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

Теория информации, Григорьев А.А., Лекция 02, 12.09.20

Чистая версия. Дата создания: ⁨03.10.2025⁩

хорошо то есть я хотел сказать плохо так где что-нибудь а вот оно это все и Студенты, напишите что-нибудь в чатик, чтобы я знал, что вы есть. Вообще, если кто-нибудь на лекцию придёт, то будет от этого лучше. И мне, и вам. А то очень не одиноко. Знать так, домашнее задание вы мне делали. Значит, надо вам ввели на буду построить график энтропии распределения, которое устроено так. имеется куточек ку из них одна имеет вероятность единицы минус п остальное p равномерно распределено по основным точкам так что их вероятности составляют по эдаку без одного вот такая картинка ну и мы с вами исследованы писали формулу для энтропии этого распределения и вы мне должны были построить график и зависимости от энтропии H от буквы P на интервале от нуля до единицы. Ну, вы блестяще справились с этой задачей, я в восторге, я получил где-то порядка 30 решений. И почти ни одно из них меня не удовлетворило немало, по той причине, что я не имел в виду, что вы мне построите этот график в вашем питоне несчастном. Ну что, это кто хочет, напишет формулу в Питоне, он а график нарисует. Поэтому придётся мне чуть-чуть поговорить об этой энтропии. Ну, в общем, в целом выводы правильные были сделаны. Смотрите, картина такая. Когда p равно нулю, вот эта точка, то всё банально. Вся вероятность сидит вот в этом месте, это единичная вероятность, а прочие вероятности нулевые. поэтому проходит от кривая вот тут через точку 0. Второе. Когда p равно 1, вот эта точка проходит на тут через 0 или нет? Ответ такой – нет. Дело в том, что в этом месте, когда p равно 1, вот эта вероятность равна 0, а все прочие составляют 1 поделить на q-1. То есть у нас имеется... q минус одна точка с равномерным распределением вероятностей на них поэтому в этом месте мы знаем какая будет энтропия она равна гагарифм вот мы брали его по основанию q от q без одного q минус одна точка и вероятности у всех одинаковые p поделить на q минус 1 энтропия этого распределения есть энтропия количество есть гагарифм количество точек так что знаю как она вот тут пройдет Ну, стало быть, уже угадывает, что это какая-то вот такая линия. Ну и, стало быть, осталось экстремум найти. Где у нее экстремум? А экстремум у нее там, где все вероятности одинаковые. Для этого надо приравнять единицу минус p, вот эту вероятность, к p делить на q без одного. Ну и откуда вы мгновенно получите, что это p равно q без одного поделить на q? Ну, то есть вот тут стоит точка. вместе q-1 поделить на q, и в ней энтропия это максимально имеет экстремум. И экстремум, естественно, единица, потому что у нас тогда получается q равновероятных точек. Все вероятности оказываются одинаковыми. q точек, значит, энтропия составляет как раз единицу. То есть она вот такая вот кривая на самом деле. Ну и дальше обсуждение такое. Многие правильный вывод сделали, что парику, стремящемся к бесконечности, вот это экстремум неограниченно приближается сюда, и различие между вот этим и единицей исчезает. То есть эта штука становится почти прямой линией. Ну и забегая сильно вперёд, я вам скажу, мы только что с вами, вот как далее окажется, я для чего эту задачу-то даю? Для того, чтобы нам дальше полегче было. А дальше нам будет совсем легко. Дело в том, что у нас это, на самом деле, мы только что доказали, что при больших размерах алфавита, когда Q стремится к бесконечности, граница Варшамова-Гильберта, она вот такая, двоичная энтропия, она приближается к прямой линии, к границе синглтона. Вот в своё время будут у нас границы для кодов, и там этот результат нам сильно понадобится. То есть то, что это вырождается в прямую линию, это весьма существенно. Ну вот так, это по поводу домашнего задания. Но знаете, я же не это имею в виду. Когда я говорю построить график, я имею в виду, что вы какое-то исследование проведёте, а не в питоне мне картинки нарисуете. Картинки красивые у вас, кстати, получились. Я был в восторге. Ну ладно. А остановились мы на том, что ввели информационную дивергенцию И выглядело это так. Вот у нас некое множество х, и на нем заданы два распределения. П есть множество П с индексом х, Пх. И второе распределение вероятности Q, которое есть множество Qх. Ну, натурально сумма по всем хам Пх... равна единице, и сумма по всем хам Qx тоже равна единице. Так вот, дивергенция информационная между распределениями PQ – это некая мера различия. Это ввёл её Кульбак, ему принадлежит это определение. И она так и прижигась, информационная дивергенция, она играет довольно важную роль в жизни и науке. Значит, дивергенция D между распределениями P и Q, вот так я ее обозначаю, дивергенция between P and Q. Это есть сумма по всем хам, в общем, почти как энтропия, на P х, на гагарифм, P х поделить на Q х. Вот такая вот штучка, информационная дивергенция, дивергенция между P и Q. Что про нее известно? Довольно очевидно следующее. То есть это как бы расстояние между P и Q, расстояние между двумя распределениями вероятностей. Но вот оно, к сожалению, на расстояние не тянет по той причине, что оно не симметрично. Вот совершенно довольно очевидно, что D между P и Q – это не то же самое, что D между Q и P. Ну, видите, тут P входит вот так, а Q только в одно место, поэтому они неэквивалентны. То есть перестановка местами Q и P меняет значение этой дивергенции. Тем не менее, значит, смотрите, картина такая. Эта самая дивергенция, она поражительная. Она поражительна, и она... Вот мы сейчас докажем, что она больше или равна нуля, и... и достигает значения 0 ровно в одном месте, когда все P равны всем Q. Если P равны Q, то есть если распределение одинаковое, то расстояние между этими распределениями нулевое, а иначе оно положительное. Вот в этом смысле она тянет на метрику. То есть свойства хорошие. То есть если расстояние между P и Q, дивергенция между P и Q нулевая, то эти по ИКУ совпадают, а иначе она положительная. Чтобы доказать, что это так, давайте воспользуемся неравенством Йенсена. Здесь стоит логарифм, это... Логарифм – это... Не очень приятно. Я люблю, когда функция выпуклая вверх, а она у нас выпуклая вниз. Поэтому вот минус логарифм меня будет интересовать. Минус логарифм — это вот такая функция, ой, это вот такая функция, к которой я люблю, чтобы входили неравенства Йенсена. Поэтому я вот запишу это, значит, так. Вот эта самая дивергенция между p и q — это есть минус... Сумма по всем х. П на логарифм Q поделить на П. Я вот тут переставил местами Q с П и поменял знак. Логарифм х равен минус логарифма единицы на х. Вот в таком виде я ее и буду изучать. Ну, стало быть, дальше я воспользуюсь... Поскольку вот эта функция логарифм вот такая, как я нарисовал, выпуклая вверх, то для нее неравенство Йенсена имеет такой вид. Матожидание φ от х больше или равно φ от мадожидания от х. Вот так. Я воспользуюсь неравенством Йенсена. То есть возьму матожидание с этой стороны. Значит, минус я вот сюда втащу и буду брать мат ожидания от минус логарифма. И у меня получится, что математическое ожидание вот это всё я запишу как мат ожидания от минус логарифма P поделить Px поделить на Q. То есть только Q уже поделить на P. Qx поделить на Px. И вот это мат ожидания берется по распределению P. То есть сумма P вошла в мат ожидания. Это согласно нашему генеральному инстуенту, больше или равно, чем минус логарифм мат ожидания того, что