Сделано в [Whisper AI](https://t.me/WhisperSummaryAI_bot)

Теория информации, Григорьев А.А., Лекция 01, 05.09.20

Чистая версия. Дата создания: ⁨03.10.2025⁩

и соблюдайте принципы нашего сообщества так студенты вы меня видите нет напишите кто-нибудь в чат мы вас видим я не чувствую себя так одиноко так меня слышно все хорошо О, видим, молодцы. Ну, хорошо, значит, так или иначе, начинаю я читать вам курс, который называется «Теория информации». Я хочу сразу, чтобы вы оценили масштаб бедствия. Значит, так, курс этот у нас семестровый. Вы выслушаете 15 моих лекций на этот курс, и потом будете сдавать экзамен. У нас с вами экзамен по этому курсу. Значит... Оцените масштаб бедствия так. Вот я вам рекомендую по этому курсу из литературы рекомендованной. Есть две книжки. Одна вот такая. Называется «Габидуллин-Пилипчук. Лекции по теории информации». Ну, у нас Рова Габидуллин. В лёд находится она, есть в библиотеке. Фамилия Габидуллин. Автор. Вот. Как видите, книжка так себе не особо толстая. То есть угроза небольшая. Вот. Вторая книжка, которая рекомендована, тот же Габидуллин, называется «Лекции по алгебраическому кодированию». Книжка ещё более тонкая. Так что совокупность того, что вам надо изучить обязательно, она вот такая. Ещё надо послушать мои лекции. А ещё есть ряд материалов, которые я вам пришлю по электронной почте. Значит, поступим так. У меня есть электронный почтовый ящик, он такой. АЛЕХ. нижнее подчёркивание, Григорий, Роман, Ирина. Гри, собака, натурально, аха.ру. Значит, если вы мне на этот адрес соберете запрос, то я вам в ответ вышлю в электронном виде необходимые материалы. Значит, эти материалы будут включать... У меня есть довольно приличная книжка по теории информации на английском языке, которую приятно почитать. Она, в отличие от этих вот изданий, ну, как вам сказать, она производит тяжелое впечатление, в ней 500 страниц. Но зато она хорошо читается. Ну, для тех, кто по-английски хорошо читает. Ну, в общем, полюбопытствуйте. Книжка действительно представляющая интерес. А самое главное стоит в следующем – Процедура сдачи экзамена будет такой. Вы получите на экзамене билет, где будет три задачи. Всего у меня составляется обычно 40 билетов по три задачи в каждом. Итого имеется 120 задач по этому курсу, по теории информации. Значит, все эти задачи, ну, по крайней мере, те, которые были в прошлом году, те, что я новые придумаю, я... пока не могу, а эти я вам пришлю. То есть от меня вы получите задачник, сборник задач, где порядка 150 задач по этому курсу. Ну и стало быть, если вы все эти задачи вот прямо сейчас сходу, вы на них посмотрите, если вы их прямо готовы решать, то, в общем, вы мой курс знаете. Сделано мною так, я просто свой курс разбил на такие... Коротенькие разделы, и по поводу каждого раздела оформил в виде задачки. Фактически это не задачи. Для того, кто владеет материалом, это просто рассказ по материалу. А для того, кто не владеет, это выглядит как некая задача. Поэтому задачи эти надо будет посмотреть. Значит, подавляющее большинство этих задач мы тут с вами решим прямо на лекциях. Мы, собственно, будем строить... Ну, то есть, они прямо по материалу лекции. Если вы лекцию слушаете, то вы видите, какие задачи я на этой лекции решаю. Каждая задача, она просто некий кусок из лекции. Некоторые из этих задач я вам буду давать в виде домашнего задания. Я вам сообщаю, что у вас по этому курсу имеется 60 часов самостоятельной работы. Это значит, что помимо срушения моей лекции 2 часа, вы должны 4 часа в неделю тратить на решение моих задач. Значит, поступаем так. На каждой лекции вы получаете некие задачи домой в виде домашнего задания. И, стало быть, их решаете, и решение присылаете мне на эту электронную почту. Можно по-всякому. Можно набрать это в каком-то редакторе, а можно просто снять на телефон листочек с решением. Важно, чтобы там была фамилия, номер группы, чтобы я идентифицировал человека. Я буду их просматривать, я не скажу проверять, буду просматривать, потому что задачи у всех будут одинаковые. Я, судя по всему, решения буду получать одинаковые, я об этом догадываюсь, но меня это не пугает. Мне важно, чтобы каждый своей рукой, по крайней мере, это решение написал, и тогда он что-то поймет. Задавальник, который я вам высылаю, он уже с решениями. есть я тут вам эту жизнь облегчил с решениями он потому что ну иначе вы просто эти задачи не решите но решение они такие я бы не назвал это даже решение это шпаргалки по поводу решений то есть если вы меня не слушаете то вам эти решения понять будет трудно потому что там ну просто изложены основные идеи выкатки я естественно там не делаю и что вот если подставить то-то туда-то, то получается вроде как так, вот такого типа. То есть всё это на полстранички, и задачи, и решения. То есть это не решение, это наводка на решение. Ну вот так действуем. Значит, слушаем лекции, решаем задачки, читаем Габидулина с Пилипчуком, Спилипчук. Ну и вот по возможности, полюбопытствуйте, я вам вышлю книжку Маккея на 500 страниц по-английски, заодно в английском поупражняетесь. Реально очень хорошая книжка по теории информации. Она не только по теории информации, она еще и по инференсу и по машин-лёнингу. Поэтому она такая толстая, но зато она очень хорошо читаемая. И там есть разделы, которые один к одному близки. В общем, я по ней учился. Мой курс построен близко к тематике и по стилю изглажения к материалу этой книжки. Книжка Габидулина тоже неплохо читается. Но поскольку она издана в нашей типографии, у нее формат неприятный, у нее буквы крупные. Короче, ее читать не слишком приятно по этим обстоятельствам. Но тем не менее, если вы ответ на большинство вопросов по этому курсу в этой книжке тоже найдете. Ну вот так будем жить. Все понятно? Ответьте в чат. Ну а тогда, если все понятно, то, в общем, ой, вопросы я изражил. Значит, когда придете на экзамен, знайте, что у вас в билете. Я билеты к экзамену вышлю вам, составлю и вышлю. Там будет три задачи из этого вот задавальника, что я вам пришлю. Так что я искренне считаю, что если человек способен решить вот эти 150 задач, то он владеет теорией информации. Ну, не в полном объеме, но в достаточной степени, чтобы получить отлично 10 богов. Поэтому вот сдача экзамена обычно трудностей не вызывает. У меня уже есть некий опыт ведения этого курса, и вот те экзамены, которые проходили, ну, оставляли у меня по уровню, который демонстрировали обучающиеся студенты, ну, в общем, терпимое, нормальное впечатление, что-то усвоено, скажем так. Ну вот такие дела. Ну а теперь начинаем, собственно, содержательную часть. Итак, адрес вы записали, адрес я стираю, это уже неинтересно. Теория информации. Это об чём? Теория информации. Наш курс. А это, в общем... Если так, с фигасовских позиций, это применение термодинамики к обработке информации. Видите, вот ситуация такая. Вы никто об этом не задумываетесь, и вас много кто об этом учит. Но ситуация такая, что вот если взять величину S, определенную как сумму большого числа величин, x1 плюс x2 плюс там и так далее, плюс xn большое, выделить на N большое. Вот такая сумма. И, в общем, эмпирический опыт, он говорит следующее, что если число N весьма велико, термодинамически велико, то есть N где-нибудь 10-23 степени числа Вагадра такого типа, то есть этих срагаемых жутко много, и вы вычисляете их в среднее значение, так сказать, сумму, делённую на N, то жизнь устроена так, что если вы эти... цифру величины иксы берете, ну скажем так, можно сказать случайно, а можно сказать от фонаря, с потолка, случайным образом, то окажется, что вот эта штука есть какая-то константа. Вот. Такие суммы сходятся к неким постоянным значениям. Определить, что значит сходится, почти невозможно. Просто так. Потому что сразу возникает вопрос. Хорошо, вы говорите, что это константа. Это какая константа? 