac{1}{P(U)} = 0$, т. е. $P(U) = 1$.

8

2

2. Понятие Hq(U) и связь H(U), U~p и Hq(U)

Рассмотрим случайную величину Q, заданную над тем же алфавитом, что и U, и опрелелим для нее $q(u)$. Для q определим функцию:

$$H_q(U) = \sum_{u \in \mathcal{U}} p(u) \log_2 \frac{1}{q(u)}$$

Обратите внимание, что это ожидаемая функция "неожиданности", но вместо неожиданности, связанной с p, это неожиданность, связанная с U, которая распределяется согласно p, но ошибочно предполагается, что она имеет распределение q. Следующий результат предполагает, что мы (в среднем) будем больше удивлены, если будем иметь в виду неправильное распределение. В этом есть интуитивный смысл!

Дальнейшие свойства энтропии

Теорема

$H(U) \leq H_q(U)$, причем равенство достигается тогда и только тогда, когда $q = p$.

Доказательство:

$$H(U) - H_q(U) = \mathbb{E}\left(\log \frac{1}{p(u)}\right) - \mathbb{E}\left(\log \frac{1}{q(u)}\right) = \mathbb{E}\left(\log \frac{q(u)}{p(u)}\right) \leq$$

$$\leq \log\left(\mathbb{E}\left(\frac{q(u)}{p(u)}\right)\right) = \log\left(\sum_{u \in \mathcal{U}} p(u) \frac{q(u)}{p(u)}\right) = \log\left(\sum_{u \in \mathcal{U}} q(u)\right) = \log 1 = 0$$

Равенство выполняется только, когда неравенство Йенсена превращается в равенство, а это достигается при постоянной $\frac{q(u)}{p(u)}$, т. е. при $q = p$.

3

3. Условная энтропия и дивергенция Кульбака-Лейбнера. Свойства условной энтропии

Условная энтропия. Дивергенция Кульбака-Лейблера

Определение

Важной мерой расстояния между вероятностными мерами является относительная энтропия или дивергенция Кульбака – Лейблера:

$$D_{KL}(p||q) = \mathbb{E}\left(\log \frac{p(u)}{q(u)}\right) = \sum_{u \in \mathcal{U}} p(u) \log \frac{p(u)}{q(u)}$$

Доказанное свойство говорит о том, что $D_{KL}(p||q)$ всегда неотрицательна с равенством нулю в случае одинаковых распределений.

Условная энтропия

Мы определили энтропию случайной величины U. Мы также видели, что когда U - это совместная случайная величина от независимых переменных, тогда $H(U)$ - это сумма энтропий. Можем ли мы сказать что-то еще об энтропии совместной случайной величины?

Определение

Условная энтропия случайных величин X и Y определяется как:

$$H(X|Y) = \mathbb{E}\left(\log \frac{1}{p(X|Y)}\right) = \sum_{x,y} P(x,y) \frac{1}{P(x|y)} = \sum_{y} P(y) \left(\sum_{x} P(x|y) \log \frac{1}{P(x|y)}\right) =$$
$$= \sum_{y} P(y) H(X|y)$$

Несколько замечаний

- Условная энтропия - это функционал от совместного распределения (X, Y)
- Эту величину можно понимать как «среднее» удивление наблюдать X, когда мы наблюдаем Y
- По Y производится усреднение
- Условная энтропия НЕ является функцией случайной величины Y (в отличие от условного математического ожидания)

14

4

4. Совместная энтропия. Энтропия системы независимых СВ. Свойства совместной энтропии.

Совместная энтропия

Определение

Совместная энтропия случайных величин X и Y определяется как:

$$H(X, Y) = \mathbb{E}\left(\log \frac{1}{p(X, Y)}\right) = \mathbb{E}\left(\log \frac{1}{p(X)p(Y|X)}\right)$$

15

Энтропия независимых СВ

Теорема

Пусть $X_1, X_2, ...X_n$ — независимые СВ, тогда

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i).$$

Если СВ зависимы, то

$$H(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} H(X_i).$$

Доказательство:

$$H(X_1, X_2, ..., X_n) = \mathbb{E}\left(\log \frac{1}{p(x_1, x_2, ..., x_n)}\right) = \mathbb{E}\left(-\log \prod_{i=1}^{n} p(x_i)\right) =$$
$$= \mathbb{E}\left(-\sum_{i=1}^{n} \log p(x_i)\right) = \sum_{i=1}^{n} \mathbb{E}\left(-\log p(x_i)\right) = \sum_{i=1}^{n} H(X_i)$$

Общий случай зависимых СВ

Теорема

Пусть $X_1, X_2, ...X_n$ — система случайных величин с заданной $p(x_1, x_2, ..., x_n)$, тогда:

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, ..., X_1)$$

5. Понятие взаимной информации.