минус логарифм мат ожидания отношения QP. А это мат ожидания, есть сумма Дальше просто порасписываю это от ожидания. Сумма по всем хам Px на Qx на Px. Вот математическое ожидание этого отношения. Сумма по всем хам с вероятностями Px. Ну и дальше сам видите, что вот это все сокращается, а сумма по хам Qx равна единице. И это все стоит под алгоритмом. Поэтому это аккуратно равно нулю. Вот мы с вами и доказали, что информационная дивергенция, она положительна. Она положительна и достигает минимального значения, тоже понятно где. Где все по ЭКУ все одинаковые. То есть неотрицательная величина принимает нулевое значение на одинаковых распределениях и, значит, иначе положительна. Так что в этом плане она годится на роль метрики. Но, если так и быть, важность этой меры, она уже отчасти у нас проявилась. Она проявилась в следующем. Мы с вами считали вероятности P от K при PNKT. Это вероятность того, что в N-испытаниях выпадает K решек или орлов, То есть рассматривается последовательность равновероятности событий, события с вероятностью p и число успехов в этой серии испытаний. Выражается такой формулой. Бинамиальный коэффициент, число сочетания из НПК на p в степени ката. Это сколько раз выпало событие с вероятностью, которое с вероятностью p. на единица на p в степени n-k, это число событий, которые с вероятностью 1-p. То есть у меня k раз выпогорел, n-k раз выпогорел, что вероятность такого события вот такая, и вот число вариантов, которыми это могло случиться. Мы с вами вот этот биномиальный коэффициент выразили, оценили через энтропию, и получилось у нас, что вот это как раз и сводится к 2 в степени минус n на вот эту самую дивергенцию между распределением вероятностей k к n. Ну, то есть вот такая вот кукатая есть k поделить на n. Вполне тянет на распределение вероятностей k и вероятностью p. Значит, как n – это величина меньше единицы, у нас с вами есть распределение вероятностей, вот оно, состоящее из вероятности k поделить на n и вероятности вот этой, единицы минус k поделить на n. Вот это, значит, буква Q, она тут и стоит, вот она. Вот такая оценка вероятности, она выражается через информационную дивергенцию. Ну и, стало быть, отсюда вытекает простой вывод, что когда отношение К к N равно P, то вероятность таких событий максимальна. Во всяком случае, она не стремится к нулю с ростом N. А если К к N отличается от P, то есть если число выпавших событий отличается от ожидания этого события, от ожидания – это как раз P, то, стало быть, у нас с вами эта величина положительная, и вот эта штука стремится к нулю. Так что в длинной серии испытаний количество выпавших решек как раз равно вероятности их относительное количество выпавших решек как раз сходится к П. Как N стремится к П. Отношение к N. Равно p. Те события, для которых это не выполняется, имеют асимптотически нулевую вероятность. Вот оценка их вероятности вот такая. И монотонно убывает по n. Кстати, та же самая информационная дивергенция годится для доказательства неравенства энтропии. Вот у нас было неравенство для энтропии такое, что h от x, распределение случайной величины х, она положительна. и меньше или равно гагарифм по основанию 2m, где m – это мощность множества х. Значит, энтропия случайно любого распределения на m точках, она не превышает вот этого и заведомо положительно. Ну вот мы доказывали, что она не превышает вот этого и достигается на равномерном распределении. Вот смотрите, как это доказать с помощью дивергенции. Смотрите, в дивергенцию совершенно ясно входит энтропия. Вот она, P гагарифм P. Ну, то есть, если этот гагарифм расписать, как гагарифм P минус гагарифм Q, то вот эта сагагаемая окажется минус энтропии. Ну, а в качестве Q давайте возьмём равномерное распределение. То есть, положим так, что у нас QX одинаковая для всех исходов и составляет единица на m. А Px это будет наше настоящее распределение. И посмотрим на дивергенцию между вот этими двумя, между P и Q. Дивергенцию между P и Q. Между распределением P вот таким и нашим и распределением Q, которое равномерно. Это в точности сумма по всем x-ам Px на гагарифм. Пх на qх, а qх это единица поделить на m. Вот так. Ну и значит, вот смотрите, тут вы узнаете минус энтропию, а тут вы узнаете гогарифм минус гогарифм единицы, ну то есть m пойдет в числитель, по p просуммируется. И получается, что это точности гогарифм m отнять аж от p. И это величина положительная, больше вероятно. нулю, как мы только что доказали, потому что это дивергенция. Но откуда следует, что энтропия меньше, чем гагарифм МЭМ? Вот еще одно доказательство этого факта. Вот эти границы для энтропии, умение их доказывать, они для нас как бы принципиально важны, поэтому я так много внимания им и уделяю. Ну вот по той причине, что мы еще будем много таких границ всяких доказывать, но пока, значит, движемся потихоньку дальше. Движение дальше произойдет вот в каком направлении. Я вам хочу про энтропию сказать еще нечто важное. Мы уже ее такие свойства важности и вероятности почти поняли. Ну вот, а теперь я вас хочу, вот одной задачи я ее дам вам решать домой, на дом. И мы ее сейчас просто поставим и обсудим, а вы мне пришлете решение. Значит, эта задача, она проливает свет на роль энтропии в больших статистических задачах. Называется это принцип максимума энтропии. И начать мне придется с выпуклого программирования. Ну, то есть... Сейчас мы с вами займемся проблемой... Там задача ведется к тому, что нам нужно будет максимизировать некий функционал при наличии ограничений. Проблема простая. Вот есть у меня пространство ИРЕНТЫ, то есть пространство блоков, множество всех возможных блоков X1, XНТ, где X вещественные числа, то есть энмерное невклидово пространство. Множество броков вещественных чисел. Это пространство РМ. И на нем заданы некие функции. f от x. Вот я этот x-вектором нарисую, чтобы не писать долго. И мне нужно эту штуку максимизировать. Максимум по ней взять. А ещё, если просто максимум, то с этим вы прекрасно разберётесь. Значит, ситуация такая. Чтобы максимизировать какую-то функцию, надлежит посчитать градиент. Градиент f от x. Ну, то есть вектор, который есть df по dx1 и так далее, df по dxn. И вот если этот градиент как вектор не равен нулю, то вы знаете, куда вам идти в сторону увеличения значения этой функции. Ну, то есть основной принцип оптимизации. В данной точке вычисляешь градиент, и он тебе указует направление, куда сместиться вдоль х, на какой вектор. Вот это, то есть идешь вдоль этого, это вектор, у него вот это... Х первая координата – это х-нт. Он показывает направление возрастания функции, наискорейшего возрастания функции. Ну и, стало быть, вы делаете шажок, приходите в следующую точку, где функции больше, и так вы доходите до стационарной точки какой-нибудь, где градиент этот равен нулю. Вот это условие… Это не условие экстремума, это слабее. Это условие стационарной точки. То есть, когда у функции f от x градиент в какой-то точке равен нулю, нулевой вектор, то, находясь в этом месте, вам идти некуда. Вам вычисление градиента ничего не даёт. То есть, вы находитесь в некой особой точке. Иногда эта особая точка оказывается экстремумом. локальным минимумом или максимумом, но не обязательно, она может быть и седловой точкой, то есть в многомерных пространствах бывают неприятности. Но, тем не менее, большинство оптимизационных задач сводят к поиску стационарных особых точек, где градиент равен нулю. Но вот в выпуклом