

Теория вероятности

Дима Трушин

Семинар 1

Случайные объекты

Случайность Во-первых, хорошо бы понять, а что такое случайный объект и как про него правильно думать. Любой случайный объект – это такой черный ящик, который мы можем спросить, в каком состоянии он сейчас находится. И этот черный ящик нам услужливо отвечает, но каждый раз на этот вопрос может давать разные ответы. В качестве примера давайте рассмотрим кубик с шестью гранями. Что мы делаем: мы берем и бросаем кубик, а потом смотрим какой стороной он выпал. Давайте поясню, как на этот процесс смотреть как на черный ящик. Бросание кубика можно рассматривать как процесс задания вопроса черному ящику, а когда на нем выпадает грань, мы на это смотрим как на ответ черного ящика.

Для изучения любого такого черного ящика для начала надо описать все возможные состояния, в котором он может находиться. Обычно множество таких состояний обозначают через Ω , а каждое состояние $\omega \in \Omega$ называют элементарным исходом. В примере с кубиком в качестве Ω можно взять множество из шести граней, нумеруя каждую соответствующим числом.

Случайные события Давайте начнем с примера. В случае кубика, мы можем задать следующее событие: «выпала четная грань» и «выпало число 7». Теперь мы задаем вопрос кубiku и смотрим, наступило требуемое событие или нет. Таким образом можно все элементарные исходы поделить на две части: те которые удовлетворяют событию и те, которые не удовлетворяют. Если $\Omega = \{1, 2, 3, 4, 5, 6\}$, то «выпала четная грань» соответствует подмножеству $\{2, 4, 6\}$, а «выпало число 7» соответствует пустому подмножеству \emptyset . То есть как мы видим из этого примера, случайные события должны соответствовать подмножеству множества элементарных исходов. Потому в общем случае это кладется в качестве определения.

Пусть Ω – множество элементарных исходов. Тогда событием (или случайным событием) называется любое подмножество $A \subseteq \Omega$. Если у нас есть два события $A, B \subseteq \Omega$, то для них определены все теоретико-множественные операции. $A \cap B$ означает, что оба события произойдут одновременно, $A \cup B$ означает, что произойдет хотя бы одно событие, $A \Delta B = (A \setminus B) \cup (B \setminus A)$ означает произойдет одно из событий, но не оба сразу. Пустое множество \emptyset означает невозможное событие (ничего не может произойти, так как ни один элементарный исход не подходит), а все множество Ω означает, что произойти может любое событие. Отрицанием события A называется событие $\bar{A} = \Omega \setminus A$, это значит произошло не A , то есть что угодно, кроме события A .

От математика не математикам Тут я хочу вставить небольшой, но важный разговор. Так часто бывает, что математика бывает сложно устроена, не потому что решаемая задача сложная, а потому что сама математика внутри устроена сложнее, чем задача, для которой она изобреталась. И вот эти самые трудности внутреннего устройства самой математики очень часто сильно переусложняют рассказываемый текст. С одной стороны, это делает строгие математические книжки просто нечитаемыми для обычных смертных, какими бы хорошими эти книжки ни были, с другой стороны, математики не могут просто игнорировать подобные явления, иначе математика внезапно станет противоречивой.¹ Однако, мы же не собираемся создавать математику, мы будем ей пользоваться. Потому я намерено буду обманывать вас относительно технических вещей связанных с теорией множеств, которые возникают в рамках теории вероятности. Говорю я это потому, что вы можете пойти и поглядеть в различную литературу определение вероятностного пространства и увидите, что после разговора о пространстве элементарных исходов Ω люди начинают строить какие-то

¹Георг Кантор, кстати, сошел с ума именно потому что придумал крутой способ мерить количество элементов в бесконечных множествах, но не сумел сделать этот способ непротиворечивым. Его идею можно легко рассказать другим, но сделать строгой и формально рабочей – задача не самая простая.

непонятные σ -алгебры событий, потом измеримые функции, интегралы Лебега и прочую хренатень. Так получилось, что без всех этих серьезных наворотов построить теорию вероятности невозможно, иначе она будет противоречивой. Грубо говоря, основная проблема следующая: когда нам дали бесконечное множество элементарных исходов Ω , то нельзя в качестве события взять произвольное подмножество, потому что, когда мы будем вводить понятие вероятности, у нас внезапно вылезет критический баг в теории. Потому надо сначала проредить правильным образом все множества. Потом надо будет прореживать функции, потом строить специальную теорию интегрирования этих функций и т.д. Но самое забавное заключается в том, что если вы попытаетесь найти те самые плохие множества, от которых мы пытались избавиться, у вас это ни за что не получится. Оказывается, что они существуют только тогда, когда вы верите в аксиому выбора², а если вы в нее не верите, то нельзя ни доказать, что таких плохих множеств нет, ни привести примера. Проще говоря, в нормальной жизни не математика они попросту не встречаются, и их вообще говоря никто никогда не видел и не увидит. Потому я при рассказе о теории вероятности вырежу всю эту трудно перевариваемую лабуду и буду рассказывать почти честно, но зато во много раз (надеюсь) понятнее.

Вероятность Пусть у нас есть какое-то пространство элементарных исходов Ω . Неформально, вероятность – это правило, которое каждому событию (то есть подмножеству) ставит в соответствие неотрицательное число (вероятность события) и ставит это некоторым хорошим способом. Давайте введем строгое определение.

Пусть Ω – пространство элементарных исходов (то есть произвольное множество), через 2^Ω обозначим множество всех подмножеств Ω .³ Тогда вероятность на Ω это отображение $P: 2^\Omega \rightarrow \mathbb{R}_+$ (из множества всех подмножеств Ω в множество неотрицательных вещественных чисел), такое что выполнены следующие правила

1. $P(\emptyset) = 0$, $P(\Omega) = 1$.
2. $P(A \cup B) = P(A) + P(B)$ при условии, что $A \cap B = \emptyset$.
3. Пусть есть последовательность непересекающихся событий $A_i \subseteq \Omega$, $A_i \cap A_j = \emptyset$, тогда

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

В этом случае значение $P(A)$ называется вероятностью события $A \subseteq \Omega$. Множество Ω вместе с вероятностью P называется вероятностным пространством. Именно эта пара полностью описывает любой случайный объект. Отображение вероятности P еще называют вероятностной мерой.

Вероятность любого события обязательно находится между нулем и единицей. Условие $P(\emptyset) = 0$ означает, что вероятность невозможного события равна нулю, а условие $P(\Omega) = 1$ означает, что вероятность, что что-нибудь да произойдет равна единице.

Третье условие – это лишь более сильная версия второго. Потому второе условие можно было бы откинуть, но оно очень полезно для понимания, что происходит. Второе условие означает, что если у вас есть два события, которые не могут случиться одновременно, то вероятность того, что случится одно из них равна сумме вероятностей. Еще полезно понимать следующие свойства

1. Если есть два события $A \subseteq B \subseteq \Omega$, то есть событие A всегда влечет событие B , то

$$P(B \setminus A) = P(B) - P(A)$$

Это значит, что если A влечет B , то вероятность, что произойдет B , но A при этом не случится, равна разности вероятностей $P(B) - P(A)$.

2. Если есть два произвольных события $A, B \subseteq \Omega$, то

$$P(B \setminus A) = P(B) - P(A \cap B)$$

Это значит, что если есть два произвольных события A и B , то вероятность того, что произойдет B , но не A , равна разности вероятности того, что произойдет B вероятности одновременного выпадения A и B .

²Это одна из очень популярных и не очевидных аксиом теории множеств. Проблема в том, что в абстрактной математике она очень полезная.

³То есть запись $A \in 2^\Omega$ означает, что $A \subseteq \Omega$.

3. Для любого события $A \subseteq \Omega$, верно

$$P(\bar{A}) = 1 - P(A)$$

То есть вероятность отрицания события A есть дополнение вероятности A до единицы.

4. Если есть два события $A \subseteq B \subseteq \Omega$, то есть A влечет B , то

$$P(A) \leq P(B)$$

То есть у не меньшего события, не меньше вероятность. Может так оказаться, что у большего события вероятность такая же, но это будет означать, что вероятность их разности нулевая.

5. Если есть два события $A, B \subseteq \Omega$, то

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Если вы знаете, что события A и B могут произойти одновременно, то вероятность того, что случится одно из них считается по формуле выше, надо сложить вероятности событий и вычесть вероятность их одновременного выпадения. Это называется формула включения исключения.

Опять пример с кубиком Давайте переключимся на наш пример с кубиком. В этом случае $\Omega = \{1, 2, 3, 4, 5, 6\}$. Откуда брать меру? Ее надо задать. Из каких соображений? Здесь надо исходить из физических условий задачи. Все зависит от того, какими свойствами обладает наш кубик. Если мы считаем, что кубик хорошо сбалансирован, то это означает, что вероятность получить любую грань одинаковая. То есть $P(i) = p \in \mathbb{R}_+$, где $1 \leq i \leq 6$. С другой стороны

$$1 = P(\Omega) = P(\{1\} \cup \{2\} \cup \{3\} \cup \{4\} \cup \{5\} \cup \{6\}) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) = 6p$$

Значит $p = \frac{1}{6}$. Но если мы сомневаемся, что кубик был правильным, то вообще говоря, чтобы задать вероятность на всех событиях мы должны задать вероятности каждого элементарного исхода произвольно $P(i) = p_i$, где $1 \leq i \leq 6$, при этом $0 \leq p_i$ и $\sum_{i=1}^6 p_i = 1$. Теоретически подходят любые наборы чисел p_i удовлетворяющие последним двум условиям. Все они будут отвечать кубикам с разной степенью балансировки граней.

Как можно было бы экспериментально измерить подобные числа p_i ? Можно было бы сделать так: берем кубик и кидаем его n раз. Считаем сколько раз выпало число 1 и получаем число $k_n(1)$, для числа 2 получаем $k_n(2)$ и т.д. Важно, что $\sum_{i=1}^6 k_n(i) = n$. Теперь посмотрим на частоту выпадения грани i после n экспериментов и получим $\frac{k_n(i)}{n}$. Если бы мы могли провести эксперимент бесконечно долго, то мы могли бы найти предел

$$p_i = \lim_{n \rightarrow \infty} \frac{k_n(i)}{n}$$

Это и были бы те самые вероятности, которые надо было бы приписать к данному конкретному кубiku.

Пример с монетками Пусть мы бросаем монетку n раз. В случае выпадения орла мы будем писать 0, а в случае выпадения решки – 1. Тогда в качестве результата нашего эксперимента (или ответа черного ящика) у нас будут последовательности из 0 и 1 длины n . Всего таких последовательностей будет 2^n . Таким образом $\Omega = \{(a_1, \dots, a_n) \mid a_i \in \{0, 1\}\}$. Вероятность зададим так: $P(a_1, \dots, a_n) = 1/2^n$. То есть все элементарные исходы равновероятны.

Конечные множества элементарных исходов Пусть у нас некоторый случайный объект описывается множеством исходов Ω и пусть это множество конечно, то есть у нас возможно всего лишь конечное число возможных состояний у нашего случайного объекта. Пусть $\Omega = \{\omega_1, \dots, \omega_n\}$. Оказывается, что чтобы задать вероятность на таком пространстве, то достаточно каждой точке ω_i приписать ее вероятность, а именно положить $P(\omega_i) = p_i$, где числа p_i удовлетворяют следующим двум условиям

$$0 \leq p_i \text{ и } \sum_{i=1}^n p_i = 1$$

Тогда вероятность $P: 2^\Omega \rightarrow \mathbb{R}_+$ задается по правилу, для любого $A \subseteq \Omega$ положим

$$P(A) = \sum_{\omega_i \in A} p_i$$

То есть, чтобы посчитать вероятность события A , надо сложить вероятности всех элементарных исходов этого события. Это самый простой способ задавать вероятности. На конечных множествах других способов задавать вероятности не существует. Потому, когда описывают случайный объект с конечным числом состояний, всегда указывают вероятности всех элементарных исходов. Еще в подобных случаях принято считать, что $p_i > 0$, так как если у элементарного исхода вероятность ноль, то его можно выкинуть из Ω ибо это бы означало, что этот исход просто не может произойти.

Примеры «непрерывных» мер Пусть у нас $\Omega = [0, 1]$ – отрезок. Тогда для любого интервала $[a, b] \subseteq [0, 1]$ положим $P([a, b]) = b - a$, то есть вероятность равна длине отрезка. Для любого подмножества $X \subseteq [0, 1]$ мы можем задать меру как

$$P(X) = \int_X 1 dx$$

Если $X = [a, b]$, то мы получим введенное выше правило – длину отрезка. Но это правило применимо для кучи других множеств. Можно взять точку, или объединение отрезков, или полуинтервалы. На самом деле можно брать любое подмножество на отрезке, но тогда придется строить правильную версию интеграла. Подобная вероятность «равномерно распределена» по отрезку $[0, 1]$. Еще обратите внимание, что вероятность попасть в точку равна нулю, то есть $P(\{a\}) = 0$ для любой $a \in [0, 1]$. Тем не менее, сама вероятностная мера P при этом не нулевая. Этот пример поясняет, почему не достаточно меру задавать только на элементарных исходах.

Условная вероятность Пусть у нас есть вероятностное пространство (Ω, P) , описывающее какой-либо случайный объект и пусть у нас есть фиксированное событие $B \subseteq \Omega$. Теперь предположим, что нас интересуют не все события случайного объекта, а только те, что случаются внутри события B . Про это можно думать так, мы игнорируем любой ответ нашего черного ящика до тех пор, пока не попался ответ из события B . То есть у такого нового черного ящика пространство всех состояний – это B . Вопрос, а как же там устроена вероятность? Эта вероятность называется условной вероятностью и считается так: пусть $A \subseteq \Omega$ – любое событие на Ω , тогда

$$P(A | B) = P_B(A) = \frac{P(A \cap B)}{P(B)}$$

Последнее выражение называется условной вероятностью A при условии B . Про это надо думать так. Так как мы теперь интересуемся только теми исходами, что лежат внутри B , то когда мы проверяем событие A у нас обязательно случается событие $A \cap B$ (то есть одновременно A и B), но $P(A \cap B)$ – это вероятность этого события в исходном эксперименте. Если мы забываем все события вне B , то вероятность B должна стать единицей. Потому мы дополнительно нормируем на вероятность B . Как видно из формулы, подобную конструкцию можно проверить только если $P(B) \neq 0$.

Еще про это можно думать так: $P(\Omega) = 1$ – это как бы доля всех возможных элементарных исходов в Ω и разумно, что все исходы дают долю 1. Тогда $P(B)$ – это доля тех событий, что лежат в B , а $P(A \cap B)$ – это доля событий одновременно в A и B . Тогда $P_B(A)$ показывает долю события $A \cap B$ в событии B .

Еще обратите внимание, что P_B – это вероятностная мера на исходном Ω , а не только на B . Просто если событие A не может наступить одновременно с B (это значит $A \cap B = \emptyset$ или что то же самое $A \subseteq \bar{B}$), то $P_B(A) = 0$. Вообще, $P_B(A) = 0$ тогда и только тогда, когда $P(A \cap B) = 0$, то есть если вероятность наступления обоих событий одновременно равна нулю.

Еще эту формулу полезно переписать так

$$P(A \cap B) = P(B)P(A | B)$$

То есть, чтобы посчитать вероятность одновременного наступления событий A и B надо умножить вероятность наступления события B на условную вероятность наступления события A при условии наступления B . Причем в этом виде нам не надо требовать, чтобы $P(B) \neq 0$. Она верна даже в случае $P(B) = 0$, в этом случае обе части равенства нулевые. В таком виде эта формула полезна в задачах.

Независимые события Пусть (Ω, P) – некоторое вероятностное пространство, описывающее какой-то случайный объект. Пусть $A, B \subseteq \Omega$ – два события. Они называются независимыми, если $P(A \cap B) = P(A)P(B)$. Давайте разберемся с этим формальным определением. Во-первых, если $P(A) = 0$, то A независимо с любым событием B и наоборот, если $P(B) = 0$, то оно независимо с любым событием A . Это важный частный случай,

который надо иметь в виду. Помимо это остается случай, когда обе вероятности не ноль. В этом случае мы можем переписать условие независимости в виде

$$P(A) = \frac{P(A \cap B)}{P(B)} = P(A | B)$$

Это равенство можно прочитать в несимметричной форме. Событие A не зависит от события B , если его условная вероятность $P(A | B)$ не зависит от B и равна обычной вероятности $P(A)$. Но заметим, что фраза « A не зависит от B » – это лишь способ прочитать формулу, реально это всегда означает симметричное условие, что оба события независимы, то есть равенство $P(A \cap B) = P(A)P(B)$.

Давайте приведем пример зависимых и независимых событий. Возьмем правильный игральный кубик, то есть $\Omega = \{1, 2, 3, 4, 5, 6\}$ и $P(i) = 1/6$. В этом случае рассмотрим события $A = \{1, 2\}$ («выпадет один или два») и $B = \{1, 3\}$ («выпадет один или три»). Тогда эти события зависимы

$$\begin{aligned} P(A \cap B) &= P(\{1, 2\} \cap \{1, 3\}) = P(\{1\}) = \frac{1}{6} \\ P(A)P(B) &= P(\{1, 2\})P(\{1, 3\}) = \left(\frac{1}{6} + \frac{1}{6}\right) \left(\frac{1}{6} + \frac{1}{6}\right) = \frac{1}{9} \end{aligned}$$

Еще полезно понимать, если у вас есть два не пересекающихся события A и B , имеющие ненулевые вероятности ($P(A) \neq 0$ и $P(B) \neq 0$), то они всегда зависимы, действительно $P(A \cap B) = P(\emptyset) = 0$, но $P(A)P(B) \neq 0$.

Теперь приведем пример независимых событий. Пусть $A = \{1, 2, 3\}$ и $B = \{1, 4\}$, тогда

$$\begin{aligned} P(A \cap B) &= P(\{1, 2, 3\} \cap \{1, 4\}) = P(\{1\}) = \frac{1}{6} \\ P(A)P(B) &= P(\{1, 2, 3\})P(\{1, 4\}) = \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6}\right) \left(\frac{1}{6} + \frac{1}{6}\right) = \frac{1}{6} \end{aligned}$$

Вообще говоря, это очень не очевидный вопрос, как, глядя в глаза двум событиям, выпытать у них независимы они или нет. Единственный способ – подставить в формулу и проверить равенство. Ничего лучше не существует. Но зато когда вы знаете, что какие-то события независимы – это обычно огромный повод для радости.

Обычная ситуация, когда получаются независимые события – когда вы проводите два разных случайных эксперимента. А именно, пусть у вас есть два черных ящика X и Y и каждый дает свой ответ. Теперь вы строите новый черный ящик Z как такой, что выдает пару ответов от двух прежних ящиков. Оказывается, что события сформулированные в терминах X и события сформулированные в терминах Y будут независимы для ящика Z . С математической точки зрения последняя процедура означает следующее. Пусть у вас есть вероятностные пространства (Ω_1, P_1) и (Ω_2, P_2) описывающие черные ящики X и Y , соответственно. Тогда если черный ящик выдает два ответа независимо, то его вероятностное пространство будет $\Omega_1 \times \Omega_2$. А какая будет вероятность? Для любых двух событий $A \subseteq \Omega_1$ и $B \subseteq \Omega_2$ мы можем положить вероятность события $A \times B \subseteq \Omega_1 \times \Omega_2$ по правилу

$$P(A \times B) = P(A)P(B)$$

потому что оно по сути означает, что X выдал ответ из A , а Y выдал ответ из B . Эта вероятность должна быть произведением вероятностей для каждого из черных ящиков в силу независимости. Таким образом мы знаем вероятность на всех «прямоугольниках» вида $A \times B$ в $\Omega_1 \times \Omega_2$. А далее из аддитивности вероятности мы можем приближать такими прямоугольниками другие произвольные события на $\Omega_1 \times \Omega_2$ и получим полноценную вероятность.

Полезным примером является следующая задача. Предположим мы равномерно кидаем точку на отрезок $[0, 1]$. Тогда $\Omega = [0, 1]$, а вероятность – длина. Если же мы кидаем равномерно на отрезок две точки независимо друг от друга, это означает, что мы кидаем точку на квадрат $[0, 1] \times [0, 1]$, а координаты этой точки – это положение первой и второй брошенной точки. А в силу независимости вероятность на квадрате становится площадью. Аналогично во всех других задачах, когда независимо кидают точки на различные фигуры в разных количествах, пространство элементарных исходов превращается в произведение пространств, а вероятность в «произведение» вероятностей в смысле выше. То есть если фигуры имели в качестве вероятности площади, длины, объемы и прочее, то итоговая вероятность будет многомерным объемом.

Независимость в совокупности Пусть (Ω, P) – наш вероятностный черный ящик и в нем заданы события $A_1, \dots, A_n \subseteq \Omega$. Тогда они называются независимыми в совокупности, если

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$$

для любого набора подмножеств A_{i_1}, \dots, A_{i_k} , где $1 \leq i_1 < \dots < i_k \leq n$ и $1 \leq k \leq n$. Я хочу обратить внимание, что если события независимы в совокупности, то любое подмножество событий A_{i_1}, \dots, A_{i_k} тоже будет независимым в совокупности. В частности все события будут попарно независимыми. Однако, из попарной независимости не следует независимость в совокупности. Давайте рассмотрим следующий пример.

Рассмотрим независимое бросание двух правильных монеток. Тогда $\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, где 0 означает выпадение орла, а 1 – выпадение решки. Вероятность каждого из этих событий полагается $1/4$. Тогда рассмотрим три события:

- $A_1 = \{(1, 0), (1, 1)\}$, что значит «у первой монеты выпала решка».
- $A_2 = \{(0, 1), (1, 1)\}$, что значит «у второй монеты выпала решка».
- $A_0 = \{(0, 1), (1, 0)\}$, что значит «монеты выпали по-разному».

Заметим, что все эти события одновременно не совместны, то есть $A_0 \cap A_1 \cap A_2 = \emptyset$. Потому равенство $P(A_0 \cap A_1 \cap A_2) = P(A_0)P(A_1)P(A_2)$ не выполняется. Но, прямые вычисления, показывают, что все эти события попарно независимы. Проверим это в случае A_1 и A_2 . Видим, что $P(A_1) = 1/2$, $P(A_2) = 1/2$ и $P(A_1 \cap A_2) = P(\{(1, 1)\}) = 1/4 = P(A_1)P(A_2)$. Аналогично делаются остальные два случая.

Думать про этот пример надо так. Для каждой пары события наступают независимо. То есть вероятность того случится A_1 или нет никак не зависит от того произошло A_0 или нет. Однако, если мы хотим сформулировать более сложное событие, например, $A_0 \cap A_1$, что означает «монеты показали разный результат и на первой монете выпала решка». В этом случае мы видим, что результат выпадения второй монеты предопределен, он обязан быть орлом, тут нет никакой независимости.

Еще раз про монетки Пусть мы n раз бросаем монетку. Тогда элементарными исходами у нас являлась последовательность из нулей и единиц длины n . Мы полагали, что $P(a_1, \dots, a_n) = 1/2^n$. Пользуясь понятием независимости, можно дать естественное объяснение тому, что это означает для нашего эксперимента. Действительно, событие (a_1, \dots, a_n) означает: «первая монета показала a_1 , вторая монета показала a_2, \dots, n -я монета показала a_n ». То есть это пересечение событий вида « i -я монета показала a_i ». Тогда видно, что мы положили нашу вероятность по правилу

$$P(a_1, \dots, a_n) = P(\text{«1-я монета показала } a_1\text{»}) \dots P(\text{«}n\text{-я монета показала } a_n\text{»}) = \frac{1}{2} \dots \frac{1}{2} = \frac{1}{2^n}$$

Таким образом предположение о равномерности распределения вероятности по всем бросаниям монеток означает, что выпадения орлов и решек у разных монет происходит независимо друг от друга. В этом случае говорят, что мы бросаем n монеток независимо друг от друга.

Формулы полной вероятности Пусть у вас есть вероятностное пространство (Ω, P) , описывающее какой-то случайный объект. Пусть у вас задан набор событий $A_1, \dots, A_n \subseteq \Omega$ такой, что $A_i \cap A_j = \emptyset$ для любых $i \neq j$ и $\Omega = A_1 \cup \dots \cup A_n$. То есть события не пересекаются и покрывают все пространство элементарных исходов. Тогда такой набор событий называется разбиением Ω . Неформально про разбиение надо думать так: вы перечисляете взаимоисключающие состояния в которых может быть система. Например, в случае кубика вы можете выбрать разбиение из двух событий: «выпало четное» и «выпало нечетное». Вы знаете, что всегда произойдет одно из двух и ничего другого быть не может. Еще можно разбить на события $A_1 = \{1, 2\}$, $A_2 = \{3, 4\}$ и $A_3 = \{5, 6\}$. Эти события сложно описать словесно, точнее можно, но будет звучать глупо. Но тем не менее, это тоже вполне себе разбиение.

