ВМК МГУ & ФУПМ МФТИ

Теория вероятности. Математическая статистика

Игорь Тао, Дима Сотников, Дима Гущин

25 августа 2019 г.

Содержание

Предисловие

Здорово! Раз вы читаете эти строки, значит вы желаете обновить или заложить свои знания в области Теории вероятности и Математической статистике.

Мы придерживались следующего порядка изложения: сначала дали неформальное объяснение всего материала, который дается в книге, затем рассказываем поверхностно о мере Лебега и переходим к теоретико-вероятностным выкладкам, которые являются инструментом в прикладных аспектах этого направления.

Часть № I. Теория вероятности

> Ты спросишь меня, кого я люблю больше: тебя или теорию вероятности. Я отвечу, что почти наверное я больше люблю тебя, и ты уйдешь, так и не узнав, что почти наверное значит с вероятностью 1.
>
> Игорян

§1. Вероятностное пространство

1.1. Стохастические ситуации

Когда можно применять теорию вероятности — поговорим об этом. Часть информации, изложенной дальше, взято из книги [1] В. Ю. Королева, так что любознательный читатель может ознакомиться с первоисточником при особом желании.

Следуя Королеву, будем говорить о *случайности* как о принципиально неустранимой неопределенности. Чтобы понять, как с ней же работать, необходимо выделить те случаи, когда мы можем сказать: «А здесь нам поможет теория вероятностей», — и не облажаться.

Методы теории вероятности работают в ситуациях, называемых *стохастическими.* Для них характерны три свойства:

❖ **1: Непредсказуемость:** исход ситуации нельзя предсказать абсолютно, то есть без какой-либо погрешности. В ином случае использовать теорию вероятностей — что по воробьям из пушки палить!

❖ **2: Воспроизводимость:** у испытателей есть возможность (хотя бы теоретическая) повторить ситуацию сколь угодно много раз при неизменных условиях.

❖ **3: Устойчивость частот:** при многократном повторении ситуации частота исхода, то есть отношение числа опытов, в которых мы получили этот исход, к их общему числу, должна колебаться около некоторого значения, приближаясь к нему все ближе и ближе. Другими словами, частота исходов должна иметь предел при стремлении числа опытов к ∞.

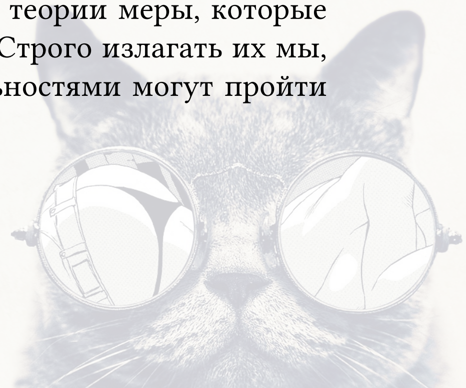
А где мы можем встретить такие ситуации, а? Правильно! В казино! Исторически теория вероятности создавалась, чтобы играть в азартные игры с опорой на науку, отсюда и типичные задачи: бросаем кости, гоняем шары в лототроне, делим ставку и т. д.

Чтобы превратить бытовую ситуацию в математическую науку, необходимо внести некий математический эквивалент словам «а ему чаще везет». И таким математическим объектом будет функция вероятности. Её область определения называется *множеством событий.* В свою очередь, событие (такое как, например, выпадание чётного числа на кубике) может являться совокупностью неких более простых событий, описывающих стохастическую ситуацию (число, выпавшее на кубике). Последнее множество называется *множеством элементарных исходов* и обозначается Ω.

Есть разные теории описывающие вероятностные пространства, но самой используемой является теория, созданная Колмогоровым в начале прошлого века. Перейдем к ее описанию, то есть к нестрогой формализации понятий: множество событий, множество элементарных исходов.

1.2. Измеримое пространство

Здесь только суть. Быстрым темпом! Спринт по началам теории меры, которые и приведут нас к определению вероятностного пространства. Строго излагать их мы, конечно, не будем. Поэтому любители обмазываться формальностями могут пройти сюда — [2].



Вероятностное пространство, которое мы дальше определим, должно содержать в себе описание трех позиций, важных в эксперименте: во-первых, оно должно содержать множество элементарных исходов Ω, чтобы было предельно понятно, что является исходом; во-вторых, на этом множестве мы будем рассматривать некоторую структуру подмножеств, которая и будет множеством событий, которые имеют для нас смысл (т.е. если мы бросаем кубик и нам важна четность выпавшего числа, то не имеет смысл рассматривать исход «выпало 1 очко» в отдельности от исходов «выпало 3 очка» и «выпало 5 очков»); в-третьих, на этом множестве мы зададим функцию, которая и будет вероятностью, то есть некоторой характеристикой, описывающей частоту того или иного исхода.

Множество событий, обозначаемое $\mathcal{A}$ (используют наравне с этим еще обозначение $\mathscr{A}$), должно обладать следующими интуитивными свойствами:

1. Отрицание события есть событие (если «пойдет дождь» событие, то «не пойдет дождь» также событие).

2. Объединение событий есть событие («пойдет дождь» или «пойдет снег»).

3. Все множество элементарных исходов является событием («что-нибудь да произойдет»).

Если мы посадим эти три свойства в одну лодку с Математикой, то получим следующее

Определение 1.1. Семейство $\mathcal{A}$ подмножеств множества Ω называется *алгеброй*, если верны три аксиомы:

(i) $\forall A \in \mathcal{A}, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$ (если объединить два множества из семейства, то обязательно получим множество из семейства),

(ii) $\forall A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$ (множество и его дополнение одновременно: либо лежат в нашем семействе, либо нет),

(iii) $\Omega \in \mathcal{A}$ (все множество точно лежит в $\mathcal{A}$).