анализе, о котором у нас, собственно, идет речь в теории информации, Особая точка такая, она всегда является локальным экстремумом, минимум или максимум. И в этом прелесть, в этом прелесть выпуклого анализа. Ну вот, а наша задача как бы разобраться со следующей проблемой. Вот проблема та же самая. Мне нужно f от х максимизировать максимум, но при наличии каких-то ограничений. φ житая от х равно константе. ограничений заданных в виде таких вот уравнений. Вот надо такую оптимизационную задачу с ограничениями. На нее имеется совершенно универсальный рецепт решения. Это он был предложен Гагранжем и в общем его не лишнее знать. Метод Гагранжа отыскания экстремума при наличии ограничений. Смотрите, сводится он к следующему. Вы выписываете так называемый лагранжиан L от x. Я сначала формулирую порядок действий, а потом мы с вами коротенько обсудим, а почему оно получает, почему оно работает. Значит, L от x – это f от x, ну вот это самая исходная функция, минус сумму по всем j по ограничениям, сколько бы их не было, их может быть 150 штук. Ну вот каких-то коэффициентов мюжитая на фи-житая от х. И дальше вы находите градиент этого самого Гранжиана, градиент L от х, и приравниваете его к нулю. И у вас получается некое уравнение для х. Ну то есть, решая это уравнение, вы находите стационарную точку. Вы находите одну или несколько точек особых, где... градиент равен нулю, и где вот этот функционал имеет нечто похожее на экстремум. Ну и значит, этот x, вообще говоря, будет зависеть от μ1, ну вот от всех этих μ, от μжитое. Так напишу, точка, точка, точка, μжитое, точка, точка, точка. Вот, то есть решение этого уравнения, если оно одно, то совсем хорошо, Но оно зависит пока от μ. А для определения этих μ пользуйтесь этими условиями. Подставляете сюда иксы ваши, и у вас уже получаются вравнения для μ. И тем самым вы определяете в конечном итоге экстремум. Давайте сначала я продемонстрирую вам этот пример, этот прием на примере еще, простом очень примере, еще одного доказательства того, что энтропия h от x Меньше или равна логарифма m. Я просто тупо займусь задачей максимизации энтропии. Вот смотрите, h от x – это есть сумма по всем x-ам px на логарифм единицы, делённые на px. Ну, или, если хотите, это минус сумма по всем x-ам px на логарифм p от px. Вот я хочу найти максимум этой самой энтропии по всем распределениям. То есть у меня переменными теперь, вот моим пространством РНТ являются переменные ПХ, то есть у меня есть П1 и так далее, ПНТ. Это мой вектор П. Вот, я хочу по нему найти экстремум, максимум, то есть решить оптимизационную задачу, при каком распределении П вот этим, при каком векторе вероятностей Ну вот, распределение энтропии принимает самое большое значение, максимальное значение. Значит, у меня будут ограничения, оно такое, одно единственное в этой задаче. Сумма по всем хам, пх равно единице. Я должен искать экстремум не по всем пброкам, п1, пнт, а только по распределениям вероятностей. То есть сумма координат должна равняться 1. Ну вот, действую дальше совершенно тупо. По рецепту. Составляю я Гагранжан. Пишу L от вектора Px по всем x. Ну, может быть так, L от вектора P. Это есть. Ну вот то, что я максимизирую. Суммы по всем x. минус px на гагарифм px, это моя энтропия. И минус какой-нибудь коэффициент мю, совершенно произвольный, постоянный, который предстоит определить на ограничение сумма по всем x px. И вот мне этот функционал, вот этот гагранжан, надо искать его стационарные точки. Для этого я должен взять градиент этого самого Лагранжана и приравнять 0. То есть DL по Dpx должно равняться 0. Считаю эту самую производную. Получается, когда эту штуку начну дифференцировать, будет минус логарифм Px. Это я по Px продифференцировал. А теперь дифференцирую логарифм. Логарифм P равен 1 на... Производная от логарифма P равна 1 на P. P у меня уходит. И получается просто минус 1. Это вот производная числитель. То есть производная вот эта. Я дифференцирую по конкретному px. Поэтому в этой сумме от px зависит только одно срыгаемое. Все остальные срыгаемые от px не зависят, поэтому производная по ним нулевая. Но от этой штуки тут тоже только одно срыгаемое зависит от px, поэтому остаётся минус мю равно нулю. Ну вот. Ну и у меня стало быть отсюда получается... условия на стационарную точку. Логарифм Px, ну, сейчас соображу, значит, это я налево увел, так, это я направо переброшу, равно минус μ плюс 1. Ну, то есть Px равно E в степени минус 1 плюс μ. Вот и всё. Это я нашёл стационарную точку. В этой стационарной точке все вероятности одинаковые и составляют вот это. Ну, чтобы найти константу μ, я должен теперь воспользоваться вот этим условием, что сумма по всем этих самых px равна единице. Ну и найду я константу μ, но этого не стоит и делать, по той причине, что она произвольная, и всегда ее можно выбрать так, чтобы... Вот смотрите, все вероятности, главное я получил, я получил следующее, что то распределение, которое максимизирует энтропию, суть равномерное распределение, у него все вероятности одинаковые, просто они еще не нормированы, их надо правильно нормировать, то есть поделить на какую-то константу, и это будут настоящие вероятности. Заниматься этой вознёй совершенно не обязательно. Ясная идея, что мы доказали, что энтропия максимизируется на равновероятном распределении. А равновероятное распределение равномерное, оно известно какое. Во всех точках вероятности одинаковы и составляют единицы на м. Так что вот это ещё одно доказательство неравенства гарифма, и оно нам понадобится для решения задачи на принцип максимума энтропии. Задача немножко громоздкая, но я её решать до конца не буду, и, как я уже сказал, я её вам отправлю в порядке домашнего задания. Подумайте. Она задача важная в том плане, что, вот понимаете, ситуация, жизнь устроена так. Вот вы... уже, наверное, к этому привыкли, что на любую сложную проблему есть какая-то ключевая задача, ключевая модель, на которой почти всё и постигается. Так вот, чтобы понять роль энтропии в науке и технике, так сказать, важно понимать, Генезис происхождения, принципы максимума энтропии. И я умею это показывать на вот такой задаче. Значит, смотрите, у нас имеется с вами один, два и так далее, J большое, районов. Это J я так написал. Живых районов, где живут люди, живых районов, И здесь живет N1 человек, здесь N2 человек, здесь N3 человек. N – это количество NJ. И вот я буду обозначать буквой NJ количество обитателей житого района. Вот тут у меня есть житый район, вот он, номер G. В нем живет NJ3 человек. Сумма этих самых NJ3 по всем G-житым. всем джей составляет n большое это общее количество обитателей нашего города имеется город в нем обыгрывается это так имеется город в нем джей большой районов в каждом районе за живет вот столько человек всего проживает н человек а вот у нас с вами рабочие места для этих товарищей их мы будем обозранина мируются от 1 до кей к и, значит, тут фигурируют числа МКАТ. МКАТ – это предприятие, на котором можно работать, а МКАТ – это количество людей, рабочих мест на этом предприятии. Ну и, стало быть, безработицы у нас с вами нет в нашей популяции, поэтому сумма по всем K МКАТ тоже аккуратно равна N. То есть у меня сколько человек живёт, Для стольких есть рабочие места. Просто проблема в том, что они теперь начнут трудоустраиваться, эти люди. Значит, вот гражданин отсюда может трудоустроиться куда-то вот сюда, а другой гражданин отсюда может трудоустроиться вот сюда. И