Пусть теперь у нас есть событие A , тогда выполняется следующая формула

$$P(A) = P(A | A_1)P(A_1) + \dots + P(A | A_n)P(A_n) = \sum_{i=1}^n P(A | A_i)P(A_i)$$

Эта формула называется формулой полной вероятности. Она бывает полезна, когда вы знаете условные вероятности $P(A | A_i)$ и знаете вероятности $P(A_i)$, с которыми ваше пространство бьется на эти события.

Формулы Байеса Предыдущая формула обычно идет в связке с еще одной популярной формулой. Она связана с тем, как пересчитать $P(A | B)$ через $P(B | A)$ и наоборот. Эти формулы называются формулами Байеса. Пусть у нас есть два события A и B в Ω , тогда

$$P(B | A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} = \frac{P(A | B)P(B)}{P(A)}$$

Итоговое равенство

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)}$$

и называется формулой Байеса.

Вероятность совместного наступления событий Пусть в вероятностном пространстве (Ω, P) заданы события $A, B, C, D \subseteq \Omega$, тогда вероятность одновременного наступления событий можно считать следующим образом

$$\begin{aligned} P(CD) &= P(C | D)P(D) \\ P(BCD) &= P(B | CD)P(C | D)P(D) \\ P(ABCD) &= P(A | BCD)P(B | CD)P(C | D)P(D) \end{aligned}$$

Семинар 1

Задачи:

1. Василий решил покрасоваться перед Василисой и утверждает, что бросив два кубика, у него обязательно выпадет шестерка. Василиса со своей стороны решила поддаться Василию и будет игнорировать все его броски, если на кубиках выпали одинаковые числа. Найдите вероятность того, что Василий произведет впечатление на Василису в этих условиях.
2. В некотором прекрасном городе «Икс» население выросло аж до 10 человек. По этому прекрасному случаю в нем открылся парк аттракционов, где можно попрыгать на батуте. Оказалось, что батут обязательно рвется, если на нем находится 5 человек, в случае 4 человек он рвется с вероятностью 0,8, в случае 3 человек с вероятностью 0,6, в случае 2 человек с вероятностью 0,4, в случае 1 человека с вероятностью 0,2 и производитель гарантирует, что их надежные и качественные батуты сами по себе не рвутся. Несмотря на праздничное событие, жители города «Икс» очень заняты, каждый из них может прийти в парк с вероятностью 0,5. Узнайте, какова вероятность того, что уже в первый день бизнес с батутом накроется.
3. В мешке лежит три шара: два черных и один белый. Мы производим следующие действия: вытаскиваем шар, записываем его цвет и обратно кладем шар другого цвета. Такие действия продлеваются три раза. Рассматривая такой эксперимент как случайный, опишите вероятностное пространство (Ω, P) , то есть пространство элементарных исходов Ω и вероятностную меру $P: 2^\Omega \rightarrow [0, 1]$ на нем. Какова вероятность получить белый шар на третьем шаге?
4. Дано множество $X = \{1, 2, 3, 4\}$. Будем равновероятно выбирать граф G на вершинах из X . Найдите следующие вероятности
 - (a) Вероятность, что граф связан.
 - (b) Вероятность, что граф является деревом.
 - (c) Вероятность, что граф дерево, при условии, что он связан.
 - (d) Вероятность, что граф связан, при условии, что он не дерево.
5. Везунчик Ермолай пришел в казино с двумя монетами. Делая ставку он выигрывает с вероятностью p и проигрывает с вероятностью q . За победу или проигрыш он получает или соответственно теряет одну монету. Ермолай решил не мелочиться и играть до полной потери денег. Найдите вероятность, что он сделает хотя бы 5 ставок.
6. В квадрат со стороной 2 равномерно бросают диск радиуса $r < 1/4$ так, что он целиком попадает в прямоугольник. Опишите вероятностное пространство и найдите вероятность того, что брошенный диск пересечет хотя бы одну диагональ квадрата.

Теория вероятности

Дима Трушин

Семинар 2

Случайные величины

Понятие случайной величины Предположим вы изучаете какой-то случайный объект. Чтобы описать такой объект, вы должны описать множество всех его состояний Ω и вероятностную меру на этом множестве P . Однако, может так случиться, что множество Ω непонятно как устроенным, ибо вы же не знаете, что внутри вашего черного ящика. Например, если работаете с элементарными частицами. Да фиг его знает, как там что у них устроено и в каких состояниях может быть частица! Но даже если вы его знаете, может случиться другая беда, оно на столько огромное, что непонятно как с ним работать. Например, если вы хотите взять случайного человека, то все возможные исходы – это все возможные люди на земле. Кто хочет работать с множеством Ω порядка 7,5 миллиарда элементов?

Однако, часто для решения конкретной задачи нам и не надо знать все Ω целиком. Вместо этого надо лишь знать что-то, что важно для задачи. Тут на помощь приходят случайные величины.

Представьте, что у вас есть некоторый черный ящик. Вместо того, чтобы задавать ему вопрос: «В каком ты состоянии?» мы можем задавать более конкретные вопросы: «какая у тебя температура?», «какой у тебя цвет?», «какое у тебя давление?» и т.д. Про это можно думать так, что к нашему черному ящику мы прикрепляем измерительный прибор, который на выходе выдает некоторое число. И когда мы задаем вопрос черному ящику, мы просто смотрим на показания нашего прибора. В результате, мы получаем информацию не в виде полного описания в каком состоянии находится наш черный ящик, а всего лишь число, показывающее значение конкретной величины.

Если наш случайный объект описывался парой (Ω, P) , то случайная величина – это правило, которое каждому элементарному исходу $\omega \in \Omega$ ставит в соответствие некоторое вещественное число. То есть, случайная величина – это отображение $\xi: \Omega \rightarrow \mathbb{R}$.

Примеры

1. Злосчастный кубик описывается парой (Ω, P) , где $\Omega = \{1, 2, 3, 4, 5, 6\}$ и $P(k) = 1/6$. Давайте рассмотрим случайную величину $\xi: \Omega \rightarrow \mathbb{R}$ по правилу $\xi(k) = k \pmod{2}$, то есть остаток от деления на 2. Тогда ξ четные переводит в ноль, а нечетные переводит в 1. Таким образом ξ моделирует нам бросание монетки $\Omega = \{0, 1\}$. С другой стороны, 0 может выпасть на событиях $\{2, 4, 6\}$, то есть вероятность выпадения 0 для ξ равна $1/2$. То есть мы получили правильную монетку с равновероятным выпадением орла и решки.
2. Пусть теперь у нас рассматривается независимое бросание n правильных монеток. Тогда $\Omega = \{(a_1, \dots, a_n) \mid a_i \in \{0, 1\}\}$ состоит из всех последовательностей из нулей и единиц длины n . Давайте рассмотрим случайную величину $\xi: \Omega \rightarrow \mathbb{R}$ равную количеству выпавших решек, то есть $\xi(a_1, \dots, a_n) = a_1 + \dots + a_n$. Тогда ξ принимает значения от 0 до n и вероятность того, что $\xi = k$ равна $C_n^k / 2^n$. Это частный случай биномиального распределения. В данном случае на ξ можно смотреть как на черный ящик с пространством $\Omega = \{0, \dots, n\}$, при этом $P(k) = C_n^k / 2^n$.

Теперь когда мы измеряем только показания прибора (нашу случайную величину), мы можем смотреть на черный ящик с прибором, как на новый случайный объект, состояния которого – это все действительные числа \mathbb{R} . Возникает вопрос, а как же на этом черном ящике устроена вероятность? По определению вероятность нового ящика устроена так:

$$P_\xi(A) = P(\xi \in A) = P(\{\omega \mid \xi(\omega) \in A\}) = P(\xi^{-1}(A)), \text{ для любого } A \subseteq \mathbb{R}$$

То есть для каждого подмножества $A \subseteq \mathbb{R}$ мы смотрим на событие, что наша случайная величина выдала ответ из множества A . Чтобы посчитать это значение, мы смотрим все элементарные исходы ω , при которых

$\xi(\omega) \in A$, сваливаем их в одно событие и измеряем его вероятность. Это и будет то, что мы только что определили.¹

Теперь, когда нас интересуют лишь показания прибора, вместо изучения (Ω, P) мы будем изучать новую пару (\mathbb{R}, P_ξ) . И нам совершенно не важно, каким было оригинальное пространство Ω и мера P . Если ответ на нашу задачу зависит только от значений ξ , то все, что нам надо знать, – это мера P_ξ на прямой. Теперь возникает резонный вопрос: а как вообще задавать вероятностные меры на прямой и какими они бывают? Я предлагаю разобраться с этим вопросом далее.

Вероятностные меры на прямой Что значит задать вероятность на прямой? Это значит, мы должны для каждого подмножества $A \subseteq \mathbb{R}$ задать число $P(A)$ так, чтобы выполнялись аксиомы для вероятности:

1. Для любого $A \in \mathbb{R}$ верно, что $0 \leq P(A) \leq 1$.
2. $P(\emptyset) = 0$, $P(\mathbb{R}) = 1$.
3. Для любой последовательности попарно непересекающихся подмножеств $A_1, A_2, \dots, A_n, \dots \subseteq \mathbb{R}$ (то есть $A_i \cap A_j = \emptyset$) выполнено $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Мягко говоря непонятно, как вообще перебрать все подмножества и приписать им нужное число, так еще и отвратительно сложное условие под третьим пунктом намекает, что мы будем мучиться. Оказывается, что достаточно научиться задавать вероятность того, что случайная величина попала в полуинтервал $(a, b]$, то есть вероятность события $P(a < \xi \leq b)$.^{2, 3} А далее есть общая процедура, которая говорит, как распространить нашу вероятность на любое подмножество прямой.

Мы не будем себя мучить этой абстрактной общей процедурой, а поясним на паре примеров, как отсюда вытащить меры отрезка и интервала. Пусть мы хотим посчитать $P([a, b])$. Тогда надо выбрать последовательность чисел $a_n < a$ таких, что $a_n \rightarrow a$ при $n \rightarrow \infty$. Тогда $P([a, b]) = \lim_{n \rightarrow \infty} P((a_n, b])$, то есть мы пересечем полуинтервалы с левыми концами чуть левее точки a . Аналогично, если мы хотим посчитать $P((a, b))$, выберем последовательность чисел $b_n < b$ такую, что $b_n \rightarrow b$ при $n \rightarrow \infty$. Тогда $P((a, b)) = \lim_{n \rightarrow \infty} P((a, b_n])$, то есть мы объединим полуинтервалы с правыми концами чуть левее b .

t.me/postypashki_old/1229

t.me/postypashki_old/1229

t.me/postypashki_old/1229

Функция распределения Чтобы научиться задавать вероятности $P(\xi \in (a, b])$ достаточно научиться измерять следующую величину

$$F_\xi(x) = P(\xi \leq x) = P(\xi \in (-\infty, x])$$

Тогда функция $F_\xi(x)$ называется функцией распределения случайной величины ξ . Действительно, в этом случае

$$P(\xi \in (a, b]) = P(\xi \in (-\infty, b]) - P(\xi \in (-\infty, a]) = F_\xi(b) - F_\xi(a)$$

То есть, чтобы задать вероятностную меру на прямой, нам достаточно задать одну единственную функцию $F_\xi(x)$ и она однозначно определит некоторую вероятность. Причем формула выше дает явный вид для вероятности попадания в полуинтервал, а в предыдущем разделе я описал как считается попадание в открытый и замкнутый интервалы.

Теперь весь вопрос в том, а любую ли функцию можно взять в качестве $F_\xi(x)$ и если не любую, то какие на нее должны быть условия? Конечно же совсем любая функция не подойдет иначе не будут выполняться аксиомы на вероятность. Но следующий список аксиом является полным описанием функций распределения.

Утверждение. Функция $F: \mathbb{R} \rightarrow \mathbb{R}$ является функцией распределения, то есть задает некоторую вероятностную меру, тогда и только тогда, когда она удовлетворяет следующим свойствам:

1. F является неубывающей функцией, то есть, если $x \leq y$, то $F(x) \leq F(y)$.
2. $\lim_{x \rightarrow -\infty} F(x) = 0$ и $\lim_{x \rightarrow \infty} F(x) = 1$.
3. Функция F непрерывна справа, то есть для любой точки $x \in \mathbb{R}$ верно

$$F(x+) = \lim_{t \rightarrow 0, t > 0} F(x+t) = F(x)$$

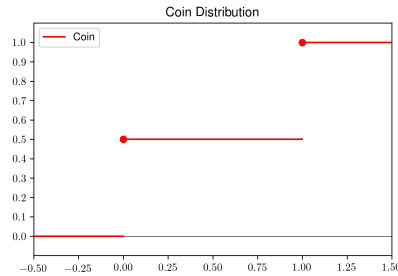
¹Ровно по этому же принципу были определены вероятностные меры в примерах выше.

²Если у вас дежавю с итераторами в стандартных библиотеках языков программирования, то это не случайно. И там и тут ренджи измеряются полуинтервалами.

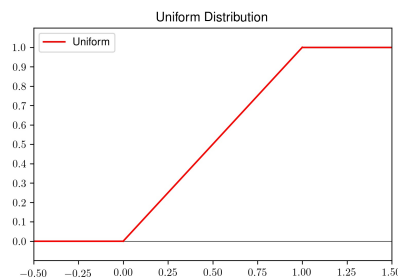
³Необходимость работы с полуинтервалами связана вот с чем. Может так оказаться, что вероятность попасть в границу интервала является ненулевым числом. Чтобы не было путаниц с правилами и вычислениями, оказывается, что полуинтервалы ведут себя сильно лучше, чем интервалы (открытые или замкнутые).

Примеры функций распределения

1. Бросание монетки.



2. Равномерное распределение на отрезке $[0, 1]$.



Какие бывают вероятностные меры на прямой — Функция распределения — это замечательно. С помощью нее можно задать абсолютно любую вероятность на прямой. Но оказывается, что если умерить аппетиты и стараться описать не все возможные вероятностные меры, можно дать еще более клевое описание. Так получилось, что в математике есть три класса вероятностных мер:

1. Дискретные.
2. Непрерывные.
3. Экзотика.

Говоря по честному, второй класс называется «Абсолютно непрерывные», а «Непрерывными» называется второй и третий класс вместе. Именно такая терминология принята в математике. Однако, так сложилось, что экзотические распределения в жизни не встречаются, это лишь плот большой фантазии математиков (хотя тут я могу и парочку примеров привести). Поэтому в инженерной и прикладной литературе третий класс вовсе не упоминают, а второй для простоты называют непрерывными. Я решил, что прикладная терминология нас вполне устроит, тем более что придется говорить меньше запутанных слов.

Дискретные случайные величины Пусть я хочу задать дискретную случайную величину ξ на прямой \mathbb{R} . Я в начале должен зафиксировать некоторые данные:

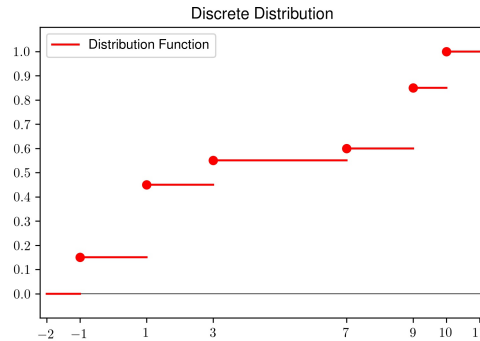
1. Последовательность чисел $a_n \in \mathbb{R}$. Это может быть конечная или бесконечная последовательность любых чисел на прямой. Это будут выделенные значения, которые разрешено принимать нашей случайной величине. Эти значения называются еще атомами.
2. Последовательность чисел $p_n \in \mathbb{R}$. Этих чисел должно быть столько же, сколько и a_n . Это будут вероятности, с которыми наша случайная величина будет принимать выделенные значения. Потому они должны удовлетворять следующим свойствам:

- (a) $0 < p_i \leq 1$.
- (b) $\sum_{i=1}^{\infty} p_i = 1$.

Тогда для произвольного $A \subseteq \mathbb{R}$ вероятность задается так

$$P_{\xi}(A) = \sum_{a_i \in A} p_i$$

то есть мы смотрим какие атомы попали в множество A и складываем соответствующие этим атомам вероятности p_i . В частности $P(\xi = a_i) = p_i$. И если число a не равно никакому из a_i , то вероятность $P(\xi = a) = 0$. Функция распределения $F_{\xi}(x)$ в этом случае будет ступенчатой



В качестве примера можно рассмотреть бросание кубика. Случайная величина ξ – число на грани. Атомами будут числа от 1 до 6, а приписанные им вероятности все равны $1/6$ если кубик сбалансированный. Еще один пример – бросание монетки. Если орел, то прибор показывает 0, а если решка, то – 1. Если монетка сбалансированная, то приписанные вероятности будут $1/2$ в обоих случаях.

Примеры

1. Распределение Бернулли. Это распределение для произвольной (необязательно правильной) монетки. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 \\ p & q \end{cases}$$

при этом $p + q = 1$ и $p, q \geq 0$.

2. Биномиальное распределение $B(n, p)$. Это распределение для числа выпавших решек в n независимых бросаниях (необязательно правильной) монетки. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 & \dots & k & \dots & n \\ q^n & C_n^1 p q^{n-1} & \dots & C_n^k p^k q^{n-k} & \dots & p^n \end{cases}$$

при этом $p + q = 1$ и $p, q \geq 0$.

3. Распределение Пуассона $P(\lambda)$, при этом $\lambda > 0$. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 & \dots & k & \dots \\ e^{-\lambda} & e^{-\lambda} \lambda & \dots & \frac{e^{-\lambda} \lambda^k}{k!} & \dots \end{cases}$$

4. Геометрическое распределение. Это распределение отвечает за вероятность появления первого успеха при бросании неправильной монетки, а именно, эта величина равна номеру первого успеха.

$$\xi \sim \begin{cases} 1 & 2 & \dots & k & \dots \\ p & qp & \dots & q^{k-1}p & \dots \end{cases}$$

Второй вариант геометрического распределения: количество неудач до первого успеха, то есть $\xi - 1$.

Непрерывные случайные величины Пусть я хочу задать непрерывную случайную величину ξ на прямой \mathbb{R} . Я в начале должен зафиксировать некоторые данные:

1. Функция $p: \mathbb{R} \rightarrow \mathbb{R}$ по правилу $x \mapsto p(x)$. Эта функция будет называться плотностью случайной величины ξ и должна удовлетворять следующим свойствам:

- (a) $0 \leq p(x)$ для всех $x \in \mathbb{R}$.
- (b) $\int_{-\infty}^{\infty} p(x) dx = 1$.

Тогда для произвольного подмножества $A \subseteq \mathbb{R}$ вероятность задается так

$$P_{\xi}(A) = \int_A p(x) dx$$

Если, например, $A = [a, b]$, полуинтервал $A = (a, b]$ или интервал $A = (a, b)$, то

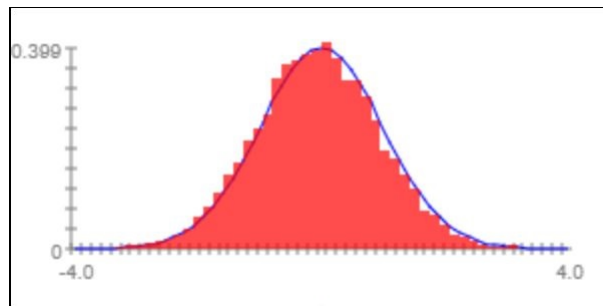
$$P_{\xi}([a, b]) = P_{\xi}((a, b]) = P_{\xi}((a, b)) = \int_a^b p(x) dx$$

То есть интеграл от a до b задает вероятность, что случайная величина выдала ответ в интервале от a до b . Еще обратите внимание, что для любого числа $a \in \mathbb{R}$ выполнено

$$P(\xi = a) = \int_a^a p(x) dx = 0$$

То есть вероятность попасть в каждую конкретную точку равна нулю, но при этом сама функция вероятности не нулевая. Еще стоит обратить внимание на связь между плотностью и функцией распределения. Если вы знаете функцию распределения $F_{\xi}(x)$, то плотность будет вычисляться как производная $p(x) = F_{\xi}(x)'$.

Частотное понимание плотности Здесь я опишу, как можно думать про плотность случайной величины. Пусть задана непрерывная случайная величина ξ с плотностью $p: \mathbb{R} \rightarrow \mathbb{R}_+$, то есть $\int_{\mathbb{R}} p(x) dx = 1$. Давайте разобьем всю прямую на интервалы равной длины, скажем длины δ . Будем думать про эти интервалы как про корзины. Теперь поставим n экспериментов с нашей случайной величиной ξ , то есть спросим ее состояние n раз. Она нам выдаст n чисел, скажем, a_1, \dots, a_n . Тогда разложим полученные ответы по корзинам. Теперь частота попадания в корзину получается равна количеству ответов в корзине деленному на n – общее количество ответов. Получится ступенчатая функция, которая при малом δ и большом n будет очень похожа на плотность. Ниже на картинке синим изображен график плотности, а красным ступеньки из частот для корзин. Сами корзины отмечены штрихами вдоль горизонтальной оси.



Функция $p(x)$ не просто так называется плотностью. Если думать про меру P не как про вероятность, а как про массу, то величина $P(A)$ показывает какая масса у куса прямой A . Тогда $p(x)$ имеет физический смысл той самой плотности к которой мы привыкли в физике. То есть она показывает как много вещества находится в данной точке пространства, а масса будет интегралом от плотности. Потому плотность вероятности показывает как бы сколько «правдоподобности» находится в данной точке, а вероятность будет интегралом от этой «правдоподобности».

Примеры

1. Равномерное распределение на отрезке $[a, b]$. Плотность имеет вид

$$p(x) = \frac{1}{b-a} \chi_{[a,b]}(x), \quad \text{где} \quad \chi_{[a,b]}(x) = \begin{cases} 1, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

2. Экспоненциальное распределение. Его плотность задается в виде

$$p(x) = \lambda e^{-\lambda x} \theta(x), \quad \text{где} \quad \theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

При этом $\lambda > 0$.

3. Нормальное распределение $N(a, \sigma)$. Его плотность задается в виде

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Здесь $a \in \mathbb{R}$ любое, а $\sigma > 0$.⁴

Характеристики случайных величин

Пусть у нас ξ – некоторая случайная величина на некотором вероятностном пространстве (Ω, P) . Формально, ξ – это некоторая функция на Ω . А что значит, что эта величина на самом деле не случайная? Это значит, что ξ принимает одно и то же значение для всех элементарных исходов, то есть ξ является константой. Таким образом мы можем рассматривать обычные числа как случайные величины, которые правда оказываются уже не такими уж и случайными. Так как с числами работать проще, чем с непонятными случайными величинами. Действительно, чтобы задать число – надо задать число⁵, а чтобы задать случайную величину нам надо задавать какую-то там функцию распределения с кучей всяких свойств, неприятно. Потому очень полезно иметь на виду некоторые числовые характеристики для случайных величин.

Математическое ожидание Философия этой характеристики следующая. Да, мы знаем, что наша случайная величина не является константой и выдает разный ответ. Но уж очень нам хочется считать ее константой. Тогда возникает вопрос, а какое число является наилучшей заменой для нашей случайной величины? То есть: какое число в среднем будет отвечать так же, как и наша случайная величина? На самом деле в математическом анализе давно известен механизм замены функции числом – замена функции ее интегралом. Раз $\xi: \Omega \rightarrow \mathbb{R}$ – это функция на Ω , то мы хотим определить математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \int_{\Omega} \xi dP = \int_{\mathbb{R}} x dP_{\xi}$$

Давайте я попытаюсь придать смысл этим формулам, не вдаваясь в детали. Равенство номер один гласит

$$\mathbb{E}(\xi) = \int_{\Omega} \xi dP$$

Надо представить себе график функции $y = \xi(\omega)$. По определению интеграл – это объем под графиком, когда за площадь на Ω отвечает P . То есть, мы хотим усреднить все значения ξ с учетом того, на сколько часто каждое значение встречается. За частоту появления значения $\xi(\omega)$ отвечает вероятность P .