↳ Из аксиом алгебры и формулы $A \cap B = \overline{\overline{A} \cup \overline{B}}$ следует, что пересечение событий является событием. Подобным образом можно показать, что и разность, и симметрическая разность двух множеств из семейства $\mathcal{A}$ тоже дадут множество из $\mathcal{A}$.

Пример. Наименьшей алгеброй является $\{\Omega, \varnothing\}$.

Косяк определения, которое мы только что дали, в том, что часто встречаются бесконечные множества элементарных исходов, поэтому полезно, чтобы множество событий было замкнуто не только относительно объединения, но и относительно счетного объединения.

Определение 1.2. Семейство $\mathcal{A}$ подмножеств множества Ω называется *σ-алгеброй* (читается как сигма-алгебра), если

(i) $\forall A_1, \ldots, A_n, \ldots \in \mathcal{A} \Rightarrow \bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{A}$,

(ii) $\forall A \in \mathcal{A} \Rightarrow \overline{A} \in \mathcal{A}$,

(iii) $\Omega \in \mathcal{A}$.

Не всегда по задаче понятно, какую сигма-алгебру надо выбрать, однако большой бонус ее аксиом заключается в том, что мы можем пересекать различные сигма-алгебры и получать снова сигма-алгебру. Из-за этого, если нам известно неполное множество событий $\mathcal{K}$, т.е. $\mathcal{K} \subseteq \mathcal{A}$, то мы можем восстановить сигма-алгебру с точностью до тех событий, которых нет в $\mathcal{K}$. Иными словами, мы можем найти такую минимальную сигма-алгебру, которая содержит в себе $\mathcal{K}$ как подмножество.

Определение 1.3. Пусть $\mathcal{K}$ — класс подмножеств Ω. σ-алгебра $\sigma(\mathcal{K})$, *порожденная классом $\mathcal{K}$* — наименьшая σ-алгебра, содержащая $\mathcal{K}$, то есть любая σ-алгебра, содержащая $\mathcal{K}$, содержит и $\sigma(\mathcal{K})$.

Пример. σ-алгеброй, порожденной $\mathcal{K} = A$ (класс состоит из одного элементарного события), будет являться $\sigma(A) = \left\{\varnothing, A, \overline{A}, \Omega\right\}$.

Сигма-алгебра является более узким понятием, нежели алгебра, то есть любая σ-алгебра является алгеброй, а обратное, вообще говоря, неверно. То есть это такая «элитка» среди всевозможных алгебр: не всем алгебрам дано быть «элиткой».

Пример. Пусть $\Omega = \mathbb{R}$, $\mathcal{A}$ содержит конечные подмножества Ω и их дополнения. Для такого множества выполнены все аксиомы алгебры: $\Omega = \overline{\varnothing} \in \mathcal{A}$, объединение конечных множеств есть конечное множество, объединение конечного множества с дополнением к конечному множеству так же является дополнением к некоторому множеству. То же можно сказать и об объединении двух дополнений. Таким образом, $\mathcal{A}$ является алгеброй. Все элементы $\mathcal{A}$ либо конечны, либо континуальны, поэтому $\mathcal{A}$ не содержит $\mathbb{N}$. Но $\mathbb{N} = \bigcup\limits_{i=1}^{\infty}\{i\}$, то есть не выполнено свойство счетной аддитивности из определения σ-алгебры. Вот видите: это не достойное нашего внимания семейство, оно не настолько элитное, чтобы называться сигма-алгеброй.

Вот мы и подошли к первому основополагающему определению.

Определение 1.4. Пара $(\Omega, \mathcal{A})$ называется *измеримым пространством*, если $\mathcal{A}$ является σ-алгеброй. Если же $\mathcal{A}$ — алгебра, то $(\Omega, \mathcal{A})$ — *измеримое пространство в широком смысле*.

По сути измеримое пространство нам говорит о том, какие у нас есть элементарные исходы и что мы считаем за событие. Но если при бросании кости ты будешь использовать кость со смещенным центром тяжести, то игроки что-то заподозрят и популярно пояснят тебе, что ты неправ (поверь, тебе это не понравится), так что имеет смысл определить не только множество событий, но и вероятность того или иного события. Мы этим и займемся дальше.

1.3. Вероятность. Вероятностное пространство

Ну чё, народ, погнали на!... На новые баррикады! Они не такие высокие, как те, которые мы уже перепрыгнули, так что следующие пару страниц вам покажутся легкой прогулкой теплым осенним вечером по желтеющему бульвару, залитому солнечными лучами заходящего солнца... Кхм-Кхм... Это из другой пьесы. Так о чем это мы? Да, теперь пора определить вероятность.

Определение 1.5. Пусть $(\Omega, \mathcal{A})$ — измеримое пространство ($\mathcal{A}$ — σ-алгебра). *Вероятностью* называется функция $\mathbb{P}\colon \mathcal{A} \to \mathbb{R}$, удовлетворяющая свойствам

1. $\forall A \in \mathcal{A} \mapsto \mathbb{P}(A) \geqslant 0$,
2. $\forall A_1, \ldots, A_n, \ldots \in \mathcal{A}, \quad A_i \cap A_j = \varnothing \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mathbb{P}(A_i)$,
3. $\mathbb{P}(\Omega) = 1$.

Второй пункт в определении вероятностной меры нельзя заменить аналогичным с конечными объединением и суммой. Однако если добавить к данному требованию так называемое свойство непрерывности вероятностной меры, т.е.

$$\forall B_1, B_2, \ldots \in \mathcal{A}, \quad B_{n+1} \subseteq B_n \Rightarrow \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(B), B = \bigcap_{n=1}^{\infty} B_n$$

то они вместе будут эквивалентны 2 из определения вероятности. Покажем это.

