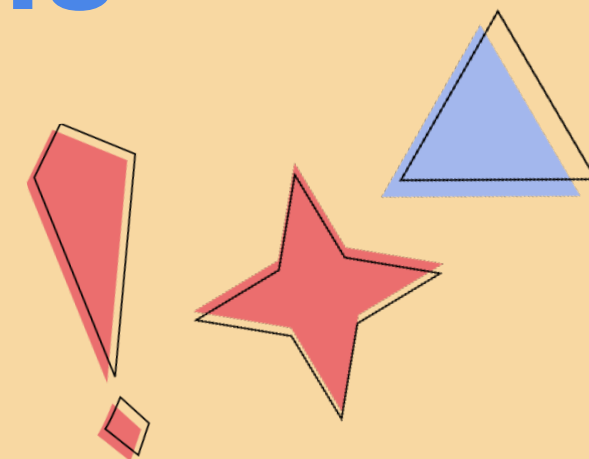




Информация, вероятность и энтропия: как наука описывает неопределённость

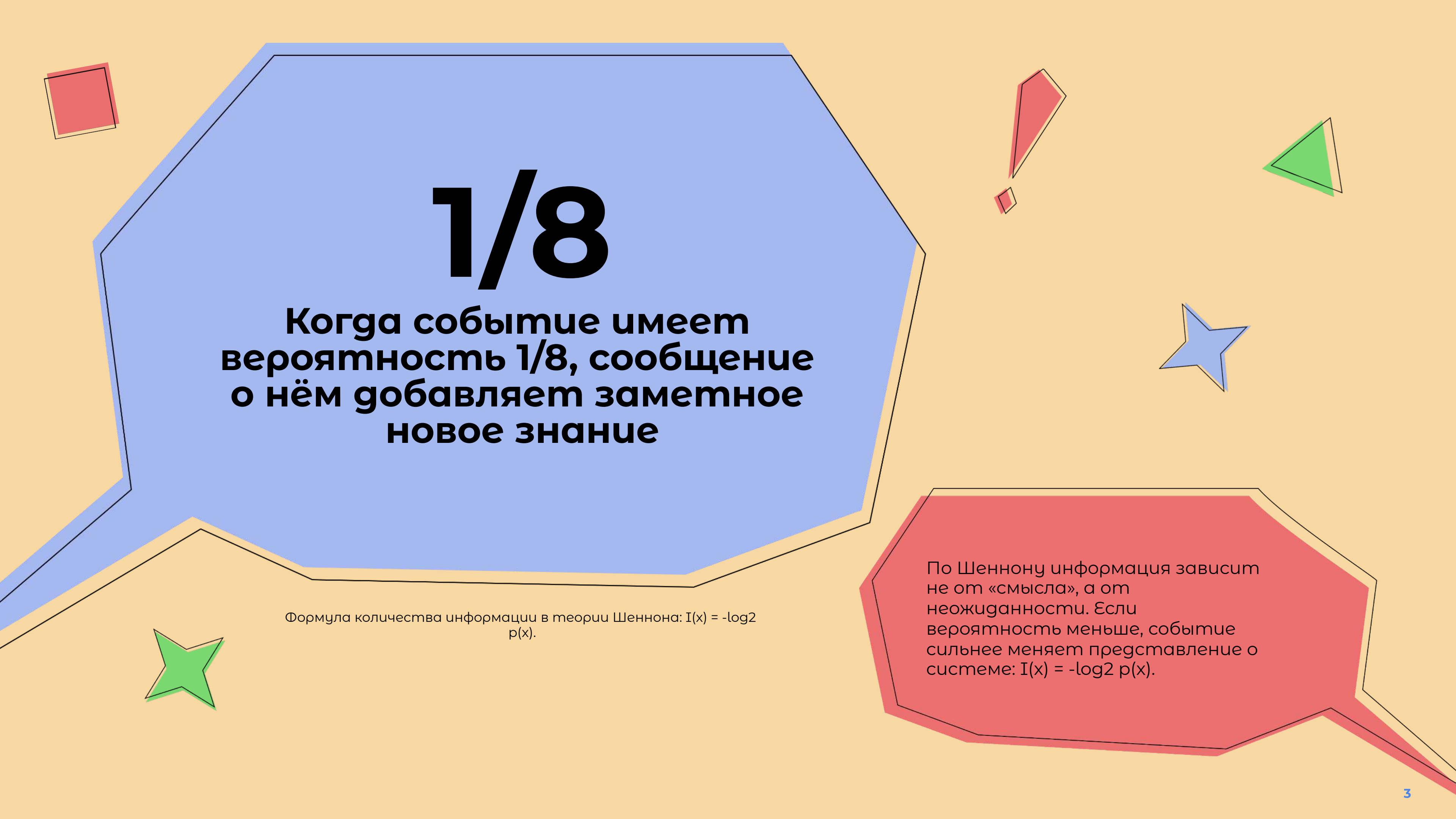
Разберём, как измеряют неожиданность, используя вероятность и энтропию в науке.



Почему эта тема входит в естествознание 11 класса

В естествознании 11 класса вероятностный язык нужен, чтобы описывать случайность в физике и информатике. Энтропия связывает теорию информации с термодинамикой и помогает количественно сравнивать сложность сообщений и систем.





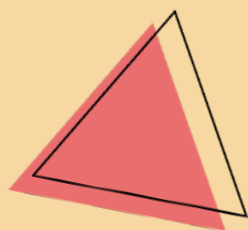
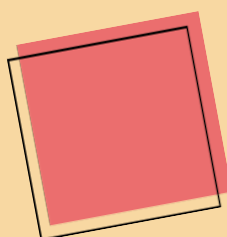
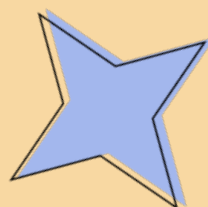
1/8

**Когда событие имеет
вероятность 1/8, сообщение
о нём добавляет заметное
новое знание**

Формула количества информации в теории Шеннона: $I(x) = -\log_2 p(x)$.

По Шеннону информация зависит не от «смысла», а от неожиданности. Если вероятность меньше, событие сильнее меняет представление о системе: $I(x) = -\log_2 p(x)$.

Вероятность события и количество информации

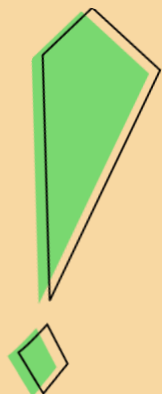
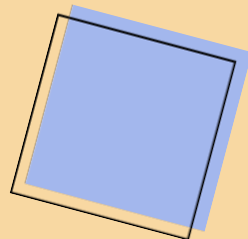
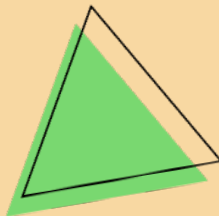


Вероятность p	Количество I , бит
$1/2$	1
$1/4$	2
$1/8$	3
$1/16$	4
1	0

Чем меньше вероятность, тем больше информации несёт сообщение о наступлении события. При каждом удвоении «редкости» прибавляется 1 бит, а при $p=1$ информация равна нулю.

Информация растёт логарифмически: редкие события информативнее частых; при $p=1$ событие полностью предсказуемо.

$I(x) = -\log_2 p(x)$, определение количества информации в теории Шеннона.



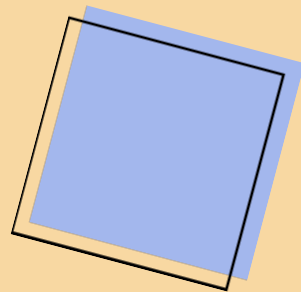
Формула Шеннона и смысл логарифма

Почему именно логарифм по основанию 2

Количество информации для одного исхода: $I(x) = -\log_2 p(x)$. Основание 2 выбирают, потому что единицей измерения служит бит, а значения удобно суммировать для независимых событий.