Давайте рассмотрим следующий простой пример. Пусть

$$\xi = \begin{cases} a, & \omega \in A \\ b, & \omega \in \bar{A} \end{cases}$$

⁴Ниже я расскажу про него чуть более подробно. Это один из самых важных классов распределения в основном благодаря Центральной предельной теореме, которая говорит, что это распределение в некотором смысле универсально.

⁵Спасибо, Кэп.

То есть случайная величина ξ принимает два значения a и b . Значение a выпадает, когда случается событие A , а значение b , когда случается \bar{A} , то есть, когда не случается A . Тогда вероятность получить a есть $P(A)$, а вероятность ответа b есть $P(\bar{A})$. Значит в среднем мы ожидаем увидеть ответ $aP(A) + bP(\bar{A})$. Хорошая интуиция получается, если думать частотно. Если мы провели n экспериментов и k раз был ответ a , то $n - k$ раз ответ b . Если мы теперь усредним все ответы, то получим

$$\frac{ak + b(n - k)}{n} = a \frac{k}{n} + b \frac{n - k}{n} = aP(A) + bP(\bar{A})$$

последнее равенство возникает из частотной интерпретации вероятностей $P(A) = k/n$ – как часто мы получим ответ из A . Теперь посмотрим на график функции ξ – это ступенчатая функция, она равна a над A и равна b над \bar{A} . По определению интеграл

$$\int_{\Omega} \xi dP$$

означает – объем под графиком функции, если за площадь на Ω отвечает мера P . Тогда объем под такой функцией будет равен сумме объемов двух цилиндров: цилиндр над A высотой a и цилиндр над \bar{A} высотой b (с учетом знаков a и b объем будет положительным или отрицательным). Но объем первого цилиндра – это его высота на площадь основания, то есть $aP(A)$, а объем второго – $bP(\bar{A})$. Таким образом взятие интеграла – это в точности усреднение всех значений случайной величины с учетом частоты их появления. Эту интуицию можно распространить на случай произвольной ξ . И формальная конструкция интеграла

$$\int_{\Omega} \xi dP$$

устроена так, чтобы эта интуиция продолжала работать.

Примеры Формула выше бывает очень полезна, когда вероятностное пространство (Ω, P) устроено «просто», то есть мы можем придать смысла этому интегралу. Давайте посмотрим на два важных случая.

1. Пусть $\Omega = [0, 1]^2$ – квадрат на плоскости. И пусть вероятность – это площадь. Тогда случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ – это произвольная функция на квадрате. Например, мы можем сказать, что $\xi(x, y) = xy + \cos(x) - \sin(y)$. И тогда математическое ожидание по определению будет равно

$$\int_{\Omega} \xi dP = \int_{[0,1]^2} (xy + \cos(x) - \sin(y)) dx dy$$

2. Другой хороший пример можно получить в случае дискретного пространства. Пусть $\Omega = \{(a_1, a_2) \mid a_i \in \{0, 1\}\}$ – пространство описывающее бросок двух монет. И мы считаем, что все исходы равновероятны, то есть мы делаем «независимые» бросания монет. Рассмотрим функцию $\xi(a_1, a_2) = a_1 + a_2$, то есть ξ – это количество решек за два броска. Тогда график функции можно изобразить таблицей

ξ	0	1	1	2
ω	(0, 0)	(0, 1)	(1, 0)	(1, 1)

Так как ξ постоянна на каждом одноточечном множестве вида $\{(a_1, a_2)\}$, то «площадь» под графиком можно считать как сумму «площадей» четырех фигур. В начале берем ступеньку над точкой $(0, 0)$. Ее высота будет 0, значит площадь будет $0 \cdot P(0, 0) = 0$. Вторая ступенька будет над точкой $(0, 1)$ и ее высота 1, значит ее площадь будет $1 \cdot P(0, 1) = 1/4$. Третья ступенька будет над точкой $(1, 0)$ и ее высота опять 1, значит и ее площадь будет $1 \cdot P(1, 0) = 1/4$. Четвертая ступенька будет над точкой $(1, 1)$ и ее высота 2. Значит ее площадь будет $2 \cdot P(1, 1) = 1/2$. В итоге математическое ожидание будет равно

$$\int_{\Omega} \xi dP = \xi(0, 0)P(0, 0) + \xi(0, 1)P(0, 1) + \xi(1, 0)P(1, 0) + \xi(1, 1)P(1, 1) = 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4}$$

По аналогии с этим случаем можно подходить к ситуации, когда вероятность распределена на прямой, лежащей на плоскости. Тогда нужно функцию с плоскости ограничить на прямую и проинтегрировать по прямой. Или другие подобные ситуации. Надеюсь, что эти два примера сделают для вас общую непонятную формулу более полезной.

Замена переменной и другая формула матожидания Теперь давайте поговорим про равенство

$$\int_{\Omega} \xi dP = \int_{\mathbb{R}} x dP_{\xi}$$

На самом деле – это замена переменной в интеграле. Смотрите что происходит. У нас есть отображение «замены переменной» $\xi: \Omega \rightarrow \mathbb{R}$. Кроме того, есть функция $f: \mathbb{R} \rightarrow \mathbb{R}$ по правилу $f(x) = x$ на прямой \mathbb{R} . Интеграл зависит от двух вещей: значений функции и способа подсчета «площади» на горизонтальной оси. То есть нам надо сделать замену для функции и для меры, при переходе от одного интеграла к другому. При такой замене переменных функция f превращается в $f \circ \xi: \Omega \rightarrow \mathbb{R}$ по правилу $\omega \mapsto \xi(\omega)$, то есть это и будет наша случайная величина ξ . По определению $P_{\xi}(A) = P(\xi^{-1}(A))$ для любого $A \subseteq \mathbb{R}$. То есть, если у нас есть кусочек прямой $A \subseteq \mathbb{R}$, то ему на Ω соответствует кусочек $\xi^{-1}(A)$ и так как они переходят друг в друга, то они должны иметь одинаковую «площадь» (то есть одинаковую вероятность). Значит при такой «замене переменной» правило подсчета «площади» P_{ξ} на прямой соответствует правилу подсчета «площади» P на Ω . В матанализе мы делали так: dx на прямой переводили в $d\xi(\omega)$ на Ω , и она совпадала с $\xi'(\omega)d\omega$. Всех этих символов у нас нет на Ω и это бессмысленная работа с бессмысленными буквами, но суть связи ровно такая же, только мы пишем dP_{ξ} вместо dx и dP вместо $d\xi(\omega) = \xi'(\omega)d\omega$.

Математическое ожидание с помощью функций распределения Пусть случайная величина ξ имеет функцию распределения $F_{\xi}(x) = P(\xi \leq x)$. Тогда

$$\mathbb{E}(\xi) = \int_{-\infty}^{\infty} x dF_{\xi}(x)$$

Чтобы эта формула работала, достаточно потребовать, чтобы F_{ξ} была дифференцируемой с непрерывной производной. Однако, есть разные другие вариации взятия интеграла вроде интеграла Римана-Стилтьеса. В реальной жизни встречаются следующие ситуации: ваша функция кусочно гладкая и между этими кусочками есть точки разрыва. Пусть $F_{\xi}(x) = F(x)$ и пусть a_1, \dots, a_n, \dots – это точки разрыва, а величина разрыва будет p_1, \dots, p_n, \dots для каждой точки (здесь $p_i = F(a_i) - F(a_i-)$). И пусть $\varphi(x)$ – это функция, в которой мы убрали все точки разрыва у функции F . Тогда для интеграла $\int_{\mathbb{R}} x dF(x)$

- Каждая точка разрыва a_i вносит вклад $a_i p_i$.
- Все остальные точки вносят вклад $\int_{\mathbb{R}} x d\varphi(x) = \int_{\mathbb{R}} x \varphi'(x) dx$.

То есть интеграл получается равен

$$\mathbb{E}(\xi) = \int_{\mathbb{R}} x dF(x) = \sum_i a_i p_i + \int_{\mathbb{R}} x \varphi'(x) dx$$

Обратите внимание, что тут мы не учитываем по сути экзотические распределения. Это не с проста, ибо это не просто. Кроме того, в случае дискретного или непрерывного распределения формулу можно сделать еще лучше.

Математическое ожидание в дискретном случае Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases}$$

Тогда определим ее математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \sum_{i=1}^{\infty} a_i p_i = \sum_{i=1}^{\infty} a_i P(\xi = a_i)$$

Если подумать, то написанное здесь говорит следующее. Мы должны каждое значение a_i взять с коэффициентом равным вероятности (частоте появления ответа в черном ящике) для этого значения. А потом все сложить. То есть мы получим какой ответ наш черный ящик дает в среднем.

Математическое ожидание в непрерывном случае Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда определим ее математическое ожидание следующей формулой

$$\mathbb{E}(\xi) = \int_{-\infty}^{\infty} xp(x) dx$$

В данном случае физический смысл такой же как и в дискретном случае. Мы усредняем все возможные значения x , которые принимает наша случайная величина, учитывая каждый x пропорционально плотности $p(x)$. Чем больше плотность в точке x , тем больше вклад внесет x в финальное значение. Опять же, математическое ожидание показывает какой ответ в среднем нам будет давать наш черный ящик.

Сравнение формул для математического ожидания Давайте вспомним непонятную формулу:

$$\mathbb{E}(\xi) = \int_{\mathbb{R}} x dP_{\xi}$$

В случае дискретной случайной величины, надо заменить знак интеграла суммой. Переменная x в этом интеграле пробегает все значения на прямой, то есть ее надо заменить на значения случайной величины, то есть на a_i . А dP_{ξ} должна отвечать за вероятность принять значение a_i . Делая замену

$$\int_{\mathbb{R}} \mapsto \sum_{i=1}^{\infty} \quad x \mapsto a_i \quad dP_{\xi} \mapsto p_i$$

мы получим формулу для дискретного случая. В непрерывном случае, знак интеграла оставляем нетронутым, переменную x оставляем нетронутой, а вот символ dP_{ξ} меняем на $p(x) dx$, то есть замена вида

$$\int_{\mathbb{R}} \mapsto \int_{-\infty}^{\infty} \quad x \mapsto x \quad dP_{\xi} \mapsto p(x) dx$$

При такой подстановке, мы получим формулу для непрерывного случая. На самом деле есть общая математическая теория, которая позволяет все формулы считать одним способом, но нам придется немного помучиться от нашего незнания. Однако, не забывайте, что от знания мучаешься еще больше! Так что это приемлемое страдание.

Свойства математического ожидания

1. Пусть ξ – некоторая случайная величина и $a, b \in \mathbb{R}$, тогда $\mathbb{E}(a\xi + b) = a\mathbb{E}(\xi) + b$.
2. Пусть ξ – некоторая случайная величина, тогда $\mathbb{E}(\xi - \mathbb{E}(\xi)) = 0$.
3. Пусть ξ и η – две случайные величины, тогда $\mathbb{E}(\xi + \eta) = \mathbb{E}(\xi) + \mathbb{E}(\eta)$.

Дисперсия Отлично, мы заменили случайную величину ξ самым ее лучшим числовым приближением $\mathbb{E}\xi$. А на сколько отличается случайная величина от своего приближения? Очень просто, разница $\xi - \mathbb{E}(\xi)$ – это новая случайная величина, которая характеризует отличие. Что, опять случайная величина? А можно число? Просто взять матожидание от $\xi - \mathbb{E}(\xi)$ – не самая лучшая идея, ибо получится ноль, что означает, что в среднем ξ и $\mathbb{E}(\xi)$ не отличаются. А вот чтобы ответить, на сколько они в среднем отличаются надо вместо разности померить расстояние между ξ и $\mathbb{E}(\xi)$. Есть разные способы это сделать, но самый популярный – квадрат разности. А потом взять математическое ожидание.

Если ξ – некоторая случайная величина, то ее дисперсией называется следующее выражение

$$\mathbb{D}(\xi) = \mathbb{E}((\xi - \mathbb{E}(\xi))^2) = \mathbb{E}(\xi^2) - (\mathbb{E}(\xi))^2$$

Величина $(\xi - \mathbb{E}(\xi))^2$ показывает квадратичное отклонение матожидания от ξ . А значит, дисперсия показывает среднее квадратичное отклонение от матожидания. То есть это число, которое характеризует, на сколько в среднем ξ отличается от своего наилучшего приближения неслучайным числом. Проблема с дисперсией

только в одном. Если ξ измерялась в метрах, то дисперсия будет измеряться в метрах в квадрате. Чтобы получить более разумную величину, определим стандартное отклонение следующим образом

$$\sigma(\xi) = \sqrt{\mathbb{D}(\xi)}$$

То есть это корень из среднеквадратичного отклонения. Вот эту величину можно рассматривать как чуть более адекватную меру отклонения от математического ожидания. Ее преимущество перед дисперсией в том, что она измеряется в тех же единицах, что и исходная случайная величина.

Дисперсия с помощью функций распределения Пусть случайная величина ξ имеет функцию распределения $F_\xi(x) = P(\xi \leq x)$. Тогда

$$\mathbb{D}(\xi) = \int_{-\infty}^{\infty} (x - a)^2 dF_\xi(x)$$

где $a = \mathbb{E}(\xi)$. И теперь это выражение считается по тем же правилам, что и математическое ожидание.

Пусть $F_\xi(x) = F(x)$ и пусть a_1, \dots, a_n, \dots – это точки разрыва, а величина разрыва будет p_1, \dots, p_n, \dots для каждой точки (здесь $p_i = F(a_i) - F(a_i -)$). И пусть $\varphi(x)$ – это функция, в которой мы убрали все точки разрыва у функции F . Тогда

$$\mathbb{D}(\xi) = \int_{\mathbb{R}} (x - a)^2 dF(x) = \sum_i (a_i - a)^2 p_i + \int_{\mathbb{R}} (x - a)^2 \varphi'(x) dx$$

Как и с матожиданием для дискретного и непрерывного случая отдельно получится сильно лучше.

Дисперсия в случае дискретного распределения Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases}$$

Тогда дисперсия считается так

$$\mathbb{D}(\xi) = \sum_{i=1}^{\infty} (a_i - m)^2 p_i, \text{ где } m = \mathbb{E}\xi = \sum_{i=1}^{\infty} a_i p_i$$

Дисперсия в случае непрерывного распределения Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда Дисперсия считается по формуле

$$\mathbb{D}(\xi) = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx, \text{ где } m = \mathbb{E}(\xi) = \int_{-\infty}^{\infty} x p(x) dx$$

Свойства дисперсии Если ξ – некоторая случайная величина, а $a, b \in \mathbb{R}$, тогда

$$\mathbb{D}(a\xi + b) = a^2 \mathbb{D}(\xi)$$

Моменты Если ξ – некоторая случайная величина, то можно определить ее k -ый момент следующим образом

$$\mathbb{M}_k(\xi) = \mathbb{E}(\xi^k)$$

То есть первый момент – это матожидание. Если матожидание нулевое, то второй момент – это дисперсия.

Моменты в дискретном случае Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases}$$

Тогда моменты считаются по формуле

$$\mathbb{M}_k(\xi) = \sum_{i=1}^{\infty} a_i^k p_i$$

Моменты в непрерывном случае Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда моменты считаются по формуле

$$\mathbb{M}_k(\xi) = \int_{-\infty}^{\infty} x^k p(x) dx$$

Замечание про функции от случайной величины Пусть у вас есть случайная величина ξ и произвольная функция $f: \mathbb{R} \rightarrow \mathbb{R}$, тогда вы можете подставить случайную величину в функцию и получить новую случайную величину $f(\xi)$. Смысл этого действия – вы обработали исходные данные своего прибора, который был воткнут в черный ящик. Если вы хотите посчитать математическое ожидание от $f(\xi)$, то можно использовать следующую картинку

$$\mathbb{E}f(\xi) = \int_{\mathbb{R}} f(x) dP_{\xi}$$

которая в дискретном и непрерывном случае превращается в следующие формулы

$$\mathbb{E}f(\xi) = \sum_{i=1}^{\infty} f(a_i) p_i \quad \text{и} \quad \mathbb{E}f(\xi) = \int_{-\infty}^{\infty} f(x) p(x) dx$$

Эти формулы получаются из предыдущей непонятной формулы заменами описанными в разделе про матожидание.

Нормальное или Гауссово распределение Есть в науке очень популярный вид распределения, который называется нормальным или гауссовым. Так получилось, что этот вид распределения моделирует много похожих процессов, например, аккуратность стрелка по мишени или случайный шум. Популярность гауссова распределения в основном связана с тем, что его математические свойства очень простые и изучены вдоль и поперек. А так как на глаз, его поведение адекватно отражает то, что нужно в приложениях, то почему бы его не использовать?

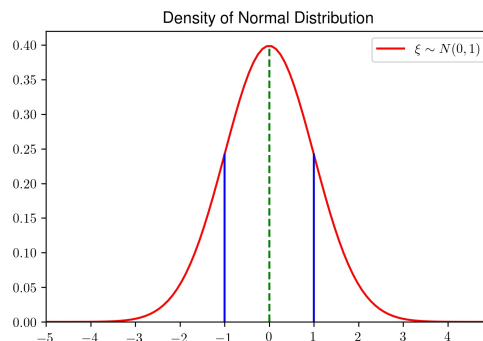
Гауссово распределение – это непрерывная случайная величина ξ на прямой \mathbb{R} задаваемая плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Когда задают гауссову случайную величину ξ с параметрами a и σ , обычно для краткости пишут $\xi \sim N(a, \sigma)$. Если посчитать числовые характеристики этого распределения, то выяснится

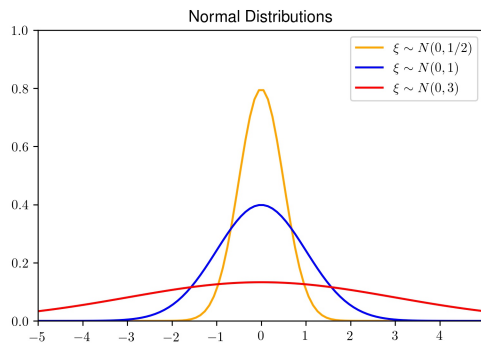
1. $\mathbb{E}(\xi) = a$.
2. $\mathbb{D}(\xi) = \sigma^2$ и в частности $\sigma(\xi) = \sigma$.

На следующей картинке изображен график плотности:



Пик плотности приходится на число a от которого плотность спадает в обе стороны экспоненциально. Думать про это надо так, мы с большей вероятностью отвечаем числа в окрестности точки a . У этого графика есть

две точки перегиба слева и справа от максимум, а именно, это точки $a - \sigma$ и $a + \sigma$. То есть среднеквадратичное отклонение указывает в точности на расстояние от максимума, до точек перегиба по бокам. Еще полезный факт, в интервале $[a - 3\sigma, a + 3\sigma]$ содержится 0,997 всей плотности, то есть вероятность $P(a - 3\sigma \leq \xi \leq a + 3\sigma) \geq 0,997$. На практике это означает, что вне этого интервала можно пренебречь значениями плотности и считать ее нулевой. Вот для ориентира еще несколько функций плотности для нормального распределения сосредоточенного в нуле, но разными дисперсиями.



Семинар 2

Задачи:

1. Пусть ξ – случайная величина такая, что $P(\xi = -1) = \frac{1}{4}$ и $P(\xi = 2) = \frac{1}{4}$ и для любых точек $a, b \in [0, 1]$ с условием $a < b$ верно $P(\xi \in [a, b]) = \frac{b-a}{2}$. Нарисуйте график функции распределения $F_\xi(x) = P(\xi \leq x)$. Найдите математическое ожидание и дисперсию ξ .
2. В лотерее на 1% билетов выпадет выигрыш в 200 рублей, на 0.01% билетов выпадает выигрыш в 1000 рублей, а остальные билеты без выигрыша. Найдите средний выигрыш в этой лотерее (в расчете на один билет), то есть среднее арифметическое выигрышей всех билетов. Подумайте, как это сформулировать в терминах случайных величин и математического ожидания.
3. В ряд расположены m предметов. Случайно выбираются k предметов, $k < m$. Случайная величина X равна количеству таких предметов i , что i выбран, а все его соседи не выбраны. Найдите математическое ожидание X .
4. В равностороннем треугольнике ABC площади 1 выбираем точку M . Найти математическое ожидание площади ABM .
5. Найти математическое ожидание и дисперсию величины ξ с плотностью $p(x) = \frac{1}{2\alpha} e^{-\frac{|x-a|}{\alpha}}$.
6. Пусть $P_\xi(x) = P(\xi = x)$ и $F_\xi(x) = P(\xi \leq x)$.
 - (а) Показать, что для $a > 0$ и $-\infty < b < \infty$

$$P_{a\xi+b}(x) = P_\xi\left(\frac{x-b}{a}\right) \quad \text{и} \quad F_{a\xi+b}(x) = F_\xi\left(\frac{x-b}{a}\right)$$

- (b) Найдите функцию распределения $F_{\xi^2}(y)$ (выразите ее через F_ξ и P_ξ).

Теория вероятности

Дима Трушин

Семинар 3

Случайные вектор

Пусть теперь на наш черный ящик, описываемый парой (Ω, P) мы хотим прицепить несколько измерительных устройств. Это означает, что мы имеем несколько случайных величин $\xi_1, \dots, \xi_n: \Omega \rightarrow \mathbb{R}$. В этом случае удобно все эти случайные величины объединить в один вектор и рассмотреть $\bar{\xi}: \Omega \rightarrow \mathbb{R}^n$ по правилу $\omega \mapsto \bar{\xi}(\omega) = (\xi_1(\omega), \dots, \xi_n(\omega))$. То есть для того, чтобы изучать несколько случайных величин вместе, достаточно изучить один случайный вектор. Тут возникает первый вопрос: а почему бы нам не изучать каждую из случайных величин отдельно? Основная проблема в том, что, изучая случайные величины отдельно, мы не можем ничего узнать о вероятностях их взаимодействия между собой. Например, если мы кидаем монетку и ξ дает 0 в случае орла и 1 в случае решки. Можно определить η , которая действует наоборот, дает 1 в случае орла и 0 в случае решки. Тогда у нас эти случайные величины не могут одновременно дать два нуля или две единицы. То есть между ними есть некое взаимодействие. Чтобы понять это взаимодействие их надо изучать одновременно, потому надо изучить случайный вектор (ξ, η) .

Как и в случае одной случайной величины, случайный вектор $\bar{\xi}: \Omega \rightarrow \mathbb{R}^n$ дает новое распределение в пространстве векторов \mathbb{R}^n . А именно, для любого подмножества $A \subseteq \mathbb{R}^n$ рассмотрим вероятность попасть в это подмножество

$$P_{\bar{\xi}}(A) = P(\bar{\xi} \in A) = P(\{\omega \mid \bar{\xi}(\omega) \in A\}) = P(\bar{\xi}^{-1}(A))$$

Вероятность $P_{\bar{\xi}}$ называется совместным распределением ξ_1, \dots, ξ_n .

Многомерная функция распределения Как и в одномерном случае, случайный вектор можно задавать некой функцией, называемой функцией распределения, но теперь это функция от нескольких переменных

$$F_{\bar{\xi}}(x_1, \dots, x_n) = P_{\bar{\xi}}((-\infty, x_1] \times \dots \times (-\infty, x_n]) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n)$$

Функция $F_{\bar{\xi}}(x_1, \dots, x_n)$ называется функцией совместного распределения случайных величин ξ_1, \dots, ξ_n . В этом случае можно восстановить вероятность попадания в произвольный прямоугольный параллелепипед $(a_1, b_1] \times \dots \times (a_n, b_n]$ по формуле включения исключения. Я продемонстрирую это на случае двух переменных. Если

$$F_{(\xi, \eta)}(x_1, x_2) = P(\xi \leq x_1, \eta \leq x_2) = f(x_1, x_2)$$

Тогда

$$P(a_1 < \xi \leq b_1, a_2 < \eta \leq b_2) = f(b_1, b_2) - f(a_1, b_2) - f(b_1, a_2) + f(a_1, a_2)$$

И далее существует некая общая процедура (которая интересна только математикам), позволяющая продлить это определение на любые подмножества в \mathbb{R}^n .