у нас с вами появляется целая числа WJK. Это количество граждан, которые живут в житом районе и трудоустроено на катом предприятии. Вот это вот WJK. Их называют в этой задаче пассажиропотоками. Ну, то есть между житым районом и катым рабочим местом будут ездить туда-сюда вот столько народу. По утрам на работу, вечером с работы. Ну, и, стало быть, совершенно очевидно, что эти числа WJK, они не просто числа, они связаны определенными закономерностями. Значит, сумма WJK по всем... J, по всем J, ну то есть это число людей, которые работают на катом предприятия. Оно составляет как раз Mk, а сумма W, J, K, по всем K, это число людей, которые живет в этом районе и работают где-нибудь. Это мы договорились, что это Nj. Вот такие ограничения на эти числах. Ну и, стало быть, требуется, не сходя с места, сказать что-нибудь внятное вот на какую тему. Пусть все эти граждане трудоустроились равновероятно, ну, случайно. Сначала поселились, а потом каждый побегал и нашел себе место работы. Каковыми будут вот эти пассажиропотоки? Такая практически значимая статистическая задача. Надо в городе определить пассажиропотоки. Как много людей будет ездить вот на этом транспортном направлении? Как много вот тут будет ездить? Ну, то есть надо определить вот эти числа. Задача плохо обусловлена. Вот по какой причине. Количество этих чисел W, J, K есть произведение J большого на K большое. Ну, то есть всего имеется вот такая туча чисел. У меня тут, скажем, 20 районов и 10 предприятий. Тогда произведение вот этих чисел 200 штук. А ограничений на них я бы выдумал всего вот этих ограничений. J штук, то есть K штук, для каждого K свое ограничение. А вот этих ограничений J штук, то есть их J плюс K штук. И вот это вот J плюс K, сильно меньше, чем J умножить на K при больших числах, во всяком случае. Поэтому условий этой задачи заведомо не хватает, чтобы определить пассажиропотоки однозначно. Ну и тогда здесь вступает в игру вот в таких задачах, когда однозначного решения нет, Вступает в игру принцип максимума энтропии. Я вам сначала его поясню происхождение, чтобы вы понимали, а потом я как бы сформулирую это как оптимизационную задачу. Значит, смотрите, мы с вами добавим еще вот к этим условиям их вот J плюс K штук условий. Дополнительное условие. Будет оно состоять в следующем... Мы введем вероятности P, JK, которые есть W, JK, поделить на N, на общее число граждан. И вот потребуем, чтобы энтропия этого распределения, H от P, вот это будет распределение вероятностей P. Вот так. Ну и, стало быть, потребуем, чтобы энтропия от P была максимальна. И тогда вот это дополнительное условие позволяет определить вот эти пассажиропотоки вполне себе однозначно. Вот. Ну, стало быть, откуда уши растут? Давайте обозначим через W большое. Вот вектор, блок, состоящий из всех возможных этих чисел. W, J, K. Ну, вот я, наверное, немножко напрасно стер картинку. Значит, у меня имеется... Живые районы в количестве 6 штук и рабочие места в количестве кей-штук. Ну и, стало быть, занимаюсь проблемами трудоустройства. Каждое конкретное трудоустройство, распределение, каждое конкретное распределение занятости, оно образует некий вектор WJK, компоненты которого удовлетворяют вот этим ограничениям. Мы будем искать... Это распределение WJK, которое максимизирует энтропию. Вот я сейчас пытаюсь объяснить, почему мы так будем искать. Ответ простой. Значит, смотрите, я сейчас посчитаю количество вариантов N, большое, от W, в котором может реализоваться вот это конкретное распределение занятости. Ну, то есть я беру какой-то конкретный набор этих чисел и буду считать, сколько вариантов, если люди будут устраиваться случайно, то сколько вариантов получить именно такое распределение существует. Считается это довольно просто. Значит, смотрите, я их всех сначала расселю, То есть все уже живут. А потом начну трудоустраивать. И посчитаю количество вариантов трудоустройства. Беру первого, устраиваю его куда хочу. Все пока свободно. На то есть N вариантов. У меня всего N рабочих мест. Следующего я устраиваю. Могу его взять где угодно тоже абсолютно. Просто для его трудоустройства будет N-1 вариант. Ну и так далее. Последнего я уже гарантированно устраиваю на одно оставшееся рабочее место. То есть количество вариантов трудоустройства составляет N факториал. Вот это N факториал. Не все эти варианты ведут к разным W ЖК. Есть у меня джитый район, катые рабочие места, между ними пассажиропоток W ЖК. Значит, если я переставлю... Вот смотрите, есть вот столько граждан, которые живут здесь и работают здесь. Если я этих граждан между собой переставлю, а таковых перестановок существует WJK факториал вариантов, то от этого пассажиропотоки не изменятся. То есть вот в этом числе N факториал на это надо поделить. Ну и, так быть, надо поделить на произведения по всем JK. Вот это, ребятки, оценка количества вариантов, которыми реализуется данное расселение. Ну а дальше логика стандартная, как мы всегда и делали. Мы берем оценку стерлинга n в степени n. Ну вот n факториал, я считаю, что это порядка n в степени n. Сейчас я мысль докончу, я вас отпущу погулять. А тут стоит произведение W. jk в степени wjk. Ну, дальше стандартно перенёс n в числитель и получил тут, значит, на n в степени сумма по всем jk wjk. Потому что сумма wjk – это n. n в n-той степени. Ну, и, стало быть, я вижу, что n в степени так… Ну, вот у меня тут получается в результате единица на произведение… WJK, деленное на N, в степени WJK. Ну, короче, дальше я беру логарифм от этой штуки, и логарифм от этой штуки, где бы мне это написать? Значит, логарифм этого... Единица на N, логарифм N от омега, Когда я стану грифмировать и поделю это всё на n, у меня как раз получится энтропия вот этого самого распределения. Есть минус сумма по всем и jk, wjk поделить на n на грифм wjk поделить на n. То есть для этого самого я имею останку такую. Это порядка 2 в степени n, на энтропию от распределения WJK поделить на N. Ну вот. Ну и отсюда становится понятно, почему надо максимизировать энтропию. Просто оказывается, что из всего количества вариантов расселения экспоненциально велико и зависит от энтропии. Там, где энтропия максимальна, распределение таких вариантов и больше, чем для любой других. Поэтому во внимание надо принимать при больших ансамблях только те распределения, которые максимизируют вот эту энтропию. Их много. Прочих распределений не так уж много. Ну ладно, погуляйте. Вот я вам просто показал, что если оценивать количество вариантов, которыми... Вот это я называю пассажиропотоки, это то, что в термодинамике называют макросостояние. А вот конкретное распределение граждан – это микросостояние. Так вот, данное макросостояние имеет много микросостояния только тогда, когда энтропия максимально возможна. То есть вот это количество подавляет всех, Ну, то есть надо искать такие распределения, которые максимизируют энтропию. А это вот отвечает вот этой моей постановке задачи. В вашем случае это ведется вот к чему. Значит, вы выпишите мне то, что вы максимизируете. Вот пишем Гогранжиан. Гогранжиан такой. L. От W житая катая. Он от всех W будет зависеть. От всего набора. Ужасная такая длинная функция от многих переменных. Он будет равен энтропии, то, что вы максимизируете, то есть минус сумма по всем JK. W, JK. Надо говорить. W, JK поделить. Давайте уже будем вероятностями. Ну ладно, W, JK надо поделить. Значит, у меня уже тут возникли W, JK. делить на n. Вот они. Давайте