Утверждение 1.6.

$$\left[\forall A_1, \ldots, A_n, \ldots \in \mathcal{A}, \quad A_i \cap A_j = \varnothing \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} A_i\right] \Leftrightarrow$$
$$\left[\forall A_1, \ldots, A_n \in \mathcal{A} \quad A_i \cap A_j = \varnothing \ (i \neq j) \Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} A_i\right] \bigwedge$$
$$\left[\forall B_1, B_2, \ldots \in \mathcal{A}, \quad B_{n+1} \subseteq B_n \Rightarrow \lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(B), B = \bigcap_{n=1}^{\infty} B_n\right].$$

Доказательство.

$\boxed{\Rightarrow}$

Обозначим $C_n = B_n \setminus B_{n+1}$. Множества $B, C_1, C_2, \ldots$ не имеют общих точек.
$\forall n \ B_n = \left(\bigcup\limits_{k=n}^{\infty} C_k\right) \cup B$. Тогда $\mathbb{P}(B_1) = \mathbb{P}(B) + \sum\limits_{k=1}^{\infty} \mathbb{P}(C_k)$. Отсюда следует, что ряд в

правой части сходится, так как имеет конечную сумму. $\mathbb{P}(B_n) = \mathbb{P}(B) + \sum\limits_{k=n}^{\infty} \mathbb{P}(C_k)$. При $n \to \infty$ сумма ряда стремится к нулю как остаточный член ряда из предыдущего выражения. В предельном переходе получаем свойство непрерывности.

$\boxed{\Leftarrow}$

Рассмотрим произвольный набор $A_1, A_2, \ldots \in \mathcal{A} \quad A_i \cap A_j = \varnothing$.

$\mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) + \mathbb{P}\left(\bigcup\limits_{i=n+1}^{\infty} A_i\right) = \sum\limits_{i=1}^{n} \mathbb{P}(A_i) + \mathbb{P}\left(\bigcup\limits_{i=n+1}^{\infty} A_i\right)$.

Обозначим $B_n = \bigcup\limits_{i=n+1}^{\infty} A_i, \quad B_{n+1} \subseteq B_n \quad \forall n, \quad \bigcap\limits_{n=1}^{\infty} B_n = \varnothing$

$\sum\limits_{i=1}^{\infty} \mathbb{P}(A_i) = \lim\limits_{n\to\infty} \left(\mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) - \mathbb{P}(B_n)\right) = \mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right) - \lim\limits_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup\limits_{i=1}^{\infty} A_i\right)$. ■

Определение 1.7. *Вероятностным пространством* $(\Omega, \mathcal{A}, \mathbb{P})$ называется измеримое пространство $(\Omega, \mathcal{A})$, с заданной на нём вероятностью $\mathbb{P}$.

Ничего сложного, если знать определения и уметь вертеть математическими понятиями и логикой. Главное не перепутайте: вертеть, а не класть.

ЧО?! ЧЁ ЭТО ТАКОЕ?! Кому вообще нужна σ-алгебра событий и зачем весь этот огород и маленькая тележка, если можно рассматривать множество всех подмножеств множества событий Ω? Когда-то давно кто-то доказал, что в случае очень большого множества элементарных исходов, например, континуального (это много, поверьте наслово), множество 2^{Ω} будет иметь такую крокодильски, или алигаторски, большую мощность, что вся теория сломается. Таким образом, алгебры нужны для того, чтобы вероятность имела хорошую область определения.

Перечислим свойства вероятности. Доказательства их можно найти в любом из классических учебников по теории вероятностей или можно их придумать самому: большинство из них тривиальны. Да, реально тривиальны, а не как на матеше в школе.

Свойства вероятности:

1. $\mathbb{P}(\varnothing) = 0$,
2. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$,
3. $A \subseteq B \Rightarrow \mathbb{P}(A) \leqslant \mathbb{P}(B)$,
4. $\mathbb{P}(A) \leqslant 1$,
5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$,
6. $\mathbb{P}(A \cup B) \leqslant \mathbb{P}(A) + \mathbb{P}(B)$,
7. $\mathbb{P}\left(\bigcup\limits_{i=1}^{n} A_i\right) = \sum\limits_{k=1}^{n} \sum\limits_{i_1 < \ldots < i_k} (-1)^{k+1} \mathbb{P}(A_{i_1} A_{i_2} \ldots A_{i_k})$,
8. $\mathbb{P}\left(\bigcap\limits_{i=1}^{n} A_i\right) \geqslant 1 - \sum\limits_{i=1}^{n} \mathbb{P}\left(\overline{A_i}\right)$ — неравенство Бонферрони.

Примите к факту, что в силу простоты этих определений, математики используют их на уровне интуиции, то есть вы можете не заметить, а при этом в паре строк уже будет использовано несколько этих свойств. В некотором роде это и достоинство теории вероятности, так как происходит подобное из-за ясной бытовой аналогии, навязанной всем распространенностью азартных игр. С другой стороны, есть отличный способ избежать возможности быть посаженным на что-нибудь за отсутствие формализма: для этого стоит каждый раз, как перед тобой будут выписаны математические выкладки, пробовать осмыслить каждый переход, который в них сделан. Если это не получается сделать, то не стремись хвататься за голову: возможно, причина в том, что используются какие-то свойства, которые автор посчитал очевидными. Тогда стоит поискать в других источниках доказательства этого факта. Вполне вероятно, что это поможет найти такое доказательство, в котором будет все предельно понятно или в тяжелых моментах будет указано, что стоит перечитать и повторить.