Виды случайных векторов Как и в случае случайной величины, можно рассматривать два вида случайных векторов: дискретные и непрерывные, то есть все координаты либо одновременно дискретные, либо одновременно непрерывные, да еще и как-то могут между собой взаимодействовать. Бывают и более интересные смеси случайных векторов, когда одна координата дискретна, а другая непрерывна. Бывает еще более интересное сочетание, когда каждая координата является смесью и того и другого. К сожалению, чем сложнее устроены координаты у случайного вектора, тем сложнее с ним взаимодействовать и приходится страдать. Чтобы страдать поменьше и понять, как все это работает я сосредоточу свое внимание в начале на «чистых» случайных векторах (все координаты дискретны или все координаты непрерывны), а потом поговорю, что делать в случае, если есть координаты обоих сортов (такие штуки встречаются в приложениях).¹

¹Математики привыкли все это делать в самом общем виде и математической теории глубоко плевать, какие именно виды случайных величин участвуют в законах, она умеет работать со всеми одинаково, только для этого приходится писать формулы,

Дискретные случайные векторы Для того, чтобы задать дискретный случайный вектор $\bar{\xi}$ в \mathbb{R}^n надо задать следующие данные

1. Набор векторов $a_1, a_2, \dots, a_n, \dots \in \mathbb{R}^n$. Он может быть конечным или бесконечным. Это будут выделенные значения, которые разрешено принимать нашему случайному вектору. Они еще будут называться атомами.
2. Набор вероятностей $p_1, p_2, \dots, p_n, \dots \in [0, 1]$, которые приписываются соответствующему значению a_i . Они должны обладать свойствами

- (a) $0 < p_k \leq 1$.
- (b) $\sum_{k=1}^{\infty} p_k = 1$.

В этом случае $P(\bar{\xi} = a_i) = p_i$ и для любого числа $a \in \mathbb{R}$ не равного ни одному из a_i мы имеем $P(\bar{\xi} = a) = 0$. Для произвольного множества $A \subseteq \mathbb{R}^n$ его вероятность считается

$$P_{\bar{\xi}}(A) = \sum_{a_i \in A} p_i$$

то есть надо сложить вероятности всех атомов, попавших в множество A .

Непрерывные случайные векторы Для того, чтобы задать непрерывный случайный вектор $\bar{\xi}$, надо задать функцию $p: \mathbb{R}^n \rightarrow \mathbb{R}$, которая называется плотностью случайного вектора, со следующими свойствами

1. $0 \leq p(x_1, \dots, x_n)$.
2. $\int_{\mathbb{R}^n} p(x_1, \dots, x_n) dx_1 \dots dx_n = 1$

Тогда для произвольного подмножества $A \subseteq \mathbb{R}^n$ вероятность попасть в него равна

$$P_{\bar{\xi}}(A) = \int_A p(x_1, \dots, x_n) dx_1 \dots dx_n$$

В этом случае функция распределения имеет вид

$$F_{\bar{\xi}}(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} p_{\bar{\xi}}(t_1, \dots, t_n) dt_1 \dots dt_n$$

То есть, зная плотность, мы посчитали функцию распределения. В обратную сторону, из функции распределения плотность находится дифференцированием по всем верхним параметрам интегралов, а именно

$$p_{\bar{\xi}}(x_1, \dots, x_n) = \frac{d}{dx_1} \dots \frac{d}{dx_n} F_{\bar{\xi}}(x_1, \dots, x_n)$$

Примеры

1. Пусть мы бросаем монетку так, что вероятность выпадения орла есть p , а вероятность выпадения решки – q (при этом $p + q = 1$). Пусть ξ – случайная величина, которая равна 0, когда выпадает орел и 1, когда выпадает решка. Тогда это дискретная случайная величина с табличкой

$$\xi \sim \begin{Bmatrix} 0 & 1 \\ p & q \end{Bmatrix}$$

Теперь рассмотрим случайную величину $\eta = 1 - \xi$, то есть у нее выпадает 0 на решку и 1 на орла. Тогда η задается табличкой

$$\eta \sim \begin{Bmatrix} 0 & 1 \\ q & p \end{Bmatrix}$$

Давайте поймем, а какое же у них будет совместное распределение? Для случайного вектора (ξ, η) получим следующую таблицу

которые содержат функции распределения. А некоторые вещи приходится определять косвенным образом (условные математические ожидания вам еще передадут привет).

Значения (ξ, η)	Вероятность выпадения
$(0, 0)$	0
$(0, 1)$	p
$(1, 0)$	q
$(1, 1)$	0

2. Пусть теперь у нас есть квадрат $[0, 1]^2$. Рассмотрим функцию плотности $p(x, y)$ такую, что $p(x, y) = 1$ при условии $(x, y) \in [0, 1]^2$ и $p(x, y) = 0$ иначе. Это дает нам $\Omega = [0, 1]^2$ и вероятность $P(A) = \int_A p(x, y) dx dy$. Тогда вектор (x, y) – будет случайным вектором на квадрате. Это равномерно распределенный на квадрате вектор в том смысле, что вероятность попасть в подмножество $A \subseteq [0, 1]^2$ в точности равно площади этого подмножества A .

Восстановление распределений координат Предположим нам задан случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_n)$ на пространстве \mathbb{R}^n и мы знаем его распределение $P_{\bar{\xi}}$. Вопрос, а как найти распределения его координат ξ_i ? Оказывается очень просто. Давайте сделаем на примере ξ_1 . Нам надо для любого множества $A \subseteq \mathbb{R}$ посчитать вероятность $P(\xi_1 \in A)$. Но

$$P(\xi_1 \in A) = P(\xi_1 \in A, \xi_2 \in \mathbb{R}, \dots, \xi_n \in \mathbb{R}) = P_{\bar{\xi}}(A \times \mathbb{R} \times \dots \times \mathbb{R})$$

С другой стороны может возникнуть желание задать другой вопрос, а как восстановить совместное распределение по распределениям координат? Тут ответ очень простой – никак. Вообще говоря совместное распределение содержит больше информации, чем просто распределение каждой координаты, оно еще и кодирует, как координаты взаимодействуют. Если мы знаем распределение координат, то прежде чем восстанавливать совместное распределение, нам нужна какая-то информация об этом самом взаимодействии между координатами. В общем виде есть только один случай, когда это возможно – при условии независимости координат. Что это такое и как восстанавливать совместное распределение в этом случае я расскажу чуть позже. А пока продемонстрирую вышесказанное на примере дискретного и непрерывного распределений.

Восстановление функций распределения координат Пусть случайный вектор $\bar{\xi}$ имеет функцию распределения $F(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n)$. Тогда функция распределения для первой координаты ξ_1 будет

$$\begin{aligned} F_1(x_1) &= P(\xi_1 \leq x_1) = P(\xi_1 \leq x_1, \xi_2 \in \mathbb{R}, \dots, \xi_n \in \mathbb{R}) = \\ &= \lim_{\substack{x_2 \rightarrow +\infty \\ \vdots \\ x_n \rightarrow +\infty}} P(\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_n \leq x_n) = \lim_{\substack{x_2 \rightarrow +\infty \\ \vdots \\ x_n \rightarrow +\infty}} F(x_1, \dots, x_n) \end{aligned}$$

Таким образом, если мы хотим найти функцию распределения i -ой координаты, надо все остальные устремить к $+\infty$.

Восстановление координат в дискретном случае Если случайный вектор $\bar{\xi}$ задавался данными

$$\xi \sim \begin{Bmatrix} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{Bmatrix} \in \mathbb{R}^n$$

Обозначим, через $(a_i)_1$ первую координату вектора a_i . Тогда значения случайной величины ξ_1 будут все числа вида $(a_i)_1$, то есть все первые координаты, которые нам встретились в таблице. А вероятность считается по формуле

$$P(\xi_1 = a) = \sum_{(a_i)_1 = a} p_i$$

То есть мы находим все векторы a_i , у которых первая координата равна a . Если таких нет, то вероятность полагается нулем. Если есть, то мы складываем все их вероятности вместе – это и будет вероятность попасть в a .

Восстановление координат в непрерывном случае Пусть некоторый случайный вектор $\bar{\xi}$ задан плотностью $p(x_1, \dots, x_n)$. Тогда посчитаем функцию распределения ξ_1 по определению

$$F_{\xi_1}(x_1) = \int_{(-\infty, x_1] \times \mathbb{R} \times \dots \times \mathbb{R}} p(x_1, \dots, x_n) dx_1 \dots dx_n = \int_{-\infty}^{x_1} \left(\int_{\mathbb{R}^{n-1}} p(x_1, x_2, \dots, x_n) dx_2 \dots dx_n \right) dx_1$$

Но плотность вычисляется, как производная от функции распределения, то есть

$$p_{\xi_1}(x_1) = F_{\bar{\xi}}(x_1)' = \int_{\mathbb{R}^{n-1}} p(x_1, x_2, \dots, x_n) dx_2 \dots dx_n$$

Таким образом, чтобы получить плотность для первой координаты, надо проинтегрировать совместную плотность по всем остальным координатам вдоль всей прямой.

Например, в случае двумерного случайного вектора (ξ, η) с плотностью $p_{\bar{\xi}}(x, y)$, плотности самих ξ и η получаются по формулам

$$p_{\xi}(x) = \int_{\mathbb{R}} p(x, y) dy \quad \text{и} \quad p_{\eta}(y) = \int_{\mathbb{R}} p(x, y) dx$$

Независимость Очень часто встречается понятие о независимых измерениях. Мы думаем о них, как об измерениях, которые никак друг на друга не влияют, никоим образом. Например, мы можем взять и независимо подбросить монетку, сначала один раз, а потом другой. Или еще лучше, взять двух разных людей и каждый из них независимо от другого подбросит монетку. Мы интуитивно чувствуем, что это должно значить, но хорошо бы иметь математическое выражение для подобного явления, чтобы им как-то пользоваться. Но для начала я хочу привести пример зависимых измерений. Скажем пример, когда ξ дает 0 в случае выпадения орла и 1 в случае выпадения решки, а η наоборот. Тут мы интуитивно понимаем, что если мы знаем ответ для ξ , то и для η ответ восстанавливается однозначно, нет никакой независимости между этими случайными величинами.

Пусть у нас есть две случайные величины ξ и η на одном и том же вероятностном пространстве (Ω, P) , то есть это два отображения $\xi, \eta: \Omega \rightarrow \mathbb{R}$. Они называются независимыми и пишут $\xi \perp \eta$, если для любых множеств $A \subseteq \mathbb{R}$ и $B \subseteq \mathbb{R}$ события $\xi^{-1}(A)$ и $\eta^{-1}(B)$ независимы. То есть событие « ξ попало в A » не зависит с событием « η попало в B » для любых двух подмножеств. Напомню, что независимость событий означает, что выполнено

$$P(\xi^{-1}(A) \cap \eta^{-1}(B)) = P(\xi^{-1}(A))P(\eta^{-1}(B))$$

или в текстовой форме

$$P(\text{«}\xi \text{ попала в } A\text{» и «}\eta \text{ попала в } B\text{»}) = P(\text{«}\xi \text{ попала в } A\text{»})P(\text{«}\eta \text{ попала в } B\text{»})$$

Давайте я немного поясню это определение. Для случайной величины $\xi: \Omega \rightarrow \mathbb{R}$ можно определить запас событий из Ω , которые выражаются в терминах этой случайной величины. Это так называемая алгебра событий ξ (или ее принято называть σ -алгебра событий). Обозначается она $\sigma\{\xi\}$ и по определению

$$\sigma\{\xi\} = \{\xi^{-1}(A) \subseteq \Omega \mid A \subseteq \mathbb{R}\}$$

Действительно, когда мы видим вместо нашего черного ящика, знающего все, что только можно, про Ω , видим только случайную величину ξ , единственная доступная нам информация – значение ξ , а это обычное число. Потому события, которые мы могли бы выразить, выражаются в терминах этого числа. Значит по сути, мы можем лишь проверить попадание значения ξ в какое-то подмножество. То есть $\xi^{-1}(A)$ для $A \subseteq \mathbb{R}$ – это все, что мы только можем прощупать в нашем черном ящике с помощью ξ . Потому определение независимости для случайных величин означает, что мы хотим, чтобы любое событие, которое мы только можем выразить в терминах одной случайной величины было независимым с любым другим событием, которое можно выразить в терминах другой случайной величины.²

²Есть и другие более приличные применения для алгебр событий, они полезны для работы с условными математическими ожиданиями.

Независимые индикаторы Пусть (Ω, P) – вероятностное пространство описывающее какой-то случайный черный ящик. Фиксируем два произвольных события $A, B \subseteq \Omega$ и рассмотрим две индикаторные случайные величины χ_A и χ_B . Напомню, что $\chi_A(\omega)$ равна 1, если $\omega \in A$, и 0 иначе. Тогда с помощью χ_A можно выразить четыре события: \emptyset, A, \bar{A} и Ω . Действительно, $A = \chi_A^{-1}(1)$ и $\bar{A} = \chi_A^{-1}(0)$, $\emptyset = \chi_A^{-1}(\emptyset)$ и $\Omega = \chi_A^{-1}(\mathbb{R})$. Ничего другого выразить через нее нельзя. Аналогично и для χ_B . Для того, чтобы проверить их независимость, надо проверить, что любое событие из списка $\{\emptyset, A, \bar{A}, \Omega\}$ независимо с любым событием из списка $\{\emptyset, B, \bar{B}, \Omega\}$. Так как \emptyset и Ω не зависят с любым событием, то остается проверить, что любое событие из $\{A, \bar{A}\}$ не зависит с любым событием из $\{B, \bar{B}\}$. В частности отсюда следует, что события A и B независимы. Наоборот, если события A и B независимы, то и события A и \bar{B} независимы (я сейчас это покажу). А значит и все остальные пары событий независимы. То есть независимость индикаторных случайных величин χ_A и χ_B равносильная независимости самих событий A и B . Теперь давайте я проверю независимость A и \bar{B} , если A и B были независимыми. Действительно,

$$P(A \cap \bar{B}) = P(A \cap (\Omega \setminus B)) = P(A \setminus (A \cap B)) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(\bar{B})$$

Независимость в терминах функций распределения Пусть заданы две случайные величины ξ и η с совместной функцией распределения $F(x, y)$. Предположим, что ξ и η независимы, тогда по определению

$$F(x, y) = P(\xi \leq x, \eta \leq y) = P(\xi \leq x)P(\eta \leq y) = F_\xi(x)F_\eta(y)$$

Равенство $F(x, y) = F_\xi(x)F_\eta(y)$ означает независимость событий вида $\xi^{-1}(-\infty, x]$ и $\eta^{-1}(-\infty, y]$. Так как эти события порождают все остальные в некотором смысле, который я обсуждал лишь неформально, то из этого следует независимость всех возможных пар событий вида $\xi^{-1}(A)$ и $\eta^{-1}(B)$ для $A, B \subseteq \mathbb{R}$.

В сухом остатке, две случайные величины ξ и η независимы тогда и только тогда, когда их функция совместного распределения есть произведение функций распределения. Здесь я хочу сделать важное замечание. Это означает, что если вы задали распределение для случайных величин ξ и η и сказали, что они независимы, то это автоматически задает их совместное распределение, а значит вы знаете всю вероятностную информацию, чтобы работать с этими случайными величинами.

Независимость в дискретном случае Пусть заданы две дискретные случайные величины ξ и η . Они будут независимы тогда и только тогда, когда

$$P(\xi = a, \eta = b) = P(\xi = a)P(\eta = b), \text{ для любых } a, b \in \mathbb{R}$$

Если же нам выдали две дискретные случайные величины

$$\xi \sim \begin{Bmatrix} a_1 & \dots & a_n & \dots \\ p_1 & \dots & p_n & \dots \end{Bmatrix} \quad \text{и} \quad \eta \sim \begin{Bmatrix} b_1 & \dots & b_n & \dots \\ q_1 & \dots & q_n & \dots \end{Bmatrix}$$

Тогда случайный вектор (ξ, η) имеет распределение

$$(\xi, \eta) \sim \begin{Bmatrix} (a_1, b_1) & \dots & (a_n, b_m) & \dots \\ p_1 q_1 & \dots & p_n q_m & \dots \end{Bmatrix}$$

Про это можно думать следующим образом. Вероятности для вектора (ξ, η) задаются с помощью матрицы A , где в i, j -ой ячейке стоит вероятность $P(\xi = a_i, \eta = b_j)$. Распределение для ξ задается вектором v , где i -я координата – это вероятность $P(\xi = a_i)$. Аналогично, распределение для η задается вектором u , где i -я координата – это вероятность $P(\eta = b_i)$. Тогда независимость означает, что $A = vu^t$. То есть матрица A является матрицей ранга 1. Это в частности показывает, что независимость это очень редкое явление. Чтобы некоторая матрица A задавала распределение дискретного вектора на плоскости, достаточно, чтобы все ее элементы были неотрицательными и сумма всех элементов матрицы была равна 1. Но среди них почти нет матриц ранга 1. Случайная такая матрица будет иметь полный ранг.

Независимость в непрерывном случае Пусть заданы две непрерывные случайные величины ξ и η с совместной плотностью $p_{\xi, \eta}(x, y)$ и каждая имеет плотность $p_\xi(x)$ и $p_\eta(y)$ соответственно. Тогда эти случайные величины независимы тогда и только тогда, когда $p_{\xi, \eta}(x, y) = p_\xi(x)p_\eta(y)$. Это дает способ проверки на независимость и способ построения вектора из двух независимых случайных величин. Кратко правило можно резюмировать следующим образом. Если у вас есть две плотности независимых случайных величин, то для построения совместной плотности, их надо перемножить, подставляя в каждую плотность свою собственную переменную.

Матожидание и дисперсия в случае независимых случайных величин Пусть ξ и η – две независимые случайные величины, тогда

1. $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$.
2. $\mathbb{D}(\xi + \eta) = \mathbb{D}(\xi) + \mathbb{D}(\eta)$.

Второе выводится из первого, пользуясь определением. А первое является предельной версией определения. Давайте дам один полезный пример. Если $\xi = \chi_A$ и $\eta = \chi_B$ – индикаторные случайные величины. Тогда $\mathbb{E}(\chi_A) = P(A)$, $\mathbb{E}(\chi_B) = P(B)$, $\mathbb{E}(\chi_A\chi_B) = \mathbb{E}(\chi_{A \cap B}) = P(A \cap B)$. Таким образом формула превращается в определение независимости $P(A \cap B) = P(A)P(B)$. Общая формула получается переходом от этой простейшей формулы к интегралу с помощью сумм и предельных переходов. Но это уже отдельная история для любителей тонкостей математики. Для нас более полезно понимать идею, почему это так работает, и уметь всем этим пользоваться.

Замечание о большом количестве случайных величин В случае, когда у нас есть несколько событий $A_1, \dots, A_n \subseteq \Omega$, то они называются независимыми, если

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k}), \quad \text{для всех возможных } 1 \leq i_1 < \dots < i_k \leq n$$

Напомню, что события могут быть попарно независимыми, но при все вместе они зависимы, то есть это понятие не сводится к проверки независимости в каждой паре. Тогда на независимость случайных величин это определение распространяется так. Если есть случайные величины $\xi_1, \dots, \xi_n: \Omega \rightarrow \mathbb{R}$, то для любых подмножеств $A_1, \dots, A_n \subseteq \mathbb{R}$, мы хотим, чтобы события $\xi_1^{-1}(A_1), \dots, \xi_n^{-1}(A_n) \subseteq \Omega$ были в совокупности независимы. Как и раньше, интуитивно это означает, что все события, которые только можно выразить в терминах наших случайных величин по отдельности не зависят в совокупности.

В дискретном случае, если ξ_1, \dots, ξ_n – дискретные случайные величины, то они независимы тогда и только тогда, когда

$$P(\xi_1 = a_1, \dots, \xi_n = a_n) = P(\xi_1 = a_1) \dots P(\xi_n = a_n), \quad \text{для любых } a_1, \dots, a_n \in \mathbb{R}$$

В непрерывном случае, если ξ_1, \dots, ξ_n – непрерывные случайные величины, то они независимы тогда и только тогда, когда $p_{\xi}(x_1, \dots, x_n) = p_{\xi_1}(x_1) \dots p_{\xi_n}(x_n)$.

С помощью функций распределения можно проверить на независимость любые наборы из случайных величин любого вида. Случайные величины ξ_1, \dots, ξ_n независимы тогда и только тогда, когда для всех x_1, \dots, x_n

$$F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n)$$

напомню, что функция распределения – это $F_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n)$.

Обратите внимание, что в случае событий для независимости мы должны проверять, что все возможные подмножества удовлетворяют формуле для произведения, а в случае случайных величин, мы всегда смотрим на весь набор случайных величин. Это кажется странным, учитывая, что независимость в совокупности для $A, B, C \subseteq \Omega$ должна быть равносильна независимости χ_A, χ_B, χ_C . Однако, не забывайте, что в случае случайных величины у нас есть дополнительный контроль. Мы можем тестировать на попадание в любое подмножество на прямой. Например,

$$\begin{aligned} P(A \cap B) &= P(\chi_A = 1, \chi_B = 1) = P(\chi_A = 1, \chi_B = 1, \chi_C \in \mathbb{R}) = \\ &= P(\chi_A = 1)P(\chi_B = 1)P(\chi_C \in \mathbb{R}) = P(\chi_A = 1)P(\chi_B = 1) = P(A)P(B) \end{aligned}$$

В итоге важно запомнить, что в случае случайных величин не надо перебирать подмножества среди самих величин, но надо проверить более сильное условие на все величины.

В общем случае для независимых в совокупности случайных величин ξ_1, \dots, ξ_n верны следующие свойства для математического ожидания и дисперсии

1. $\mathbb{E}(\xi_1 \dots \xi_n) = \mathbb{E}(\xi_1) \dots \mathbb{E}(\xi_n)$.
2. $\mathbb{D}(\xi_1 + \dots + \xi_n) = \mathbb{D}(\xi_1) + \dots + \mathbb{D}(\xi_n)$.

Первая формула говорит о том, что на случайные величины можно смотреть как на «нечеткие события», а их математическое ожидание – это как бы «вероятности» этих нечетких событий. То есть событие – это такое событие с кратностями для элементарных исходов (причем мы можем считать, что элементарный исход не принадлежит «событию» с какой-то кратностью при отрицательном значении). Правда тогда, вот эта «вероятность» в виде математического ожидания становится странной, она может быть и положительной и отрицательной, выражая нашу степень уверенности принадлежности «нечеткому событию». После такой аналогии первая формула превращается в привычную формулу независимости для событий. Надо трактовать эту формулу как «независимость в среднем». Конечно же это равенство слабее честной независимости случайных величин.

Построение независимых случайных величин Пусть на случайном черном ящике (Ω, P) задан набор независимых в совокупности случайных величин $\xi_1, \dots, \xi_n: \Omega \rightarrow \mathbb{R}$. Давайте разобьем их на две не пересекающиеся части ξ_1, \dots, ξ_m и ξ_{m+1}, \dots, ξ_n . Теперь пусть $f: \mathbb{R}^m \rightarrow \mathbb{R}$ и $g: \mathbb{R}^{n-m} \rightarrow \mathbb{R}$ – две произвольные функции. Тогда случайные величины $f(\xi_1, \dots, \xi_m)$ и $g(\xi_{m+1}, \dots, \xi_n)$ будут независимыми. Аналогичный трюк работает, если мы разобьем исходное множество случайных величин ξ_1, \dots, ξ_n на произвольное число не пересекающихся подмножеств, элементы каждого подмножества подставим в свою функцию, тогда полученный набор случайных величин так же окажется независимым в совокупности. Поэтому, когда вы стартовали с чего-то независимого, то вы можете очень естественным образом строить целые семейства независимых величин исходя из наблюдения выше.

Характеристики случайных векторов

Если на вероятностном пространстве (Ω, P) определен случайный вектор $\bar{\xi}: \Omega \rightarrow \mathbb{R}^n$, то для него определены аналоги математического ожидания и дисперсии. Математическое ожидание будет некоторым вектором, который в среднем приближает наш случайный, а вот дисперсия будет матрицей квадратичной формы, отвечающей за случайные отклонения от этого среднего значения.

Математическое ожидание Математическое ожидание опять же определяется мутной формулой для интегрирования

$$\mathbb{E}(\bar{\xi}) = \int_{\Omega} \bar{\xi} dP = \int_{\mathbb{R}^n} \bar{x} dP_{\bar{\xi}}$$

Здесь $\bar{x} = (x_1, \dots, x_n)$ – вектор из переменных. Давайте придадим ей смысл в тех ситуациях в которых мы умеем работать со случайными векторами.