их как-нибудь обозначим, потому что ясная идея, что оптимизировать придется по вероятностям. То есть назовем Pjk величину, которая есть Wjk, поделить на n. Это распределение вероятностей, потому что сумма по всем jk пассажиропотоков как раз равна n. Я суммирую пока, получаю j, nj, а потом суммирую по J и получаю N. Такая двойная сумма, она равна нулю. Поэтому здесь на самом деле от этих вероятностей будет всё зависеть. То есть я... Стандартная формула для энтропии. P, J, T, K, T, многогарифм P, J, T, K, T, вот тут с минусом. А потом у меня будет чёртова туча ограничений. Я их как-то вот тут выписывал. Ограничения теперь будут такие. терминах вероятности. Они еще были написаны, я их только что стер. Теперь я их выпишу. Ну, прям вот, значит, сумма p, j, k по всем j, это есть, значит, вот было так, сумма w, j, k по всем j, это было m, k, t, а сумма m, k, t равно n. Ну, вот это равно, значит, m, k, t поделить на n. Мы все на n поделили. А сумма p, j, k по всем k это n житая поделить на n. Вот у нас такой вот набор ограничений. Хотите, не хотите. Их надо все присуммировать к агренжану. То есть минус сумма по всем j. Лямбда с индексом j. p, j, k. Минус сумма По всем k, это я вот эти ограничения теперь учитываю. Mu k. А вот тут еще сумма по k, естественно. Вот это ограничение надо поставить. Сумма по j. Так, у меня k там сверху, значит, лям-то k таким образом. Сумма по всем j, p, j, k. Это вот это ограничение. Ну и, значит, еще надо минус, сумма по всем k, по всем k, μk, какие-то произвольные, постоянные, ну а вот это ограничение kt, то есть сумма по всем j, по всем k, тут я с индексами запутался. Значит, у меня имеются вот Вот такие ограничения, они с индексом k, а сумма в них идет по j. Значит, поэтому λ k и сумма по j. Все правильно. А здесь у меня ограничения. Вот они. Сумма идет по k, а сами они с j. Поэтому μ j надо написать. Ну и сумма по всем k. Вот она, p, j, k. И вот этот длинный гагранжан. Значит, его надо честно продифференцировать. Взять градиент. DL по D PJK. Тут окажется, что почти все уйдет в ноль, потому что PJK входит сюда один раз конкретное, сюда тоже один раз и сюда тоже один раз. То есть войдет производная от энтропии и два вот этих коэффициента, житые и катые, JK. Ну и вы получите систему уравнений, решая которую вы найдете, приравняв это дело к нулю, вы получите формулки для нахождения в ЭДЖК. Система очень большая, но она простая, она симметричная, поэтому легко решается. Вот формулы для ПЖК находятся. И, в общем, оказывается следующее, что эти самые ПЖК оптимальны. Итогом решения этой оптимизационной задачи оказывается... Вот такой вывод, такой ответ. Я вам ответ скажу. Ответ он такой. Ответ. Эти самые P, J, K. Есть просто произведение N житых умножить, поделить на N. Это умножить на M катых, поделить на N. Смотрите, NGT на N, это на самом деле у меня вот были количество чисел тут NGT, и сумма NGT равна N. Это, в общем, вероятность поселиться в житом районе. Вот это NGT на N. А это вероятность строиться на катаре работы. Вот это у меня было NKT. В общем, ответ такой, что оптимальное распределение, оно вот такое. Но, соответственно, WGT, а пассажиропоток, он равен NGT умножить на NKT, Просто тупо поделить на это. Ну и вот получилось следующее, что, смотрите, опора на принцип максимума энтропии позволила плохо определенную задачу, в которой пассажиропотоки были никак, нельзя было указать однозначно, потому что существовали всякие расселения, где были самые разные пассажиропотоки. Принцип максимума энтропии позволяет эту задачу решить однозначно и выдать некий ответ. И что самое главное, вы можете быть уверены, что если речь идет о городе-миллионнике, где много народу, где оно велико, то вот это ваше решение, максимизирующее энтропию, будет правильным. Оно единственно возможно на самом деле. Просто все остальные варианты в силу закона больших чисел никогда не реализуются. Это похоже на то, как если бы сейчас все молитвы в воздух собрались в одной половине этого зала, Но такого просто в принципе никогда не бывает. Ну и вот поэтому вот этот вариант, он оказывается надёжным. Ну и, стало быть, вся теория информации, о которой я веду речь, она, собственно, вот кронится к этому, что у нас с вами работает закон больших чисел, и мы на этом основании делаем некие выводы такого термодинамического характера, о задачах обработки информации. Значит, жду эту задачку с решением. Ну, а стало быть, в принципе, в том тексте, в тех задачах, что я вам прислал, она решена в каком-то виде. Там есть решение, так что можете посмотреть, разобраться. Она довольно простая, реально простая. После того, как Эгренжиан продефекцируешь, там получается совершенно примитивное уравнение. Но это видно по форме ответа. На самом деле можно думать так, что у меня есть распределение вероятностей поселиться и распределение вероятностей трудоустроиться. И вероятности вот эти совместные, поселиться и трудоустроиться, они просто такие, как если бы я селился и трудоустраивался независимо. Это независимое произведение и распределение вероятности одного и на другое. Так что получается так, что селимся и распределяемся, и возродустраиваемся независимо от другого. И такое распределение оказывается оптимальным. Ну ладно, а теперь мы с вами уже в своё время занимались законом больших чисел, слабым законом, где мы оценивали дисперсию суммы. То есть было вот так. Мы рассматривали последовательность случайных величин x1, xn, и рассматривали сумму sn, которая определена как сумма этих самых x-ов по j от 1 до n поделить на n. И мы пришли к выводу, что при довольно общих предпоражениях на эти x-ы, эта сумма, в общем, ее дисперсия стремится к нулю. Это так называемый слабый закон больших чисел, с этого я начал. Нам для нашей дальнейшей деятельности нужны чуть посильнее закона больших чисел. Вот сейчас мы займемся усиленным законом чисел. Сходить будем из простого, простой вещи. Сегодня речь пойдет в оставшееся время о неравенствах границах Чебышева для вероятностей больших уклонений. границы Чебышева. Их много. И нам о них что-то надо знать. Значит, смотрите, речь идет вот о чем. Вот у вас есть случайная величина х и какое-то распределение вероятностей P от х. Я нарисовал непрерывное распределение, и тогда вот это плотность вероятности. Ну и, стало быть, всех вот интересует... Вероятность большого уклонения, вероятность события, состоящего в том, что х больше а, где а – это вот это число, то есть вероятность вот в этом хвосте. Вот, собственно, Чебышевские оценки, они связаны с этой задачей. Вот, ну и, стало быть, вот самое простое. И, смотрите, есть некая самая оба. Значит, если посмотреть в литературе неравенства Чебышева, границы Чебышева, то существует довольно много их всяких выводов. Вот я предлагаю такой, значит, есть некая общая граница Чебышева, которая как бы включает в себя всё возможное. Вот она строится так. Берётся некая функция φ от х и строится мат ожидания, мат ожидания, average, от этой самой функции. Всё с этого начинается. Ну, то есть сумма по всем х, px на φ от х. Вот. Теперь некие предположения по поводу этой функции. Значит, функция φ от х должна быть положительна. φ от х должна принимать только положительные значения. Пока она абсолютно любая, но вот я накладываю первое ограничение. Она положительна. И мне надо это от ожидания оценить. Я буду его оценивать так. Коль скоро функция φ от х положительная, то я могу от суммы оставить часть. Поскольку