В заключение вырежем клинышком в этой глиняной массе

Теорема 1.8 (Каратеодори)**.** Пусть $(\Omega, \mathcal{A})$ — измеримое пространство в широком смысле, а некоторая функция $\mathbb{P}$ обладает свойствами вероятностной меры. Тогда на измеримом пространстве $(\Omega, \sigma(\mathcal{A}))$

$$\exists!\ \mathbb{P}' : \forall A \in \mathcal{A} \mapsto \mathbb{P}(A) = \mathbb{P}'(A).$$

Доказательство. Ушло и не вернулось. Последний раз его видели здесь [2]. ■

Зочем? Теорема Каратеодори говорит о том, что любую вероятностную меру, заданную на алгебре, можно однозначно продолжить на σ—алгебру, то есть расширить область ее определения. При этом значения функции на алгебре не изменятся. Теорема будет использоваться при определении интеграла Лебега.

§2. Понятие меры и интеграла Лебега

2.1. Мера Лебега

Не пугайтесь. Сейчас мы введем меру Лебега, но не так быстро, как вводили вероятностное пространство. Представьте, что мера Лебега — это несколько воздушных шариков, надутых внутри друг друга. Самый маленький шарик — это полуинтервалы. Средний — борелевские множества (о них будет дальше). Большой — измеримые множества на прямой.

И задается мера итеративно, то есть сначала для маленького множества объектов, а потом все больше и больше.

Дадим определение семейства, которое нас будет сильно интересовать дальше.

Определение 2.1. *Борелевской σ-алгеброй* называется минимальная σ-алгебра, содержащая все открытые подмножества топологического пространства. Элементы борелевской σ-алгебры называются *борелевскими множествами.*

ЧО?! ЧЁ ЭТО ТАКОЕ?! Мы будем рассматривать только топологическое пространство $\mathbb{R}$, так что это стремное словосочетание можно прямо сейчас забыть и понимать открытое множество как открытое множество из матана (все точки внутренние).

Пример. Покажем, что все «хорошие» множества являются Борелевскими.

1. Все открытые интревалы входят по определению.
2. Отрезок вида $[a, b]$ входит как $\overline{(-\infty, a) \cup (b, +\infty)}$.
3. Точка ходит как вырожденный отрезок $[a, a]$.
4. Счетное объединение таких множеств входит по определению.

Определение 2.2. Теперь вспомним сначала аксиомы *счетно-аддитивной меры* μ из функана (злодейский смех): это функция на множестве B, $\mu\colon B \to \mathbb{R}$, со свойствами

1. $\forall A \in B \mapsto \mu(A) \geqslant 0$ (неотрицательность),
2. $\forall A_1, A_2, \ldots \in B\colon A_i \cap A_j = \varnothing \quad (i \neq j) \Rightarrow \mu\left(\bigcup\limits_{i=1}^{\infty} A_i\right) = \sum\limits_{i=1}^{\infty} \mu(A_i)$ (σ-аддитивность).

Начнем использовать это определение в самом простом случае. Введем сначала меру на борелевских множествах полуинтервала $[0, 1)$, то есть на $\mathscr{B}_{[0,1)}$. Итак,

- На интервалах (a, b) введем меру как $\mu_{[0,1)}((a, b)) = b - a$, то есть примерно если бы ваш сосед пришел бы с рулеткой к прямой и, померив, грубым голос добавил: «Здесь на вскидку $b - a$ получается.»
- Мера одной точки равна нулю, и мы можем не обращать внимание на концы множеств. Это так пусть будет по определению.

- На любых множествах, которые можно представить как конечное объединение интервалов и какого-то числа точек, мы зададим мера как сумму мер каждого из составляющих. Это как если бы вам надо было добраться по проспекту от точки A до точки B: прошли до остановки метров 100, сели в автобус, проехали пару километров, вышли, пересели в другой автобус, проехали на нем, вышли, и прошли до места назначения метров 200. Тогда множество точек проспекта (прямой), где вы шли или ждали автобус, будет иметь меру (длину) равную 300 метрам:
$$300 = 100 + 0 + 200.$$

- Теперь осталось определить меру на бесконечных объединениях/пересечениях. Для этого воспользуемся теоремой Каратеодори (1.8), согласно которой можно продолжить меру на σ-алгебру.

Аналогичным образом определим меру $\mu_{[k,k+1)}$ на всех полуинтервалах с целыми концами $[k, k+1)$, $(k \in \mathbb{Z})$. Осталось доопределить меру на всей $\mathscr{B}$.

$$\forall A \in \mathscr{B} \quad \mu(A) = \sum_{i=-\infty}^{\infty} \mu_{[i,i+1)}(A \cap [i, i+1)).$$

2.2. Интеграл Лебега

Теперь введем *интеграл Лебега* (интеграл по мере Лебега). В отличие от интеграла Римана, где происходит разбиение области определения и выбирается значение функции из образа элемента разбиения, в интеграле Лебега разбивается область значений (то есть Oy), и некоторое значение из элемента разбиения умножается на меру прообраза этого элемента, затем все благополучно складывается. Введем теперь это формально.

Будем рассматривать функции $f\colon \mathbb{R} \to \mathbb{R}$ такие, что $\forall c \in \mathbb{R}$ множество $A_c \, \{x\colon f(x) < c\}$ борелевское, то есть прообразы полупрямых являются борелевскими множествами.

Введем сначала понятие *индикатора*. Индикатор события A — это случайная величина, принимающая значение 1, если событие произошло, и 0 в противном случае. Таким образом
$$\mathbb{1}(x) = \begin{cases} 1, x \in A, \\ 0, x \notin A. \end{cases}$$

Определим интеграл Лебега на полуинтервале $[0, 1)$

- Пусть функция имеет вид $f(x) = \sum\limits_{i=1}^{k} y_i \mathbb{1}(A_i)$, где A_i не пересекаются и покрывают весь полуинтервал $[0, 1)$. Функции такого вида называются *финитными*. В этом случае
$$\int\limits_0^1 f(x)\mu(dx) = \sum_{i=1}^{k} y_i \mu(A_i).$$

  То есть берется мера кусочка, на котором функция принимает значение y_i, и умножается на это значение. Получается нечто, напоминающее площадь под графиком.