Взаимная информация

Определение

Пусть заданы случайные величины X, Y и известна функция $P(x, y)$. Тогда взаимной информацией между X и Y называется величина:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) = D_{KL}\left(P_{x,y} || P_x \times P_y\right)$$

Взаимная информация — это каноническая мера количества информации, передаваемой одной случайной величиной о другой. Эту величину можно понимать как уменьшение "среднего удивления" при наблюдении коррелированной случайной величины. Взаимная информация является функционалом совместного распределения пары (X, Y). Ее также можно рассматривать как относительную энтропию между совместным распределением и произведением (сверткой) вероятностей каждого распределения.

Графическая интерпретация

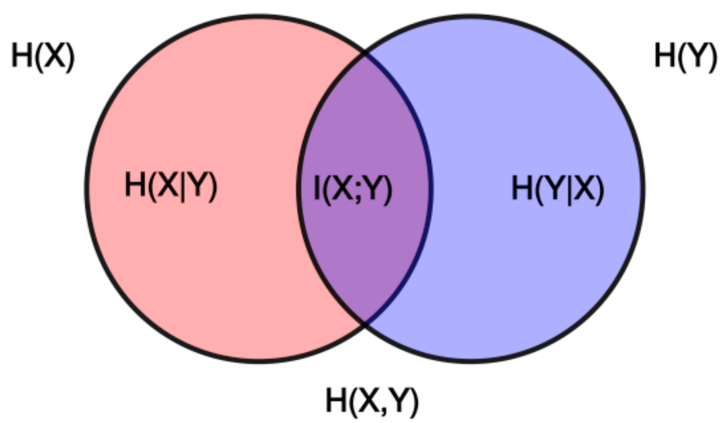


Рисунок 1: Соотношения между основными величинами

6

6. Базовые свойства взаимной информации.

Базовые свойства взаимной информации

Теорема

1. $I(X;Y) \geq 0$
2. $I(X;Y) \leq \min\{H(X), H(Y)\}$, причем равенство достигается когда $X = f(Y)$ или $Y = \phi(X)$.

21

7

7. Выпуклость дивергенций Кульбака-Лейбнера

Условная дивергенция КЛ

Определение

Для совместных вероятностей $p(x,y)$ и $q(x,y)$ условная относительная дивергенция $D_{KL}(p(y|x)||q(y|x))$ вычисляется как:

$$D_{KL}(p(y|x)||q(y|x)) = \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)}$$

26

Выпуклость дивергенции

Теорема

$D_{KL}(p||q)$ выпукла на паре (p, q); то есть, если (p_1, q_1) и (p_2, q_2) - две пары функций вероятности и $0 \le \lambda \le 1$, то:

$$D_{KL}\left(\lambda p_1 + (1-\lambda)p_2 || \lambda q_1 + (1-\lambda)q_2\right) \le \lambda D_{KL}(p_1||q_1) + (1-\lambda)D_{KL}(p_2||q_2)$$

Доказательство:

Непосредственно следует из выпуклости функции $f(x) = x \log x$

28

8

8. Закон больших чисел

Закон больших чисел

Пусть $X_1, X_2, \dots X_n$ - iid, $X_i \sim p(x)$, тогда

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X,\ n \to \infty$$

Или более формально: $\forall\, \varepsilon > 0\ \exists\, \delta > 0$:

$$Pr\left(|\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X| < \epsilon\right) > 1 - \delta$$

И в самом общем виде: $\forall\, \varepsilon > 0\ \exists\, \delta > 0$:

$$Pr\left(|\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}f(X)| < \epsilon\right) > 1 - \delta$$

9

9

9. AEP теорема

AEP-теорема

Теорема 1

Если $X_1,\ X_2,\ \dots, X_n$ — iid, $X_i \sim p(x)$, то имеет место сходимость по вероятности:

$$-\frac{1}{n}\log p(X_1, X_2, \dots, X_n) \to H(X)$$

Доказательство: $-\frac{1}{n}\log p(X_1, X_2, \dots, X_n) = -\frac{1}{n}\sum_{i=1}^{n}\log X_i \to -\mathbb{E}\log p(X) = H(X)$.

Осталось применить ЗБЧ для $f(X) = \log X$

15

10

10. Определение типичного множества. Основные свойства типичных множеств

Типичное множество

Определение

Пусть задана последовательность случайных величин $X_1,\ X_2,\ \dots, X_n$ — iid, $X_i \sim p(x)$, заданная над алфавитом $\mathcal{X}^n$, тогда множество реализаций $(x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ назовем $A_\varepsilon^{(n)}$-типичным множеством, если $\forall x = (x_1, x_2, \dots, x_n) \in A_\varepsilon^{(n)}$:

$$2^{-n(H(X)+\varepsilon)} \le p(x_1, x_2, \dots, x_n) \le 2^{-n(H(X)-\varepsilon)}$$

16

Основные свойства $A_\varepsilon^{(n)}$ - I

Теорема

Если $x_1, x_2, ..., x_n) \in A_\varepsilon^{(n)}$, то

$$H(X) - \varepsilon \le -\frac{1}{n}\log p(X_1, X_2, ..., X_n) \le H(X) + \varepsilon$$

Доказательство.