вероятности положительные и сама функция положительная, то я могу написать так, что больше или равно сумма по х большим а. П от х на φ от х. Решает неравенство... Никаких особых проблем не вызывает вопросов. Я просто от суммы положительных чисел отбросил нечто, отбросил все слагаемые, у которых х меньше а. Ну, то есть вот у меня было распределение х, ось х, там были всякие х, а я взял только точку а и оставил только вот эти х. Вот это я выбросил. И поэтому вот эта штука заведомо уменьшилась от этого, то есть больше или равно. Вот. А следующий этап состоит в следующем. Значит, на функцию φ от х у меня пока одно ограничение. Она положительна. Ну, не отрицательна. Больше или равна нуля. Вот. А второе ограничение, там мне понадобится такое, что из условия х больше а, ну, то есть она монотонная, следует, что φ от х больше, чем фи от а. Ну и тогда смотрите сами. В этой сумме идет сумма по х больше а, и тут стоят фи от х. Я их могу заменить на фи от а. Поскольку х больше а, то фи от х больше, чем фи от а. И поэтому я могу это неравенство продолжить так. Больше или равно Сумма по х больше а, пх на фи от а. А это и есть то, что мне нужно. Вот. Ну и отсюда вот вытекает общая граница Чебышева. Она выглядит так. Значит, вероятность события, состоящая в том, что случайная величина х превышает а. Это есть сумма по всем хам, Большим А, ПХ. Ну, то есть ровно вот эта штука. Она меньше или равна, верхняя оценка, на мат ожидания ФИАТ-Х, мат ожидания ФИАТ-Х, поделить на вот это самое число А, на ФИАТ-А. Вот и всё. Эта граница верна. Для любой функции Фиат-Х, которая А не отрицательна, Б монотонна. Неравенство Чебышева прошу любить и жаловать. Вот так. С него все начинается в нашей науке. Значит, Фиат-А меньше, чем Фиат-При. Если А меньше, ну да, все у меня правильно. Ну, то есть два условия. Одно условие — неотрицательность. Фиат-Х позволило мне вот тут перейти. И второе, поскольку тут сумма по большим иксам, а при больших иксах эта штука больше, чем значение. Ну, то есть она так вот себя ведет. Вот если вернуться к моей картинке, вот есть у x, вот я поставил точку А, а Фиат-Х, она какая-то вот такая, неотрицательная. И для больших иксов вот это значение Фиата а вот это значение φ от x. Ну и стало быть, φ от x больше, чем φ от А, поэтому вот это неравенство верно. Дальше я просто вот тут скобочку поставил, φ от А это константа, и написал, что эта сумма меньше, чем вот это поделить на φ от А. Вот это неравенство Чебашева в общем виде. Прелесть этого всего заключается, ну вот как вам сказать, Это, конечно, некое трюкачество. Меня привлекает в этом всём простота этого доказательства, простота и изящество. По-другому тоже можно, но как-то логически сложнее. А здесь всё очень стройно и просто. Вот доказательство в две строчки, и мы получаем фундаментальную границу Чебышева, из которой все остальные границы как-то просто выводятся. Значит, вот смотрите, какие варианты Fiat X мы можем взять? Первое, что публика делает, берет в качестве fiat x xa равно x. Ну вот, вот эту ось x, вот эту fiat x, ну и стало быть вот такой x. Выполнены наши требования на fiat x? Ответ такой – нет, не выполнены. Дело в том, что x нельзя брать по той причине, что вот тут fiat x отрицательна, она должна быть положительна у нас. Но вот тут выход такой. Мы говорим такие слова, что случайная величина х принимает только положительные значения, то есть она определена вот здесь. Вот тогда х годится. Так что если случайная величина х принимает только положительные значения, больше или равно нуля, то мы можем в качестве хи от х воспользоваться функцией х. Она годится. Она монотонна и везде положительна. Ну и тогда я, смотрите, пишу такие вещи, что если вероятность того, что это самое х больше а, ну как сумма по х больше а по х, меньше или равна мат ожидания х, среднего значения х, поделить на эту самую величину а. Ну, у меня Фи от Х равно Х. Поэтому тут стоит мат ожидания Х, а здесь Фи от А равно А. Вот и всё. Вот это, ну, как бы такой ослабленный вариант границы Чебышева. Значит, вероятность большого уклонения, вероятность того, что неотрицательная случайная величина превышает А, она не превышает мат ожидания этой случайной величины, делённой на А. Ну и, стало быть, с ростом А, вот смотрите, в чём логика. Когда А большой, у меня идёт речь о больших отклонениях, то эта штука, вероятность падает. Проверим силу этой границы. Значит, вот сейчас я вас немного разочарую. Значит, смотрите, граница производит гнусное впечатление. Вот доподлинно известно, что средний рост человека на планете Земля составляет 2 метра. Ну, то есть, вот возьму за случайную величину рост человека. Х – это рост. Значит, мат ожидания Е от х составляет 2 метра. Ну, условно, я не хочу точными цифрами оперировать. Тогда эта граница Шебышева говорит следующее, что вероятность того, что я встречу человека высотой х больше 20 метров, а равно 20 метров. она, видите ли, не превосходит х среднего 2. Поделить нам от ожидания на число А, на 20. Одна десятая. Ну, то есть, вот такая слабая оценка получается с виду. Вероятность встретить 20-метрового клиента не более одной десятой. Об этом говорит граница Щебышева. Вот плохо это или хорошо? Хорошо. Значит, ребята, вот создается впечатление, что это очень плохо, никаких 20-метровых индивидов по улицам не гуляет. Вот, поэтому... Но дело в том, что на самом деле это хорошо, эта граница достигается, и она неугудшаема. Проблема в ней такая, она, видите ли, сформулирована для любого распределения вероятностей P от X. На P от X вообще никаких ограничений, кроме существования мат ожидания вот этого. Вот и всё. Вот, не сходя с места, привожу пример, когда граница Чебышева достигается. Значит, представим себе, что у нас с вероятностью p 1 десятая рост человека составляет 20 метров. Ну, то есть из каждых 10 граждан 1 десятая имеет рост 2h равно, вот, x равно 20 метров. А все остальные 9 десятых имеют рост нулевой. Ну, то есть, этот Гулливер, а все остальные релипуты. Ну, вот тогда получается следующее, что вероятность встретить Гулливера 1 десятая как раз, да? Вот она, 1 десятая. А средняя высота человека как раз составляет 2 метра. 9 десятых умножить на 0 плюс 20 метров умножить на 1 десятую. То есть, х средняя, мат ожидания х, а, составляет 2 метра. Это значит 1 десятая умножить на 20 плюс 9 десятых умножить на 0. Ну то есть двойка, 2 метра. Ну то есть вот на этом примере границ Чешево просто достигается. Вероятность встретить вот в этом ансамбле, так заданном, с таким распределением вероятность третьих человек высотой больше 20 метров аккуратно равна 1 10 то есть границы достигается с равенством это говорит о том что границы это не улучшаем от слова совсем и поэтому считать ее плохой как-то противоестественно чтобы улучшить ну надо как-то ограничивать класс распределения вероятностей на х. Если по х допускается любое распределение вероятности, например, вот такое идиотское, то она достигается на нём. Вот, поэтому граница Чебышева. Ну, а стало быть следующее, что можно, это посмотреть улучшенная граница Чебышева, так сказать. Следующий вариант, φ от х равно х квадрата. Совершенно очевидно. Ну и стало быть, смотрите, я тут поведу речь уже не о случайной величине х, а о случайной величине модуля х. Вот такая вот случайная... m – это мотожидание х, то есть отклонение от мотожидания. Ну то есть... m это мат ожиданием случайной величины х, мат ожидания от х. Я взял центрированную случайную величину с нулевым