В начале надо сказать, что интегрировать вектор – это то же самое, что интегрировать каждую его координату. Потому формула выше означает вычисление n интегралов для каждой из координат. А значит, для вектора $\bar{\xi} = (\xi_1, \dots, \xi_n)$, верна формула

$$\mathbb{E}(\bar{\xi}) = (\mathbb{E}(\xi_1), \dots, \mathbb{E}(\xi_n))$$

То есть математическое ожидание можно вычислять по координатам. Для начала надо вычислить распределение каждой из координаты, а потом посчитать матожидание для каждой координаты.

Сказанное выше является чуть ли не ультимативным знанием про математическое ожидание вектора, которое не требует больше вообще никаких знаний. Тем не менее, давайте обсудим явные формулы в трех случаях.

Математическое ожидание в дискретном случае Пусть случайный вектор $\bar{\xi}$ имеет дискретное распределение

$$\xi \sim \begin{cases} a_1 & \dots & a_n & \dots & \in \mathbb{R}^n \\ p_1 & \dots & p_n & \dots & \in [0, 1] \end{cases}$$

Тогда

$$\mathbb{E}(\bar{x}) = \sum_{i=1}^{\infty} a_i p_i = \sum_{i=1}^{\infty} a_i P(\bar{\xi} = a_i)$$

Математическое ожидание в непрерывном случае Пусть теперь $\bar{\xi}$ имеет непрерывное распределение с плотностью $p(x_1, \dots, x_n)$. Тогда

$$\mathbb{E}(\bar{\xi}) = \int_{\mathbb{R}^n} \bar{x} p(x_1, \dots, x_n) dx_1 \dots dx_n$$

Обратите внимание, что тут интегрируется векторная функция $\bar{x}p(x_1, \dots, x_n)$, где $\bar{x} = (x_1, \dots, x_n)$.

Математическое ожидание и функция распределения Пусть случайный вектор $\bar{\xi}$ на пространстве \mathbb{R}^n имеет функцию распределения $F(x_1, \dots, x_n)$, тогда

$$\mathbb{E}(\bar{\xi}) = \int_{\mathbb{R}^n} \bar{x} dF(x_1, \dots, x_n)$$

Пока проще в понимании этого интеграла нам не стало. Однако, тут можно прибегнуть к такому трюку. Надо найти все атомы функции распределения и превратить ее в функцию без разрывов (аналогично одномерному случаю) $\varphi(x_1, \dots, x_n)$. Тогда атомы дадут часть математического ожидания по формулам для дискретного случая. А если φ оказалась дифференцируемой нужное количество раз, то по ней можно найти плотность непрерывной части для распределения, а именно, надо $\varphi(x_1, \dots, x_n)$ продифференцировать по каждому аргументу один раз. Это пока лучшее, что я могу сказать про общий случай.

Матрица ковариации В случае нескольких случайных величин мы будем определять не только среднее отклонение от среднего, но и коррелированность каждой пары случайных величин. По определению ковариацией случайных величин ξ и η называется следующее выражение

$$\text{cov}(\xi, \eta) = \mathbb{E}((\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)) = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta)$$

Обратите внимание, что $\mathbb{D}(\xi) = \text{cov}(\xi, \xi)$. Кроме того, равенство нулю ковариации равносильно равенству $(\xi\eta) = (\xi)(\eta)$. Как я говорил выше, про это можно думать, как про «независимость в среднем» для наших случайных величин. То есть зависимость вообще говоря есть, но при большом количестве экспериментов для усредненных данных мы ее не наблюдаем.

Если для двух случайных величин ξ и η мы имеем $\text{cov}(\xi, \eta) = 0$, то говорят, что случайные величины некоррелированы. Отметим, что если случайные величины независимы, то они и некоррелированы, однако, обратное вообще говоря не верно.

Если $\bar{\xi} = (\xi_1, \dots, \xi_n)$ – случайный вектор на некотором вероятностном пространстве, то матрицей ковариации этого вектора называется

$$\Sigma(\bar{\xi}) = \begin{pmatrix} \text{cov}(\xi_1, \xi_1) & \dots & \text{cov}(\xi_1, \xi_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(\xi_n, \xi_1) & \dots & \text{cov}(\xi_n, \xi_n) \end{pmatrix}$$

Важное замечание. Эта матрица всегда является симметричной и неотрицательно определенной. А значит, в случае, если она не вырождена, то она задает некоторое скалярное произведение на \mathbb{R}^n , а именно $\beta(x, y) = x^t \Sigma(\bar{\xi})^{-1} y$.³ Мы знаем, что можно представить матрицу $\Sigma(\bar{\xi})$ в следующем виде

$$\Sigma(\bar{\xi}) = UDU^t, \text{ где } U = (u_1 | \dots | u_n) \text{ – ортогональная матрица и } D = \text{diag}(d_1, \dots, d_n)$$

В этом случае давайте посмотрим на поверхность $x^t \Sigma(\bar{\xi})^{-1} x = 1$ – это поверхность, показывающая границу среднего отклонения вектора $\bar{\xi}$ от его математического ожидания. Эта граница образует эллипс. Векторы u_1, \dots, u_n – будут осями этого эллипса, а числа d_1, \dots, d_n характеризуют на сколько этот эллипс вытянут вдоль каждой из осей u_1, \dots, u_n (корни из d_i показывают максимальное значение x_i по модулю).

Свойства ковариации Для ковариации случайных величины выполнены следующие свойства

1. Если ξ и η – произвольные случайные величины, то

$$\text{cov}(\xi, \eta) = \text{cov}(\eta, \xi)$$

³Обратите внимание на обратную степень!

2. Если ξ_1, ξ_2, η – произвольные случайные величины, то

$$\text{cov}(\xi_1 + \xi_2, \eta) = \text{cov}(\xi_1, \eta) + \text{cov}(\xi_2, \eta)$$

3. Для любых случайных величин ξ и η и числа $a \in \mathbb{R}$ верно

$$\text{cov}(a\xi, \eta) = a \text{cov}(\xi, \eta) = \text{cov}(\xi, a\eta)$$

4. Для любых случайных величин ξ и η и числа $a \in \mathbb{R}$ верно

$$\text{cov}(\xi - a, \eta) = \text{cov}(\xi, \eta)$$

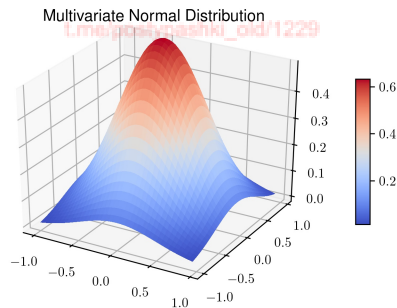
Многомерное гауссово распределение Я хочу задать аналог гауссова распределения в пространстве \mathbb{R}^n . Для этого зафиксируем некоторый вектор $a \in \mathbb{R}^n$ и матрицу $\Sigma \in M_n(\mathbb{R})$ такую, что Σ симметрична и положительно определена. Тогда можно задать непрерывное распределение со следующей плотностью

$$p_{\bar{\xi}}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-a)^t \Sigma^{-1}(x-a)}$$

В этом случае получаем

1. $\mathbb{E}(\bar{\xi}) = a$.
2. $\text{cov}(\xi_i, \xi_j) = \Sigma_{i,j}$.

Ниже на картинке нарисована плотность нормального распределения со средним 0 и матрицей ковариации $\frac{1}{2}E$:



Важно, что многомерное гауссово распределение полностью определяется своим математическим ожиданием и матрицей ковариации на столько, что его координаты независимы тогда и только тогда, когда они некоррелированы.

Геометрия гауссова распределения Пусть у нас задан случайный гауссов вектор ξ с плотностью

$$p_{\xi}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-a)^t \Sigma^{-1}(x-a)}$$

Тогда выражение под экспонентой $(x-a)^t \Sigma^{-1}(x-a)$ является квадратичной формой относительно вектора $x-a$. Так как матрица Σ является положительно определенной, то уравнение $(x-a)^t \Sigma^{-1}(x-a) = 1$ задает эллипсоид с центром в точке a . Точка a является математическим ожиданием нашей случайной величины ξ , потому мы будем ожидать, что ответы ξ будут копиться близко к этой точке. А указанный эллипсоид показывает границу среднего отклонения значения ξ от математического ожидания.

Мы знаем, что существует ортонормированный базис v_1, \dots, v_n такой, что матрица Σ будет диагональна $D = \text{diag}(d_1, \dots, d_n)$. Пусть $C = (v_1 | \dots | v_n)$ – матрица перехода от стандартного базиса к базису v_1, \dots, v_n . Тогда матрица C ортогональна, то есть $C^t C = E$. В этом случае $D = C^t \Sigma C$.

Давайте сдвинем начало координат в a и выберем координатные оси вдоль векторов v_1, \dots, v_n . Тогда в новом базисе v_1, \dots, v_n и новым началом координат случайный вектор ξ имеет вид $\eta = C^t(\xi - a)$. Вектор η

будет нормально распределенным вектором с центром в точке 0 и матрицей ковариации $D = \text{diag}(d_1, \dots, d_n)$. То есть

$$p_\eta(x) = \frac{1}{\sqrt{(2\pi)^n d_1 \dots d_n}} e^{-\frac{1}{2} \left(\frac{x_1^2}{d_1} + \dots + \frac{x_n^2}{d_n} \right)} = \frac{1}{\sqrt{2\pi d_1}} e^{-\frac{x_1^2}{2d_1}} \dots \frac{1}{\sqrt{2\pi d_n}} e^{-\frac{x_n^2}{2d_n}}$$

Значит d_1 будет дисперсией первой координаты для η , d_2 для второй и т.д. Кроме того, координаты вектора η оказались независимыми и нормально распределены. Потому мы знаем, что i -я координата в среднем меняется от $-\sqrt{d_i}$ до $\sqrt{d_i}$, то есть в среднем значения вектора будут заключены в прямоугольник. Однако, этот результат можно уточнить, значения будут реже бывать в углах этого прямоугольника и в итоге будут сосредоточены в среднем внутри упомянутого выше эллипсоида.

Семинар 3

Задачи:

1. На станцию приходят в случайное время две электрички. Времена их приходов независимы и имеют экспоненциальное распределение с плотностью $e^{-x} \cdot \theta(x)$. Студент приходит на станцию в момент времени 2. Найдите

- (а) вероятность того, что он сможет уехать хотя бы на одной электричке;
- (б) математическое ожидание времени ожидания студентом ближайшей электрички (считаем, что время ожидания равно нулю, если студент опоздал на обе электрички).

Здесь

$$\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

2. Пусть (ξ, η) – нормальный вектор с математическим ожиданием $0 \in \mathbb{R}^2$ и матрицей ковариации $\Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. Найдите плотности для его координат ξ и η . В этой задаче надо использовать следующий факт

$$\int_{\mathbb{R}} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$$

3. Пусть ξ_1, \dots, ξ_n – независимые одинаково распределенные непрерывные случайные величин с плотностью $p(x)$. Найдите плотность случайных величин: $\max(\xi_1, \dots, \xi_n)$ и $\min(\xi_1, \dots, \xi_n)$.

4. Пусть ξ и η – две случайные величины. Докажите, что

- (а) ξ и η независимы тогда и только тогда, когда $a\xi + b$ и η независимы. Здесь $a, b \in \mathbb{R}$ причем $a \neq 0$.

- (б) Если каждая из случайных величин ξ и η принимает только два значения, то они независимы тогда и только тогда, когда $\text{cov}(\xi, \eta) = 0$.

Теория вероятности

Дима Трушин

Семинар 4

Условные вероятности и условные математические ожидания

Дискретный случай Пусть даны две дискретные случайные величины ξ и η

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases} \quad \text{и} \quad \eta \sim \begin{cases} b_1 & b_2 & \dots & b_n & \dots \\ q_1 & q_2 & \dots & q_n & \dots \end{cases}$$

Тогда определены условные вероятности¹

$$P(\xi = a \mid \eta = b) = \frac{P(\xi = a, \eta = b)}{P(\eta = b)}$$

В случае, если $b \neq b_i$ ни для какого i (то есть при условии $P(\eta = b) = 0$), то мы полагаем эту вероятность равной нулю. В этом случае можно определить условное математическое ожидание

$$\mathbb{E}(\xi \mid \eta = b) = \sum_{i=1}^{\infty} a_i P(\xi = a_i \mid \eta = b)$$

t.me/postypashki_old/1229

t.me/postypashki_old/1229

t.me/postypashki_old/1229

Непрерывный случай Пусть теперь ξ и η – непрерывные случайные величины с плотностями $p_\xi(x)$ и $p_\eta(y)$ и совместной плотностью $p_{\xi,\eta}(x, y)$. Тогда можно определить условную плотность

$$p_{\xi|\eta=y}(x) = p(\xi = x \mid \eta = y) = \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)}$$

В этом случае можно определить условную вероятность следующим образом

$$P(\xi \in A \mid \eta = y) = \int_A p_{\xi|\eta=y}(x) dx = \int_A \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)} dx$$

В этом случае условное математическое ожидание определяется так

$$\mathbb{E}(\xi \mid \eta = y) = \int_{\mathbb{R}} x p_{\xi|\eta=y}(x) dx = \int_{\mathbb{R}} x \frac{p_{\xi,\eta}(x, y)}{p_\eta(y)} dx$$

Смысл условного распределения $P(\xi \in A \mid \eta = y)$ и в дискретном и в непрерывном случае одинаковый. Оно показывает, какова вероятность, что ξ покажет результат из множества A , при условии, что η показало значение y . Важно, что для дискретных случайных величин условное распределение тоже будет дискретным, а для непрерывных – непрерывным. Про условное математическое ожидание $\mathbb{E}(\xi \mid \eta = y)$ надо думать так. Оно показывает, какое в среднем значение будет показывать ξ среди тех попыток, когда η показывает y . Обратите внимание, что $g(y) = \mathbb{E}(\xi \mid \eta = y)$ можно рассматривать как случайную величину на \mathbb{R} с распределением η . Тогда ее математическое ожидание будет²

$$\mathbb{E}_y(g(y)) = \mathbb{E}(\mathbb{E}(\xi \mid \eta = y)) = \mathbb{E}(\xi)$$

¹Чтобы в этой формуле посчитать числитель, не достаточно знать распределение каждой из случайных величин. Надо полностью знать их совместное распределение. Однако, я не стал его вводить явно.

²Индекс y в математическом ожидании указывает по какой из случайных величин мы берем математическое ожидание. Это очень часто встречаемое обозначение особенно полезно во втором равенстве, чтобы понять, какого черта вообще происходит.

Если бы случайные величины ξ и η были бы независимы, то $\mathbb{E}(\xi \mid \eta = y) = \mathbb{E}(\xi)$. То есть мы варьируем значение η с помощью y и при этом среднее для ξ остается одним и тем же. Однако, если случайные величины зависимы, то при разных значениях η случайная величина ξ будет иметь разные средние $\mathbb{E}(\xi \mid \eta = y)$. Меняя y мы будем менять среднее для ξ при условии, что $\eta = y$. Но если мы возьмем и усредним все эти средние по η (с учетом меры для η) то мы в итоге получим обычное среднее для ξ . Именно эта последняя мысль и заключается в формуле для математического ожидания от условного ожидания. Еще на эту формулу можно смотреть как на интегрирование по частям.

Условное математическое ожидание относительно индикатора Если $\xi: \Omega \rightarrow \mathbb{R}$ – произвольная случайная величина и $A \subseteq \Omega$ – некоторое событие, то можно рассмотреть условное математическое ожидание ξ относительно χ_A , а именно функцию $\mathbb{E}(\xi \mid \chi_A = y)$. Математическое ожидание $\mathbb{E}(\xi \mid \chi_A = 1)$ совпадает с математическим ожиданием ξ , но по вероятностной мере $P_A(*) = P(* \mid A)$, то есть $\mathbb{E}(\xi \mid \chi_A = 1) = \int_{\Omega} \xi dP_A$. Таким образом это будет усреднение значений ξ по всем исходам, которые встречаются в A . Кроме того, $\mathbb{E}(\xi \mid \chi_A = 0)$ совпадает с $\mathbb{E}(\xi \mid \chi_{\bar{A}} = 1)$, а значит является усреднением по всем значениям ξ на элементарных исходах не из A .

Связь с кратными интегралами Предположим у нас есть непрерывный вектор (ξ, η) на каком-то вероятностном пространстве с плотностью $p_{\xi, \eta}(x, y)$. Пусть $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ – некоторая функция и мы хотим посчитать $\mathbb{E}f(\xi, \eta)$. Тогда это задача по вычислению следующего интеграла

$$\int_{\mathbb{R}^2} f(x, y) p_{\xi, \eta}(x, y) dx dy$$

Тогда его можно было бы посчитать по частям следующим образом

$$\int_{\mathbb{R}^2} f(x, y) p_{\xi, \eta}(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) p_{\xi, \eta}(x, y) dy \right) dx$$

Однако, у промежуточных интегралов нет никакого вероятностного смысла. Вместо этого, можно сначала сделать перенормировку.

$$\int_{\mathbb{R}^2} f(x, y) p_{\xi, \eta}(x, y) dx dy = \int_{\mathbb{R}^2} f(x, y) \frac{p_{\xi, \eta}(x, y)}{p_{\xi}(x)} p_{\xi}(x) dx dy$$

Конечно же она возможна там, где $p_{\xi}(x) \neq 0$, но давайте отложим сейчас этот формальный разговор.³ А вот теперь интегрирование по частям дает

$$\int_{\mathbb{R}^2} f(x, y) p_{\xi, \eta}(x, y) dx dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) \frac{p_{\xi, \eta}(x, y)}{p_{\xi}(x)} dy \right) p_{\xi}(x) dx$$

Тогда внутреннее выражение дает

$$\int_{\mathbb{R}} f(x, y) \frac{p_{\xi, \eta}(x, y)}{p_{\xi}(x)} dy = \int_{\mathbb{R}} f(x, y) p_{\eta \mid \xi=x}(y) dy = \mathbb{E}(f(\xi, \eta) \mid \xi = x) = \varphi(x)$$

И получаем некоторую функцию от x , потому что мы для каждого значения x считаем математическое ожидание $f(\xi, \eta) = f(x, \eta)$ только по той части Ω , где $\xi = x$. Формально последние слова не понятно, что значат, но интуитивно, мы усредняем значения η по линиям уровня ξ . И если представить совместную плотность $p_{\xi, \eta}(x, y)$, то внутренний кратный интеграл

$$\int_{\mathbb{R}} f(x, y) p_{\xi, \eta}(x, y) dy$$

³На самом деле надо сделать так: выкинуть из прямой кусок, где $p_{\xi}(x) = 0$ и писать интеграл только по оставшейся части. В этом случае совместная плотность тоже будет нулевой всюду если она была непрерывна и почти всюду в общем случае (что бы это ни значило).

Означает, что мы усредняем значение $f(x, y)$ по y на срезке, когда x – константа. Но для придания вероятностного смысла, мы бы хотели, чтобы $p_{\xi, \eta}(x_0, y)$ задавала вероятность для фиксированного x_0 , а потому можем ренормировать, поделив на $\int_{\mathbb{R}} p_{\xi, \eta}(x_0, y) dy = p_{\xi}(x_0)$. Потому по сути разницы с кратным интегралом нет. Если же посмотреть на наши вычисления дальше, то увидим, что

$$\mathbb{E}f(\xi, \eta) = \int_{\mathbb{R}} \mathbb{E}(f(\xi, \eta) \mid \xi = x) p_{\xi}(x) dx = \mathbb{E}_x(\mathbb{E}(f(\xi, \eta) \mid \xi = x))$$

То есть мы доинтегрируем $\varphi(x) = \mathbb{E}(f(\xi, \eta) \mid \xi = x)$ по оставшейся плотности для координаты x . А по координате x у нас жила ξ , вот по ее плотности и доинтегрируем. Последнее равенство – это жаргон для «удобства». Я пишу в кавычках, потому что это: во-первых, не удобно, во-вторых, математики так не делают. Так делают инженеры, а значит нам надо понимать их жаргон и что они имеют в виду, когда такое пишут.

Применение УМО

Давайте разберем следующую задачу.

Задача. Игральную кость с n гранями (и числами от 1 до n на этих гранях) подбрасывают до тех пор, пока сумма выпавших очков не станет больше либо равна n . Все грани кости выпадают с одинаковой вероятностью. Найдите математическое ожидание числа бросков.

Решение. Давайте введем следующие случайные величины: ξ_i – значение, выпавшее на i -ом бросании кости и η_k – минимальное количество бросков, которое нужно сделать, чтобы выпало хотя бы k очков суммарно. Тогда нам надо посчитать $\mathbb{E}(\eta_n)$. Давайте решим более общую задачу и посчитаем все числа $a_k = \mathbb{E}(\eta_k)$ при $k \leq n$. Заметим, что $\eta_1 = 1$, чтобы получить хотя бы одно очко достаточно кинуть кость 1 раз. Так же для удобства можно положить, что $\eta_0 = 0$, то есть чтобы получить 0 очков надо бросить кость 0 раз. Аналогично для отрицательных чисел $\eta_{-1} = 0$ и т.д.

Теперь посчитаем математическое ожидание с помощью условного математического ожидания.

$$\mathbb{E}(\eta_k) = \mathbb{E}_m(\mathbb{E}(\eta_k \mid \xi_1 = m)) = \mathbb{E}(\eta_k \mid \xi_1 = 1)P(\xi_1 = 1) + \dots + \mathbb{E}(\eta_k \mid \xi_1 = n)P(\xi_1 = n)$$

Мы знаем, что $P(\xi_1 = m) = 1/n$ для любого m . Теперь надо посчитать $\mathbb{E}(\eta_k \mid \xi_1 = m)$ для фиксированного m . Если $\xi_1 = m$ то мы уже сделали 1 бросание и теперь нам осталось набрать $k - m$ очков. То есть при условии, что $\xi_1 = m$ случайная величина η_k превращается в $1 + \eta_{k-m}$. При этом η_{k-m} будет величиной от ξ_2, \dots, ξ_k , а то есть не зависит от ξ_1 . Потому

$$\mathbb{E}(\eta_k \mid \xi_1 = m) = \mathbb{E}(1 + \eta_{k-m} \mid \xi_1 = m) = \mathbb{E}(1 + \eta_{k-m})$$

Таким образом получили, что

$$\mathbb{E}(\eta_k) = \frac{\mathbb{E}(1 + \eta_{k-1})}{n} + \dots + \frac{\mathbb{E}(1 + \eta_{k-n})}{n}$$

Так как мы считаем значения при $k \leq n$ и величины $\eta_k = 0$ при не положительных k , то последнюю формулу можно переписать так

$$\mathbb{E}(\eta_k) = \frac{\mathbb{E}(1 + \eta_{k-1})}{n} + \dots + \frac{\mathbb{E}(1 + \eta_1)}{n}$$

или, раскрыв все скобки, так

$$\mathbb{E}(\eta_k) = 1 + \frac{\mathbb{E}(\eta_{k-1})}{n} + \dots + \frac{\mathbb{E}(\eta_1)}{n}$$

То есть на нашу числовую последовательность у нас есть следующие соотношения

$$a_k = 1 + \frac{a_1 + \dots + a_{k-1}}{n}$$

При этом $a_1 = 1$. Вычтем из этого соотношения соотношение при $k - 1$, получим

$$a_k - a_{k-1} = \frac{a_{k-1}}{n} \Rightarrow a_k = a_{k-1} \left(1 + \frac{1}{n}\right) \Rightarrow a_k = a_1 \left(1 + \frac{1}{n}\right)^{k-1} = \left(1 + \frac{1}{n}\right)^{k-1}$$

Отсюда получаем, что

$$\mathbb{E}(\eta_n) = a_n = \left(1 + \frac{1}{n}\right)^{n-1}$$

Так как это число близко к e , то за два-три броска в среднем мы будем справляться. \square

Формулы Байеса и полной вероятности в непрерывном случае

Прежде чем формулировать формулы я хочу обратиться к жаргону. Когда у вас есть набор случайных величин ξ_1, \dots, ξ_n на Ω с совместной плотностью $p(x_1, \dots, x_n)$, то люди часто меняют вероятностное пространство (Ω, P) на новое (\mathbb{R}^n, P_ξ) . В этом случае наши случайные величины становятся просто координатными функциями. То есть вместо $\xi_1: \Omega \rightarrow \mathbb{R}$, мы можем изучать функцию $x_1: \mathbb{R}^n \rightarrow \mathbb{R}$ по правилу $(x_1, \dots, x_n) \mapsto x_1$. Ровно в этом и суть смены пространства, что ничего не изменится с вероятностной точки зрения, если мы хотим изучать только случайные величины ξ_1, \dots, ξ_n и ничего больше про Ω знать не хотим.