- Пусть теперь $f(x) \geqslant 0$ на $[0,1)$. Будем приближать функцию финитной. Возьмем отрезок области значений $[0,n], n \in \mathbb{N}$. Разобьем этот отрезок на $n2^n$ кусочков $[\frac{k-1}{2^n}, \frac{k}{2^n}), k = 1, 2, \ldots, n2^n$. На каждом кусочке скажем, что значение функции равно $\frac{k-1}{2^n}$. Осталось приблизить $[n, +\infty)$. Будем считать это одной частью, на которой функция принимает значение n. Устремим n к бесконечности: тогда, так как размер каждого кусочка первой части равен $\frac{1}{2^n}$, их длина станет бесконечно малой, а та часть, которую мы «обрубили» сверху (то есть $[n, +\infty)$), уйдет в бесконечность. Итак,

$$\int\limits_0^1 f(x)\mu(dx) = \lim_{n\to\infty}\left[\sum_{k=1}^{n2^n}\frac{k-1}{2^n}\mu\left(\left\{x\colon \frac{k-1}{2^n} \leqslant f(x) < \frac{k}{2^n}\right\}\right) + n\mu(\{x\colon f(x) \geqslant n\})\right].$$

- Остался случай произвольной функции $f(x)$. Разобьем ее на две: одна функция совпадает с $f(x)$ там, где та положительна, и равняется 0 в остальных случаях, другая равна $|f(x)|$ там, где $f(x)$ отрицательна, и 0 иначе.

$$f^+(x) = \max\{0, f(x)\}, \quad f^-(x) = -\min\{0, f(x)\}.$$

  Для каждой из этих функций интеграл определен по предыдущему пункту. Пользуясь тем, что $f(x) = f^+(x) - f^-(x)$, определим интеграл так:

$$\int\limits_0^1 f(x)\mu(dx) = \int\limits_0^1 f^+(x)\mu(dx) - \int\limits_0^1 f^-(x)\mu(dx).$$

  В случае, если оба интеграла в правой части расходятся, интеграл от $f(x)$ не определен. Так как $|f(x)| = f^+(x) + f^-(x)$, сходимость интеграла Лебега от модуля функции (абсолютная сходимость) эквивалента сходимости интеграла от самой функции (то есть обычной сходимости). Это следует из того, что интеграл сходится только в случае конечности обоих интегралов правой части (иначе он не определен), откуда следует конечность их суммы и разности. Таким образом, для интеграла Лебега не существует условно сходящихся функций.

Аналогично вводим интеграл на каждом полуинтервале $[i, i+1), i \in \mathbb{Z}$. Тогда на всей прямой интеграл по множеству $A \subset \mathbb{R}$ будет определяться так:

$$\int\limits_A f(x)\mu(dx) = \sum_{i=-\infty}^{+\infty}\int\limits_{A\bigcap[i,i+1)} f(x)\mu(dx).$$

Пример. С помощью интеграла Лебега можно считать интегралы от функций, об интегрировании которых раньше было страшно даже подумать. Например, от функции Дирихле:

$$D_{[0,1)}(x) = \begin{cases} 1, x \in [0,1) \setminus \mathbb{Q}, \\ 0, x \in \mathbb{Q}. \end{cases}$$

Данная функция является финитной, а именно $D_{[0,1)}(x) = \mathbb{1}([0,1) \setminus \mathbb{Q})$. Поэтому по первому пункту

$$\int\limits_0^1 D_{[0,1)}(x)\mu(dx) = 1.$$

В дальнейшем $\mu(dx)$ будет опускаться обозначаться просто как dx или dy чтобы подчеркнуть, что считается именно интеграл Лебега.

§3. Связь между событиями

3.1. Условная вероятность

Условно мы будем тут работать с вероятностью нескольких событий. Ведь бывают разные игры: и в карты, и в кости. Есть и такие, в которых придется не единожды просить судьбу «выдать козырь», поэтому выгодно понимать, сколько в действительности у тебя шансов дёрнуть судьбу так, как тебя тогда на рынке дёрнули.

Рассмотрим произвольное событие $B \in \mathcal{A}$: $\mathbb{P}(B) > 0$.

Определение 3.1. *Условной вероятностью* события $A \in \mathcal{A}$ при условии B называется

$$\frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \overset{\text{def}}{=} \mathbb{P}(A|B) = \mathbb{P}_B(A).$$

Что это означает на пальцах? Спокойно, сейчас поясним. Условная вероятность $\mathbb{P}(A|B)$ — это вероятность того, что произойдет событие A, если мы точно знаем, что произошло событие B.

Пример. Представь, что ты бросаешь 2 кубика. Событие A — у тебя выпало суммарно меньше 5 очков. А B, например, означает, что у тебя выпало 6 очков на первом кубике. Очевидно, что если на одном выпало 6, то суммарно у тебя меньше 6 быть не может, так что вероятность тоже должна быть равна нулю, что согласуется с определением:

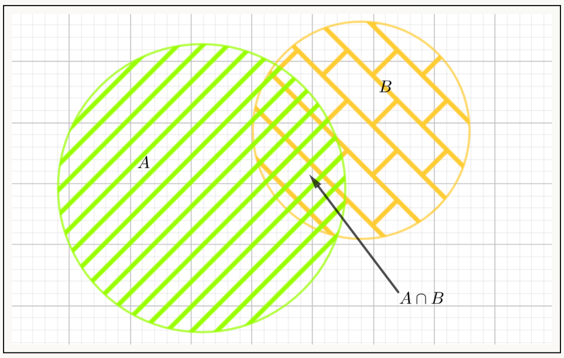
$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{0}{P(B)} = 0.$$

Допустим теперь, что B — у тебя выпало на первом кубике 1 очко. Тогда понятно, что вероятность события A — будет равна $1/2$, так как при выпадении $1, 2, 3$ на втором кубике событие произойдет, а при выпадении других чисел — нет. Вычислим теперь по определению.