математическим ожиданием. Дело в том, что мат ожидания от квадрата этой штуки это и есть дисперсия. Поскольку я беру функцию х квадрат, тайный смысл-то какой? У меня там войдет в числителе мат ожидания φ от х. Вот мат ожидания. от этой самой функции φ от x в общей границе Чебышева. А х у меня, если взять x минус m, то это мат ожидания, а φ от x от x квадрата, то это будет мат ожидания от x минус m, ну, модуль тут для порядка надо написать, в квадрате. Вот мат ожидания вот тут. Я вспомнил, что φ от х у меня равно х квадрата, в качестве х подставил х минус m. То есть у меня в числителе неравенствовать границу Чебышева появится вот это e от φ от х, и я пытаюсь понять, чем оно будет. Оно будет просто дисперсией этой самой случайной величины. Это так и называют, неравенство Чебышева через дисперсию или для дисперсии. То, что тут модуль, это несущественно. Я теперь могу, поскольку это в квадрате, Ну, модуль x в квадрате равен просто x в квадрате. Поэтому угловые скобки эти можно заменить на обычные. А это и есть σ в квадрат. Так что у меня будет получится так. Смотрите. Давайте я где-нибудь вот этот пример сотру теперь же. Ну и, стало быть, напишу границу Чебышева. Значит, меня будет интересовать вероятность того, что P, вероятность того, что X минус A, вот модуль X минус A этой самой случайной величины, ой, то есть минус M, этой самой случайной величины, превышает обычную границу Чебышева, как она есть. Вот, это меньше или равно, согласно Чебышеву, нам от ожидания, у меня сейчас FiatX, есть х квадрат. Вот, ну и стало быть, вот гляжу туда, в числителе стоит мат ожидания φ от х. Я с ним уже как бы разобрался, это дисперсия случайной величины моей. Сигма квадрат. Мат ожидания, сигма квадрат, то самое, которое математическое ожидание от х минус n в квадрате. Вот, а в знаменателе стоит, ну, более-менее понятно, что, а квадрат. Вон там стоит Фиат А, а Фиат Х это у меня Х квадрат. Вот так выглядит неравенство Чебышева для дисперсии. Прошу любить и жаловать. И, собственно, не надо ничего доказывать. Я в общую границу в качестве Фиат Х поставил Х квадрат и получил оценку для укронения случайной величины от ее мат-ожидания, от ее среднего значения. Получается, что если это укронение велико, то вероятность этого события, ну, в общем, при больших А становится никакой, причем квадратично никакой. Вот тут А2, это существенно. В той границе, что была выписана, в знаменателе было А, а здесь А2. Поэтому вот эта граница, она как бы порезче зависит от А. Хотя сама случайная величина вполне линейная, вполне себе линейная. В общем, фокус тут в том, что модуль x минуса в квадрате условия модуль x минус m больше а это то же самое что условие x минус m в квадрате больше а квадрат ну просто возвел обе части в квадрат и можно вот через это эту границу получать тогда квадрат вот естественным образом получается а так это весьма существенно в обычном В обычной границе Чебышева у нас была вероятность того, что неотрицательная случайная величина больше А, меньше х от ожидания х поделить на просто А. А здесь на А квадраты это существенно, оно бывает квадратично. Так, все у меня что ли? Нет? А, люди из прошлого. Откуда вы такие взялись? Ну ладно. Хорошо, значит, таким образом вот две границы Чебрчева. Так, ну и, стало быть, у нас там как со временем-то? Хотел бы я сориентироваться. Так, сейчас у нас 17... 15 минут еще есть, хорошо. Так, ну стало быть, как-то мне дальше двигаться вот в этом направлении не хочется, потому что мы не успеем. Я расскажу один момент, который я просто упустил. Я хотел это рассказывать, но как-то это у меня за постановкой задачи вылетело из головы. Значит, я вам хочу рассказать, чтобы вы стали понимать, откуда берется эта гранжевая оптимизация, потому что я ей часто буду пользоваться, мне это все важно. Мне важно уметь максимизировать всякие энтропии по всяким переменам. Значит, смотрите, я вот рассмотрю случай двух переменных. Итак, вот f от x, y мне надо максимизировать при условии, что φ от x, y есть некая константа. И мы с вами все поймем на этом простом примере. Выглядит это так. У меня имеется проскость x, y. Это как бы фазовое пространство моей задачи, а в общем случае это нмерное пространство. Здесь вот я смотрю самый простой пример. Итак, вот у меня х, вот у меня у, и в каждой точке определено некая функция. Я ищу ее экстремум. Значит, смотрите, у функции f от x, y, если написать условие f от x, y равно а, Поэтому то, что будет, это называется линия уровня. Это то, что на типографских картах наносится. Вот какая-то линия, и на этой линии у нас с вами f от x, y принимает вполне конкретное значение. Вот тут она принимает значение побольше, а вот тут еще побольше. И кроме того, в каждой точке вы можете посчитать градиент этой функции. Ну и, так как быть, он всегда ортогонален линии уровня. Вот это вектор градиента f. тот самый, который имеет координаты df по dx и df по dy. Вот такая вот картина. И вот в стандартной задаче оптимизации без ограничений вы поступаете так. Чтобы двигаться к стационарной точке, вы просто идёте в сторону, куда показывает градиент, и всё время переходите на более и более высокие уровни, то есть движетесь на горку, так сказать. А если против градиента, то под горку. А теперь у нас появилось вот это ограничение φ от x, y. И оно нам накладывает некое жёсткое условие. Мы двигаться можем не как угодно, а только там, где φ от x, y есть величина постоянная. То есть наряду с линиями уровня функции f, где она принимает постоянное значение, у нас появились линии уровня функции φ. И вот одна из них, скажем, вот такая. Вот это φ от xy равно константе. Вот она. Здесь удовлетворены наши условия. Так вот, смотрите, стационарная точка имеет место, точка подозрительная на экстремум, она имеет место там, где вот эти линии уровня касаются. одна и другая, тогда это означает, что гуляя вот здесь, я не могу пойти в сторону градиента, вот этой максимизируемой функции. Градиент вот сюда смотрит. А гулять я должен вдоль линии уровня, которая идет по касательной. Поэтому если я сижу в этой точке, то у меня как бы не двигается вдоль этой линии, значение вот этой функции не меняется. Это особая точка. Ну и стало быть условие на особую точку совершенно простое. Вот смотрите, в этом месте у нас должны быть коллинеарные градиенты. Вот у нас имеется линия уровня функции f от x, y и линия уровня функции φ от x, y. Вот они. И они вот в этом месте касаются друг друга. Это означает, что там, где они касаются, у нас градиент f... градиент F, и градиент F, сонаправленные, они ортогональны одной и той же линии, и градиент F коллинеарные. То есть условие такое, градиент F равен μ на градиент F. Вот это и есть условие на экстремум в случае с ограничениями. Но, стало быть, обыгрывается это так, это записывается в градиент от f минус мюфи равно нулю это вот объявляется гагранджианом и градиент гагранджиан равен нулю вот что означает коэффициент это коэффициент от коэффициент пропорциональности между коллинеарными градиентами, максимизируемые функции и функции ограничений. Они должны лежать в одном пространстве, быть коллинеарны. Тогда эта точка претендует на точку экстремума какого-то. То есть, двигаясь вдоль линии ограничения в окрестности этой точки, я не меняю... максимизируемую функцию. Она безразлична в этом положении. Ну вот так. А в общем случае у нас картина такая. У нас есть 50 ограничений, много может быть, φjt от x, φj от x, и у каждого этого