Как «читать» шкалу редкости

Логарифм превращает вероятности в шкалу редкости: частые исходы дают малую информацию, редкие — большую. Важно помнить: формула описывает вероятностную модель, а не эмоциональную «значимость» события.



Как растёт информация при уменьшении вероятности

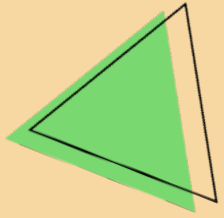
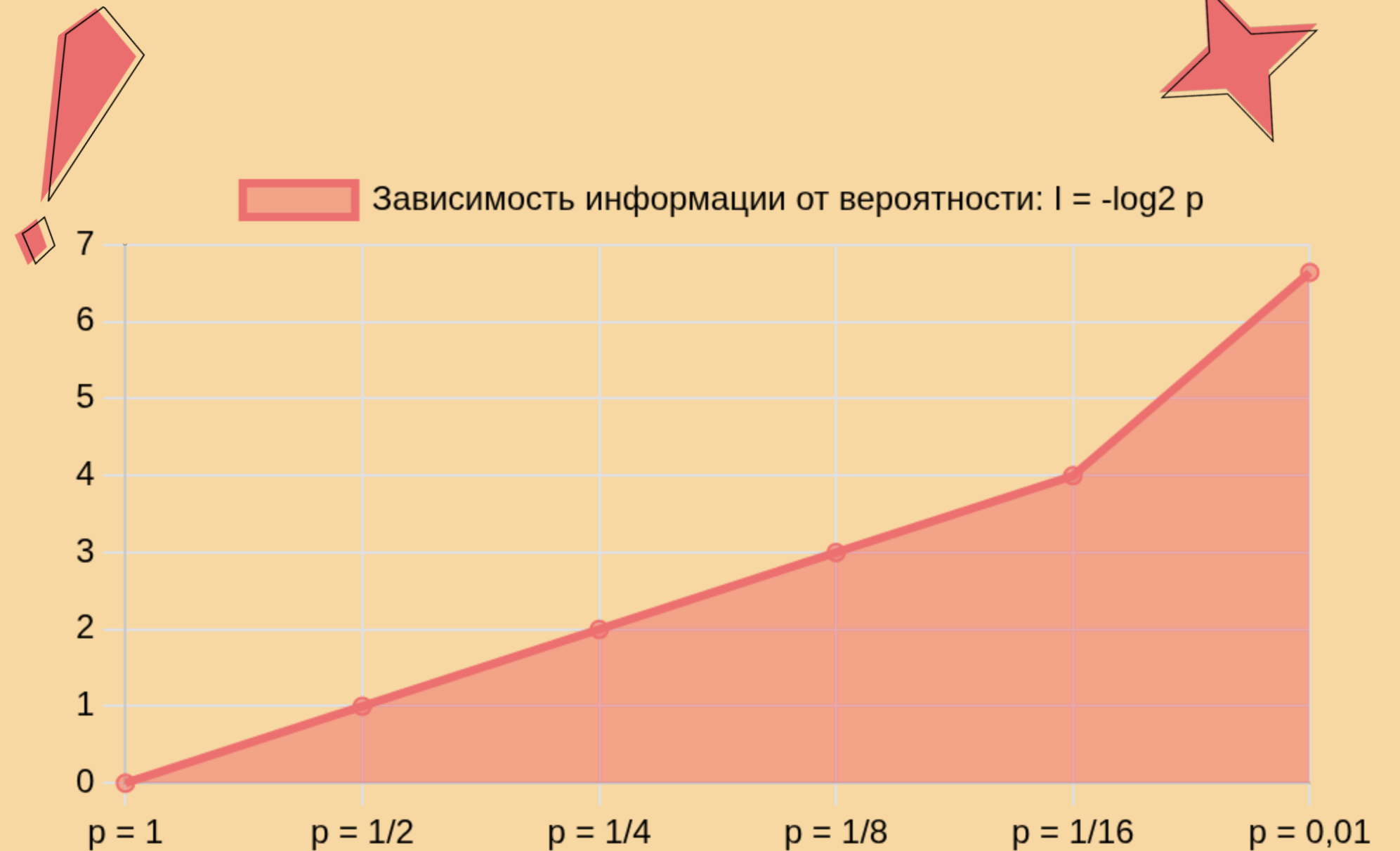
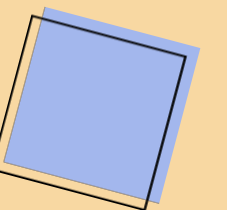
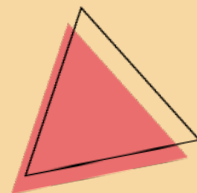
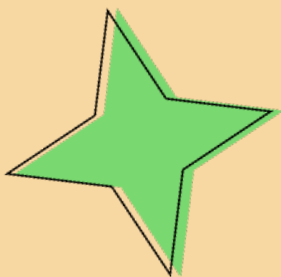


График подчёркивает неравномерность: падение вероятности сильнее «поднимает» I при малых p , поэтому редкие случаи особенно информативны.

Почти у нуля значения растут быстрее всего: небольшая дополнительная «непредсказуемость» даёт большую информационную прибавку.



Построено по формуле $I = -\log_2 p$; опорные точки: $p = 1, 1/2, 1/4, 1/8, 1/16$ и $p = 0,01$.





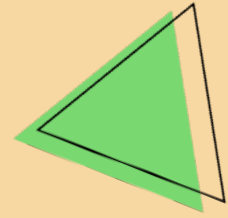
2,585

**Для идеального
шестигранного кубика
энтропия близка к 2,585 бита,
потому что шесть исходов
равновероятны**

Энтропия равномерного распределения из 6 исходов: $H = \log_2 6 \approx 2,585$.

