1. Понятие энтропии случайной величины. Основные свойства энтропии

Энтропия H(U) – это средняя «неожиданность» (или «сюрприз») при наблюдении исходов случайной величины U.

- Представьте, что вы отгадываете, какой из N цветных шариков вытащат из коробки. Чем более равномерно распределены шансы — тем больше «неожиданности» в каждом вытягивании.
- Математически:

$$H(U) = \sum_u p(u)\,\log_2 \frac{1}{p(u)}.$$

- Свойства:
  - **Ненегативность**: H(U) ≥ 0;
  - **Максимум при равновероятном распределении**: если все p(u)=1/|U|, то H(U)=\log_2|U|;
  - **Аддитивность для независимых**: если X и Y независимы, H(X,Y)=H(X)+H(Y).

---

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

Здесь $H_n(U)$ обычно означает энтропию при «неправильном» распределении q, то есть

$$H_q(U) = -\sum_u p(u)\,\log_2 q(u).$$

- Это «перекрёстная энтропия», которая показывает, сколько бит понадобится, если кодировать с распределением q вместо истинного p.
- Связь:

$$H_q(U) = H(U) + D_{\mathrm{KL}}(p\|q),$$

  где $D_{kl}(p\|q)$ – дивергенция Кульбака–Лейблера (см. вопрос 3).

---

3. Условная энтропия и дивергенция Кульбака–Лейблера. Свойства условной энтропии

- **Условная энтропия H(X|Y)** – средняя неопределённость X при известном Y:

$$H(X|Y) = \sum_y p(y)\, H(X|Y=y).$$

- **Дивергенция $D_{kl}(p\|q)$** измеряет «расстояние» между распределениями p и q:

$$D_{\mathrm{KL}}(p\|q) = \sum_x p(x) \log_2 \frac{p(x)}{q(x)}.$$

- Свойства H(X|Y):
  - **Ненегативность**: H(X|Y) ≥ 0;
  - **H(X|Y) ≤ H(X)**: знание Y не увеличивает неопределённость про X;
  - **Аддитивность**: H(X,Y)=H(Y)+H(X|Y).

---

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

- **Совместная энтропия H(X,Y)** – мера «сюрприза» при паре (X,Y):

$$H(X,Y) = -\sum_{x,y} p(x,y) \log_2 p(x,y).$$

- Если X и Y независимы, то p(x,y)=p(x)p(y) ⇒ H(X,Y)=H(X)+H(Y).
- Свойства: симметрия H(X,Y)=H(Y,X), и связь с условной: H(X,Y)=H(Y)+H(X|Y).

---

5. Понятие взаимной информации

I(X;Y) показывает, сколько информации об X даёт наблюдение Y:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

- Аналог: перекрёстное пересечение двух кругов «информации» X и Y.
- I(X;Y)≥0 и равно нулю, если X и Y независимы.

---

6. Базовые свойства взаимной информации

- **Ненегативность**: $I(X;Y) \geq 0$.
- **Симметрия**: $I(X;Y)=I(Y;X)$.
- **Аддитивность для независимых пар**: если $(X_1,Y_1) \perp\!\!\!\perp (X_2,Y_2)$, то $I((X_1,X_2);(Y_1,Y_2))=I(X_1;Y_1)+I(X_2;Y_2)$.

---

7. Выпуклость дивергенции Кульбака–Лейблера

Функция $(p,q) \mapsto D_{KL}(p\|q)$ выпукла по своей паре распределений:

$$D_{\mathrm{KL}}(\lambda p_1 + (1-\lambda)p_2 \| \lambda q_1 + (1-\lambda)q_2) \leq \lambda D_{\mathrm{KL}}(p_1\|q_1) + (1-\lambda) D_{\mathrm{KL}}(p_2\|q_2).$$

Интуиция: «смешанный» подход не даёт больше расхождения, чем смешение отдельных дивергенций.

---

8. Закон больших чисел

Если $X_1, \ldots, X_n$ — iid с математическим ожиданием μ, то

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu,$$

то есть среднее по выборке сходится к истинному среднему при $n \to \infty$.

---

9. AEP-теорема

Асимптотическая равномерная вероятность (Asymptotic Equipartition Property) говорит, что для большого n подавляющее число последовательностей длины n имеют вероятность примерно 2^{-nH}, где H — энтропия источника. Тем самым «типичное множество» содержит почти всю массу вероятности.

---

10. Определение типичного множества. Основные свойства типичных множеств

- **Типичное множество A_n^ε** – все последовательности x^n, чья эмпирическая энтропия близка к H:

$$\left| -\frac{1}{n}\log p(x^n) - H \right| < \varepsilon.$$

- Свойства:

- **Высокая вероятность**: $P(X^n \in A_n^\varepsilon) \to 1$;
- **Мощность ≈ 2^{nH}**;
- **Почти равные вероятности внутри A_n^ε**.