Вероятность выпадения 1 на первом кубике, притом чтобы сумма была меньше пяти, равна $1/12$, так как всего 36 комбинаций на кубиках, а подходит только 3: $(1,1), (1,2), (1,3)$. Вероятность выпадения любого из чисел на кубике — $1/6$, так что

$$P_B(A) = \frac{P(AB)}{P(B)} = \frac{1/12}{1/6} = \frac{1}{2}.$$

Опа! Америка-Европа! Все сошлось с интуитивными предположениями.



Графически это означает, что, когда произошло событие B, мы оказались в круге B. Тогда формула

$$\frac{\mathbb{P}(AB)}{\mathbb{P}(B)}$$

есть просто вероятность попасть в $A \cap B$.

Из определения следует так называемый «Закон умножения вероятностей»:

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(AB).$$

Легко проверяется, что $(B, \mathcal{A}_B, \mathbb{P}_B)$, где $\mathcal{A}_B = \{A \cap B \colon A \in \mathcal{A}\}$, так же является вероятностным пространством.

3.2. Независимость событий. Независимость в совокупности

Да, и в теории вероятности есть независимость! Но это уже другая независимость, а не та, про которую так много новостей нынче.

Определение 3.2. События $A, B \in \mathcal{A}$ называются *независимыми*, если

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$$

.

Для независимых событий

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

Пример. Являются ли несовместные события ($A \cap B = \varnothing$) независимыми? Ответ: нет, пусть $A, B \in \mathcal{A}\colon \mathbb{P}(A) > 0,\ \mathbb{P}(B) > 0$. Тогда $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B) = 0$, что является противоречием. По-простому, если произошло одно из несовместных событий, то второе уже не может произойти, и его условная вероятность равна 0, а не вероятности самого события, что требуется для независимости.

Обобщим понятие независимости на произвольное количество событий.

Определение 3.3. События $A_1, A_2, \ldots, A_n$ называются *независимыми в совокупности*, если

$$\forall\, m = 2, \ldots, n \quad \forall\, 1 \leqslant j_1 < \ldots < j_m \leqslant n \mapsto \mathbb{P}\left(\bigcap_{k=1}^{m} A_{j_k}\right) = \prod_{k=1}^{m} \mathbb{P}\left(A_{j_k}\right).$$

Пример. На примере тетраэдра Бернштейна можно убедиться в том, что попарной независимости событий недостаточно для независимости в совокупности. Рассмотрим тетраэдр, у которого три стороны покрашены в красный, синий и зеленый, а четвертая содержит все три цвета. События «выпадет красный» = {К}, «выпадет синий» = {С}, «выпадет зеленый» = {З} попарно независимы (например, вероятность события {С}∩{К} равна вероятности выпадения четвертой грани, т. е. $1/4$, в то время как вероятность выпадения каждого цвета равна $1/2$). Однако

$$\mathbb{P}(\{\text{С}\} \cap \{\text{К}\} \cap \{\text{З}\}) = \frac{1}{4} \neq \left(\frac{1}{2}\right)^3.$$

3.3. Формула полной вероятности. Формула Байеса

Чуть выше мы рассмотрели условную вероятность, то есть поговорили о том, что некоторые события могут пересекаться и когда в таких случаях можно сказать, что они независимы. Часто встречается следующая группа событий.

Определение 3.4. $B_1, \ldots, B_n$ образуют *полную группу*, если выполнены следующие условия:

1. $\mathbb{P}(B_i) > 0 \quad \forall i = 1, \ldots, n$,
2. $B_i B_j = \varnothing \quad (i \neq j)$,
3. $\bigcup\limits_{i=1}^{n} B_i = \Omega$.

Теорема 3.5. Пусть $B_1, \ldots, B_n$ образуют полную группу. Вероятность события $A \in \mathcal{A}$ можно вычислить по *формуле полной вероятности*:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Доказательство.

$$A = \bigcup_{i=1}^{n} AB_i, \quad AB_i \cap AB_j = \varnothing \quad (i \neq j),$$

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(AB_i) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

Последний переход следует из закона умножения вероятностей. ∎

Первое требование определения полной группы необходимо для возможности определить условную вероятность, второе позволяет разбить множество A на непересекающиеся части. Третье требование, вообще говоря, можно ослабить, потребовав, чтобы $A \subseteq \bigcup\limits_{i=1}^{n} B_i$. Доказательство при этом не изменится.

Пример. Проиллюстрировать формулу полной вероятности можно обычным экзаменом: A — {студент сдал экзамен}, B_i — {студент попал к преподавателю i}. Как и в любой другой лотерее, можно оценить вероятность попадания к преподавателю i, то есть $\mathbb{P}(B_i)$, а трезво оценивая свои силы можно прикинуть и вероятность сдать тому или иному преподавателю $\mathbb{P}(A|B_i)$. Зная все вышеперечисленное, несложно по формуле вычислить вероятность успешной сдачи.

Формула полной вероятности используется для вычисления априорной вероятности, т.е. вероятности события, которое еще не произошло. Пусть теперь $\mathbb{P}(A) > 0$. Тогда

$\mathbb{P}(B_i|A) = \mathbb{P}(AB_i)/\mathbb{P}(A)$. Используя формулу полной вероятности, получаем *формулу Байеса*:

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum\limits_{j=1}^{n} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

Формула Байеса используется для вычисления апостериорной вероятности. То есть уже известно, что произошло некоторое событие A, и нужно вычислить вероятность того, что произошло некоторое B_i. В примере с экзаменом, например, может быть известно, что студент не сдал экзамен, и хочется вычислить вероятность того, что он сдавал преподавателю «P».