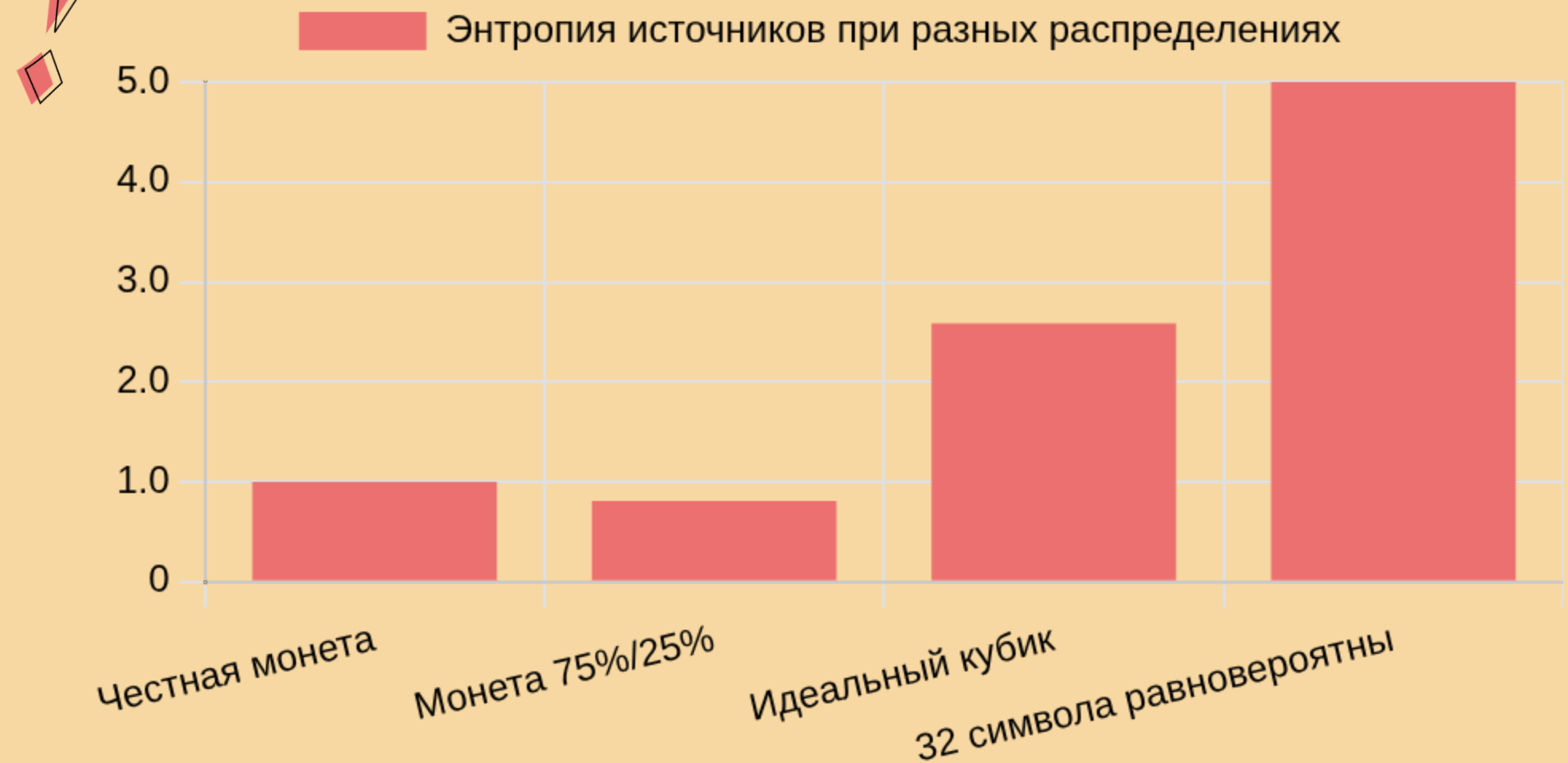
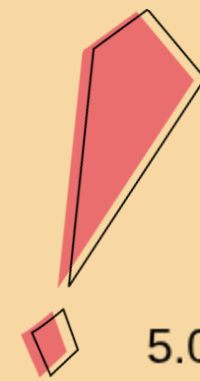
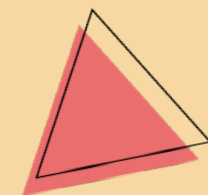
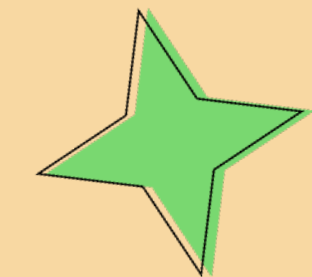
Энтропия $H = -\sum p(x) \log_2 p(x)$
задаёт среднее количество
информации на одно сообщение.
Чем равномернее распределение,
тем выше H и тем меньше
предсказуемость.

Сравнение энтропии разных источников

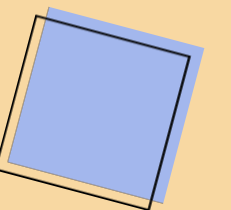


При одинаковом числе исходов энтропия определяется не «количеством» вариантов само по себе, а тем, насколько вероятности различаются.

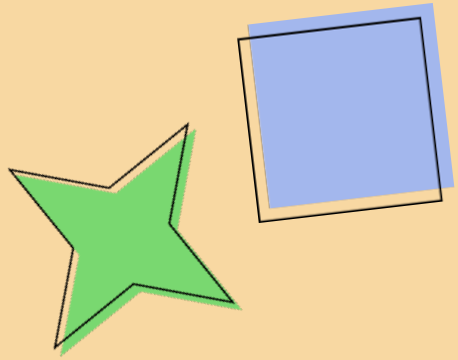
Равномерное распределение даёт максимум энтропии; при перекосе вероятностей средняя неопределённость уменьшается.



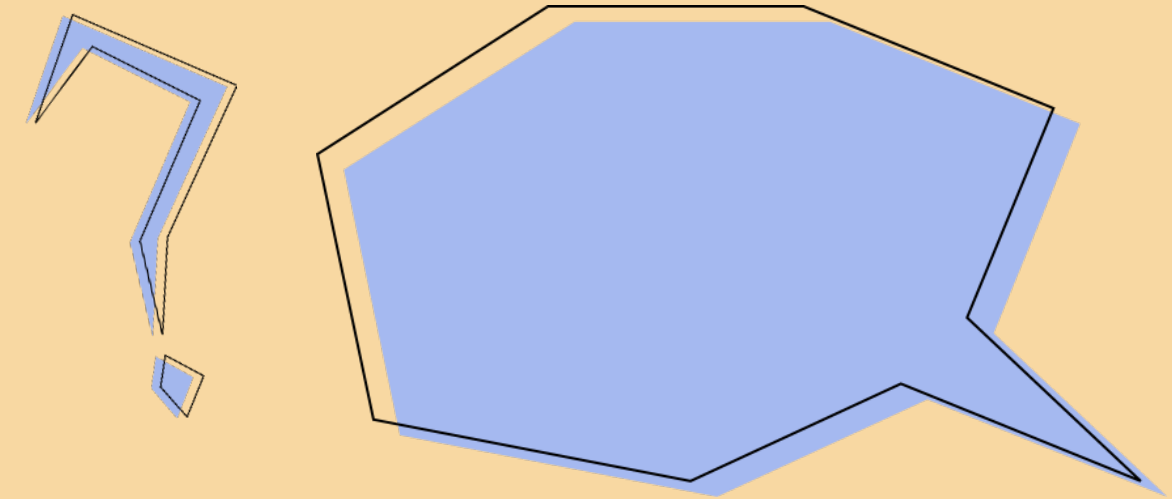
Значения рассчитаны по $H = -\sum p(x) \log_2 p(x)$ для указанных распределений (включая $\log_2 32 = 5$).



Максимальная и минимальная энтропия

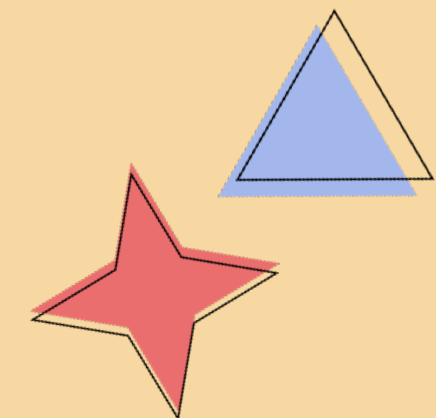
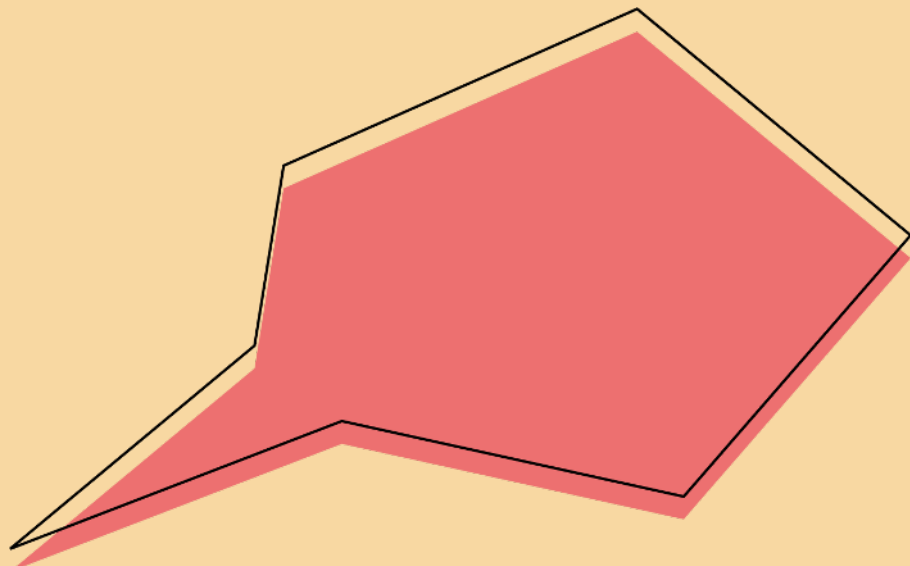


Максимальная энтропия достигается, когда все исходы равновероятны: тогда результат труднее предсказать и средняя неопределённость максимальна.