5 или 3? Вот это зависит от того, как вы выбрасываете эти вот фонаряшные числа. И поэтому жизнь показывает, что вот это все хорошо выполняется. А как оно выполняется и как вот такие величины сходятся к этому постоянному значению, Это надо обсуждать. Значит, вот с этим мы сталкиваемся с вами в повседневной жизни. Вот я гуляю тут, на меня сейчас отказывается, я живу при какой-то температуре, и на меня оказывается некое давление. Это самые температуры и давление, это ровно вот такие вещи. То есть, почему имеет место температура? Температура – это на самом деле средняя энергия движения молекулы в газе. То есть она определена как сумма большого числа энергии, усреднённых по всем атомам, которые тут сейчас в помещении существуют. То же самое с давлением П. Вот есть стенка. Об нее ударяются молекулы газа, отражаются, в момент удара возникает некий импульс силы, изменение импульса этой частицы при отражении вызывает некую силу, и совокупность большого числа этих всплесков сил мы воспринимаем как давление. Вот такие величины называются агрегатами. То есть у нас есть как бы микромир, где живут вот эти иксы, и есть макромир, наблюдаемая величина, которую мы воспринимаем как некую физическую сущность, давление, температура и так далее. То же самое имеет место в обработке информации. Идея в том, что когда вы столкнетесь с большими информационными потоками, то количество битовой информации, с которыми вы будете иметь, оно становится также термодинамически большим. То есть у вас возникают терабайты информации, это где-то 10-12 степени битов. При обработке таких больших объемов данных возникают вот эти самые термодинамические эффекты, связанные с агрегированием. Ну и, стало быть, одним из таких агрегатов, с которым мы сегодня же и столкнемся, является вот эта самая фундаментальная для теории информации категория энтропия. Это типичный агрегат. То есть он определяется по большому какому-то ансамблю каких-то случайных величин, по случайной величине с большим ансамблем исходов, скажем так, либо по большому количеству случайных величин. Ну и, собственно, теория информации связана с тем, что мы обсуждаем вот этот агрегат, энтропию, и ее последствия. Значит, чтобы вот обосновать... необходимость вот таких сумм к постоянным значениям, на самом деле человечество ничего лучшего не придумало, как теория вероятностей и закон больших чисел в этой самой теории вероятностей. Вот давайте я вам сейчас изражу слабый закон больших чисел. Он состоит в следующем. Если эти х и х Есть случайные величины, распределенные с плотностью вероятности или распределением вероятностей. Это может быть дискретная случайная величина, принимающая конечное количество значений, либо непрерывная случайная величина, тогда это плотность вероятностей. Короче, от нам требуется только одно, чтобы эта случайная величина обладала конечным средним. То есть, чтобы было так или иначе, определено математическое ожидание. E от x, average, среднее значение случайной величины x. Ну вот определяется как интеграл от x на p от x по dx. Это если эта случайная величина непрерывная. Если она дискретная, то это будет сумма по всем x, x умножить на p от x. Вот, то есть имеется средняя, и пусть эта средняя составляет m. Ну, то есть беру случайную величину со средним значением m. Ну, и, стало быть, еще мы введем среднюю дисперсию этой случайной величины, как матожидание, матожидание, скобочки квадратные, от x минус m, Квадрате. Среднеквадратичное отклонение. Это вот сигма квадрат будет. Ну, то есть случайная величина, которая как угодно распределена, но имеет матожидание, математическое ожидание, и математическое ожидание вот этой центрированной разности квадрата, отклонения. Это называется среднеквадратичное отклонение от среднего значения или дисперсия. Ну и вот давайте рассмотрим сумму большого количества таких величин. То есть сумму СНТ, определенную, как вот здесь стоит, сумма Х житых по всем J от 1 до N большого делить на N. Что мы можем сказать о статистических свойствах этой суммы? Первое. Каково будет математическое ожидание этой суммы СНТ? Вот математическое ожидание СНТ. Но оно просто надо взять мат ожиданий от каждого х, от каждой этой случайной величины. Это будет сумма по всем j мат ожиданий от х, житая, поделить на n. Каждое из этих мат ожиданий составляет число m. Таковых чисел имеется n штук, так что это равно m в точности. То есть мат ожиданий этой случайной величины – есть в просто равном от ожидания каждой отдельной случайной величины, входящей в эту сумму. Ну а теперь займемся дисперсией этой случайной величины. То есть я от S-N-T должен теперь отнять Ем, возвести это дело в квадрат и посмотреть на среднее значение. Вот такая вот передо мной задача стоит. Ну вот, стало быть, давайте посчитаем. Это самое S-N-T есть... единица на n, в числителе стоит сумма хитых, поделить на n, отнять m. Всё это возведено в квадрат. Вот, значит, вот смотрите, я вот тут займусь тем, что у меня в скобочке находится. Значит, я это n вот сюда перенесу, у меня получится тут m в количестве n штук, и я напишу это как сумму, В числителе у меня будет х, j минус m в квадрате. Нет, х, m минус j просто. Сейчас перепишу. Значит так, х, j отнять m в скобочках и поделить все это дело на n. И вот это все развести в квадрат. Вот так. Я умножил n на m, и дальше эти n-m раскинул по разным слагаемым. Вот получилось так, xj минус m. Здесь сумма по j от 1 до n, так что просуммируешь вот эти m, получится m умножить на n, и n сократится. Так что оно вот так. Ну и стало быть, теперь надо возводить в квадрат здесь. Значит, при возведении в квадрат у меня появится срогаемое, ну, это сведется к двойной сумме. Я должен такую скобку умножить на такую же скобку. Это будет, в знаменателе будет n квадрат, а в числителе будет сумма по j и по k, ну, вот xj минус m на xk минус m. И вот с такой суммой нам предстоит разбираться. От нее придется посчитать математическое ожидание, average. Ну и, так как бы тут окажется, что в этой сумме есть компоненты, у которых j равно k, тогда мы тут имеем вот эту величину в квадрате. И есть компоненты, где j не равно k. Их мат ожидания равно нулю. Вот что я сейчас и покажу. Значит, мат ожидания от xj минус m на xk минус m. Мы сейчас будем предпорагать, что эти случайные величины xj и xk статистически независимы, а поэтому мат ожидания от произведения xj на xk равно произведению мат ожиданий e от xj на e от xk. А каждая из этих величин составляет m квадрат, как мы договорились, так что это m квадрат. Ну вот, теперь если рассматривать это мат-ожидание, при j не равно k, то вы получите, что это есть мат-ожидание на xj умножить на xk плюс m квадрат, это я вот это умножил на это, и минус m на мат-ожидание xj плюс мат-ожидание xk. Ну вот так. Ну и стало быть видно, что это аккуратно равно нулю. Вот эта штука равно m квадрат, а тут у меня вот этот равно m, а это равно и это равно m, а тут два раза m квадрат. У меня плюс 2m квадрат и минус 2m квадрат. Аккуратненько равно нулю. Так что те компоненты с j не равно k выпадают, а те, которые с j равно k, это вот это от ожидания равно дисперсии. Так что тут получается n умножить на σ2 делить на n2. Ну вот, после того, как сократишь на n, получается, что это есть σ2 делить на n. Итак, вот окончательный вывод из этой теоремы. Получается следующее, что если у меня есть случайная величина Sn, которая образована как сумма большого числа случайных статистически независимых значений, и мат ожидания х составляет m, а дисперсия х от х минус m в квадрате составляет σ2, то вот эта случайная величина имеет, обладает такими свойствами. Ее мат ожидания, мат ожидания Sn, составляет ровно m, так же, как и у каждой случайной