---

11. Мощность типичного множества

$|A_n^\varepsilon| \approx 2^{nH}$. То есть количество «типичных» последовательностей растёт экспоненциально с n, с показателем H.

---

12. Теорема о вероятности типичного множества

$P(X^n \in A_n^\varepsilon) \geq 1-\delta$ для любых $\varepsilon > 0$ и достаточно большого n. Это гарантирует, что почти все наблюдаемые последовательности будут «типичными».

---

13. Теорема Шеннона о кодировании источника

Для дискретного источника U с энтропией H(U):

- **Прямая часть**: можно построить код длины ≈nH бит/символ с малой вероятностью ошибки декодирования.
- **Обратная часть**: нельзя сделать среднюю длину меньше H(U).

---

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

Высоковероятное множество – любой набор X^n с $P(X^n) \geq 1-\varepsilon$. Типичное множество – пример высоковероятного с дополнительно почти равными вероятностями внутри.

---

15. Понятие префиксного и однозначно-декодируемого кода

- **Префиксный код**: ни один кодовыйword не является началом другого (пример: коды Хаффмана).
- **Однозначно-декодируемый**: любую последовательность кодов можно разбить лишь одним способом. Префиксность ⇒ однозначность.

---

16. Кодирование источника с диадическим распределением

Диадическое распределение: p(u)=2^{-k} для целых k. Тогда можно строить простые двоичные коды, длина которых целочисленна.

---

17. Свойства диадического распределения

- Длины кода $l(u) = -\log_2 p(u)$ целочисленны;
- Средняя длина равна энтропии: EL=H(U).

---

18. Коды Шеннона

Шеннон предложил код, где каждому символу u даётся длина $\lceil -\log_2 p(u) \rceil$. Это гарантирует EL<H(U)+1.

---

19. Неравенство Крафта

Для префиксных кодов длины $l_1,\ldots,l_m$ должно выполняться

$$\sum_{i=1}^{m} 2^{-l_i} \leq 1.$$

Обратное также верно: любое множество длин, удовлетворяющее этому неравенству, задаёт префиксный код.

---

20. Коды Хаффмана. Оптимальность кодов Хаффмана

Хаффман строит префиксный код, минимизирующий среднюю длину для заданных p(u). Он объединяет наименее вероятные символы рекурсивно. Оптимальность гарантируется жадным алгоритмом.

---

21. Понятие кодовой схемы. Достижимая скорость передачи кодовой схемы

Кодовая схема = выбор кодов для блоков символов + правила кодирования/декодирования. Достижимая скорость R = (число информационных бит)/(число переданных символов).

---

22. Пропускная способность канала. Примеры

C – максимальная скорость передачи через канал при произвольно малой ошибке.

- Для двоичного симметричного канала (BSC с вероятностью ошибки τ):

$$C = 1 - H_2(\tau),$$

  где H_2 – бинарная энтропия.

---

23. Пропускная способность двоичного симметричного канала

См. предыдущий пункт: $C_{BSC}(\tau)=1-H_2(\tau)$.

---

24. Пропускная способность двоичного стирающего канала

Для BEC (вероятность «стирания» ε):

$$C = 1 - \varepsilon,$$

потому что при стирании бит просто исчезает и требует повторной передачи.

---

25. Дифференциальная энтропия и взаимная информация. Их свойства

- **Дифференциальная энтропия h(X)** для непрерывной X с плотностью f(x):

$$h(X) = -\int f(x)\log_2 f(x)\,dx.$$

- **Взаимная информация** $I(X;Y)=h(X)-h(X|Y)$ сохраняет те же свойства (ненегативность, симметрию).

---

26. Энтропия нормального распределения

Для $X\sim N(\mu,\sigma^2)$:

$$h(X) = \tfrac{1}{2}\log_2(2\pi e\sigma^2).$$

Это максимальная дифференциальная энтропия при данном разбросе σ^2.

27) Совместные типичные последовательности и их свойства

Представьте, что у вас есть две игрушки — мячик **X** и машинка **Y**, и они прыгают и едут вместе. Совместные типичные последовательности — это такие «пары движений» (мячик вверх/машинка вперёд) длины n, которые «обычно» случаются вместе и очень вероятны, если наблюдать много раз.

- **Высокая вероятность**: почти все наблюдаемые пары попадают в это множество.
- **Размер ≈ 2^{nH(X,Y)}**: число таких «обычных» пар растёт экспоненциально.
- **Почти равные вероятности**: внутри множества каждая пара движений почти одинаково вероятна.

---

28) Прямая теорема Шеннона

Шеннон сказал: если ты хочешь передать буквы алфавита по трубе с шумом, и скорость передачи **R** меньше пропускной способности **C**, то можно придумать способ кодировать так, что **ошибок почти не будет**.

- При **R < C** существует код, где вероятность ошибки ≤ ε (сколько угодно малое).
- Это — «прямая» (achievability): как построить хороший код.