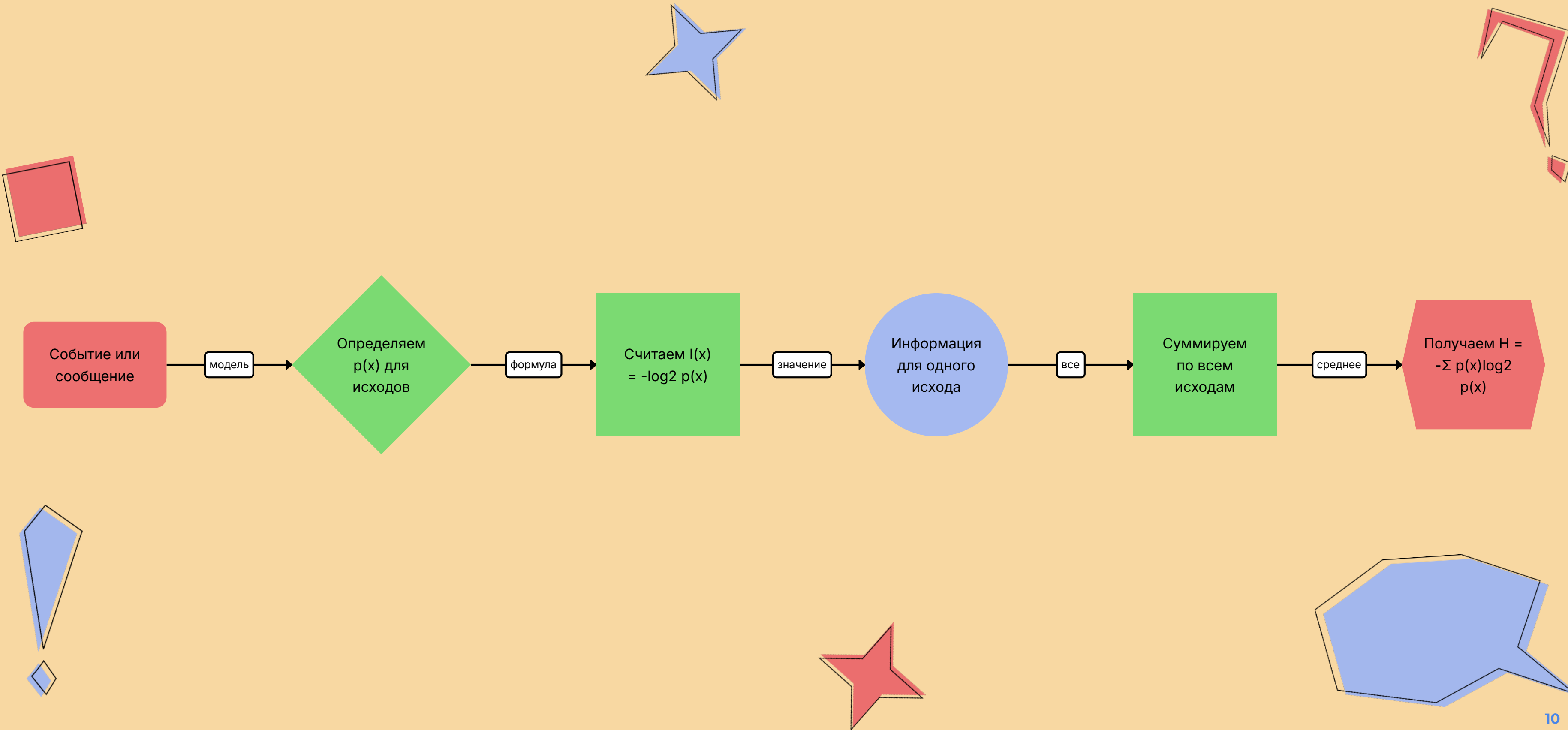
Минимальная энтропия равна нулю в полностью предсказуемой системе, где возможен только один исход: сообщение не добавляет новой информации.

При сравнении «более случайных» и «более упорядоченных» сообщений видно: чем больше повторяемости и неравномерности, тем ниже энтропия и ниже средняя информационная неопределённость.



Как вычисляются информация и энтропия

Схема расчёта по вероятностной модели распределения исходов



Почему редкие события особенно заметны



«Ливень после затяжной паузы»

Когда долгое время держится сухая погода, дождь воспринимается как сильное отклонение от ожиданий. По вероятностной модели событие с малой $p(x)$ приносит больше информации: оно заметно меняет наше представление о ситуации. Поэтому новость «начался дождь» ощущается информационно насыщенной.



Монета, которая «почти поверила в себя»

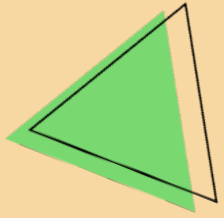
Если монета десять раз подряд даёт один и тот же результат, мозг начинает подмечать закономерность: серия «выглядит необычной» в следующем испытании. Хотя отдельный опыт имеет свою реализацию, в вероятностной логике снижается ожидаемость привычного паттерна. Именно разрыв между ожиданием и результатом делает сообщение заметным.



Поворот сюжета в случайной системе

Представьте поиск нужного файла в огромной папке: находка при первой попытке редко предсказуема. Когда вероятность исхода мала, каждое подтверждение становится информационным событием. Даже один факт может существенно сократить неопределённость — поэтому редкость часто «высвечивает» значимость сообщения.

Неравномерное распределение вероятностей и его влияние на энтропию



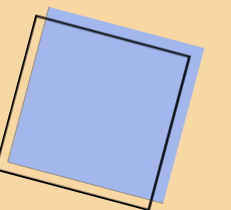
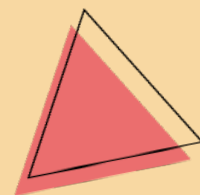
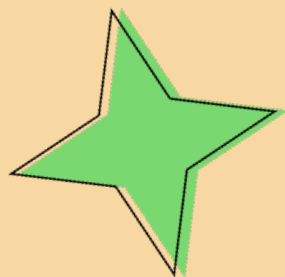
Когда один исход встречается заметно чаще, неопределённость падает: знание о наиболее вероятном варианте экономит «информационный ресурс» на последующих сообщениях.

При разности вероятностей энтропия ниже, чем в равномерном случае четырёх исходов: ожидаемый результат чаще, а значит, средняя информационная неопределённость уменьшается.

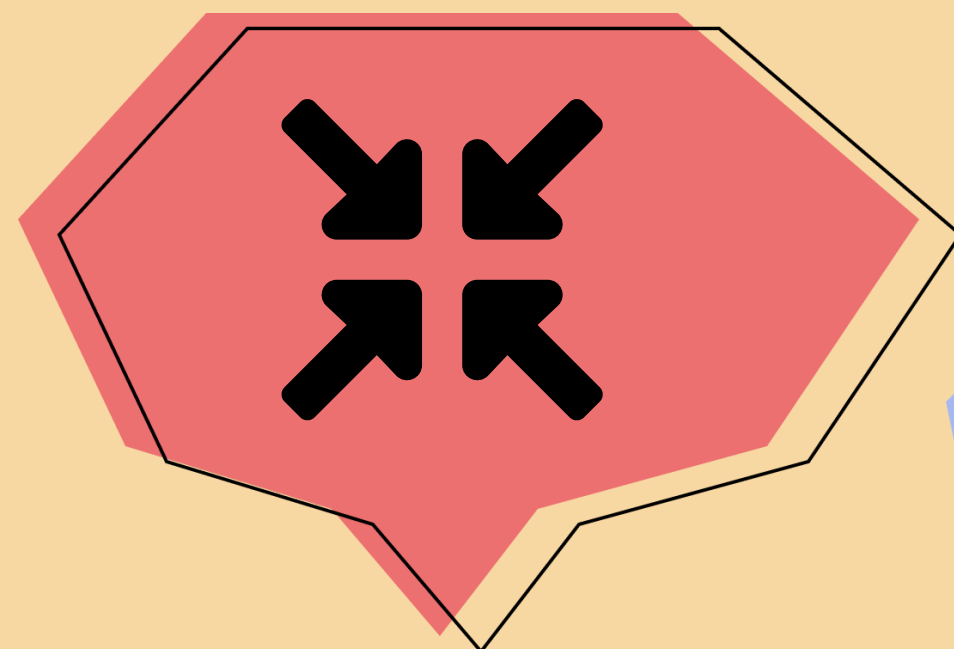
A B C D



Набор вероятностей для иллюстрации: $A=0,5$; $B=0,25$; $C=0,125$; $D=0,125$ (расчёт по формуле энтропии Шеннона, без внешних данных).