величины, а вот дисперсия, average от Sn-M, Sn-T, большое, минус m в квадрате, составляет σ квадрат, деленное на n. Ну и, стало быть, это стремится к нулю при n, стремящемся к бесконечности. То есть вот такая сумма агрегат, Она ведет себя так. Когда n становится термодинамически велико, она принимает вот это значение почти заведомо. То есть разброс этих значений, отклонение, средний квадрат отклонения реального значения от мат ожидания, становится нулевым. То есть она сходится, случайная величина Sn сходится к m в среднеквадратичном. то есть Sn-m в квадрате, среднее значение, оказывается нулевым. Вот это и демонстрирует факт образования агрегата. Вы будете удивляться, но, так сказать, почти вся физика, с которой мы имеем дело, физическая реальность, она так или иначе основана на проявлении этих закономерностей. Я даже больше скажу, вот... Вы изучаете теорию поля, в которой вас учат тому, что есть электрическое поле Е, есть магнитное поле H. На самом деле это такие же агрегаты, как давление и температура. По той причине, что с позиции квантовой теории поля никаких электромагнитных полей нет, а есть заряженные частицы, которые обмениваются в процессе взаимодействия переносчиками этого самого взаимодействия, фотонами. И вот представьте себе, что вы начинаете решать задачу о простом конденсаторе с этих позиций квантовой теории поля. Тогда у вас есть атомы вот тут, есть атомы вот тут, вот тут летают кванты электромагнитного поля, их много, и вот усреднение всей этой картины и дает вам электрическое поле внутри конденсатора. Вот попытка решать эту задачу с позиции квантовой теории поля, не прибегая к агрегатам, она, в общем, напоминает попытку анализировать работу паровой машины с позиции молекулярно-кинетической теории, что вот есть цилиндр, есть поршень, тут летают молекулы газа и ударяются об этот поршень, его толкают. Значит, в принципе, можно пытаться это смоделировать и что-то понять, но это безумно сложно. Поэтому применение получается так, что мы хорошо понимаем, что есть молекулярная кинетическая теория и как она все устроена на самом деле, но при обсуждении термодинамических эффектов мы переходим к агрегированным параметрам, к этим агрегатам. И, стало быть, получается некая теория, которая называется термодинамика. И когда мы обсуждаем макроскопические явления в электромагнитных полях, то мы тоже забываем про фотоны, забываем про заряженные частицы, а работаем с агрегатами. Электрическое и магнитное поле — это, в общем, такие же агрегаты, как давление и температура. Ну и вот и в термодинамике нам придется работать с этими агрегатами, то есть в теории информации уже теперь. Нам придется работать с этими агрегатами, и это нам позволит... Ну, кое-чего достигать. Мы будем получать некие предельные оценки для всякого рода процессов обработки информации. Как то? Сжатие. Мы будем рассматривать сжатие данных, будем рассматривать передачу данных по каналам связи с ошибками, по шумящим каналам связи. Вот это все, пока занимаешься передачей отдельного бита или двух-трех битов, оно не работает. А когда начинаешь заниматься термодинамическими количествами битов, большими, тогда, стало быть, обнаруживаются некие предельные эффекты и некие теоремы. Это, в общем, фактически применение термодинамических категорий к таким задачам, как обработка информации. Ну и отсюда вся эта теория информации, которой мы с вами будем заниматься, она опирается на теорию вероятностей. То есть это вероятностная статистическая теория. И начнем мы с простого. Попытаемся начать с определения энтропии. Итак, великая категория энтропии, которая энтропия либо количество информации. Значит, начинается все со следующего. Существует безумное количество попыток определения, обоснования понятие энтропия или количество информации, содержащейся в источнике. Самый такой простой тривиальный подход состоит в следующем. Пусть у меня имеется некая случайная величина Х, и она может принимать M, большое, различных значений. Вот значение номер один, вот значение номер два, вот значение номер М. Эта величина случайно выпадает, то есть кто-то бросает кость, и выпадает одно из m возможных значений. И я вам должен сообщить о том, какая кость выпала. Что вот при данном испытании выбора кость номер J. Я вам должен об этом крикнуть, выразить звуки и передать вам эту информацию. Значит, вот что тут получается? Вот смотрите, если таковых исходов m всего 2, то есть вот в случае m равно 2, то хватит мне передать вам один бит, ноль либо единицу. Мы с вами договоримся так, что вот, m равно, то есть, единицу я передаю нулевым битом, а двойку я передаю единичным битом. Это мы заранее договорились. И когда есть всего два исхода, m равно 2, то я наблюдаю, какой исход произошел, и согласно ранее оговоренной нами кодировке, вам передаю по каналу связи один бит данных, либо 0, либо 1. И вы узнаете, что на этот раз выпала двойка, если вы получили 1. Ну а далее так, если это самое m равно 4, то я вам должен уже буду передавать не один бит, 0, либо 1, а 2 бита, потому что 4 варианта, вот 4 варианта, 1, 2, 3, 4. Я закодирую так. 0, 0, 0, 1, 1, 0, 1, 1. Всё исчерпано. Других двоичных комбинаций нет. Но вот с помощью двух битов я могу передать вам информацию об одном из четырёх исходов. Ну и так далее. Значит, если таковых исходов любая степень двойки, 2 в n-той степени, то для того, чтобы вам передать информацию о том, каким был исход, мне достаточно n битов. Вот, если я имею m равно 1024, что составляет как раз 2 в десятой степени, то мне хватит вам передать, вот, бит 0, бит 9. Десятибитовый блок. Значит, я эти 1024 варианта занумерую 10-битовыми блоками и передам вам этот блок, И стало быть, когда число исходов равно 2 вентой, хватает для передачи 10 битов. Ну так вот, что количество информации, которое необходимо передать для сообщения сведений об одном из m возможных исходов, можно определить так. h – это есть логарифм по основанию 2 от числа исходов m. Это называется энтропия Хартли. Хартлей. Вот она так вводится. Дворичный эргарифм от числа исходов. Вполне все понятно. Недоумение может вызвать только следующее. Когда m не равно степени двойки, то эта энтропия принимает нецелое значение. Когда m заключено больше 2 в 10, но меньше 2 в 11, то h принимает значение больше 10, но меньше 11, но нецелое. То есть вот у меня есть ось h, Стоит тут точка 10, стоит тут точка 11, а энтропия вот эта, она принимает значение где-то вот здесь. То есть она не целая. В этом нет ничего удивительного. Когда мы будем заниматься кодированием источника, я вас научу, как изготовить битовый поток, поток нулей единиц, 0,1, 0,1, 0,1, так, чтобы, скажем, количество информации на бит составляло 1 треть. Это всего лишь надо бит данных передавать с помощью трёх символов. Вот разобью этот поток на трёшки, и у меня тогда каждая такая трёшка будет переносить один бит информации. То есть количество информации на символ будет составлять 1 треть. Ну так что не целое значение количества информации – это вполне нормальная величина. Ну а дальше выясняется следующее, что... Мера Хартли рассчитана на то, что все исходы, вот у меня случайная величина, вот m ее возможных значений, и пока считается неявно, что все эти значения равновероятны. Вот тогда количество информации, которое можно передать с помощью битов, это и есть логарифм по основанию 2m. h есть логарифм по основанию 2 от этого самого числа m. И тут, в этой постановке, ни о каком сжатии данных речи идти не может. То есть я никак не смогу вам сообщить меньше, чем вот это количество битов, если мы хотим эти точки идентифицировать однозначно. Потому что для того, чтобы каждой этой точке присвоить уникальную битовую комбинацию, нужно вот ровно столько битов. Ну а на самом деле картина такая, что вот эти исходы