В этом случае люди начинают использовать для x_1, \dots, x_n вероятностные обозначения. Например, если я хочу написать условную плотность, скажем ξ_1 при условии ξ_2 , то я могу написать так

$$p(x_1 | x_2) = p_{\xi_1|\xi_2}(x_1, x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$

Напомню, что в этом случае

$$p(x_1) = \int_{\mathbb{R}} p(x_1, x_2) dx_2 \quad \text{и} \quad p(x_2) = \int_{\mathbb{R}} p(x_1, x_2) dx_1$$

Я предпочитаю думать про плотность, как про «правдоподобие». Потому тут написана формула для условного правдоподобия x_1 при условии x_2 . А значит по аналогии можно записать формулы Байеса

$$p(x_2 | x_1) = \frac{p(x_1 | x_2)p(x_2)}{p(x_1)}$$

Теперь давайте поговорим про аналог формулы полной вероятности, которая тоже должна писаться для плотности. По определению имеем

$$p(x_1) = \int_{\mathbb{R}} p(x_1, x_2) dx_2$$

Теперь давайте распишем плотность внутри через условную плотность, получим

$$p(x_1) = \int_{\mathbb{R}} \frac{p(x_1, x_2)}{p(x_2)} p(x_2) dx_2 = \int_{\mathbb{R}} p(x_1 | x_2) p(x_2) dx_2$$

На эту формулу можно смотреть так. События вида $\{\omega \in \Omega \mid \eta(\omega) = x_2\}$ – дают полную систему событий (они не пересекаются и покрывают все). Беда в том, что их более чем счетное множество и они все имеют нулевую вероятность (из-за непрерывности η). Но интегрирование решает эту проблему. Потому мы, считая «правдоподобие» $p(x_1)$ можем посчитать его через интегрирование⁴ условных «правдоподобий» относительно x_2 с весами в виде «правдоподобия» $p(x_2)$.

И последний аналог, который я хочу затронуть – «правдоподобие» совместного наступления событий. Проще говоря, как связана совместная плотность с условными, когда у вас много переменных. В этом случае я хочу уточнить определение

$$p(x_1, \dots, x_k \mid x_{k+1}, \dots, x_n) = \frac{p(x_1, \dots, x_k, x_{k+1}, \dots, x_n)}{p(x_{k+1}, \dots, x_n)}$$

Тогда смотрите, что получится

$$\begin{aligned} p(x_3, x_4) &= p(x_3 \mid x_4) p(x_4) \\ p(x_2, x_3, x_4) &= p(x_2 \mid x_3, x_4) p(x_3, x_4) = p(x_2 \mid x_3, x_4) p(x_3 \mid x_4) p(x_4) \\ p(x_1, x_2, x_3, x_4) &= p(x_1 \mid x_2, x_3, x_4) p(x_2 \mid x_3, x_4) p(x_3 \mid x_4) p(x_4) \end{aligned}$$

Замечание про общий подход Скажу по секрету, что есть общий подход к условному математическому ожиданию. Тогда можно определить случайную величину $\mathbb{E}(\xi \mid \eta)$, которая является усреднением ξ на всех событиях, выражаемых в терминах η . Потому сохраняется более общая формула

$$\mathbb{E}(\mathbb{E}(\xi \mid \eta) \chi_A) = \mathbb{E}(\xi \chi_A), \quad \text{для любого } A = \eta^{-1}(B), \quad \text{где } B \subseteq \mathbb{R}$$

⁴Вместо суммирования.

Плотность композиции

Плотность композиции для случайной величины Пусть у нас есть случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ на некотором вероятностном пространстве (Ω, P) . Предположим, что она имеет плотность $p_\xi(x)$. Пусть теперь задано отображение $f: \mathbb{R} \rightarrow \mathbb{R}$ и мы хотим найти распределение случайной величины $f(\xi)$. Обратим внимание на то, что в общем случае $f(\xi)$ не обязана иметь плотность, даже если ξ имела. Например, f может быть постоянной функцией $f(x) = a$, тогда $f(\xi)$ будет атомом в точке a веса 1. Потому нам нужны разумные условия на f . Чтобы не было атомов, надо потребовать, чтобы линии уровня f были нулевой вероятности относительно меры P_ξ . Но в этом случае вмешивается всякая экзотика (непрерывные меры без плотностей) и функциональный анализ для написания явных формул ортогопроекции в функциональных пространствах. Давайте предположим очень простое условие: $f: \mathbb{R} \rightarrow \mathbb{R}$ является строго монотонной гладкой функцией. В этом случае

$$F_{f(\xi)}(x) = P(f(\xi) \leq x) = P(\xi \leq f^{-1}(x)) = \int_{-\infty}^{f^{-1}(x)} p(t) dt$$

Теперь продифференцируем по x , чтобы получить плотность для $f(\xi)$, получим

$$p_{f(\xi)}(x) = \frac{p_\xi(f^{-1}(x))}{f'(f^{-1}(x))} \quad \text{или} \quad p_\xi(y) = p_{f(\xi)}(f(y))f'(y)$$

Эту формулу можно получить еще следующим образом. Для любого подмножества $A \subseteq \mathbb{R}$ должно быть выполнено $P_{f(\xi)}(A) = P_\xi(f^{-1}(A))$. Если обе меры имеют плотность, то

$$\int_A p_{f(\xi)}(y) dy = \int_{f^{-1}(A)} p_\xi(x) dx$$

Теперь сделаем замену переменных $y = f(x)$ в левой части, тогда получим

$$\int_{f^{-1}(A)} p_{f(\xi)}(f(x))f'(x) dx = \int_{f^{-1}(A)} p_\xi(x) dx$$

Так как f биективно и A пробегает все подмножества, то это означает, что $f^{-1}(A)$ пробегает все подмножества. А отсюда следует, что подинтегральные функции равны, что и дает одну из формул выше.

Плотность композиции для случайного вектора Пусть $\bar{\xi}: \Omega \rightarrow \mathbb{R}^n$ – случайный вектор с плотностью $p_{\bar{\xi}}(x_1, \dots, x_n)$ и пусть $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^n$ – обратимое линейное отображение заданное по правилу $\phi(x) = Ax$, где A – обратимая квадратная матрица размера n . Тогда мы хотим найти плотность $A\bar{\xi}$. Пойдем вторым путем выше. Пусть $X \subseteq \mathbb{R}^n$ – произвольное множество, тогда $P_{A\bar{\xi}}(X) = P_{\bar{\xi}}(A^{-1}(X))$. Так как обе части имеют плотность, имеем

$$\int_X p_{A\bar{\xi}}(y_1, \dots, y_n) dy_1 \dots dy_n = \int_{A^{-1}X} p_{\bar{\xi}}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Теперь сделаем замену $y = Ax$ и получим

$$\int_{A^{-1}X} p_{A\bar{\xi}}(Ax) |\det(A)| dx_1 \dots dx_n = \int_{A^{-1}X} p_{\bar{\xi}}(x_1, \dots, x_n) dx_1 \dots dx_n$$

Аналогично из произвольности $X \subseteq \mathbb{R}^n$ и биективности ϕ мы получаем, что

$$p_{A\bar{\xi}}(Ax) |\det(A)| = p_{\bar{\xi}}(x)$$

для любого $x \in \mathbb{R}^n$, в частности

$$p_{A\bar{\xi}}(y) = \frac{p_{\bar{\xi}}(A^{-1}y)}{|\det(A)|}$$

Распределение суммы независимых случайных величин

Пусть на некотором вероятностном пространстве (Ω, P) даны две независимые случайные величины ξ и η . Предположим, что мы знаем распределение каждой из них, то есть знаем меры P_ξ и P_η на прямой \mathbb{R} . Наша задача найти меру для суммы $P_{\xi+\eta}$.

Дискретный случай. Если обе случайные величины ξ и η дискретны, то и сумма их будет дискретна. Пусть исходные случайные величины заданы следующим образом

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots \\ p_1 & p_2 & \dots & p_n & \dots \end{cases} \quad \text{и} \quad \eta \sim \begin{cases} b_1 & b_2 & \dots & b_n & \dots \\ q_1 & q_2 & \dots & q_n & \dots \end{cases}$$

Тогда все возможные значения для суммы $\xi + \eta$ имеют вид $a_i + b_j$. А соответствующая мера считается так

$$P(\xi + \eta = x) = \sum_{a_i + b_j = x} p_i q_j$$

Непрерывный случай. Свертка. Теперь предположим, что ξ непрерывна и задана плотностью $p(x)$, а η имеет плотность $q(y)$. Утверждается, что случайная величина $\xi + \eta$ так же будет непрерывной с плотностью

$$p_{\xi+\eta}(x) = \int_{\mathbb{R}} p(t)q(x-t) dt = \int_{\mathbb{R}} p(x-t)q(t) dt$$

Действительно, давайте для начала вычислим функцию распределения

$$F_{\xi+\eta}(x) = P(\xi + \eta \leq x) = \int_{u+v \leq x} p(u)q(v) dudv$$

Теперь можно сделать две замены переменных (одну для получения первого выражения, другую для – второго).

$$\begin{cases} s = u + v \\ t = u \end{cases} \quad \text{или} \quad \begin{cases} s = u + v \\ t = v \end{cases}$$

В матричном виде эти замены выглядят

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v \\ u \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

Обратная замена

$$\begin{pmatrix} v \\ u \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix}$$

Теперь пользуемся формулой

$$dudv = |\det(C)|dsdt \quad \text{где} \quad \begin{pmatrix} u \\ v \end{pmatrix} = C \begin{pmatrix} s \\ t \end{pmatrix}$$

Получаем

$$F_{\xi+\eta}(x) = \int_{s \leq x} p(t)q(s-t) dsdt = \int_{-\infty}^x \int_{-\infty}^{\infty} p(t)q(s-t) dt ds = \int_{-\infty}^x \left(\int_{-\infty}^{\infty} p(t)q(s-t) dt \right) ds$$

Аналогично для второй замены

$$F_{\xi+\eta}(x) = \int_{s \leq x} p(s-t)q(t) dsdt = \int_{-\infty}^x \int_{-\infty}^{\infty} p(s-t)q(t) dt ds = \int_{-\infty}^x \left(\int_{-\infty}^{\infty} p(s-t)q(t) dt \right) ds$$

Так как плотность – это производная от функции распределения, то нужный ответ получается дифференцированием по x .

Интеграл вида

$$\int_{\mathbb{R}} p(t)q(x-t) dt = \int_{\mathbb{R}} p(x-t)q(t) dt$$

Называется сверткой функций p и q . То есть мы проверили, что для независимых непрерывных величин, плотность суммы есть свертка плотностей слагаемых.

Характеристические функции

Вы когда-нибудь слышали про преобразование Фурье? Отлично, сейчас еще раз услышите. Есть общая концепция того, что такое преобразование Фурье и в рамках этой концепции можно сделать преобразование Фурье для случайной величины, а точнее для ее меры на прямой. Зачем это нужно? Есть несколько причин для его введения. Оказывается, что изучение разных видов сходимости можно проводить с помощью вычисления интегралов от всех возможных непрерывных функций. Однако, достаточно взять только «хорошие» функции. Это некое хорошее плотное множество с замечательными алгебраическими свойствами. Самое важное в этом семействе функций, что их достаточно, чтобы прощупать любое свойство меры, а все семейство можно параметризовать вещественным числом и свести изучение сложных сходимостей к свойствам одной функции, которая и называется характеристической. С другой стороны, само преобразование Фурье нетривиальным образом перетасовывает свойства объектов, что тоже может быть очень полезно.

Пусть $\xi: \Omega \rightarrow \mathbb{R}$ – некоторая случайная величина. Тогда характеристической функцией ξ называется

$$\varphi_{\xi}(t) = \mathbb{E}e^{it\xi}$$

Обратите внимание, что $\varphi: \mathbb{R} \rightarrow \mathbb{C}$ – комплекснозначная функция. Здесь $e^{i\phi} = \cos(\phi) + i\sin(\phi)$ по определению.

Характеристическая функция и функция распределения Если мы знаем $F_{\xi}(x)$ функцию распределения для ξ , то мы можем считать мат ожидания от любых функций от ξ с помощью интеграла по прямой следующим образом

$$\varphi_{\xi}(t) = \mathbb{E}e^{it\xi} = \int_{\mathbb{R}} e^{itx} dF_{\xi}(x)$$

Таким образом случайная величина $e^{it\xi}$ соответствует тригонометрической функции e^{itx} на прямой после смены Ω на \mathbb{R} . Теперь важное замечание: множество функций e^{itx} (здесь t – параметр, а x – аргумент функции) является «плотным» среди всех непрерывных функций, в том смысле, что мы можем на любом отрезке равномерно приблизить любую непрерывную функцию с помощью линейной комбинации тригонометрических функций. Это были хорошие свойства плотности. Если же говорить про алгебраические свойства, то обратите внимание, что наше семейство $\{e^{itx}\}_{t \in \mathbb{R}}$ замкнуто по умножению (произведение функций такого вида, тоже функция такого вида) и переводит сумму аргументов в произведение значений функций. Это два основных технических механизма, которые используются для работы с характеристическими функциями.

Дискретный случай Пусть у нас задана дискретная случайная величина

$$\xi \sim \begin{cases} a_1 & a_2 & \dots & a_n & \dots & \in \mathbb{R} \\ p_1 & p_2 & \dots & p_n & \dots & \in [0, 1] \end{cases}$$

Тогда хар функция будет считаться так

$$\varphi_{\xi}(t) = \sum_k e^{ita_k} p_k$$

Непрерывный случай Пусть у нас задана непрерывная случайная величина ξ с плотностью $p(x)$. Тогда хар функция считается так

$$\varphi_{\xi}(t) = \int_{\mathbb{R}} e^{itx} p(x) dx$$

То есть это классическое преобразование Фурье от плотности распределения.

Примеры характеристических функций

1. Самый простой пример – не случайная случайная величина! $\xi(\omega) = a \in \mathbb{R}$ для любого $\omega \in \Omega$. Тогда

$$\varphi_{\xi}(t) = e^{ita}$$

То есть для константы хар функция является гармоникой с частотой a .

2. Для распределения Бернулли

$$\xi \sim \begin{cases} 1 & 0 & \in \mathbb{R} \\ p & 1-p & \in [0, 1] \end{cases}$$

Хар функция будет

$$\varphi_{\xi}(t) = 1 - p + pe^{it}$$

3. Пусть у нас дано распределение Пуассона $P(\lambda)$, при этом $\lambda > 0$. Эта случайная величина задана таблицей

$$\xi \sim \begin{cases} 0 & 1 & \dots & k & \dots \\ e^{-\lambda} & e^{-\lambda}\lambda & \dots & \frac{e^{-\lambda}\lambda^k}{k!} & \dots \end{cases}$$

Тогда хар функция будет

$$\varphi_{\xi}(t) = e^{\lambda(e^{it}-1)}$$

4. Для экспоненциального распределения с плотностью $p(x) = \lambda e^{-\lambda x} \theta(x)$ при $\lambda > 0$, хар функция будет

$$\varphi_{\xi}(t) = \frac{1}{1 - \frac{it}{\lambda}}$$

5. Для нормального распределения $\mathcal{N}(a, \sigma^2)$, то есть непрерывной случайной величины с плотностью

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

хар функция будет иметь вид

$$\varphi_{\xi}(t) = e^{iat - \frac{\sigma^2 t^2}{2}}$$

Свойства характеристических функций

t.me/postypashki_old/1229

t.me/postypashki_old/1229

t.me/postypashki_old/1229

1. Хар функция является равномерно непрерывной на всей прямой. Это значит, что для любого ε можно найти δ , что как только $|t_1 - t_2| < \delta$, то $|\varphi_{\xi}(t_1) - \varphi_{\xi}(t_2)| < \varepsilon$. По простому, ваша функция не просто непрерывно, но ее «не колбасит вверх и вниз как попало». Она одинаково колеблется вне зависимости от куса прямой, где вы ее рассмотрели. Это важное техническое свойство для доказательства разных фактов, но не особенно полезное в противном случае.
2. Мы кое-что можем сказать о значениях функции. Например, $\varphi(0) = 1$ или даже, что $|\varphi(t)| \leq 1$ на всей прямой. Кроме того, $\varphi(-t) = \overline{\varphi(t)}$.
3. Хар функция однозначно задает распределение. Если две случайные величины имеют одну и ту же хар функцию, то они имеют одинаковую меру на прямой.
4. Через хар функцию можно считать моменты случайной величины как производные

$$\mathbb{E}(\xi^n) = i^{-n} \varphi_{\xi}^{(n)}(0)$$

Это свойство берется из дифференцирования интеграла по параметру.

5. Для независимых случайных величин ξ_1, \dots, ξ_n выполнено равенство

$$\varphi_{\xi_1 + \dots + \xi_n}(t) = \varphi_{\xi_1}(t) \cdot \dots \cdot \varphi_{\xi_n}(t)$$

Опять же, это еще один плюс в копилку хороших свойств независимости. Зачем это нужно, станет понятно ниже, когда будем обсуждать сходимость по распределению, а потом и ЦПТ в будущем.

Характеристическая функция случайного вектора Пусть $\bar{\xi}: \Omega \rightarrow \mathbb{R}^n$ – случайный вектор со случайными координатами ξ_1, \dots, ξ_n . Тогда для него так же определена характеристическая функция, а именно

$$\varphi_{\bar{\xi}}(t) = \mathbb{E}e^{i(t, \bar{\xi})} = \mathbb{E}e^{it^t \bar{\xi}}$$

Здесь $t \in \mathbb{R}^n$ – вектор, $(t, \bar{\xi}) = t^t \bar{\xi} = t_1 \xi_1 + \dots + t_n \xi_n$ означает стандартное скалярное произведение. Или явно, как функция от нескольких переменных

$$\varphi_{\xi_1, \dots, \xi_n}(t_1, \dots, t_n) = \mathbb{E}e^{i(t_1 \xi_1 + \dots + t_n \xi_n)}$$

Хар функция от вектора так же однозначно определяет распределение вектора и обладает кучей замечательных свойств. Но самое главное – оно будет отвечать за сходимость по распределению, о которой речь пойдет ниже.

В качестве примера давайте рассмотрим нормальное многомерное распределение, то есть случайный вектор $\bar{\xi}$ с плотностью

$$p_{\xi}(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-a)^t \Sigma^{-1}(x-a)}$$

где $a \in \mathbb{R}^n$ и $\Sigma \in M_n(\mathbb{R})$ – симметричная положительно определенная матрица (с положительным спектром). Тогда хар функция для нормального распределения будет иметь вид

$$\varphi_{\bar{\xi}}(t) = e^{i(t, a) - \frac{1}{2}(t, \Sigma t)} \quad \text{где } t \in \mathbb{R}^n$$

И вот это определение нормального распределения работает даже для вырожденной матрицы Σ .

Задача математической статистики

В теории вероятностей, изучая какой-нибудь черный ящик, всегда предполагается, что за ним стоит некоторое вероятностное пространство (Ω, P) . Кроме того, обычно к черному ящику цепляют какой-нибудь измерительный прибор, то есть рассматривают случайную величину $\xi: \Omega \rightarrow \mathbb{R}$. В этом случае главный интерес для изучения – это мера P_{ξ} , которая полностью описывает все свойства прибора ξ на нашем черном ящике. Типичная задача теории вероятности звучит так: мы знаем меру P_{ξ} , предсказать, как будет отвечать прибор на наши запросы. В математической статистике по сути решается обратная задача. Мы не знаем как устроено (Ω, P) , мы понятия не имеем, как выглядит мера P_{ξ} , но мы знаем, что прибор уже успел дать несколько ответов $x_1, \dots, x_n \in \mathbb{R}$. Задача – вытащить как можно больше информации про неизвестные компоненты черного ящика. Как не трудно догадаться, раз мы опрашиваем не сам черный ящик, а лишь висящий на нем прибор, то (Ω, P) находится вне нашей досягаемости. Самое лучше, что мы можем попытаться узнать – это как устроена случайная величина $\xi: \Omega \rightarrow \mathbb{R}$. Но раз мы не знаем, что из себя представляет Ω , лучшее, что мы можем узнать – это как устроена мера P_{ξ} . С другой стороны, если мы знаем вероятностную меру P_{ξ} для случайной величины ξ , то мы знаем про нее все. Таким образом общая задача математической статистики выглядит так.

Задача Пусть на некотором вероятностном пространстве (Ω, P) задана случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ и пусть заданы измерения случайной величины $x_1, \dots, x_n \in \mathbb{R}$, мы хотим восстановить P_{ξ} или какие-то другие характеристики ξ .

Как надо думать про измерения? Когда мы измеряли, скажем, значение x_1 , наш черный ящик находился в каком-то состоянии ω_1 , потому $x_1 = \xi(\omega_1)$. Потом, когда мы измеряли x_2 , наш черный ящик находился в состоянии ω_2 , то есть $x_2 = \xi(\omega_2)$ и т.д. Таким образом, можно думать, что x_1, \dots, x_n – это последовательность $\xi(\omega_1), \dots, \xi(\omega_n)$. Значит про эту задачу можно думать, как про задачу восстановления функции $\xi: \Omega \rightarrow \mathbb{R}$ по значению в конечном числе точек. Однако, такая модель не самая лучшая. Основная проблема в том, что мы не знаем, как именно прибор переключается между состояниями $\omega_1, \dots, \omega_n$, где именно в пространстве Ω они сгруппированы. Кроме того, не ясно, влияет ли первое измерение на второе и если влияет, то как учитывать подобное в данной модели? По-хорошему эту модель надо дополнить некоторой информацией о переходе из одного состояния в другое при замерах. Но это автоматически означает, что нам надо знать какие есть состояния у черного ящика, чтобы описать подобный механизм. А такая информация может быть просто недоступна нам. Оказывается, можно закодировать эту информацию по-другому, так что необходимость знать Ω полностью пропадает.

Прежде чем перейти к следующей модели, давайте проделаем такой эксперимент. Будем думать, что сами элементарные исходы $\omega_1, \dots, \omega_n$ являются случайными. То есть у нас существует какое-то другое вероятностное пространство (Ω^*, P^*) и $\bar{\omega}_i: \Omega^* \rightarrow \Omega$ такие отображения, что $\bar{\omega}_i$ сохраняют вероятность, то есть для любого $A \subseteq \Omega$ выполнено $P(A) = P^*(\bar{\omega}_i^{-1}(A))$. Тогда мы будем считать, что наши $\omega_i \in \Omega$ получились так $\omega_1 = \bar{\omega}_1(s), \dots, \omega_n = \bar{\omega}_n(s)$ для некоторого $s \in \Omega^*$. В этом случае мы можем рассмотреть случайные величины $\xi_1 = \xi \circ \bar{\omega}_1, \dots, \xi_n = \xi \circ \bar{\omega}_n$. Кроме того, мы видим, что $x_i = \xi_i(s)$. То есть вместо того, чтобы изучать один прибор ξ и как-то мучить исходы ω_i , мы выбор этих исходов записали внутри ξ и получили несколько копий⁵ нашего прибора ξ_1, \dots, ξ_n . И все семплы получены не последовательным измерением ξ в разных точках, а единовременным измерением копий нашего прибора в одной точке. Так как теперь ξ_1, \dots, ξ_n – это случайные величины, то мы можем говорить об их зависимости или независимости. Про пространство (Ω^*, P^*) нужно думать как про «мироздание». То есть мироздание находилось в состоянии $s \in \Omega^*$ и мы замерили на этом состоянии «мироздания» копии нашего прибора ξ_1, \dots, ξ_n и получили наши семплы. Сразу скажу, если не удалось осознать и переварить то это потому, что тут было много дичи. Но здесь я немного намекнул на то, как именно возникает несколько копий приборов вместо одного. Ниже, описывая вторую версию модели, я постараюсь подойти к этому вопросу с другой стороны и, надеюсь, что в сочетании эти два объяснения помогут лучше осознать происходящее.

Основная модель математической статистики

Как и в наивной модели, мы должны предположить, что у нас есть некоторый изучаемый черный ящик, за которым стоит некоторое неизвестное нам вероятностное пространство (Ω, P) . Разница будет в том, как мы будем моделировать наш измеряющий прибор.