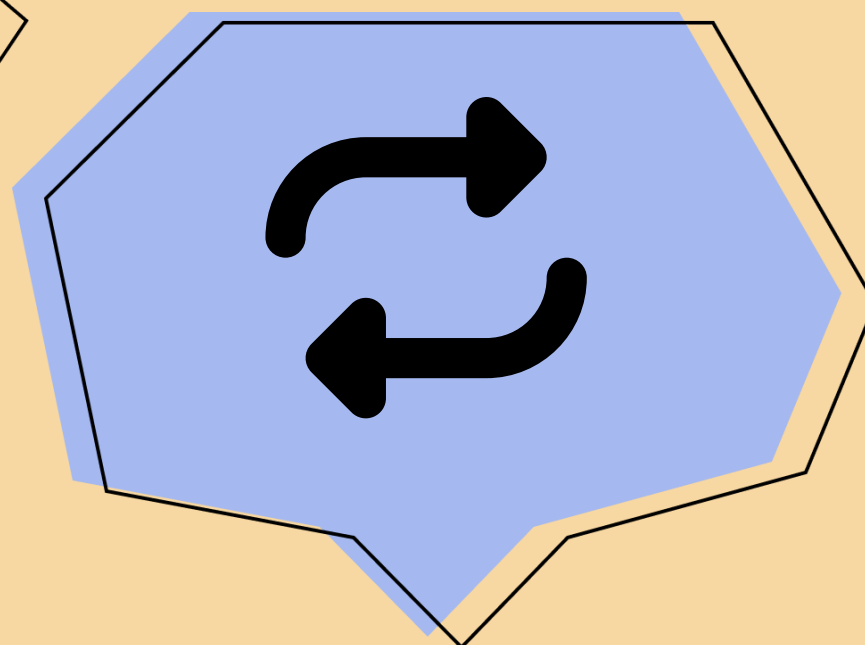
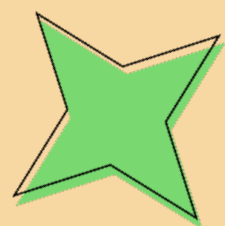


Энтропия и сжатие данных



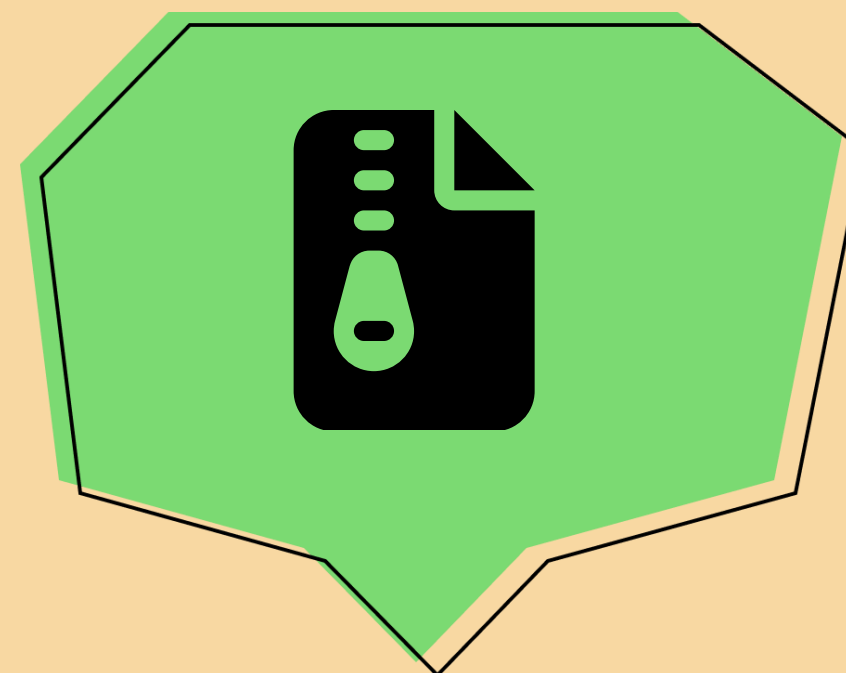
Насколько «сжимаемо» сообщение

Высокая энтропия означает, что в данных мало предсказуемости и повторяемость слабая. Такое сообщение труднее представить короче без потерь, поскольку нечего «угадать» заранее.



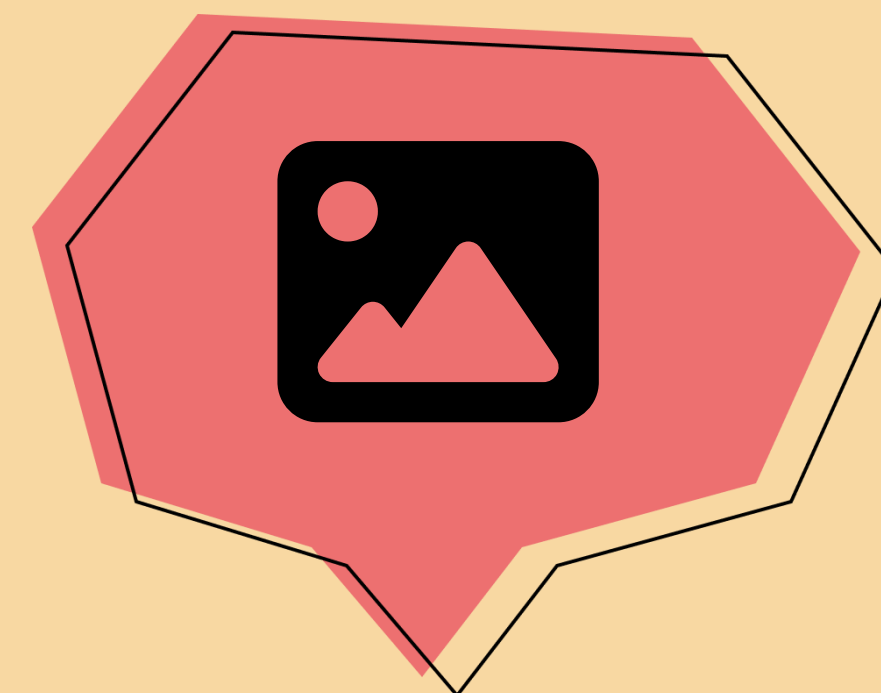
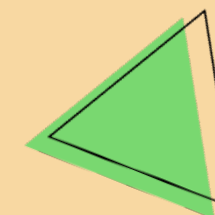
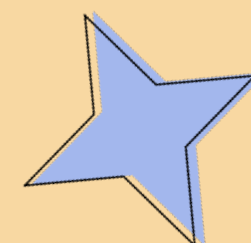
Повторы уменьшают неопределённость

Если встречается много одинаковых символов или устойчивых фрагментов, вероятности становятся неравномерными. Тогда энтропия ниже, а кодирование может использовать более короткие обозначения для частых элементов.