могут быть неравновероятными. То есть мы с вами теперь будем называть исход буквы х маленькая и введём на этом множестве исходов некую вероятностную меру. То есть каждому х поставим соответствующую вероятность p от х, так что сумма по всем х вероятностей этих как раз аккуратненько равна 1. То есть я задал распределение вероятностей на этих символах. Ну и тогда возникает такая проблема, что те исходы, которые маловероятны, их получение переносит больше информации, чем те исходы, которые почти достоверные. Поэтому поступают так, вводят некую меру информации, которую поставляет каждый исход, содержит в себе каждый исход х. И таковой мерой берут минус логарифм по основанию 2p от х. Это как бы, говорим такие слова, что столько информации содержит в себе исход, вероятность которого составляет p. Ну и энтропию определяем как среднее значение этой информации по всем исходам. То есть энтропия h, Это просто математическое ожидание вот этой самой случайной величины. Минус логарифм по основанию 2 p от x. Ну, иначе говоря, это сумма по всем x минус сумма по всем x p от x на логарифм по основанию 2 на p от x. Вот. Вот это и называется энтропией случайной величины, заданной с помощью, определенной с помощью некого распределения вероятности p от x. Вот так вводится эта энтропия. Значит, она совпадает с энтропией Хартли, В случае, когда все эти самые P от X, все X равновероятны. Тогда P от X не зависит от X и составляет 1 на M. Это чтобы сумма этих самых P от X была равна 1. P от X их ровно M штук, и каждая из них одинаковая. Значит, здесь M на P, короче говоря, и P равно 1 на M. Если вот сюда это все подставить, то получится 1 на M, на логарифм единица на m. Когда по х это просуммируешь, вот это уйдет, и вот с этим минусом это как раз-то даст логарифм m. Так что энтропия вот так введенная, она совпадает с энтропией Хартли, ну вот в случае, когда все события равновероятны. А иначе она меньше энтропии Хартли. То есть в произвольном случае, когда есть какое-то распределение, Тогда h от x, вот это уже теорема, что h от x меньше или равно логарифму по основанию 2m. Вот, ну равно я сказал когда. Равно это когда все они равновероятны. Ну еще можно установить нижнюю границу. Нижняя граница вот как реализуется. Если в этом ансамбле один элемент появляется с вероятностью единица, а все остальные с вероятностью нуль, ну тогда энтропия равна нулю. Дело в том, что вот эта функция минус p на гагарифм p, ну то есть вот f от p, так определённо, ну иначе её можно писать как p на гагарифм единица на p. если вот этого знака загнать под гогарифом. У нее график вот такой. Значит, вот это у нас с вами P, а это P на гогарифм единица на P. Единица на P. Значит, есть точка 0, нулевая вероятность, есть точка единицы, единичная вероятность. Когда P равно 1, имею нуля, когда P равно нулю, тоже имею нуля. То есть она вот такая, эта функция. Она проходит через 0 здесь и здесь, минус p вега рифм p. Причем забавно, вот в этом месте у нее бесконечная производная. То есть она стремится сюда очень круто. Вот эта функция, собственно, и положена в основу определения энтропии. Ну и, стало быть, нижняя граница следует из того, что когда у вас... одна вероятность равна единице, а все остальные вероятности равны нулю, то все сгагаемые в этой сумме нулевые, потому что нулю подаем либо в этого нуля, либо в этого нуля. И поэтому тогда энтропия равна нулю. То есть больше или равна нулю. Если хотя бы одна из точек имеет отличную от нуля и единицы вероятность, то есть если вы попадаете, хотя бы одна из точек лежит где-то здесь с вероятностью p, отличная от нуля и единицы, то у вас вот в этой сумме появляется положительный вклад. p на гарифм единица минус p на гарифм p – это всегда положительная величина в этом интервале от нуля до единицы. И поэтому энтропия заведомо больше нуля. То есть равенство нулю достигается, если и только если все случайные величины детерминированы. То есть какое-то значение, все равно какое, принимает с вероятностью единицы, а все остальные имеют вероятность нуль. Во всех других случаях она отлична от нуля, но не превышает рогарифмы m. Вот это мы сейчас с вами попытаемся подоказывать. Значит... Значит, смотрите, я буду использовать так называемое неравенство логарифма. Значит, логарифм натуральный x меньше или равен x без одного. Это называется неравенство логарифма. Но логарифм натуральный. Значит, следует из простого графика. Рисую вот эту ось x, рисую график логарифма. Он вот такой вот. Это ln x. А вот у меня график линии х без одного. Это прямая линия, которая проходит через точку единицы и точку минус единицы. Она вот так проходит. Она вот тут является касательной к гагарифму. И поэтому, видите, гагарифм х везде. Вот у меня ведь гагарифмы х, а вот у меня ведь х без одного. Совершенно отчетливо, что гагарифм х везде меньше х. чем х без одного за исключением одного случая, вот когда х равен единице. Туда и гогарифм нулевой, и х минус один нулевой. У нас это неравенство интересует вот в этой области. Мы его будем применять к вероятностям, которые всегда меньше единицы. Мы вот сюда, собственно, не полезем никогда. И так вот имеет место быть неравенство гогарифма. Оно, к сожалению, тут возня с тем, что оно определено для гогарифма по натуральному основанию, по основанию Е. А тут я использовал основание 2 для гагарифма. Это не принципиально, потому что на самом деле переход от одного к основанию к другому — это просто умножение гагарифмов на константу. Только и всего. Ну, давайте прервемся на 5 минут. Значит, всё, приступили. Значит, тут можно, если быть занудой, то надо переходить здесь... Короче говоря, смотрите, поскольку меня интересует вот это неравенство, то я вполне могу от логарифма по основанию 2 в обеих частях перейти по логарифму по основанию E. Дело в том, что переход от замены основания — это умножение логарифмы на константу. если я умножу на константу эту и эту часть. Так что давайте я буду доказывать это для срочи, когда используется гарифм натуральный. Мне всё равно на самом деле. Ну, стало быть, а дальше что это значит? Мне надо доказать, что вот эта штука меньше вот этой штуки. Я, значит, из этой отниму вот это и покажу, что это отрицательно, меньше или равно нуля. А для этого я использую вот это неравенство. Ну, стало быть, начинаю писать. Слева стоит энтропия. Это сумма по всем Px, ну, по всем x, на логарифм единицы, делённые на Px. Ну, и отнять от этого дело только логарифм уже натуральный я хочу писать, ln. И отнять от этого дела надо логарифм m. ln, логарифм натуральный от буквы m. Ну, стало быть, Для порядка, чтобы это свернуть, я тут возьму и припишу Px к каждому гагарифму. Держу, что сумма по Px равна 1, поэтому, поскольку это подсуммой, то можно вот так написать. И получится, что это равно сумму по всем x. Px на гагарифм натуральный теперь уже. Ну вот, 1 на m и Px. Вот так. Минус логарифм m я заменил на логарифм минус единица на m, логарифм свернул. px я приписал, потому что сумма по x равна единице. Только и всего. Ну а дальше вот гляжу на доску и вижу неравенство логарифма. Вот эту штуку, вот это x у меня, можно заменить на x без одного. Значит, получится сумма по всем x, px, а в скобочке будет вот этот x без одного. Единица на m на px без одного. Вот. Ну и, стало быть, великий фокус применения этого неравенства состоит в следующем. Я поделюсь с вами, открою тайну. Тайность в следующем. Надо, чтобы px сократилось. Господи. Ну и тогда все упрощается. Смотрите. Поскольку вот тут px сокращается, тут сумма единиц на m, констант, в количестве m штук. Ну, то есть, получается, сумма по всем x-ам Единица на m минус сумма по всем хам, px. И та, и другая сумма аккуратно равна единице. Здесь я, вот у меня количество слагаемых в этой сумме как раз равно m. То