Из

$$2^{-n(H(X)+\varepsilon)} \le p(x_1, x_2, ..., x_n) \le 2^{-n(H(X)-\varepsilon)}$$

логарифмированием и последующим делением на $-1/n$ получаем

$$-n(H(X)+\varepsilon) \le \log p(x_1, x_2, ..., x_n) \le -n(H(X)-\varepsilon),$$

$$H(X) - \varepsilon \le -\frac{1}{n}\log p(X_1, X_2, ..., X_n) \le H(X) + \varepsilon.$$

Основные свойства $A_\varepsilon^{(n)}$ - II

Теорема

Если $(x_1, x_2, ..., x_n) \in \mathcal{X}^n$, $X_i \sim p(X)$ то для достаточно больших n

$$Pr\left((x_1, x_2, ..., x_n) \in A_\varepsilon^{(n)}\right) > 1 - \varepsilon$$

Доказательство.

$$Pr\left((x_1, x_2, ..., x_n) \in A_\varepsilon^{(n)}\right) = Pr\left(H(X) - \varepsilon \le -\frac{1}{n}\log p(X_1, X_2, ..., X_n) \le H(X) + \varepsilon\right) =$$

$$= Pr\left(|-\frac{1}{n}\log p(X_1, X_2, ..., X_n) - H(X)| \le \varepsilon\right)$$

$$Pr\left((x_1, x_2, ..., x_n) \in A_\varepsilon^{(n)}\right) > 1-\varepsilon \iff Pr\left(|-\frac{1}{n}\log p(X_1, X_2, ..., X_n) - H(X)| \le \varepsilon\right) < 1-\varepsilon$$

Основные свойства $A_\varepsilon^{(n)}$ - III

Теорема

$$|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$

Доказательство.

$$1 = \sum_{x \in \mathcal{X}^n} p(x) \geq \sum_{x \in A_\varepsilon^{(n)}} p(x) \geq \sum_{x \in A_\varepsilon^{(n)}} 2^{-n(H(X)+\varepsilon)} = 2^{-n(H(X)+\varepsilon)} |A_\varepsilon^{(n)}|,$$

откуда сразу

$$|A_\varepsilon^{(n)}| \leq 2^{n(H(X)+\varepsilon)}$$

Основные свойства $A_\varepsilon^{(n)}$ - IV

Теорема

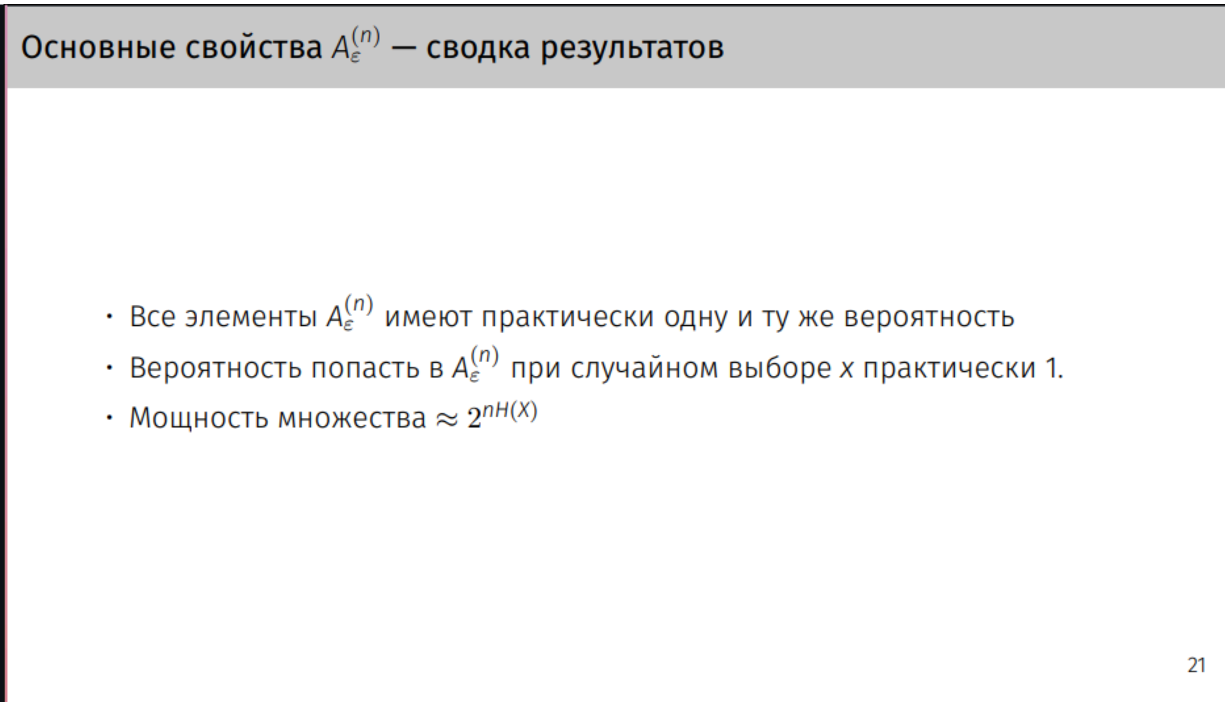
$$|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$$