ZIP: экономия на закономерностях

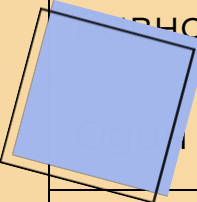
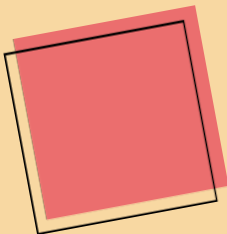
Алгоритмы типа ZIP извлекают из текста статистические структуры: когда одни фрагменты повторяются чаще, их можно описывать компактнее. Сжатие особенно хорошо работает при низкой энтропии.



PNG и предсказуемые пиксели

Для изображений PNG стремится кодировать повторяющиеся участки и предсказуемые изменения яркости. Чем регулярнее структура, тем меньше энтропия и тем эффективнее конечный размер файла.

Избыточность и эффективность кодирования



Ситуация	Пример вероятностей	Эффективная длина кода	Комментарий
Равномерный алфавит	8 символов	3 бита/символ	все равновероятны
Один символ чаще	один в 2 раза чаще	код короче	для частого символа
Неравномерность	вероятности различны	средняя длина меньше	экономия на частых
Алгоритмы	Шеннона–Фано	построение кодов	экономные префиксные
Алгоритмы	Хаффмана	построение кодов	очень эффективные

Обобщение школьных свойств оптимальных префиксных кодов и оценок средней длины по вероятностям (качественные выводы; без внешних экспериментальных датасетов).

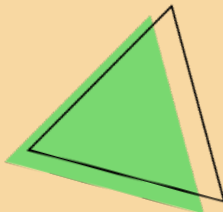
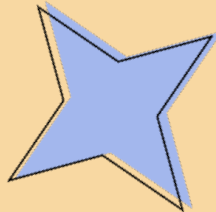


Таблица показывает, как длина кода зависит от вероятности символов: более вероятные символы кодируются короче. Отсюда следует связь между неравномерностью распределения и средней длиной сообщения.



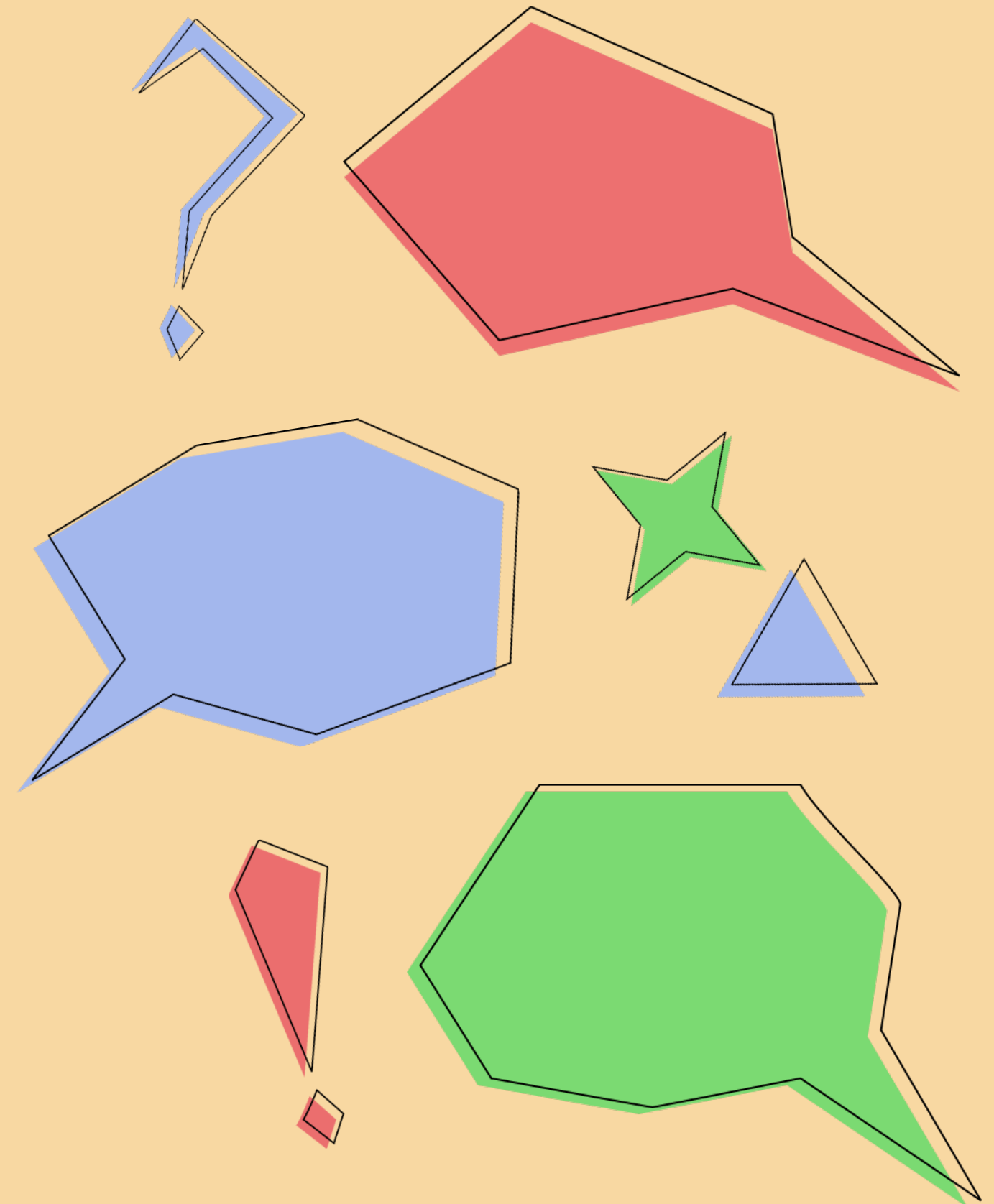
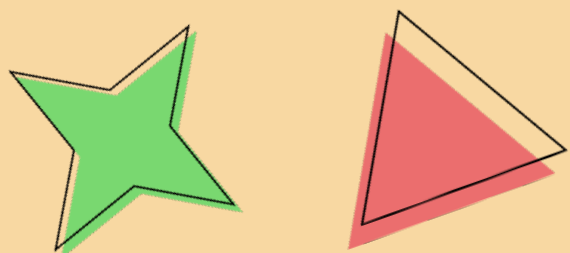
Чем выше вероятность символа, тем меньше может быть длина его кодового слова. При неравномерном распределении средняя длина сообщения сокращается.

Связь с физикой: энтропия в природе

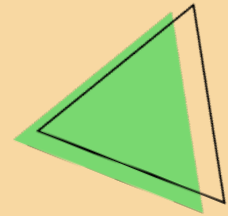
В физике энтропия описывает не «хаос в голове», а число возможных микросостояний системы. Чем больше вариантов внутреннего состояния, тем сильнее неопределённость: система меньше поддаётся точному предсказанию.

Формула Больцмана $S = k \ln W$ показывает: при увеличении W энтропия растёт. Здесь W — количество микросостояний, а k — постоянная Больцмана. Логарифм делает рост умеренным и согласует масштабы.

В термодинамике энтропия связана с направленностью процессов: переход от более упорядоченного к менее упорядоченному состоянию обычно сопровождается ростом числа доступных вариантов. Информационный смысл усиливает эту картину — больше вариантов значит больше неопределённости.