ограничения есть градиент. Градиент φjt от x. Это вполне конкретный вектор. Вот у вас есть точка конкретная x, И в ней для каждого ограничения есть вектор, который называется градиент этого самого Фи-Жи-Т от Х. Его координаты – это просто частная производная от Фи по всем Х. Ну и таковых векторов много. Для каждого ограничения свой вектор. То есть есть градиент Фи-1 от Х, есть градиент Фи-2 от Х и так далее. Градиент Фи-НТ от Х. от x вот их в этой точке торчит много векторов мы сидим не на плоскости мы сидим в многомерном пространстве и стало быть вот на эти вектор а натянута некое подпространство это подпространство запрещенных движений туда ходить не ходи ну то есть каждое каждый У меня же ограничение тут проходит, и сместиться в сторону градиента ограничение не даёт. Вот когда у меня есть ограничение, φ от x, y есть константа, то я обязан ходить вдоль этого ограничения, а градиент направлен сюда. Движения в сторону градиентов этих запрещены. Поэтому вывод тут такой. Особая какая-то точка на максимизируемой поверхности стационарна, если градиент в этой точке лежит в подпространстве, натянутом на градиенты. В пространстве запрещенных движений. Ну, то есть условие такое. Градиент F есть линейная комбинация по J от 1 до N. Каких-то коэффициентов на J градиент фи j от x ну вот на градиент фи j ну то есть не могу я сдвинуться никак в сторону градиента по той причине что все градиент лежит в линейном подпространстве к попадании в которые запрещено ну вот это условие гагранжа для общего случая значит обыгрывается это так записывается что когда у вас много ограничений то градиент максимизируемой функции минус сумма нюжитых на фижитах, все это от х и от х, равно нулю. Вот это ищется. На самом деле это условие того, что градиент максимизируемой функции лежит под пространством, натянутым на градиенты ограничений. Ну вот так работает Лагранж. Ну и дальше все просто. Значит, вы раскрываете Вот это называете Гагранджианом L от x. И пишите условие, что градиент L от x равен 0. Это даёт вам какие-то решения x. Зная эти x, эти выражаются через эти произвольные коэффициенты мюжитые. Зная эти x, вы их подставляете в ваше ограничение и получаете уравнение для мюжитые. удивительно но это работает во всяком случае выпуклому анализе это работает тут точно ну то есть задача оптимизации она не очень простая вот изменение не рассматривать выпуклые функции а мы как раз вот находимся с вами в области выпуклого анализа и поэтому задача максимизации у нас проходит всегда и ну а значит вот совсем короткий экспорс выпуклый анализ. Значит, смотрите, вот множество распределений, вот возьмем точное распределение P1 PNT и обозначим его вектором P. Значит, сумма P житых равна единице, естественно. Вот это некий вектор, вот точка, вот начало координат, и вот это вектор P в мерном пространстве. Пусть есть другой вектор в инмерном пространстве, вектор Q. Вектор Q, который есть сумма Q1, QnT. Значит, интересно то, что это тоже распределение вероятностей, то есть сумма Qжитов тоже равна 1. Отсюда можно сделать далеко идущий вывод, что множество всех возможных распределений вероятности суть выпуклое множество. Потому что при любом лямбда... от нуля до единицы, вот такой вектор, лямбда на p вектор, плюс единица минус лямбда на q вектор. Ну, то есть, вот у меня есть вектор p, вот у меня есть вектор q, я провел через них отрезок прямую, соединив их отрезком. При лямбда, равном нулю, я попадаю в q точку, вот эта точка лямбда равно нулю, вот эта точка лямбда равно единице. То есть, я гуляю вот тут по прямой. И любая эта точка при любом лямбда является распределением вероятностей. Потому что если я вот все эти вероятности просуммирую вот так, сумма по всем g, ну то есть сумма лямбда на p житая плюс единица минус лямбда на q житая, сумма по всем g, это вот одна из вероятностей, входящая в этот вектор. И я по всем таким вероятностям просуммировал, Ну, поскольку сумма P равна 1 и сумма Q равна 1, то это λ плюс 1 минус λ и равно 1. Так что это тоже распределение вероятностей. Так что множество распределений вероятностей образует выводное подножество. В том смысле, что если две точки являются распределениями, то... Отрезок, соединяющий их, входит в это множество. Это одно из определений выпуклого множества. Ну а дальше вот появляются границы понятия выпуклой функции. Значит, функция Фиат-Х является выпуклой, вот такой вот, выпуклый вверх Фиат-Х. Если выполнено такое условие, что вот если взять две точки, точку 1 и точку 2, ну, в нашем случае это будут распределение P и распределение Q, и посчитать значение функции в средней точке, ну, вот я это сформулирую для энтропии. Значит, энтропию H от распределения λP плюс 1 минус λ на Q Ну, то есть энтропия посчитана, и вот я соединю точки P и Q прямой линии и сегу вот в эту точку с конкретным лямбда. Так вот, энтропия в этом месте больше, чем среднее значение энтропии в этом и в этом местах. Это больше или равно, чем лямбда на H от P. Вот, лямбда на H от P. Прибавить единица минус лямбда на H от Q. Вот в этом смысле энтропия является выпуклой функцией. Ну вот, в дополнение к той задаче, что я вам уже дал, попробуйте это доказать, я в следующий раз с этого начну. Значит, простое очень доказательство. Значит, приносишь всё налево и пытаешься воспользоваться неравенством вот этим границей информационной дивергенции. Короче, вот Это энтропия минус... Ну, тут когда распишешь, что получится λp на сумму вот этих вероятностей, и он скомбинируется с суммой λp на... Ну, вот смотрите тут как. Идею я подскажу. Идея совершенно простая. Вот я рассматриваю энтропию, формулу для энтропии слева и справа. Слева будет так. Сумма по всем j λp на pj плюс... единица минус лямбда на QJ на гагарифм. Ну вот того же самого, что вот тут написано. Не в этом суть. А справа будет сумма по J лямбда-множителем PJ на гагарифм PJ и сумма с коэффициентом единицы минус лямбда PJ на гагарифм PJ. Куджей, гарифм Куна, единица минус Ку. Вот так структура этой штуки. Но и стало видно, как надо объединять. Смотрите, П с П, вот этих, объединяем воедино, лямбда множитель единый. И у нас получаются информационные дивергенции между распределением П и вот этим всем распределением. Она положительная. А тут объединяем Куи. компонент из Q. Тут у нас входит вот эта сумма вероятности λp плюс единица минус λq. А тут у нас входит просто распределение Q. Вот и всё. И опять же получается информационная дивергенция между вот этим распределением и вот этим распределением. Обе эти дивергенции положительны, поэтому получается неравенство для выпуклости энтропии. Такая вот нехитрая задача. Значит, если знать информационную дивергенцию, то доказывает совсем просто. Если не знать, то тоже просто. Тогда можно через неравенство гагарифмов. То есть тут туча существует методов доказательства. Но в общем, знайте, что... Множество вероятностей, распределение вероятности – это выпуклое множество, а энтропия, с которой мы имеем дело, это выпуклая функция на этом множестве. И поэтому задача максимизации всякого рода энтропии проходит очень хорошо. Видите, выпуклые функции так устроены, что у них не может быть локальных экстремумов. Если какой-то экстремум имеет место быть, вот стационарная точка, то он глобален. Поэтому задача оптимизации выпуклой сильно проще, чем просто задача оптимизации произвольных функций. И поэтому это хорошо. Это позволяет нам в теории информации жить более-менее простой жизнью. Но главное, все.