есть я m раз беру по m минус 1. Короче, это равно m поделить на m равно единице. Ну а это сумма вероятности, она тоже равна единице, поэтому это равно нулю. Вот и все доказательство. То есть доказательство проходит в одну строчку, и оно дает нам границы для энтропии. Итак, энтропия случайной величины не превышает энтропии Хартли, вот этой грифм по основанию 2M, и всегда положительная. И мы знаем, когда эти границы достигаются. Вот эта граница, это когда Px равно 1 на m, все равновероятные. А вот эта граница, когда Px равно 1 для какого-нибудь х и нулю для всех остальных. То есть это детерминированная случайная величина, а это равномерно распределенная случайная величина. А все остальное, оно где-то вот в этих интервалах. Значит, ребятки, вот эти энтропийные границы, они настолько... будут востребованы в моем этом курсе, это что ли основной способ доказательства всех фактов. Значит, вы увидите с годами, что теория, которую я буду изглагать, она вообще довольно скучная. Она вся опирается на всякие границы для энтропии. Мы в основном первые лекции будем этим заниматься, что будем строить всякие границы оценки для энтропии, условных энтропии всего на свете. Так построена эта наука, тут ничего не сделаешь. Поэтому в построении этих границ нужно приобрести некий скилл, научиться чему-то. Значит, смотрите, очень полезно для построения этих границ знать неравенство Йенсена из выпуклого анализа. Значит, гасит оно следующее, что если есть некая функция φ от х, и она выпуклая кверху, выпуклая вверх, вот такая. Вот это у меня ось х, а это функция φ от х. Выпуклый вверх, значит, в нее можно налить воды, и она не будет скатываться. Бывают функции выпуклые вниз. Это я вам сейчас рассказываю зачатки так называемого выпуклого анализа, где рассматриваются выпуклые множества и выпуклые функции на них. Так вот, значит, если х некая случайная величина... то и φ от х тоже некая случайная величина. То есть, если природа случайно выбрасывает х, то получается, что мы вычисляем случайный φ от х. Ну так вот, мат ожидания average от этого самого φ от х для выпуклой функции больше или равно, чем φ от мат ожидания average. Вот это великое неравенство Йенсена. Тут надо скобочки квадратные поставить. Его значение в выпуклом анализе трудно переоценить, потому что там основной инструмент. И для нас это тоже будет важный инструмент. Значит, мат ожидания выпуклой функции больше или равно значение функции в точке мат ожидания этой самой переменной х. Самое главное, что это необыкновенно просто доказывается. Значит, Вы можете встретить чертову тучу доказательства этого факта разной степени сложности. Я вот знаю самое простое доказательство. Давайте так, возьмем некую точку х0. Вот она. И в этой точке проведем что-то типа касательной. Если это гладкая функция, то это будет касательная. А если не гладкая, а вот такая, допустим, недифференцируемая в этом месте, то понятие касательной не определено. Тогда ее можно как угодно провести. Мне нужно, чтобы вся эта выпуклая функция была с одной стороны от этой касательной. Это называется в выпуклом анализе опорная линия. Вот у вас есть выпуклое множество. Оно когда выпуклое? Оно тогда выпуклое, когда для каждой точки можно провести опорную линию так, что все это множество с одной стороны, вот в этой полуплоскости. То есть проведем какую-нибудь опорную линию, фонаряшно, все равно как. Ну и, что бы быть, мы с вами увидим, что значение функции вот в этой, если взять любую точку x больше x0, то φ от x больше вот это, вот φ от x, это вот эта точка, оно больше вот этого значения. Ну, то есть неравенство вот так изначальное выглядит. φ от x для любого x, for any x, φ от x больше или равно φ от x0 плюс α, некая константа, на x минус x0. Альфа – это коэффициент наклона этой линии просто-напросто. Вот я отступил отсюда на х-х0, и вот альфа на х-х0 – это вот этот отрезок. Это я написал уравнение вот этой прямой, ничего более. Это прямая, которая проходит через точку вот х0, фи от х0, по другому альфа. Ну вот, каким-то. Альфа может быть касательной. А может быть, производные в точке, а может не быть производной. Это не имеет значения для нас сейчас. Ну и, так же быть, начнем тупо от этого неравенства, от обеих частей, брать мат ожидания. Справа, слева, уже получим то, что надо. Мат ожидания и фиат х. И оно больше или равно. Фиат х0, от него мат ожидания брать не надо. Это константа. Плюс альфа. А в скобочках останется мат ожидания от х минус х0. От х0 опять же мат ожидания брать не надо. Ну и что осталось? Осталось взять х0. Х0 взято от фонаря, произвольно, как хотел. Я взял эту точку совершенно свободно. Я могу взять х0, равным мат ожиданию от х. Ну тогда вот это всё выпадет. И я получу неравенство Йенсена. E от φ от x, мат ожидания от φ от x больше или равно мат ожидания от... Значит, вот это φ от x0. Это будет φ функцией φ от мат ожидания от x. От average от x. Вот, x0 есть, average от x. Ну, все было правильно написано. Только скобки тут квадратные надо поставить. Вот, правильный факт вот такой. Ну, вот оно, неравенство Йенсена. Вот в таком виде оно совершенно примитивно доказывается и, стало быть, хорошо работает. Вот, чтобы доказать наше неравенство, у нас логарифм. Минус логарифм-то это функция... Вот он, он выпуклый вниз, значит, надо его переделать в логарифм единицы на p, сделать его выпуклым вверх. Значит, логарифм, ну вот, логарифм натуральный единицы на p. Это вот такая вот функция. Вот она, пристойно выпуклая вверх, лежит выше этой линии. Значит, логарифм единицы на p можно оценить сверху неравенством Йенсена. Ну и, стало быть, получится так. Мат ожидания... Логарифмы единицы на p, которые и есть наша любимая энтропия, она меньше или равно, по теореме Йенсена, логарифмы от мат-ожидания случайной величины, которая принимает значение единицы на p. Вот эта штука есть сумма по всем px, по всем x, px на единицы, деленные на px. Каждое такое сгагаемое равно одному, поэтому эта сумма равна m. Ну вот меньше или равно, таким образом, логарифму натурального m. Вот вам второе доказательство этого неравенства с опорой на неравенство Йенсена. Неравенством Йенсена мы будем вынуждены дальше пользоваться, поэтому прошу на него обратить внимание. Это важная вещь для нас. Ну а теперь пара полезных функций. Значит... В теории информации как основная функция, с которой приходится жить и работать, есть двоичная энтропия H от P. H от P. Это энтропия распределения двухточечной случайной величины 0 и 1, у которой одно значение принимается с вероятностью P, а второе с вероятностью 1 минус P. Вот такая энтропия вот этой двухточечной случайной величины. Это называется основная для нас энтропийная функция. H от P. Ее мы вынуждены будем хорошо знать. Формула для нее такая. H от P. Это есть минус P на логарифм P. Значит, ребята, логарифм все равно на самом деле по какому основанию. Если по основанию 2, то вы мерите энтропию в битах. Если по основанию E, то в натах. Можете по десятичному основанию, тогда вы просто используете десятичный способ представления чисел. Не битовый, не двоичный, а десятичный. Это все равно, это отличается на константу, поэтому этому особо внимания уделять не надо. Я даже никогда не пишу, по какому основанию я беру грифм, потому что это обычно ясно из контекста. Вот это называется двоичная энтропия. Значит, каждый студент обязан уметь нарисовать график двоичной энтропии. Сейчас я его вам нарисую. Значит, картинка выглядит так. Вот по этой оси у нас с вами H от P, по этой оси у нас с вами P. P натурально принимает значение из интервала от нуля до единицы. Далее эта функция не представляет интереса. P — это вероятность. Ну и, что бы быть, равна она нулю в нуле, КАП равно нулю, то вот из-за этого 0 имеет место быть, а здесь получается единица