Предположим, что мы взяли некоторый прибор $\xi: \Omega \rightarrow \mathbb{R}$ и начинаем делать измерения. Сначала сделали первое измерение, получили x_1 . Потом взяли тот же самый эксперимент повторили с нуля и получается независимо сделали второе измерение и получили x_2 и т.д. На этот процесс можно смотреть вот каким образом. Мы можем использовать ξ как эталонный прибор, а для n независимых измерений сделать n его независимых копий: $\xi_1, \dots, \xi_n: \Omega \rightarrow \mathbb{R}$. Тогда вместо того, чтобы делать n измерений подряд с помощью одного прибора, мы просто разом сделаем n измерений с помощью n независимых приборов. То есть мы можем считать, что

$$x_1 = \xi_1(\omega), x_2 = \xi_2(\omega), \dots, x_n = \xi_n(\omega)$$

В таком подходе мы думаем, что за каждое измерение отвечает своя случайная величина. Но все эти величины являются копиями одной величины ξ . Математически это означает, что $P_{\xi_i} = P_\xi$. Заметим, что мы при описании процесса предполагали, что будем делать измерения независимыми друг от друга. Это означает, что все случайные величины ξ_1, \dots, ξ_n должны быть в совокупности независимыми. Однако теперь, когда за каждое измерение отвечает своя случайная величина, мы можем (если захотим) отбросить условие независимости и считать, что они могут быть зависимыми. Кроме того за свойство зависимости или независимости будет отвечать совместное распределение $P_{(\xi_1, \dots, \xi_n)}$ в \mathbb{R}^n . Таким образом мы избавились от необходимости знать Ω , чтобы понять, какая взаимосвязь между измерениями. Я бы еще добавил, что предположение о независимости измерений часто является очень естественным и потому популярно в приложениях. Один из самых популярных способов описания зависимостей являются Марковские цепи, которых мы касаться не будем, но слышать эти слова надо. Давайте теперь формально опишем задачу и основную модель математической статистики.

Задача Пусть на некотором вероятностном пространстве (Ω, P) заданы случайные величины ξ_1, \dots, ξ_n такие, что они одинаково распределены (то есть $P_{\xi_i} = P_{\xi_j}$) и, например, независимы. Пусть нам задан результат измерений

$$x_1 = \xi_1(\omega), \dots, x_n = \xi_n(\omega) \in \mathbb{R}$$

мы хотим восстановить $P_\xi = P_{\xi_i}$ на прямой или какие-то численные характеристики этой вероятностной меры.

Замечания

⁵Тот факт, что ξ_i именно копии, то есть распределены так же, как ξ следует из того, что $\bar{\omega}_i$ сохраняют меру. Это более или менее по определению, но я не собираюсь копаться в этом формализме и проверять что-либо, я лишь хотел пояснить откуда берутся основания для такого заявления и как понимать слово «копия».

1. Чтобы описать семейство одинаково распределенных случайных величин обычно делают так. Говорят, пусть ξ – некоторая случайная величина с интересующем нас распределением P_ξ . А после этого говорят, что пусть нам заданы случайные величины ξ_1, \dots, ξ_n распределенные так же как и ξ , то есть $P_{\xi_i} = P_\xi$. Обычно этот факт пишут так $\xi_i \sim \xi$, читается « ξ_i распределена так же как ξ ». И тогда говорят уже про восстановление распределения для ξ или характеристик распределения для ξ .
2. Мы во всех задачах (но не во всех) в дальнейшем будем предполагать, что случайные величины ξ_1, \dots, ξ_n независимы в совокупности. Это одна из самых популярных гипотез и она очень сильно упрощает рассуждения, мы же не хотим себя мучить.
3. В подобных задачах, часто предполагается из каких-нибудь теоретических соображений, что ξ вообще говоря распределена не как попало, а подчиняется какому-нибудь закону, у которого нам надо восстановить параметры. Например, может быть известно, что ξ – дискретная случайная величина, принимающая 0 и 1, но мы не знаем с какими вероятностями и их надо восстановить. Или может быть известно, что ξ – непрерывная случайная величина и ее плотность имеет вид $p(x, \theta_1, \dots, \theta_n)$, где θ_i – какие-то настраиваемые параметры, и нам надо установить какие эти параметры были на самом деле. Скажем, может оказаться, что ξ является нормально распределенной случайной величиной с неизвестными параметрами математического ожидания и дисперсии и нам надо их определить.

Оценки

Пусть ξ – некоторая случайная величина и $x_1, \dots, x_n \in \mathbb{R}$ – ее независимые измерения. Мы хотим восстановить математическое ожидание и дисперсию случайной величины ξ . Для того, чтобы решить эту задачу, нам надо написать какую-нибудь формулу $\mathbb{E}\xi = \phi(x_1, \dots, x_n)$. Вообще говоря, это может быть абсолютно любое выражение от переменных x_1, \dots, x_n . Подобная формула $\phi(x_1, \dots, x_n)$ называется оценкой (в данном случае оценкой математического ожидания). Однако, если мы напишем совсем дурацкую формулу, например, $\phi(x_1, \dots, x_n) = 0$, то получим дурацкий ответ, ничего не говорящий о нашей случайной величине, а хотелось бы что-то про нее да узнать. Для этого есть несколько характеристик, которые можно проверить у оценки. Давайте разберем все эти премудрости на примере математического ожидания и дисперсии.

Выборочное математическое ожидание Самая популярная формула для оценки математического ожидания – среднееарифметическое из измеренных значений. А именно

$$\bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Чтобы понять на сколько это хорошая формула, давайте вспомним, что в рамках нашей модели, мы считаем, что

$$x_1 = \xi_1(\omega), \dots, x_n = \xi_n(\omega)$$

Тогда наша оценка считает

$$\eta(\omega) = \frac{\xi_1(\omega) + \dots + \xi_n(\omega)}{n}$$

Но мы не знаем, в какой именно точке ω мы были. Потому вообще говоря, то что мы получили будет случайная величина

$$\eta = \frac{\xi_1 + \dots + \xi_n}{n}$$

Несмещенность Первый важный вопрос такой: а попадаем ли мы с помощью этой оценки в математическое ожидание хотя бы в среднем? То есть верно ли, что $\mathbb{E}\eta = \mathbb{E}\xi$? Если посчитать

$$\mathbb{E}\eta = \mathbb{E} \left(\frac{\xi_1 + \dots + \xi_n}{n} \right) = \frac{\mathbb{E}\xi_1 + \dots + \mathbb{E}\xi_n}{n} = \frac{\mathbb{E}\xi + \dots + \mathbb{E}\xi}{n} = \mathbb{E}\xi$$

Подобная характеристика называется несмещенностью.

Состоятельность Второй вопрос, который можем задать такой: а приближаемся ли мы к ответу с увеличением n хотя бы в среднем? Как это интерпретировать. Давайте вспомним, что η – случайная величина сосредоточенная вокруг $\mathbb{E}\eta$. А величину разброса η вокруг ее мат ожидания можно оценить с помощью среднеквадратичного отклонения (или дисперсии). Потому нас интересует, падает ли дисперсия η с увеличением n . Давайте проверим

$$\mathbb{D}\eta = \mathbb{D}\left(\frac{\xi_1 + \dots + \xi_n}{n}\right) = \frac{\mathbb{D}(\xi_1 + \dots + \xi_n)}{n^2} = \frac{\mathbb{D}\xi_1 + \dots + \mathbb{D}\xi_n}{n^2} = \frac{\mathbb{D}\xi}{n}$$

Мы видим, что дисперсия с увеличением n стремится к нулю. Такая характеристика называется состоятельностью оценки.

Семинар 4

Задачи:

1. Пусть ξ, η – независимые случайные величины имеющие показательное распределение, то есть

$$p_{\xi}(x) = p_{\eta}(x) = \lambda e^{-\lambda x} \theta(x)$$

где

$$\theta(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Найти следующие плотности

$$p_{\xi, \xi+\eta}(x, y) \quad \text{и} \quad p_{\xi|\xi+\eta=z}(x)$$

Важно: для совместной плотности укажите область в которой она равна нулю.

2. Используя характеристические функции и их свойства, найти распределение суммы двух независимых нормальных случайных величин с параметрами (a_1, σ_1^2) и (a_2, σ_2^2) .
3. Пусть ξ и η независимые случайные величины с распределением $\mathcal{N}(0, 1)$. Найти плотность величины $\chi = \xi^2 + \eta^2$.
4. Случайные величины X и Y независимы. X имеет распределение Лапласа с плотностью $\frac{1}{2}e^{-|x|}$, а Y – равномерное на отрезке $[1, 2]$.
- (а) Найдите плотность распределения случайной величины $-2Y$.
- (б) Найдите плотность распределения случайной величины $X - 2Y$.
5. Игрок приходит в казино с k монетами. Каждый кон он выигрывает с вероятностью p и получает одну монету и проигрывает с вероятностью $q = 1 - p$ и отдает одну монету. Игра заканчивается если игрок потеряет все монеты или если он наберет предельную сумму выигрыша N . Найдите среднюю продолжительность игры в зависимости от k .

Теория вероятности

Дима Трушин

Семинар 5

Виды сходимости

Когда мы изучаем оценки вида $\phi(x_1, \dots, x_n)$ то обычно у нас есть оценка для любого натурального n (или как минимум сколь угодно большого). Тогда интересным является такой вопрос: а к чему стремится выражение $\phi(\xi_1, \dots, \xi_n)$ при $n \rightarrow \infty$? Чтобы на него ответить, для начала надо понять, а в каком смысле вообще случайные величины могут стремиться друг к другу. Оказывается, что существует несколько разных видов сходимости на случайных величинах, о них и пойдет сейчас речь.

Случайные величины – это функции на пространстве Ω . Потому если мы говорим о сходимости случайных величин, мы просто говорим о сходимости функций. Давайте перечислим самые популярные виды сходимости на функциях. Пусть $\xi, \xi_n: \Omega \rightarrow \mathbb{R}$ – случайные величины.

1. **Равномерная сходимость.** Последовательность ξ_n равномерно сходится к ξ , пишут $\xi_n \rightrightarrows \xi$, при $n \rightarrow \infty$, если $\sup_{\omega \in \Omega} |\xi_n(\omega) - \xi(\omega)| \rightarrow 0$.

Это значит, что во всех элементарных исходах ξ_n стремится к ξ с некой гарантированной общей скоростью. Это самый сильный вид сходимости и одновременно самый бессмысленный для случайных величин, ибо вероятностное пространство Ω вообще говоря определено с точностью до события вероятности 0. Подробно об этом ниже.

2. **Поточечная сходимость (сходимость всюду).** Последовательность ξ_n сходится к ξ поточечно (всюду), пишут $\xi_n \rightarrow \xi$, если для любого $\omega \in \Omega$ последовательность $\xi_n(\omega)$ сходится к $\xi(\omega)$.

Это самый простой вид сходимости и вряд ли нуждается в дополнительном пояснении. Мы просто хотим на любом исходе сходиться. Опять же не сильно полезный для случайных величин.

3. **Сходимость почти всюду (почти наверное).** Последовательность ξ_n сходится к ξ почти всюду (почти наверное), пишут $\xi_n \xrightarrow{\text{a.s.}} \xi$, если существует подмножество $\Omega_0 \subseteq \Omega$ такое, что $P(\Omega_0) = 1$ и для любого $\omega \in \Omega_0$ последовательность $\xi_n(\omega)$ сходится к $\xi(\omega)$.

Это самый популярный вид сходимости. Дело в том, что в теории вероятности все определено с точностью до события вероятности ноль. Мы можем изменить вероятностное пространство на множество вероятности ноль и ничего не изменится. Потому формально мы не можем гарантировать ничего в отдельно взятом исходе. Потому поточечная сходимость (как и равномерная) не имеют особенного смысла для случайных величин. Чтобы исправить ситуацию, пользуются сходимостью почти всюду (или как принято в теории вероятностей почти наверное). Эта та же самая поточечная сходимость, но только на множестве вероятности 1. То есть последовательность ξ_n может не сходиться к ξ но только вероятность такого события равна нулю, формально $P(\{\omega \mid \xi_n(\omega) \not\rightarrow \xi(\omega)\}) = 0$.

4. **Сходимость в среднем.** Фиксируем число $1 < p < \infty$. Тогда ξ_n сходится в $L_p(\Omega)$ к ξ , пишут $\xi_n \xrightarrow{L_p} \xi$, если $\mathbb{E}((\xi_n - \xi)^p) = \int_{\Omega} (\xi_n - \xi)^p dP \rightarrow 0$. Самые популярные сходимости в среднем: сходимость в среднем квадратичном, то есть в $L_2(\Omega)$ и сходимость в $L_1(\Omega)$.

Про сходимость в среднем можно думать так: среднее от модуля разности (может быть в какой-то степени p) от случайных величин идет к нулю. То есть площадь между графиками функций ξ_n и ξ идет к нулю. То есть мы может быть и не сходимся поточечно, но зазор между графиками уменьшается.

5. **Сходимость по мере (по вероятности).** Последовательность ξ_n сходится к ξ по мере (по вероятности), пишут $\xi_n \xrightarrow{P} \xi$, если для любого $\varepsilon > 0$ последовательность $P(\{\omega \mid |\xi_n(\omega) - \xi(\omega)| > \varepsilon\})$ сходится к нулю.

Неформально это означает, что мера тех точек, в которых ξ_n дальше от ξ на ε стремится к нулю. То есть как только n все больше и больше, у нас все больше и больше точек, в которых мы все ближе и ближе к пределу.

Сходимость по распределению. Кроме того, каждая случайная величина $\xi: \Omega \rightarrow \mathbb{R}$ дает меру P_ξ на прямой \mathbb{R} . Потому можно говорить о сходимости мер. Есть такая замечательная наука – функциональный анализ. Она придумала как можно проверять сходимость мер друг к другу. Этих видов сходимостей тоже очень много, но мы остановимся только на одной из них: *-слабая сходимость (сходимость по распределению).

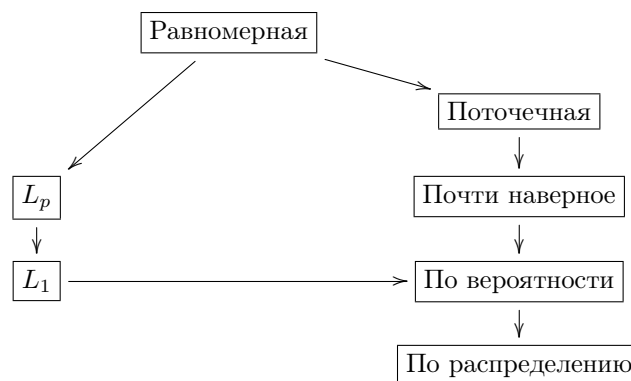
Пусть ξ_n, ξ – случайные величины. Следующие условия эквивалентны:

1. Для любой непрерывной ограниченной функции $f: \mathbb{R} \rightarrow \mathbb{R}$ верно, что $\mathbb{E}(f(\xi_n)) \rightarrow \mathbb{E}(f(\xi))$.
2. Для любой точки $x \in \mathbb{R}$, в которой функция распределения $F_\xi(x)$ непрерывна, верно, что $F_{\xi_n}(x)$ сходится к $F_\xi(x)$.
3. Для любого отрезка $[a, b]$ такого, что $P(\xi = a) = P(\xi = b) = 0$, верно $P(\xi_n \in [a, b]) \rightarrow P(\xi \in [a, b])$.
4. Хар функция ξ_n поточечно сходится к хар функции ξ , то есть $\varphi_{\xi_n}(t) \rightarrow \varphi_\xi(t)$ для любого $t \in \mathbb{R}$.

В этом случае будем говорить, что последовательность ξ_n *-слабо сходится к ξ или что то же самое сходится по распределению и будем писать $\xi_n \xrightarrow{D} \xi$. Важно понимать, что предел по распределению не единственный! В том смысле, что случайная величина ξ определена не однозначно, а вот ее распределение P_ξ на прямой будет определено однозначно.

Обратите внимание, что вообще говоря, проверять сходимость по распределению – то еще удовольствие. Самый важный случай, когда такая сходимость возникает – это Центральная предельная теорема. Есть много разных ее аналогов, но это обычно явления, которые говорят, что если у вас много случайных независимых явлений распределенных одинаково (или очень похоже), то некоторые их характеристики становятся распределены по весьма конкретным законам, независимо от распределения исходных явлений. Кроме ЦПТ можно вспомнить распределение Уишарта для собственных значений случайных матриц. Так вот, чтобы подобрать к таким теоремам, люди пользуются характеристическими функциями. Как показать, что ξ_n сходится по распределению, а если сходится, то к какому распределению? Очень просто, берете $\varphi_{\xi_n}(t)$ и смотрите на ее поточечный предел. Если его нет хотя бы для одной точки, то все, значит по распределению не сходится. А если есть и он совпал с $\varphi_\xi(t)$ для некоторой ξ , то вот вы и доказали сходимость.¹ Это чуть ли ни единственный метод бороться с этим видом сходимости. Вот за это мы и любим хар функции, что для проверки сходимости по распределению надо всего лишь работать с одной функцией.

Связь между разными видами сходимости На диаграмме ниже показано, какие виды сходимостей какие влекут:



Держите эту картинку перед глазами, когда надо вспоминать, какая сходимость влечет какую. И еще не забывайте, что для теории вероятности в ней верхние два вида сходимости не интересны.

¹ Учтите, что не любая функция является хар функцией для какой-то случайной величины ξ . Потому, даже если предел $\varphi_{\xi_n}(t)$ существует в любой точке, то это еще не означает сходимость по распределению, так как предел может не соответствовать никакой случайной величине. Для проверки такого соответствия есть критерии. Но не то что бы они безумно полезные. Проверять с помощью них можно, но это боль и унижение.

Сходимость и качество оценок Давайте еще раз посмотрим на нашу оценку

$$\bar{x} = \frac{\xi_1(\omega) + \dots + \xi_n(\omega)}{n}$$

Теперь мы знаем, что в среднем мы попадем в математическое ожидание. Но вдруг для одних точек ω подобное среднее будет далеко от матожидания, а для других близко? Тогда мы вообще говоря не можем ничего гарантировать. Однако, тут на помощь приходят различные асимптотические законы. Вот пример закона больших чисел:

Утверждение (Закон больших чисел). Пусть $\xi_1, \dots, \xi_n, \dots \sim \xi$ – последовательность одинаково распределенных независимых случайных величин. Тогда последовательность

$$\frac{\xi_1 + \dots + \xi_n}{n}$$

сходится к $\mathbb{E}\xi$ почти наверное.

Давайте расшифруем, что здесь сказано. Здесь говорится, что если я возьму почти любое $\omega \in \Omega$, то имеется сходимость

$$\frac{\xi_1(\omega) + \dots + \xi_n(\omega)}{n} \rightarrow \mathbb{E}\xi, \quad n \rightarrow \infty$$

Фразу «почти любое» надо понимать так: пусть $X \subseteq \Omega$ – множество тех ω для которых есть такая сходимость, тогда $P(X) = 1$. Это значит, что если вероятность попасть в ω для которого такой сходимости нет равна нулю. На практике это означает, что мы всегда попадаем в ω , для которого выполнена подобная сходимость.

Таким образом, этот асимптотический закон говорит о том, что не только в среднем наша формула дает математическое ожидание, но вообще всегда (формально почти всегда) результат по этой формуле будет сколь угодно близко стремиться к настоящему математическому ожиданию.

Надо сказать, что у этого закона есть куча разных вариаций. Есть вариация, которая даже оценивает скорость сходимости. А именно, формулировка приближительного такого вида

$$P\left(\left|\frac{\xi_1 + \dots + \xi_n}{n} - \mathbb{E}\xi\right| > \varepsilon\right) < \delta(\varepsilon, n)$$

где $\delta(\varepsilon, n)$ – некоторая известная функция. Подобные законы говорят, что выражение $\frac{\xi_1(\omega) + \dots + \xi_n(\omega)}{n}$ отклоняется от матожидания $\mathbb{E}\xi$ на величину большую ε с вероятностью меньше $\delta(\varepsilon, n)$. То есть мы можем гарантировать близость к математическому ожиданию с некоторой вероятностью при достаточно большом количестве измерений. Я не привожу точную формулировку этого утверждения, просто потому что есть куча разных формулировок. Все они основаны на неравенстве Чебышева.

Утверждение (Неравенство Чебышева). Пусть ξ – случайная величина, такая что $\mathbb{E}\xi = a$ и $\mathbb{D}\xi = \sigma^2 > 0$, тогда для любого числа $t > 0$ верно

$$P(|\xi - a| \geq t\sigma) \leq \frac{1}{t^2}$$

Выборочная дисперсия Для оценки дисперсии случайной величины первая формула, которая приходит на ум – это

$$S = S(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{где } \bar{x} = \frac{x_1 + \dots + x_n}{n}$$

Чтобы проверить несмещенность этой формулы, надо заменить каждое вхождение x_i на независимую $\xi_i \sim \xi$. В этом случае получим, что

$$\mathbb{E}S(\xi_1, \dots, \xi_n) = \frac{n-1}{n} \mathbb{D}\xi$$

Оказывается, что в этой наивной формуле мы в среднем промахиваемся мимо математического ожидания. Потому вводится понятие несмещенной оценки для выборочной дисперсии

$$S_0 = S_0(x_1, \dots, x_n) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{где } \bar{x} = \frac{x_1 + \dots + x_n}{n}$$

С другой стороны, всякие вариации на тему закона больших чисел, показывают, что обе оценки для дисперсии стремятся к самой дисперсии. Я не буду расписывать этот факт более подробно по нескольким причинам. Во-первых, строгие формулировки важны только для математики внутри математики. Во-вторых, от нестрогой формулировки надо лишь знать, что в каком-то смысле подобная формула в пределе стремится к желаемой оцениваемой величине. В-третьих, надо понимать, что на практике нам все равно ни холодно ни жарко от разновидностей пределов в теории вероятности, важно лишь, что есть какой-то вид сходимости, а раз он есть, то мы просто надеемся, что наших измерений достаточно, чтобы быть близкими к желаемому пределу (сто процентных гарантий все равно нет).

Векторные измерения Пусть у нас теперь есть некоторый неизвестный случайный вектор $\bar{\xi} = (\xi_1, \dots, \xi_m)$ на \mathbb{R}^m и пусть у нас есть n независимых измерений $x_1, \dots, x_n \in \mathbb{R}^m$. Обратите внимание, что мы теперь измеряем вектор, а не число, то есть каждый вектор x_i имеет свои координаты $x_i = (x_{i1}, \dots, x_{in})$.

Как можно думать про подобные эксперименты. Пусть у вас на черном ящике есть несколько приборов ξ_1, \dots, ξ_m , вообще говоря зависящих между собой. И вы делаете такой эксперимент: снимаете показания с первого прибора, потом со второго и так далее до последнего m -го прибора и записываете все m чисел в вектор x_1 . Потом сбрасываете эксперимент в начальное состояние и независимо проводите новую серию таких же измерений и записываете результат в вектор x_2 . Потом опять сбрасываете все в начальное состояние и независимо от предыдущих попыток проводите следующий эксперимент и получаете вектор x_3 и т.д. Таким образом внутри i -ой серии, когда вы замеряли координаты вектора x_i , измерения могут быть зависимыми, но между сериями у вас нет зависимостей.

Математическое ожидание векторной величины оценивается так же, как и в скалярном случае

$$\bar{x} = \frac{x_1 + \dots + x_n}{n} \in \mathbb{R}^m$$

Выборочная матрица ковариации В этом случае самая популярная формула для оценки матрицы ковариации следующая

$$\Sigma = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t \in M_m(\mathbb{R}), \text{ где } \bar{x} = \frac{x_1 + \dots + x_n}{n} \in \mathbb{R}^m$$

Аналогично дисперсии, если подставить вместо x_i векторы $\bar{\xi}_i$, то окажется, что математическое ожидание от Σ будет промахиваться мимо матрицы ковариации². Потому вводится несмещенная оценка

$$\Sigma_0 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t \in M_m(\mathbb{R}), \text{ где } \bar{x} = \frac{x_1 + \dots + x_n}{n} \in \mathbb{R}^m$$

Оценка корреляции двух случайных величин Пусть есть две случайные величины ξ и η . Напомню, что коэффициентом корреляции между ξ и η называется величина

$$\frac{\text{cov}(\xi, \eta)}{\sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta}} = \frac{\mathbb{E}(\xi - \mathbb{E}\xi)(\eta - \mathbb{E}\eta)}{\sqrt{\mathbb{D}\xi}\sqrt{\mathbb{D}\eta}}$$

Дисперсии мы уже оценивать умеем, а чтобы оценить ковариацию, надо проделать серию измерений следующего вида: мы измеряем ξ и η одновременно, получаем пару чисел (x_1, y_1) . Потом независимо измеряем их еще раз и получаем пару чисел (x_2, y_2) и т.д. В результате мы получим набор векторов

$$v_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \dots, v_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \in \mathbb{R}^2$$

Значит мы можем оценить матожидание и матрицу ковариации случайного вектора (ξ, η) . Тогда оценка для $\text{cov}(\xi, \eta)$ – будет коэффициент Σ_{12} , а именно, оценочная ковариация равна

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \in \mathbb{R}, \text{ где } \bar{x} = \frac{x_1 + \dots + x_n}{n} \text{ и } \bar{y} = \frac{y_1 + \dots + y_n}{n}$$

И несмещенная версия оценочной ковариации выглядит так

$$\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \in \mathbb{R}, \text{ где } \bar{x} = \frac{x_1 + \dots + x_n}{n} \text{ и } \bar{y} = \frac{y_1 + \dots + y_n}{n}$$

²А именно, $\mathbb{E}\Sigma$ будет матрица ковариации умноженная на $\frac{n-1}{n}$.