Доказательство.

$$1 - \varepsilon < Pr\left((x_1, x_2, ..., x_n) \in A_\varepsilon^{(n)}\right) \leq \sum_{x \in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} = 2^{-n(H(X)-\varepsilon)} |A_\varepsilon^{(n)}|,$$

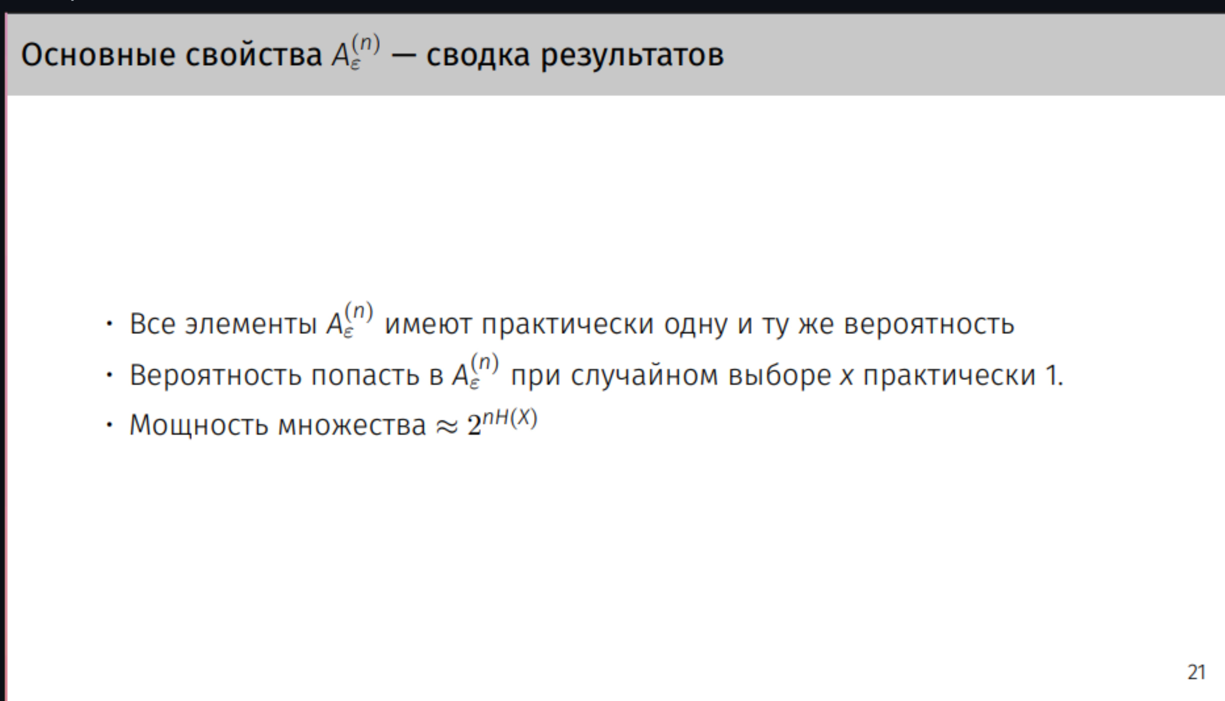
откуда сразу

$$|A_\varepsilon^{(n)}| \geq (1-\varepsilon) 2^{n(H(X)-\varepsilon)}$$



11

11. Мощность типичного множества



12

12. Теорема о вероятности типичного множества

13

13. Теорема Шеннона о кодировании источника

14

14. Понятие высоковероятного множества. Связь типичного множества и высоковероятного множества.

15

15. Понятие префиксного и однозначно-декодируемого кода. Связь между этими понятиями

16

16. Кодирование источника с диадическим распределением

17

17. Свойства диадического распределения

18

18. Коды Шеннона

19

19. Неравенство Крафта

20

20. Коды Хаффмана. Оптимальность кодов Хаффмана

21

21. Понятие кодовой схемы. Достижимая скорость передачи кодовой схемы.

22

22. Пропускная способность канала. Примеры