под гогарифмом. Когда П равно единице, то наоборот, у нас вот эта штука обращается в 0, а П равно единице, поэтому гогарифм обращается в 0. Так что проходим через 0 здесь и здесь, и есть вот такая вот симметричная кривая. Очевидно, что... Вот эта функция инвариантна относительно замены p на 1 минус p. Это очевидно. В одном слигаемом p, во втором 1 минус p. Поэтому вот здесь и здесь она ведет себя одинаково симметрично, а вот здесь она принимает значение 1. Значит, 1 — это энтропия, равномерно распределенная, с чего мы и начали. Значит, если у нас случайное значение... равновероятно каждая по 1, 2, 1, 2, 1, 2, то энтропия равна 1 биту такой случайной величины, 1 бит на символ. А иначе она только меньше. Вот это двоичная энтропия. Кроме того, в теории информации, спрошу рядом, нам придется использовать еще одну полезную энтропийную функцию, Она вот как определяется. Значит, имеется Q значений и случайной величины. Значит, значение 1, значение 2, значение 3, ну и вот значение Q последнее. То есть их Q штук. То есть мы имеем Q-ичный алфавит. Не двоичный алфавит, а Q-ичный. И тогда вводится такая случайная величина. Она принимает значение 1 минус P в каком-то месте, А все остальные значения она принимает тогда с вероятностью p. И эти значения равновероятны. То есть все остальные значения имеют вероятность p на 1 на q без 1. Значит, смотрите, есть одно значение, в котором вот эта вероятность. Оставшаяся вероятность, которая приходится на все остальные Q-1 штуку, это P. Я ее поделил поровну. То есть это вот такая случайная вероятность. Здесь у нее вероятность P, а здесь все вероятности одинаковые, вот такие. И составляют P, поделенное на Q-1. Вот эта энтропия называется Q-ичная энтропия. Это двоичная энтропия, а это будет у нас Q-ичная энтропия. HQ3 от P. Значит, она вот так вот будет выглядеть. h от p. Считаю энтропию этой случайной величины. Это минус 1 минус p на логарифм. Ну вот по основанию 2 сейчас. Я ее как двоичную пока посчитаю на 1 минус p. Чудесно. Это вот я вот это срыгаемое. Плюс все остальные срыгаемые. А всех остальных срыгаемых имеется q минус 1 штука. Значит, тут надо написать, что их Q минус одна штука, и каждая даст вклад P, делённое на Q без одного, на логарифм, опять же, по основанию 2, P на Q без одного. Вот такая вот энтропия. Ну и счастье в том, что на Q минус 1 благополучно сокращается, после чего P на логарифм P выносится отдельно, и получается минус H2 от P. Дворичная энтропия. Вот она. 1 минус P, 1 минус P, P, P. И отнять и прибавить к этому делу логарифм P на логарифм по основанию 2 от Q. От Q без одного только. От Q без одного. Значит... Это еще не совсем куичная энтропия. Когда рассматривают куичную энтропию, то разумно перейти к алгоритмам по основанию Q. А для этого надо это все поделить на алгоритм по основанию 2 от Q. То есть перейти во всех алгоритмах к основанию Q. И тогда получится то, что публика называет настоящей куичной энтропией. HQT от P. Это есть минус P. на гагарифм по основанию Q от P, минус единица минус P на гагарифм по основанию Q от единицы минус P, ну и вот та самая добавка, плюс P на гагарифм по основанию Q от Q без одного. Вот. Вот это настоящая, праведная куличная энтропия. Значит так. Глушите меня внимательно. Это первое ваше сегодняшнее домашнее задание. Я хочу, чтобы вы изучили график. Попробуйте построить график этой самой куличной энтропии и прислать мне набросок этого графика на бумажке. Он демонстрирует очень забавное поведение. Не хочу открывать как. То есть от вас требуется мне на листочке бумажки прислать график. Вот тут H-ку от P, а тут P. от нуля до единицы. Функция выписана, все правильно. Ну и, строго быть, вот задача на построение графиков. Ребятки, это хорошая для вас задача. Дело в том, что мы часто будем пользоваться этой функцией. Вы должны обе эти функции хорошо знать, иначе они будут для нас основными. Вот когда мы начнем заниматься всякими оценками для для кодов, они там просто широко используются, без них никуда. Поэтому вот эти две функции, ну вот, знайте, что это, во-первых, полезные примеры вам на освоение понятия энтропии, во-вторых, они имеют значение в теории. Поэтому задачи эти не просто так. Ну вот, всякий раз, когда вводишь понятие энтропии, у публики возникает вопрос, откуда берется гогорифм. Ну, стало быть, связано это самом деле на тому можно дать массу обоснований связано это сугубо прикладными вещами вот мы с вами вот сейчас давайте я поговорю пять минут о кодировании источника вот теперь вот вернемся то столько картинки с которой я начинал у меня имеется ансамбль из м возможных исходов я хочу вам передавать информацию об этом исходе я желаю сэкономить то есть понизить скорость передачи. Если никаких мер не предпринимать, а передавать просто тупой исход, не думая о вероятностях, то получится так, что на каждый исход я должен вам передать логарифм по основанию m 2m битов. И все. Никакой экономии достичь невозможно. А я хочу передавать число битов поменьше этого. Это, ребятки, подход к сжатию данных. Ну вот, вот тогда, значит, мы с вами этим будем подробно заниматься. Я просто в порядке анонса, так сказать. Ну и картинка такая. Есть два подхода к сжатию. Самый простой, значит, надо как-то уменьшить число передаваемых точек. Это значит, что мы должны с вами как-то договориться о том, что близкие точки в определенном смысле близкие, ну, я солью в одну точку. Если я передаю изображение, то я его могу немножко покорежить и сказать, вы согласны с таким изменением? Скорее всего, если это изменение невелико, вы его просто не заметите. И поэтому вот некий кластер можно... передавать как одну точку. И таких кластеров тогда будет сильно поменьше, чем М. Вот это передача с искажениями. Это так называемая теория rate distortion. Distortion – это уровень искажений. Мы договариваемся о допустимых искажениях. Вот что вместо сообщения я вам сообщаю сообщение с ошибками, и вы сообщение с этим соглашаетесь. Тогда я могу повысить скорость передачи. И вот удивительная вещь такая. Давайте попробуем подумать над следующим. Вот я попытаюсь порисовать график. По этой оси я отложу rate, скорость. За единицу я тут обозначу rate, я его понормирую на гагарифм по основанию 2m. Так что скорость заведомо не будет превышать единицу. То есть если я передаю все, то скорость у меня равна одному, а если я передаю, ну, а иначе я могу передавать вот поменьше. Если скорость вот такая, то, скорее всего, я подверг сжатию свои данные. А вот по этой оси у меня будет D, distortion, искажение. Ну, какая-то условная мера искажения. Вот она от нуля тоже до единицы. Совсем не искаженный сигнал, искаженный до безобразия. Так что я вам в любом случае, что бы ни было, передаю 1 бит, например. Ну, то есть мера искажения вот такая высокая. Вот какова ожидаем, вот так сказать, на взгляд непосвященного, какова должна быть зависимость рейта, скорости от дистошена? Всякий человек, которому предъявляешь вот эту задачу навскидку, он говорит такие слова, что это должна быть какая-то вот такая кривая. Ну, смотрите, при нулевых искажениях скорость единичная, если я допускаю искажения, то скорость становится поменьше, и будешь иметь вот такую кривую. Так вот, самый удивительный вывод теории информации в том, что это совсем не так. Кривая будет совсем не такая, она бывает вот такая. И вот этот уровень равен h от x. Энтропия источника. Ну, то есть, если вы соглашаетесь хоть на какие-нибудь искажения, то вы можете понизить скорость до энтропии источника. И, значит, за счет высоких искажений вы ничего не выигрываете. Только вот здесь где-то, вот в этих областях. То есть, в широком области уровня искажения у вас... Ну, естественно, это при М, все это при М стремящемся к бесконечности. Это асимптотика. Когда вот этот ансамбль