Метод максимального правдоподобия

Предположим, что у нас есть непрерывная случайная величина ξ , и ее плотность $p(x, \theta_1, \dots, \theta_k)$ зависит от параметров, которые мы не знаем, но очень хотим узнать. При этом у нас есть n независимых измерений $x_1, \dots, x_n \in \mathbb{R}$. Вопрос: как найти $\theta_1, \dots, \theta_k$? Можно воспользоваться следующим соображением. Мы знаем, что измерения пришли из независимых измерений, значит есть такие независимые случайные величины $\xi_1, \dots, \xi_n \sim \xi$, что $x_1 = \xi_1(\omega), \dots, x_n = \xi_n(\omega)$. Из-за независимости, у случайного вектора $\bar{\xi} = (\xi_1, \dots, \xi_n)$ плотность будет задаваться по правилу

$$p_{\bar{\xi}}(z_1, \dots, z_n, \theta_1, \dots, \theta_k) = p(z_1, \theta_1, \dots, \theta_k) \dots p(z_n, \theta_1, \dots, \theta_k)$$

Здесь набор $\theta_1, \dots, \theta_k$ – это параметры нашего распределения, а для каждой ξ_i соответствующая координата в пространстве обозначается z_i . С одной стороны, эта функция задает плотность $\bar{\xi}$, а с другой, нам выпал вектор $(x_1, \dots, x_n) = \bar{\xi}(\omega)$. Как вы думаете, если нам задана плотность, то какая точка вероятнее всего выпадет в одном эксперименте? Конечно же максимум плотности, просто по ее сельскохозяйственному смыслу – это самая вероятная точка, которая должна выпасть. Значит, нам надо подобрать параметры $\theta_1, \dots, \theta_k$ так, чтобы точка (x_1, \dots, x_n) оказалась максимумом функции $p_{\bar{\xi}}(z, \theta)$. То есть надо решить следующую оптимизационную задачу

$$y = p(z_1, \theta_1, \dots, \theta_k) \dots p(z_n, \theta_1, \dots, \theta_k) \rightarrow \max_{\theta_1, \dots, \theta_k}$$

Заметим, что максимизировать некоторую функцию $\phi(\theta)$, принимающую только положительные значения – это то же самое, что максимизировать функцию $\ln(\phi(\theta))$. Так как \ln монотонно возрастает, то на такую процедуру можно смотреть, как на замену координаты y для значения функции, то есть мы просто растягиваем по вертикали график функции $\phi(\theta)$ и не меняем ее точек максимума и минимума, но меняем сами значения в этих точках. Тогда задача оптимизации превращается в следующую

$$\sum_{i=1}^n \ln p(z_i, \theta_1, \dots, \theta_k) \rightarrow \max_{\theta_1, \dots, \theta_k}$$

И часто удобно поменять знак у этого выражения и минимизировать его.

Давайте продемонстрируем, как работает метод на примере гауссова распределения. Пусть ξ имеет распределение с плотностью

$$p(x, a, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}$$

Но мы не знаем конкретных значений для $a \in \mathbb{R}$ и $\sigma > 0$, зато знаем несколько независимых измерений этой случайной величины: $x_1, \dots, x_n \in \mathbb{R}$. Тогда нам надо максимизировать следующее выражение

$$\phi(a, \sigma) = \sum_{i=1}^n \left(-\frac{(x_i - a)^2}{2\sigma^2} - \ln(\sigma) - \ln(\sqrt{2\pi}) \right)$$

Во-первых, мы можем изменить знак у этого выражения и минимизировать полученную функцию. Во-вторых, мы можем отбросить константное слагаемое $\ln(\sqrt{2\pi})$, так как прибавление константы не меняет минимумов функции. Потому будем минимизировать выражение

$$\phi(a, \sigma) = \sum_{i=1}^n \left(\frac{(x_i - a)^2}{2\sigma^2} + \ln(\sigma) \right) \rightarrow \min_{a, \sigma}$$

Найдем критические точки

$$\begin{aligned} \frac{\partial \phi}{\partial a} &= \sum_{i=1}^n -\frac{x_i - a}{\sigma^2} = 0 \\ \frac{\partial \phi}{\partial \sigma} &= \sum_{i=1}^n \left(-\frac{(x_i - a)^2}{\sigma^3} + \frac{1}{\sigma} \right) = 0 \end{aligned}$$

Из первого уравнения получаем выражение для a и, подставив его во второе, получаем выражение для σ . Выражения будут иметь вид

$$\begin{aligned} a &= \frac{x_1 + \dots + x_n}{n} = \bar{x} \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \end{aligned}$$

Полученные оценки и называются оценками максимального правдоподобия для гауссова распределения.

Давайте проверим Гессиан, что действительно получилась точка минимума. Вторые частные производные будут

$$\begin{aligned}\frac{\partial^2 \phi}{\partial a^2} &= \frac{n}{\sigma^2} \\ \frac{\partial^2 \phi}{\partial a \partial \sigma} &= 2 \sum_{i=1}^n \frac{x_i - a}{\sigma^3} \\ \frac{\partial^2 \phi}{\partial \sigma^2} &= 3 \sum_{i=1}^n \frac{(x_i - a)^2}{\sigma^4} - \frac{n}{\sigma^2}\end{aligned}$$

Теперь надо вместо a и σ подставить выражения из оценок максимального правдоподобия, потому что мы считаем Гессиан не в произвольной точке, а в той, которую хотим проверить на минимум. Тогда получим

$$\begin{aligned}\frac{\partial^2 \phi}{\partial a^2} &= \frac{n}{\sigma^2} \\ \frac{\partial^2 \phi}{\partial a \partial \sigma} &= 0 \quad \text{то есть} \quad H(\phi) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix} \\ \frac{\partial^2 \phi}{\partial \sigma^2} &= \frac{2n}{\sigma^2}\end{aligned}$$

А тут легко видеть, что это положительно определенная матрица (здесь вместо σ^2 стоит выражение из оценки максимального правдоподобия, но для краткости я его оставил в виде σ^2).

Метод максимального правдоподобия в векторной форме

Давайте хоть раз пальцем из наших супер пушек по электронным воробьям. В предыдущей задаче мы разобрались с МНК для нормальной случайной величины. Если же мы хотим сделать то же самое в векторном случае, то считать градиенты в лоб – то еще извращение. И тут на помощь нам придут матричные дифференцирования. Это будет хороший пример их применения. Поехали.

Пусть у меня есть нормальный вектор ξ с распределением $\mathcal{N}(a, S)$, где $a \in \mathbb{R}^n$ и $S \in M_n(\mathbb{R})$ – симметричная положительно определенная матрица.³ Плотность для ξ задается следующим образом

$$p_\xi(x) = \frac{1}{\sqrt{(2\pi)^n \det S}} e^{-\frac{1}{2}(S^{-1}(x-a), x-a)}$$

где $(u, v) = u^t v$ – стандартное скалярное произведение.

Пусть мы намерили сэмплов $x_1, \dots, x_m \in \mathbb{R}^n$ и хотим восстановить $a \in \mathbb{R}^n$ и $S \in M_n(\mathbb{R})$ симметричную и положительно определенную. Тогда мы думаем, что у нас есть последовательность независимых векторов x_1, \dots, x_n распределенных $\mathcal{N}(a, S)$. Тогда «правдоподобие» выпадения набора x_1, \dots, x_n будет

$$\phi(a, S) = \prod_{i=1}^m \frac{1}{\sqrt{(2\pi)^n \det S}} e^{-\frac{1}{2}(S^{-1}(x_i-a), x_i-a)}$$

и мы хотим решить задачу максимизации $\phi(a, S)$. Для этого опять рассмотрим $\ln \phi(a, S)$ так как это не изменит точки максимума функции, получим

$$\psi(a, S) = \sum_{i=1}^m \left(-\frac{1}{2}(S^{-1}(x_i - a), x_i - a) - \frac{1}{2} \ln \det S - \frac{1}{2} \ln((2\pi)^n) \right) \rightarrow \max_{a, S}$$

Теперь можно откинуть независимые от a и S слагаемые, поменять знак и умножить на $\frac{2}{m}$, получим

$$\varphi(a, S) = \frac{1}{m} \sum_{i=1}^m (S^{-1}(x_i - a), x_i - a) + \ln \det S \rightarrow \min_{a, S}$$

³Я выбрал не каноничное обозначение S , чтобы не путать со знаком суммирования в выкладках.

Сделаем для удобства замену $Q = S^{-1}$

$$\varphi(a, Q) = \frac{1}{m} \sum_{i=1}^m (Q(x_i - a), x_i - a) - \ln \det Q \rightarrow \min_{a, Q}$$

Вот теперь мы готовы к поиску производной. Для этого надо найти дифференциалы по a и по Q и приравнять их к нулю. Получим

$$\begin{cases} d_{a,Q}\varphi(\dot{a}) = \frac{1}{m} \sum_{i=1}^m ((Q(-\dot{a}), x_i - a) + (Q(x_i - a), -\dot{a})) = 0 \\ d_{a,Q}\varphi(\dot{S}) = \frac{1}{m} \sum_{i=1}^m (\dot{Q}(x_i - a), x_i - a) - \text{tr}(Q^{-1}\dot{Q}) = 0 \end{cases}$$

Равенство нулю означает, что дифференциал принимает нулевые значения для любого значения аргумента. В первом случае для любого \dot{a} , во втором, для любого \dot{S} . Давайте решим первое уравнение в начале

$$\begin{aligned} d_{a,Q}\varphi(\dot{a}) &= \frac{1}{m} \sum_{i=1}^m ((Q(-\dot{a}), x_i - a) + (Q(x_i - a), -\dot{a})) = \frac{2}{m} \sum_{i=1}^m (Q(x_i - a), -\dot{a}) = \\ &= \left(\frac{2}{m} \sum_{i=1}^m Q(x_i - a), -\dot{a} \right) = -2 \left(Q \left(\frac{1}{m} \sum_{i=1}^m x_i - a \right), \dot{a} \right) = 0 \quad \text{для любого } \dot{a} \in \mathbb{R}^n \end{aligned}$$

Но последнее равенство для любого $\dot{a} \in \mathbb{R}^n$ возможно лишь при условии

$$Q \left(\frac{1}{m} \sum_{i=1}^m x_i - a \right) = 0$$

И так как Q у нас обратима, то это равносильно

$$\frac{1}{m} \sum_{i=1}^m x_i - a = 0$$

А значит

$$a = \frac{1}{m} \sum_{i=1}^m x_i = \bar{x}$$

То есть это выборочное среднее по сэмплам. Теперь давайте решим второе уравнение. Не забывайте, что Q у нас еще и симметричная, потому я не буду ставить на нее знак транспонирования.

$$\begin{aligned} d_{a,Q}\varphi(\dot{S}) &= \frac{1}{m} \sum_{i=1}^m (\dot{Q}(x_i - a), x_i - a) - \text{tr}(Q^{-1}\dot{Q}) = \frac{1}{m} \sum_{i=1}^m (x_i - a)^t \dot{Q}(x_i - a) - \text{tr}(Q^{-1}\dot{Q}) = \\ &= \frac{1}{m} \sum_{i=1}^m \text{tr} \left((x_i - a)(x_i - a)^t \dot{Q} \right) - \text{tr}(Q^{-1}\dot{Q}) = \text{tr} \left(\frac{1}{m} \sum_{i=1}^m (x_i - a)(x_i - a)^t \dot{Q} \right) - \text{tr}(Q^{-1}\dot{Q}) = \\ &= \text{tr} \left(\left(\frac{1}{m} \sum_{i=1}^m (x_i - a)(x_i - a)^t - Q^{-1} \right) \dot{Q} \right) = 0 \quad \text{для любой симметричной } \dot{Q} \in M_n(\mathbb{R}) \end{aligned}$$

Так как матрица

$$\frac{1}{m} \sum_{i=1}^m (x_i - a)(x_i - a)^t - Q^{-1}$$

симметрична, то последнее равенство означает, что в пространстве симметричных матриц скалярное произведение этой матрицы с любой⁴ другой равно нулю, а это значит, что

$$\frac{1}{m} \sum_{i=1}^m (x_i - a)(x_i - a)^t - Q^{-1} = 0$$

⁴Если Q была симметричная и положительно определена, то \dot{Q} может быть любой симметричной, потому что мы можем «отступить» от положительной симметричной в любую сторону, а разность между двумя положительными симметричными не обязательно положительна, но симметрична. Потому и производная будет вообще говоря только симметричной. И на самом деле любую симметричную мы получим. Но это если вы вдруг захотели видеть здесь хоть какое-то пояснение этому факту.

Кроме того, вспомним, что $S = Q^{-1}$ и $a = \bar{x}$, получаем

$$S = \frac{1}{m} \sum_{i=1}^m (x_i - \bar{x})(x_i - \bar{x})^t$$

то есть это выборочная матрица ковариации (со смещением). По хорошему тут надо бы еще проверить вторую производную для того, чтобы убедиться, что мы получили точку минимума, но я не буду насиловать ваше сознание. Кроме того в большой науке часто используют выпуклость или геодезическую выпуклость (простите меня математики, что я произношу эти инженерные слова, но так уж повелась жизнь, что у всех своя терминология). И использование соображений выпуклости позволяет провести подобную проверку с минимальными потерями для психики и ваших ручек.

PCA и SVD

Пусть у нас задан случайный вектор ξ в пространстве \mathbb{R}^m и мы для него намерели независимо несколько сэмплов $x_1, \dots, x_n \in \mathbb{R}^m$. Первым делом мы можем оценить математическое ожидание и матрицу ковариации для ξ по формулам

$$\mathbb{E}\xi = \bar{x} = \frac{x_1 + \dots + x_n}{n} \quad \Sigma_0 = \frac{1}{n-1} (x_i - \bar{x})(x_i - \bar{x})^t$$

Давайте составим матрицу из всех сэмплов $X = (x_1 | \dots | x_n)$ и сдвинем каждый сэмпл на выборочное математическое ожидание $Y = (x_1 - \bar{x} | \dots | x_n - \bar{x}) = X - (\bar{x} | \dots | \bar{x})$.

А теперь представьте, что вы почувствовали непреодолимое желание сделать SVD для матрицы Y . Давайте посмотрим, к чему это может привести. На помню, что алгоритм для нахождения SVD для матрицы Y начинается с того, что надо диагонализировать матрицу YY^t . Но если мы применим к ней блочные формулы, то получим, что $YY^t = (x_1 - \bar{x})(x_1 - \bar{x})^t + \dots + (x_n - \bar{x})(x_n - \bar{x})^t$, что с точностью до коэффициента совпадает с матрицей выборочной ковариации. А значит диагонализация выборочной ковариации – это первый шаг в SVD для Y .

Напомним, что под SVD для Y мы понимаем разложение $Y = UDV^t$, где $U \in M_m(\mathbb{R})$, $V \in M_n(\mathbb{R})$ и обе ортогональные (то есть $UU^t = E$ и $VV^t = E$) а матрица $D \in M_{m,n}(\mathbb{R})$ и на ее главной диагонали стоят числа $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ и быть может нули. В этом случае $YY^t = UDD^tU^t$, где DD^t будет диагональной матрицей размера m на m с числами $\sigma_1^2 \geq \dots \geq \sigma_r^2 > 0$ на диагонали и быть может нулями. В этом случае числа $\{\sigma_1^2, \dots, \sigma_r^2\}$ будут ненулевыми точками спектра YY^t (то есть не нулевыми собственными значениями этой матрицы, которые совпадают с ненулевыми корнями ее хар. многочлена или минимального многочлена). А матрица $U = (u_1 | \dots | u_m)$ будет состоять из собственных векторов для матрицы YY^t причем они все будут ортогональны и длины один относительно стандартного скалярного произведения $(x, y) = x^t y$. При этом u_1 будет собственным для σ_1^2 и так далее до u_r , остальные u_i будут собственными для 0.

На SVD разложение можно посмотреть вот как $D = U^t Y V$. То есть мы стартовали с матрицы из несмещенных сэмплов Y и теперь с помощью матрицы U пытаемся поменять координаты в пространстве \mathbb{R}^m , где живут наши сэмплы, а с помощью матрицы V пытаемся комбинировать наши сэмплы между собой. Часто операция комбинирования сэмплов между собой не очень физически осмысленна, потому давайте ее проигнорируем и рассмотрим равенство $U^t Y = DV^t$. Тогда вот какой смысл у написанного. Мы стартовали с Y . Потом выбрали для выборочной матрицы ковариации Σ_0 собственные векторы (длины один и взаимно-ортогональные) и сложили их в матрицу $U = (u_1 | \dots | u_m)$. Далее привели матрицу Σ_0 к главным осям, то есть сделали замену стандартного базиса на базис u_1, \dots, u_m . При этом координаты наших сэмплов как раз

изменяться по правилу $Y \mapsto U^t Y = DV^t$. Давайте внимательно посмотрим на правую часть этого равенства

$$U^t Y = DV^t = \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \vdots \\ & & \sigma_r & & 0 \\ & & & 0 & \vdots \\ & & & & 0 \\ & & & & \vdots \\ & & & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & \dots & v_{n1} \\ \vdots & \dots & \dots & \vdots \\ v_{1m} & \dots & \dots & v_{nm} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_1 & & & & \\ & \ddots & & & \\ & & \sigma_r & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \dots & \dots & v_{n1} \\ \vdots & \dots & \dots & \vdots \\ v_{1m} & \dots & \dots & v_{nm} \end{pmatrix}$$

То есть мы у ортогональной матрицы V^t отрезали верхнюю часть. Так как матрица V была ортогональна, то интуитивно, ее координаты вносят одинаковый вклад в итоговую матрицу. Однако, после этого мы каждую координату домножили на весовой коэффициент σ_i , а последние даже на 0. Таким образом после смены координат для несмещенных сэмплов $Y \mapsto U^t Y$ мы отсортировали координаты сверху вниз по их важности, где σ_i означает вес важности координаты. При этом мы видим, что последние координаты мы вообще можем проигнорировать, ибо они стали нулевыми. Кроме того, мы еще можем проигнорировать координаты с малыми σ_i .

По другому еще можно сказать так, мы нашли, что все наши сэмплы реально жили в подпространстве \mathbb{R}^r и после поворота пространства, отрезав лишние координаты, мы можем считать, что наши сэмплы имеют вид

$$\begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix} \begin{pmatrix} v_{11} & \dots & \dots & v_{n1} \\ \vdots & \dots & \dots & \vdots \\ v_{1r} & \dots & \dots & v_{nr} \end{pmatrix}$$

А откинув еще и малые σ_i мы можем понизить размерность задачи еще сильнее.

Центральная предельная теорема

Оказывается, что нормальное распределение обладает одним удивительным асимптотическим свойством. В некотором смысле, если много раз делать один и тот же случайный эксперимент с измерением одной и той же случайной величины, то при правильной нормировке, со временем все сэмплы будут вести себя так, как будто они взялись из нормального распределения. Это означает, что при большом количестве сэмплов, нам вообще говоря плевать на распределение, все можно считать нормальным. Звучит слишком уж круто, чтобы быть правдой, тем не менее, я постараюсь придать строгий смысл столь серьезному заявлению. Строгая формулировка этого явления называется Центральной предельной теоремой.⁵

Утверждение (Центральная предельная теорема). Пусть $\xi_1, \dots, \xi_n, \dots$ – последовательность независимых одинаково распределенных случайных величин с математическим ожиданием μ и дисперсией σ^2 . Тогда

$$\frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}} \xrightarrow{\mathcal{D}} N(0, 1)$$

Из закона больших чисел следует, что если мы рассмотрим дробь⁶

$$\frac{\sum_{k=1}^n (\xi_k - \mu)}{n} = \bar{x}_n - \mu$$

⁵На самом деле, это не единственный результат такого рода, но один из первых и уж точно один из важнейших. Есть куча других асимптотических результатов аналогичных ЦПТ и в них фигурируют не только нормальное распределение.

⁶Здесь $\bar{x}_n = \frac{1}{n} \sum_{k=1}^n x_k$ содержит индекс n , чтобы подчеркнуть зависимость от количества сэмплов.

то она обязательно сойдется к нулю, так как средне арифметическое одинаково распределенных независимых случайных величин сходится к математическому ожиданию. А в центральной предельной теореме мы оцениваем на сколько быстро \bar{x} приближается к μ . Если нормировать не так агрессивной и положить

$$\frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}} = \sqrt{n} \frac{\bar{x}_n - \mu}{\sigma}$$

то полученное выражение будет себя вести как нормально распределенная случайная величина. Еще неформально это можно запомнить так, при больших n выполнено

$$\bar{x}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Таким образом выборочное математическое ожидание будет раскидано вокруг настоящего математического ожидания по нормальному закону.

Если расшифровать ЦПТ и вспомнить третье эквивалентное определение сходимости по распределению, то получается, что

$$P\left(\frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}} \in [a, b]\right) \rightarrow \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

То есть при больших значениях n вероятность для случайной величины

$$\frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}}$$

считается по формулам для нормального распределения, в частности вероятность попасть в любое подмножество. В примере выше в качестве такого подмножества был выбран интервал $[a, b]$.

Если дополнительно для случайных величин ξ_k существует $\mathbb{E}|\xi_k^3| = \rho$ (оно не зависит от k , так как все случайные величины одинаково распределены), то можно оценить скорость сходимости к нормальному распределению. Скорость будем оценивать на функции распределения. Пусть

$$\eta_n = \frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}} \quad \text{и} \quad \zeta \sim N(0, 1)$$

Тогда

$$|P(\eta_n \leq x) - P(\zeta \leq x)| \leq \frac{\rho}{2\sigma^3 \sqrt{n}}$$

Но написанное слева есть ни что иное, как функции распределения

$$|F_{\eta_n}(x) - F_{\zeta}(x)| \leq \frac{\rho}{2\sigma^3 \sqrt{n}}$$

Константу $1/2$ можно чуть улучшить в правой части неравенство, которое называется неравенством Берри-Эссеена. Благодаря нему мы не только знаем, что η_n сойдется к нормальному распределению, но и для каждого конкретного интервала оценить на сколько близко соответствующая вероятность подойдет к вероятности для нормального распределения, а именно

$$\left| P\left(\frac{\sum_{k=1}^n (\xi_k - \mu)}{\sigma \sqrt{n}} \in [a, b]\right) - \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \right| \leq \frac{\rho}{\sigma^3 \sqrt{n}}$$

Семинар 5

Задачи:

1. Пусть $\xi_n \xrightarrow{P} \xi$ и $\eta_n \xrightarrow{P} \eta$. Покажите, что $\xi_n + \eta_n \xrightarrow{P} \xi + \eta$
2. Пусть $\Omega = [0, 1]$ и вероятность P – это длина. Рассмотрим последовательность случайных величин ξ_n на Ω заданную следующим образом

$$\xi_{2k+1}(x) = \left(\frac{1}{4} - \frac{1}{k}\right)x + \frac{3}{8} + \frac{1}{2k} \quad \text{и} \quad \xi_{2k}(x) = \left(\frac{1}{k} - \frac{1}{4}\right)x + \frac{5}{8} - \frac{1}{2k}$$

- (a) Покажите, что эта последовательность сходится по распределению и найдите предельное распределение.
 - (b) Покажите, что эта последовательность не сходится ни в каком другом виде сходимости.
3. Пусть ξ_n – случайная величина имеющая биномиальное распределение $B(n, p)$, то есть вероятность успеха p , а n – количество бросаний. Покажите, что если $p = \frac{\lambda}{n}$ для некоторой константы λ , то при $n \rightarrow \infty$ характеристическая функция для ξ_n сходится поточечно к характеристической функции распределения Пуассона с параметром λ .