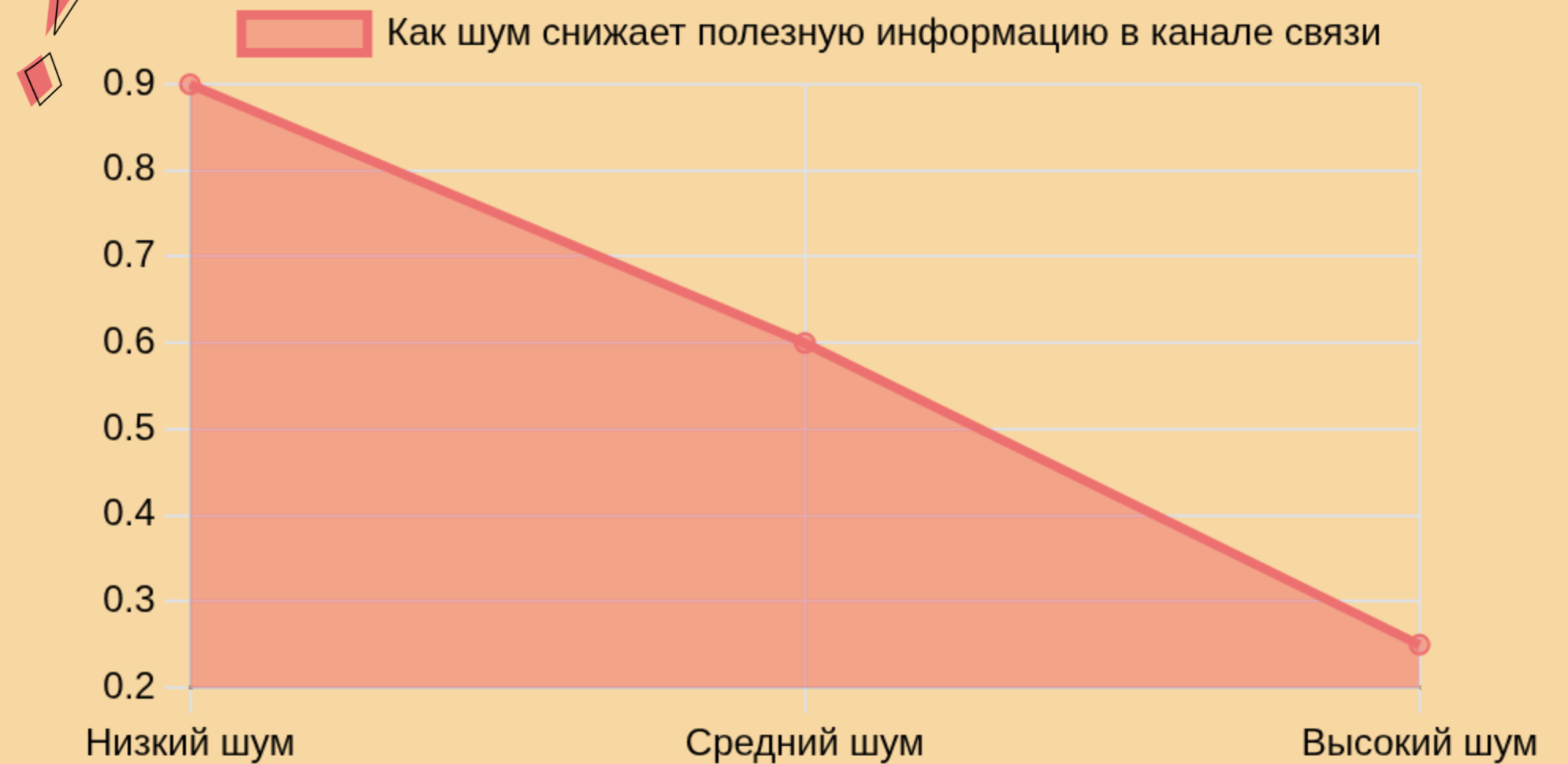
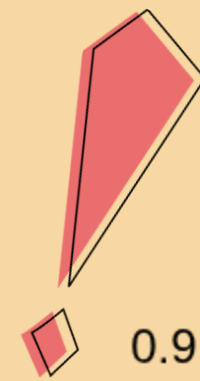
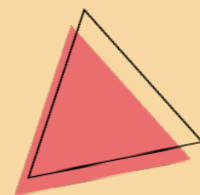
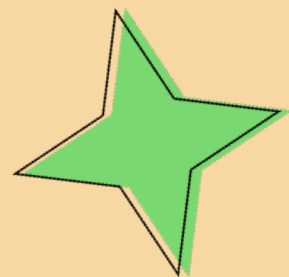


Информация, шум и надёжность связи

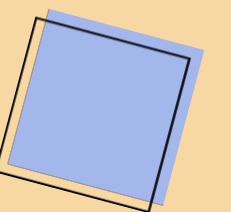


При росте шума наблюдаемый результат всё чаще отличается от исходного. Тогда растёт вероятность ошибок и уменьшается доля полезной информации, извлекаемой из сообщений.

Сравнение показывает: надёжность передачи определяется не одной мощностью сигнала, а тем, насколько сигнал различим на фоне шума; иначе информация «теряется».

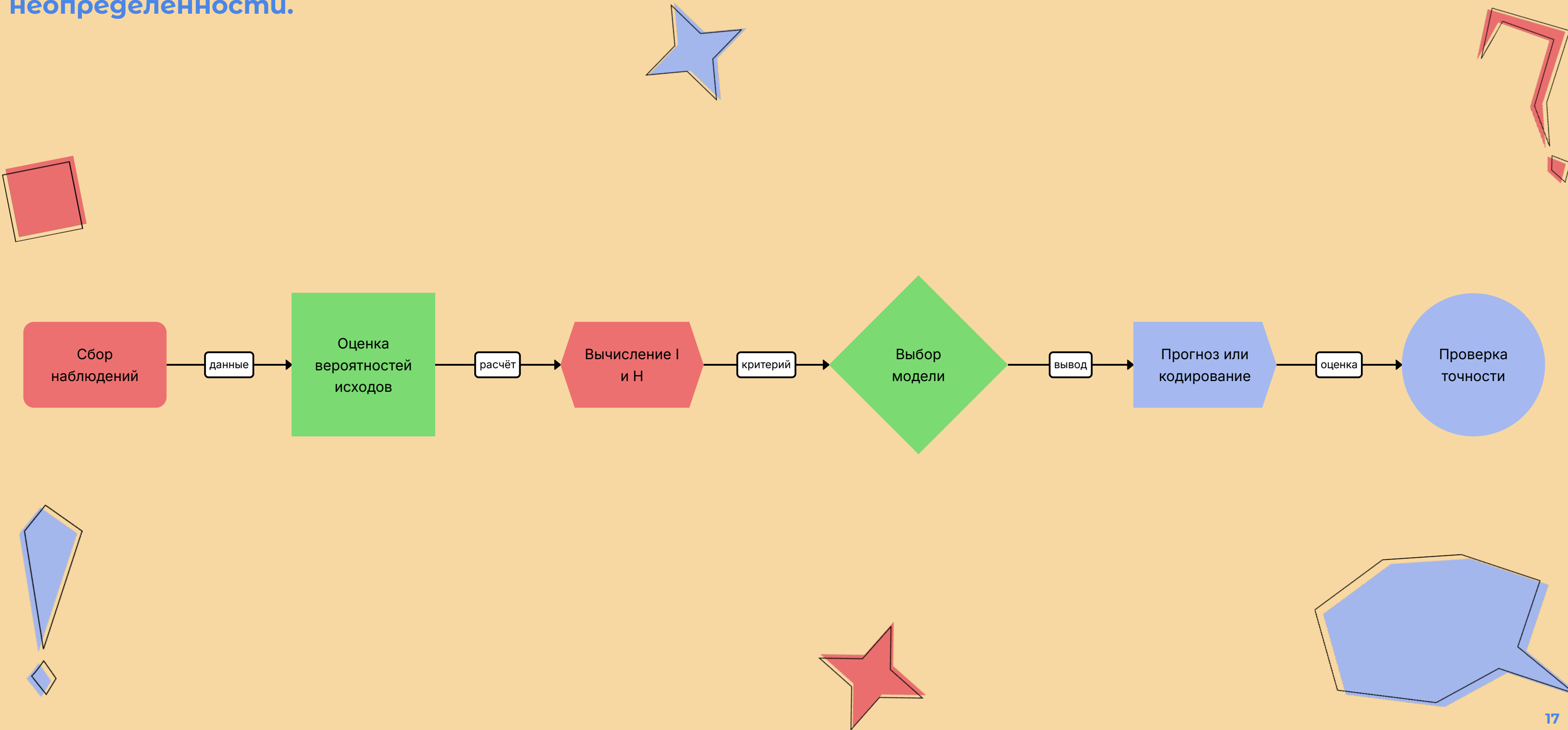


Иллюстративная схема без привязки к конкретному эксперименту: три уровня шума и относительное изменение «полезной доли». В реальности значения зависят от модели канала.