ужасно большой. Вот тогда вот в пределе получается вот такая кривая. И вот это, ребятки, это основа теории сжатия данных. То есть мы умеем сжимать данные вплоть до скорости равной энтропии источника. Этим мы будем далее подробно заниматься. В порядке обоснования энтропии я люблю ссылаться именно на эти теоретические факты. То есть начинаешь понимать, зачем нужна энтропия, почему именно энтропия. Ответ вот такой. Потому что она получается при решении таких асимптотических задач. А вторая обстановка такая. Вот у вас есть тут точки, и какие-нибудь маловероятные из них. маленькая, а какие-нибудь более вероятные, P большое. Ну вот давай так напишу, P маленькое, P большое. Тогда, передавая эти точки по каналу битовыми потоками, разумно поступить так, наверное. Те точки, которые часто появляются, вероятные, им присвоить малые длины кодовых, битовых комбинаций. А те точки, которые малые, более вероятные, им присвоить длины кодовых комбинаций побольше. И тогда нас будет интересовать средняя длина этого самого слова двоичного, которое мы передаем в обид по канаву. Ну вот L средняя, то самое, которое сумма по X, по всем X, P от X, PX, на L от X. длина битовой комбинации, количество битов, которые я запланировал до передачи вот этого самого х. Хочется придумать систему кодирования, которая дает наименьшую среднюю длину. Вот здесь мы тоже этим будем заниматься, и окажется, что вот эта наименьшая средняя длина как раз равна энтропия h от х. То есть если энтропия считается по основанию 2, Это количество битов, и среднее количество битов, которое нужно для передачи символов, оказывается как раз вот таким. Это теорема кодирования символьного источника без потерь. Тут речь идет о средней длине. Если говорить об счет, то смотрите, наука развивается давно, и эта наука ей уже почти 100 лет. И люди очень много чего сделали в плане придумывания других Мер количества информации и мер энтропии, кроме вот этих шиноновских мер, о которых я вам рассказываю. Значит, дело заключается в следующем. Каждый волен придумать свою меру качества вот этого. Но рассматривать не среднюю длину кодового слова, а, скажем, средний квадрат длину кодового слова или что-нибудь L в степени P, единица на P, какую-нибудь P-норму от длины кодового слова. И тогда можно придумать энтропийные меры, отличные вот этой гарифмической, которые как раз нужны для решения таких задач с иными метриками, с иным определением эффективной длины, ну, то есть критерия качества. Ну, то есть тут появляется некая зависимость критерий качества и вероятностные меры. Хорошо. Значит, второе обоснование гогарифма, оно кроется вот в чем. Оно кроется на самом деле в бинамиальном распределении. Вот смотрите, в полости у меня имеется случайная величина, которая, вероятно, принимает два значения. 0 и 1 с вероятностью 1,2. Я хочу... Короче, меня сейчас интересует динамиальный коэффициент. Вот тогда вероятность выпадения блока х0, хn-1, блока длины n, у которого k... этих самых единиц и n-k нулей, она такая, число сочетаний из n по k, ну вот, деленное на единицы на 2 в n-той степени. Все комбинации равновероятны, потому что буквы равновероятны. Меня интересует сейчас бинамиальный коэффициент. Число сочетаний из n по k. Вот я им займусь. Это самое число сочетаний. Тут проявляется энтропия. Вот она как проявляется. n факториал на k факториал на n минус k факториал. Я так полагаю, что это формула из третьего класса школы знает. Число сочетания из m по k. Мы сейчас попытаемся связать эту величину с энтропией. Вот как это делается. На то существует оценка Стирлинга для f факториала. Значит, смотрите. n факториал, как известно... хорошо очень аппоксимируется такой величиной. Корень из 2πe, натуральный галагарифм, на n в степени n, на e в степени минус n. Это очень классная оценка для факториала. Она получается методом перевала, она приводится во многих книжках, мы ее сейчас воспользуемся. Значит, на 2πe я не буду обращать внимания, я буду писать, что это порядка, n в степени n на e в степени минус n. Это чепуха, она для нас не будет играть роли. И вот я этой оценкой попытаюсь воспользоваться для оценки биномиального коэффициента. Получится это так, у меня так, что вот это порядка. n факториал я заменю на n в n-той степени на e в степени минус n. Вот. Здесь у меня будут знаменатели k в степени k, на e в степени минус k. Это я оценю k! n-k! Значит, будет n-k в степени n-k на e в степени n-k. Ну вот, такая оценка. Значит, смотрите, вмиг выясняется, что e в степени n вообще ни при чём, оно сокращается. Е в степени минус К и Е в степени минус Н минус К. Это Е в степени минус Н. С этим сокращается. Остается Н в нтой степени на все, что осталось. А на Н в нтой степени я с Божьей помощью поделю. И она получит, что это так. Единица на К к Н в степени К. Я использовал. Вот у меня в числителе Н в степени Н. Я Н в степени К уже использовал. А тут я использую Н в степени Н минус К. И получится 1 минус k к n в степени n минус k. n минус k. Ну вот, пока все. Теперь давайте прикинем, что у нас получится, если я возьму логарифм от этой величины. Уже дальше понятно, что я хочу сделать. Я хочу получить оценку вот для чего. Единица на n на логарифм миномиального коэффициента n по k. То есть я возьму от этого дела логарифм и все это дело поделю на n. Вот у меня получится так, что это есть минус k, деленное на n, на логарифм k к n, и минус 1 минус k на n на логарифм в единице минус k к n. Вот такие дела. То есть у нас с вами получается, что... Единица на n, многогрифм бенемиального коэффициента, это ровно наша двоичная энтропийная функция h, вот это вот h, от k поделить на n. Ну, отсюда получается такая оценка для бенемиального коэффициента, что число сочетания из n по k, оно порядка 2 в степени n на энтропию двоичную от k к n. Вот такая великая формула. И так вот она как-то проявляет роль этой самой энтропии в оценивании бенумиальных коэффициентов. Вот откуда уши растут. Видите ли, бенумиальный коэффициент растет экспоненциально по n вот с этим показателем. Ну и стало быть, получается как? Что вот если посчитать долю двоичных комбинаций с заданным k, Вот смотрите как, у меня вот этот k, число единиц в блоке, в н-блоке, случайном н-блоке, у меня имеет место быть случайный н-блок, вот у него длина n. А k – это количество единиц в нём, а n-k – это количество нулей. Это вот агрегат. Этот агрегат как-то вот распределён. Я имею закон распределения этого агрегата, вот каких-то агрегатов много, каких-то мало. Если посчитать долю, то есть выделить число сочетаний из n по k на 2 в степени n, то получится оценка вот какая. Это будет 2 в степени минус n на 1 минус h, вот эта самая двоичная энтропия от k к n, отношение k к n. Вот так. Ну и, стало быть, отсюда следует, что если энтропия h от k к n отлично от единицы, а она бывает только меньше единицы. Вот она как двоичная интерфейсия ведёт h от p. Вот так. И вот тут она принимает значение единицы. То есть если k равно n пополам, половина, то есть половина единиц, половина нулей, то энтропия равна единице, и я не имею, в общем, никакой оценки для вот этого отношения. Если же энтропия отлично от нуля, не равна меньше единицы, то вот тут у меня положительная величина, и доля таких вот блоков, она стремится к нулю. Вот смотрите, в результате как получается, что если вы делаете случайные блоки из черных и белых шаров, длинные, с большим N, то у вас с вероятностью единицы будут получаться блоки, у которых k равно n пополам. То есть количество белых шаров и количество черных одинаково. Вот такие блоки будут выходить у вас почти на верное, потому что вероятность выпадения любого другого блока с k, неравного n пополам, она стремится к нулю при n стремящемся к бесконечности, в то время как вероятность выпадения этого блока но