---

29) Неравенство Фано

Когда вы стараетесь отгадать, что нарисовано на картинке (X) по подсказке (Y), Фано говорит, что если ошибки $\mathbf{P_e}$ велики, то условная энтропия **H(X|Y)** тоже большая, то есть

$$H(X|Y) \leq H_2(P_\text{e}) + P_\text{e} \log_2(|X| - 1).$$

Интуитивно: чем чаще вы ошибаетесь, тем более неопределённым остаётся X после Y.

---

30) Обратная теорема Шеннона

Это «конверсе»: если **R > C**, то никакие хитрые коды не помогут — **ошибки не уйдут**.

---

31) Понятие линейного кода. Базисные матрицы линейного кода

Линейный код — это набор бинарных слов (кодов), которые могут складываться как векторы.

- **Генераторная матрица G**: каждая строка — базис, и любая комбинация строк даёт кодовое слово.
- **Проверочная матрица H**: если умножить кодовое слово на H^T, получится ноль — это гарантия, что слово «правильное».

---

32) Конструкция полярного кода

Полярный код «разделяет» канал на хорошие и плохие-точки:

1. Берём n=2^m битов,
2. Применяем особое преобразование (точка-байточный XOR-плетёнка),
3. «Хорошие» биты передаём информацией, «плохие» — фиксируем нулями.

---

33) Кодирование полярного кода

Чтобы закодировать:

1. Собираем в вектор u длины n: информационные биты на «хороших» позициях, нули на «плохих».
2. Умножаем u на матрицу конструктора (Kronecker-произведение базиса).

---

34) SC-декодирование полярного кода

SC = successive cancellation. Декодируем биты один за другим, используя уже принятые ранее решения, двигаясь слева направо.

---

35) SCL-декодирование полярного кода

SCL = successive cancellation list. Похож на SC, но держит несколько (L) самых «правдоподобных» вариантов промежуточных решений, чтобы снизить количество ошибок.

---

36) Сжимающий код, понятие R(D), D®, Rate–Distortion

Когда мы кодируем картинки с потерями, появится «искажение» D и скорость R (бит/символ).

- **R(D)** — минимальная скорость, при которой среднее искажение ≤ D.
- **D®** — минимально достижимое искажение при скорости R.

---

37) Вывод формулы R(D) для X~Ber§

Для случайного бита с P(1)=p и H_2 — бинарная энтропия:

$$R(D) = H_2(p) - H_2(D),$$

для 0≤D≤min(p,1−p).

---

38) Вывод формулы R(D) для нормальной СВ

Для X~N(0,σ²) и среднеквадратичного искажения D:

$$R(D) = \tfrac{1}{2}\log_2 \frac{\sigma^2}{D}, \quad 0 < D < \sigma^2.$$

---

39) Вывод R(D) для векторной нормальной СВ

Если X — вектор из k независимых N(0,σ_i²), то

$$R(D) = \frac{1}{2}\sum_{i=1}^{k} \log_2 \frac{\sigma_i^2}{D_i},$$

где ∑D_i = D, оптимальное распределение искажений.

---

40) Теорема о выпуклости R(D)

R(D) — выпуклая функция D: смешивая две схемы с (R_1,D_1) и (R_2,D_2), можно получить скорость $\le \lambda R_1+(1-\lambda)R_2$ при искажении $\lambda D_1+(1-\lambda)D_2$.

---

41) Доказательство обратной теоремы Шеннона для сжимающего кода

Конверсе: если R<R(D), то при любом коде среднее искажение **E[d(X,\hat X)]** будет > D. Это доказывается через оценки энтропии и KL-дивергенцию.

---

42) Понятие ε-типичных искажений и множеств, их основные свойства

Как типичное множество для сообщений, но для пар (x,ŷ):

- В нём находятся «обычные» пары, где среднее искажение близко к D.
- **Высокая вероятность**: P((X^n,Ŷ^n)∈A^n_ε)→1.
- **Мощность ≈2^{nR(D)}**.

---

43) Доказательство прямой теоремы Шеннона для сжимающего кода

Achievability: берём случайную «книгу кодов» из типичных Ŷ^n, находим для каждого x^n пару с искажением ≤ D. Вероятность успеха → 1, и нужно ≈2^{nR(D)} кодов.

---

44) Сильно типичные последовательности и их свойства

Сильная типичность следит не только за общей вероятностью, но и за **частотами** каждого символа:

- Для каждого символа a частота ≈ p(a) с погрешностью ε.
- Множество сильной типичности тоже имеет P→1 и размер ≈2^{nH}.

---

45) Доказательство достижимости R(D) для сильно типичных последовательностей

Как и в обычном кодировании с потерями, но подставляем сильную типичность, чтобы жёстко контролировать частоты искажения. Это делает доказательство более строгим, сохраняя нужный объём кодов ≈2^{nR(D)}.