§4. Случайные величины

4.1. Случайные величины: определение

Случайные события — это хорошо, но события типа «на монетке выпал герб» плохо формализуемы, а мы хотим формальности и математичности. Поэтому вместо всяких событий хочется работать с числами. Вот этим и займемся. При рассмотрении случайных событий мы ввели вероятностное пространство, которое выглядит так:

$$(\Omega, \mathcal{A}, \mathbb{P}),$$

где Ω — множество элементарных событий, $\mathcal{A}$ — σ-алгебра подмножеств множества элементарных событий, а $\mathbb{P}$ — вероятность. Мы же будем рассматривать теперь тройку

$$(\mathbb{R}, \mathscr{B}, \mathbb{P}),$$

где $\mathbb{R}$ — действительная прямая, $\mathscr{B}$ — борелевская σ-алгебра, а $\mathbb{P}$ — вероятность.

Теперь формально введем понятие случайной величины (может использоваться сокращение с.в.).

Определение 4.1. Пусть $(\Omega, \mathcal{A}, \mathbb{P})$ — вероятностное пространство. Тогда *случайной величиной* ξ называется функция $\xi : \Omega \to \mathbb{R}$, измеримая относительно $\mathcal{A}$ и $\mathscr{B}$. По-другому, ξ — случайная величина, если

$$\forall B \in \mathscr{B} \quad \xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathcal{A}.$$

ЧО?! ЧЁ ЭТО ТАКОЕ?! Таким финтом ушами мы, по сути, сопоставили каждому событию какое-то «хорошее» множество на числовой прямой, и можем рассматривать не вероятности событий, а вероятности попадания в эти «хорошие» подмножества числовой прямой.

4.2. Функция распределения

Введем еще несколько «полезных» определений, которые в дальнейшем использоваться не будут, но знать их не вредно.

Определение 4.2. С каждой случайной величиной свяжем два вероятностных пространства: первое — $(\Omega, \mathcal{A}_\xi, \mathbb{P})$ — вероятностное пространство, *порожденное* ξ. Здесь $\mathcal{A}_\xi$ - наименьшая σ-алгебра, для которой выполняется свойство измеримости. Второе — $(\mathbb{R}, \mathscr{B}, \mathbb{P}_\xi)$, где $\mathbb{P}_\xi(B) = \mathbb{P}(\xi^{-1}(B)) \quad \forall B \in \mathscr{B}$ и называется *распределением вероятностей* ξ.

Идем дальше в сторону упрощения работы со случайностями. Вместо того чтобы рассматривать произвольные борелевские множества, мы будем рассматривать только множества вида $(-\infty, x)$. Действительно, интервал (a, b) получается из полупрямых так: $(a, b) = (-\infty, b) \setminus (-\infty, a]$. Таким образом, мы можем рассматривать случайные величины только на таких множествах. Здесь имеется в виду, что для удовлетворения определению случайной величины достаточно измеримости только на полупрямых, что следует из следующих свойств полного прообраза: прообраз объединения есть объединение прообразов, прообраз пересечения есть пересечение прообразов, прообраз отрицания есть отрицание прообраза. Выше показано, что из полупрямых с помощью этих операций можно получить интервалы, а из интервалов и все $\mathscr{B}$.

Теперь несколько полезных утверждений. Пусть ξ — случайная величина. Тогда $-\xi$ также случайная величина, так как её прообраз от любой полупрямой является прообразом ξ от симметричной полупрямой, то есть лежит в $\mathcal{A}$. Величина $\xi + c$ также будет случайной величиной, поскольку ее прообразом для любой полупрямой будет прообраз ξ для полупрямой, сдвинутой на c, то есть будет лежать в $\mathcal{A}$.

Утверждение 4.3. Пусть ξ, η — случайные величины. Тогда множество $\{\omega \colon \xi(\omega) < \eta(\omega)\}$ является событием.

Доказательство. $\{\omega \colon \xi(\omega) < \eta(\omega)\} = \bigcup\limits_{r \in \mathbb{Q}} \{\omega \colon \xi(\omega) < r, \eta(\omega) > r\}$. Заметим, что $\{\omega \colon \xi(\omega) < r\}$ является событием. Аналогично для η. Выражение, написанное выше, является счетным объединением пересечений двух событий, то есть событием. ■

Похожими махинациями, а также с использованием этого утверждения, доказывается, что $\xi^2, \xi + \eta, \xi\eta$ являются случайными величинами. Более того, если $\xi_1, \ldots, \xi_n$ — с.в., а функция $\varphi(x_1, \ldots, x_n)$ является непрерывной на множестве их значений, то $\varphi(\xi_1, \ldots, \xi_n)$ будет случайной величиной.

Определение 4.4. Рассмотрим вероятностное пространство $(\Omega, \mathcal{A}, \mathbb{P})$ и определенную на нем случайную величину ξ. Тогда её *функцией распределения $F_\xi(x)$* называется функция $F_\xi : \mathbb{R} \to \mathbb{R}$

$$F_\xi(x) = \mathbb{P}(\omega : \xi(\omega) < x) = \mathbb{P}(\xi < x)$$

.

Запись $\mathbb{P}(\xi < x)$ является в некотором смысле жаргонной, так как аргументов вероятности должно быть событие из $\mathcal{A}$. Но $\xi < x$ мы в дальнейшем будем отождествлять с объединением элементарных событий, образ которых меньше x. Из определения случайной величины получаем, что это объединение является событием, поэтому применение к нему функции вероятности корректно.

Функция распределения (сокращение ф.р.) является очень полезной штукой, поскольку имеет достаточно простой вид и несет в себе всю информацию о распределении, то есть однозначно определяет $\mathbb{P}_\xi$.