она приближается к единице, потому что она так к нулю не стремится. Вот такие дела. То есть, вот смотрите, вот это типичное проявление закона больших чисел, о котором я, так сказать, вот как умею, так рассказываю. Получается так, если вы изготовили фаб... Мы изготовили фабрику по производству вот таких тросточек из белых и черных шариков. И белые и черные шарики выпадают у вас с равной вероятностью. Ну и вы, стало быть, эти тросточки складываете на склад, а публика, которая приходит у вас эти тросточки покупать, она желает покупать определенную, то есть она интересуется вот этим макроскопическим параметром. Тогда выяснится, что если этот их интерес не совпадает с половиной, со средним, ваше производство впустую на схват никто не возьмет ваши тросточки с к равным n пополам они только они у вас и будут получаться в чем дело то есть вам надо изменить стратегию и начинать как-то выдавать стросточки производить с иной вероятностью ну вот это вот такая вторая важная оценка которая с этим связана это оценка для вероятности бинамиального распределения вот теперь у нас уже будет n блок блок длины n, и стало быть тут нулей и единиц, либо черные и белые шары, и все равно что с вероятностью p. Тогда вероятность этого блока, в котором k единиц и n минус k нулей, или черных и белых шаров, она такая. pk т. Есть число сочетания из n по k на p в k той степени. Значит, k раз выпало p, и 1 минус p выпало n минус k раз. Вот, то есть распределение по k, оно вот такое. Но вот если с ним проделать то же самое, значит, биномиальный коэффициент аппроксимировать как... Вот как мы это делали. Биномиальный коэффициент аппроксимируем как k, k деленное k на n в степени k, И тут 1 минус k на n в степени n минус k. Ну, а то, что останется, это p в степени k. Вот в числителе напишу p в степени k на 1 минус p в степени n минус k. Ну, а дальше вот стандартно. Значит, беру логарифм от pк и нормирую ее на n. Когда нормирую на n, все вот тут на n поделится, когда я логарифмы посчитаю. Короче, вот... На этой половине доски я позволю себе написать эту формулу. Значит, она выглядит так. Значит, логарифм ПКТ поделить на n. Это есть. Ну вот, значит, беру логарифм от этого сначала и делю на n. Это будет минус... k, делённое на n, на логарифм k, делённое на n, это вот от церковь я взял логарифм, минус, и поделим на n, вот когда я поделил на n, k поделилось на n, вот тут вот. Минус, значит, единица минус k, ну это уже то, что было фактически, k на n на логарифм k к n, вот. Ну а то, что остаётся, это вот надо взять логарифм от этого и тоже поделить на n. плюс, это положительная будет добавка, плюс kk на гогорифм p и плюс единица минус kk на гогорифм единица минус p. Вот. Ну и все, пока. Сейчас я введу важную величину. Значит, смотрите, чтобы причесать это дело и свести его к какому-то компактному виду, нам понадобится так называемая информационная дивергенция. Вот пусть у меня есть на ансамбле X с исходами X и с вероятностями PX две случайные величины PX и QX, по-разному распределенные. Тогда информационная дивергенция, дивердженсы, Между P и Q распределениями, вот так, она не симметрична, переставлять местами нельзя. P и Q, она определяется вот так. Это сумма по всем x на P, Px, на логарифм Px, делённое на Qx. это информационная дивергенция ну и стало быть тут у нас с вами вот видно что у нас одно распределение вот оно какое по единице минус p а второе распределение как n и единица минус как n это вот п и q наши значит это как раз информацию это минус информационная дивергенция вот то что у нас получилось минус d от распределения как n распределение и распределение p вот так ну и стало быть Поэтому P, K, T, оно на самом деле порядка 2 в степени минус N на информационную дивергенцию между K, N и P. Эта самая информационная дивергенция замечательна тем, что она положительна. Больше или равна нуля. Доказательство элементарное. Вот смотрите, я буду доказывать, что минус информационная дивергенция, минус D, которая есть сумма по иксам, на гогарифм q, поделенное на p, отрицательное, меньше или равна нуля. Для этого я воспользуюсь неравенством гогарифма, например. Но, стало быть, гогарифм заменю на x-1. Это меньше или равно, чем сумма по всем p. По всем p. Тут стоит p, а тут стоит q, деленное на p, без одного. Ну и, стало быть, фокус тот же срабатывает. Вот эти p сокращаются, и я получаю сумму q минус сумма p. Это равно нулю. Так что информационная дивергенция всегда положительна, и она равна нулю. Распределение вероятностей бинумиальное с параметром P выражается через информационную дивергенцию. Вот смотрите, какая картина. Когда P равно как N, то есть относительное число шариков равно P, то информационная дивергенция равна нулю, и у меня нет стремления к нулю этой вероятности. Вот эта штука равна нулю. То есть она константой остается. А вот если P не равно как N, то значит эта штука положительная, и вероятность стремится к нулю. Так что если вы будете изготавливать тросточки, выбрасывая шарики черные с вероятностью P, то у вас доля черных шариков, как N, будет равна P, Заведомо при достаточно большой длине. Ну вот это еще одно проявление, заодно познакомились с информационной дивергенцией. Все у меня на этом сегодня. Ну что, значит так, жду решения задач. Значит так, как бы это... Значит так, пришлите мне на почту привет, я вам отправлю методические материалы. Раз. Ну и во-вторых, присылайте решение задач. Поскольку писем будет много, хотелось бы, чтобы они в один... один ящик упали научимся как-то их надо одинаково называть давайте называть так задание 1 вот если в теме будет указано задание то они все лягут ко мне в один ящик и я все 150 этих самых сообщений оттуда достану не буду искать по всей почте Задание о фамилии вы уже приводите на том листочке, который пришлёте. Мне важно разобрать это на фамилии и номера групп, потому что я строгий учёт буду вести орехов, и в конце будут... Ага. Так, Егор, значит, условия задач... Я сегодня вам дам одну задачу на исследование этой самой коечной энтропии. Можно ещё раз условия... Значит... В том тексте, в котором я... Значит, у вас функция будет... Речь вот о построении графика. Ещё раз. Значит, график вы должны мне построить такой. h с индексом q от x от p, куичная энтропия, это есть минус p на логарифм по основанию q от p, минус 1 минус p на логарифм по основанию q от 1 минус p, Плюс, плюс, то есть P. На гагарифм по основанию Q, P без одного. Ой, гагарифм по основанию Q, Q без одного, натурально. Это энтропия Q точечного распределения, то есть у вас имеются точки 1, 2, чтобы вы не вспоминали, 3Q, у которых одна вероятность равна 1 минус P, а прочие все одинаковые и составляют P на Q без одного. Вот такая обстановка. Я хочу график этой функции. Как она себя ведет при больших Q, при Q, стремящемся к бесконечности? Вот попробуйте, это довольно забавно. Ну, то есть, вот, значит, смотрите, вот это верно всегда, оно верно и для двоичной энтропии. Если вы тут выпишите H2 от P, подставите вместо Q двойку, то у вас все галагарифмы становятся по основанию 2, а эта добавка исчезает, обратите внимание. При Q равном 2 у вас вот это выпадает, и вы имеете обычную двоичную энтропию. Вот так она становится куичной, вот с этой добавкой. И вот надо понять, как эта добавка влияет на график этой функции. Двоичная энтропия, она вот такая. Ну, всё вот. Ау! Тоже по письму задание получим. Ну, не задание, задание, я сказал. Ну, таск. Ну, стало быть, литературу получите легко. Ну, всё тогда. Григорь, спасибо и тебе. Ну, всё хорошо у меня, на этом всё закончилось. Мне Григорий Роман Ирина, собака, а-инч-а, точка, ру, мальчик.ру. У меня очень простой сервер, уж слишком простой, проще некогда. Аха! Ну, ладно, значит, узнаете вы мою почту, её один ленивый на фестихе не знает.