От вероятности к решению в анализе данных

Схема обобщает типичный вероятностный рабочий процесс из наблюдений и расчётов неопределённости.

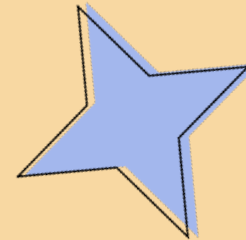


Где используются вероятность и энтропия сегодня



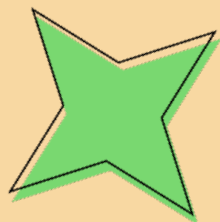
Сжатие файлов и изображений

При кодировании учитывают распределение символов и фрагментов. Энтропия помогает понять, есть ли в данных избыточность и насколько эффективно можно уменьшить объём без потерь.



Распознавание речи

Речь не фиксирована: произношение меняется. Вероятностные методы подбирают наиболее правдоподобную последовательность слов по статистике звуков и контексту, снижая неопределённость.



Биоинформатика и генетика

Вероятностные модели помогают сравнивать ДНК-последовательности и искать участки с похожими функциями. Неравномерность встречаемости нуклеотидов отражает биологические ограничения и контекст.



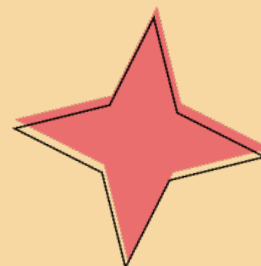
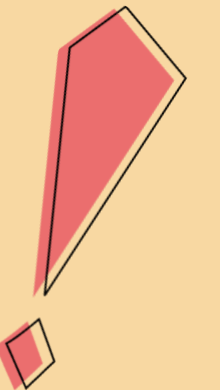
Мобильная связь и передача по каналам

Помехи и ограниченная пропускная способность заставляют использовать вероятностные модели ошибок. Энтропия связана с тем, сколько данных можно надёжно передать и где нужны исправления.



Поисковые системы и анализ данных

Статистика помогает ранжировать результаты и оценивать вероятность гипотез. Энтропия служит индикатором сложности данных: чем выше неопределённость, тем осторожнее выводы.



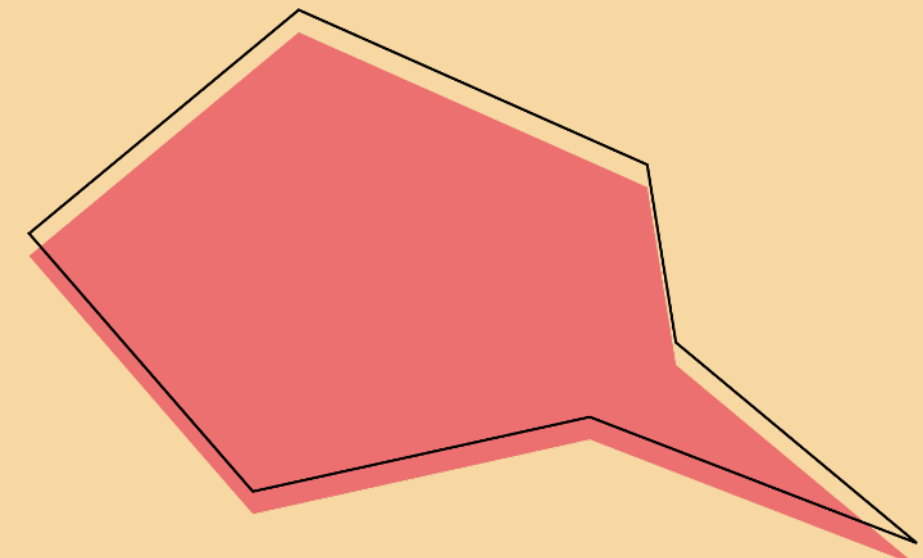
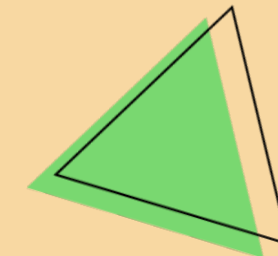
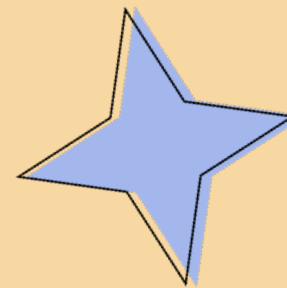
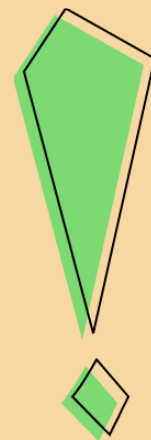
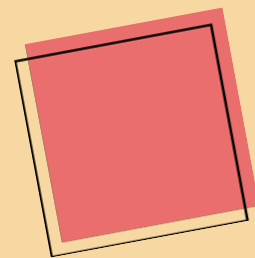
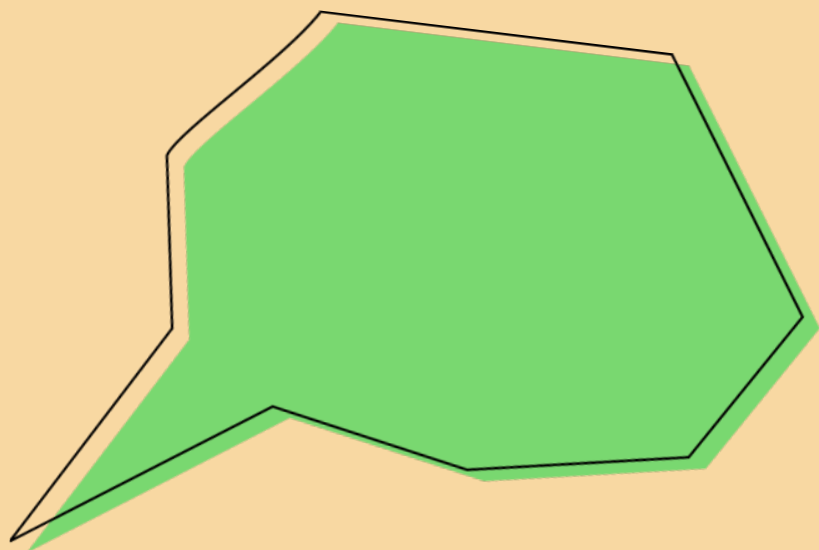
Типичные ошибки в понимании темы

Информация не равна смыслу сообщения. В теории Шеннона важна вероятность события: одинаково «понятный» текст может быть информационным по-разному в зависимости от того, насколько он предсказуем в модели.

Энтропия не равна бытовому «хаосу». Это строго определённая математическая мера средней неопределённости. Два процесса могут казаться хаотичными по ощущениям, но иметь разную энтропию при корректном расчёте.

Вероятность не гарантирует конкретный исход. Низкая вероятность события не отменяется в отдельном опыте, а высокая — не означает, что событие обязательно произойдёт в следующий раз.

Единичный эксперимент нельзя путать с закономерностью серии наблюдений. На практике вероятность описывает, как часто исход появляется в долгом повторении, поэтому оценки информации требуют статистической опоры.





Итог: один язык для случайности, данных и природы

**Вероятность описывает ожидание события,
информация измеряет неожиданность результата, а
энтропия показывает среднюю неопределённость
системы. Эти идеи объединяют математику, физику
и информатику, помогая кодировать, сравнивать и
анализировать данные.**