Рассмотрим основные свойства функции распределения:

1. $F_\xi(x) \in [0, 1]$
2. $\lim\limits_{x \to -\infty} F_\xi(x) = 0$
3. $\lim\limits_{x \to +\infty} F_\xi(x) = 1$
4. $F_\xi(x)$ монотонно не убывает.
5. $F_\xi(x)$ непрерывна слева.

Вероятность попадания с.в. в полуинтервал $\mathbb{P}_\xi[a, b) = F_\xi(b) - F_\xi(a)$. При стремлении $b \to a$ получим $\mathbb{P}(\xi = a) = F_\xi(a + 0) - F_\xi(a)$, то есть вероятность попадания в точку равна скачку функции распределения в этой точке.

Определение 4.5. Точка x_0 называется *точкой роста* $F_\xi(x)$, если $\forall \varepsilon > 0$ $\mathbb{P}(x_0 - \varepsilon \leqslant \xi < x_0 + \varepsilon) > 0$

Пример. Это очень полезный пример, который будет использоваться в матстате и который очень любят спрашивать. Пусть ξ — случайная величина. $\eta = F_\xi(\xi)$. Чему равна $F_\eta(x)$? По определению

$$F_\eta(x) = \mathbb{P}(\eta < x) = \mathbb{P}(F_\xi(\xi) < x) = \mathbb{P}(\xi < F_\xi^{-1}(x)) = F_\xi(F_\xi^{-1}(x)) = x \qquad (4.1)$$

Вообще, тут было бы неплохо сказать, что F_ξ непрерывна и строго монотонна, чтобы со спокойной совестью использовать обратную функцию. Таким образом η имеет равномерное распределение.

4.3. Виды распределений

Распределения случайных величин можно разделить на 3 типа: непрерывные, дискретные и сингулярные.

Определение 4.6. Случайная величина ξ называется *абсолютно непрерывной*, если существует интегрируемая функция $p_\xi(x) \geqslant 0$, $x \in \mathbb{R}$ такая, что функция распределения ξ является почти всюду (за исключением не более, чем счетного числа точек) дифференцируемой функцией и представима в виде

$$F_\xi(x) = \int\limits_{-\infty}^{x} p_\xi(y)dy$$

Отсюда следует, что функция распределения непрерывна на $\mathbb{R}$. $p_\xi(x)$ называется *плотностью распределения*, и почти всюду выполнено $p_\xi(x) = F'_\xi(x)$. Плотность, вообще говоря, определена не однозначно.

Определение 4.7. Случайная величина ξ называется *дискретной*, если множество точек роста не более, чем счетно, но распределение не является сингулярным, или, другими словами $\exists B = \{x_1, x_2, \ldots\} \colon \mathbb{P}(\xi \in B) = 1$.

Определение 4.8. Случайная величина ξ называется *сингулярной*, если F_ξ непрерывна, и $\exists B \in \mathscr{B} \colon \mu(B) = 0$, $\mathbb{P}(\xi \in B) = 1$, то есть множество значений случайной величины имеет меру 0, но вероятность попасть в каждую точку этого множества так же нулевая.

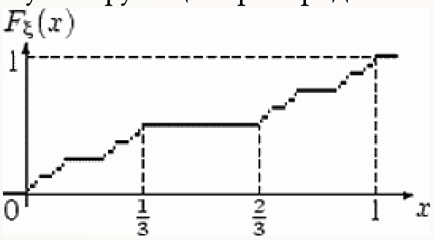
Пара слов о жизненном смысле определений: непрерывная случайная величина имеет областью значений континуальное множество, при этом вероятность попасть в отдельно взятую точку нулевая. Пример: равномерное распределение по отрезку.

Плотность же отражает вероятность попасть в ту или иную область: интеграл по области равен этой вероятности. Дискретная случайная величина принимает конечное или счетное множество значений, вследствие этого имеет ступенчатую функцию распределения, например, бросок монетки имеет дискретное распределение. Сингулярное распределение — это крокодил, который не встречается в жизни и будет рассмотрен отдельно.

Утверждение 4.9. Дискретная случайная величина имеет не более, чем счетное число скачков.

Доказательство. Из свойств функции распределения следует, что дискретная величина имеет не больше двух скачков величины больше $\frac{1}{2}$. Аналогично, скачков величины больше $\frac{1}{3}$ не больше 3. То есть скачков величины больше $\frac{1}{n}$ не более n. Для любого скачка можно указать $n \in \mathbb{N}$ такое, что величина, этого скачка больше $\frac{1}{n}$. Значит,каждому скачку можно поставить в соответствие n, множество которых счетно. При этом для каждого n существует не более чем счетное число скачков, ему соответствующих (величины $> \frac{1}{n}$). А так как объединение не более, чем счетного числа не более, чем счетных множеств, не более, чем счетно, получаем требуемое. ■

Пример. Для полного счастья приведем пример сингулярной случайной величины. Пусть функция распределения — так называемая лестница Кантора (см. рисунок).



Посчитаем меру множества, на котором функция константа, то есть точки этого множества не будут точками роста: сначала это одна ступенька длины $1/3$, потом две длины $1/9$, и т. д.

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} = \frac{1}{3}\sum_{k=1}^{\infty}(\frac{2}{3})^{k-1} = 1.$$

Тогда множество точек роста имеет меру 0 в силу свойства аддитивности меры.

Вообще говоря, существуют менее изысканные примеры сингулярных распределений. Например, при стрельбе из лука в круглую мишень распределение будет сингулярным, если стрелок попадает только в точки одной прямой. В самом деле, двумерная мера прямой равна 0, как и вероятность попасть в каждую отдельную точку.

Теорема 4.10 (Лебега). Любую случайную величину можно представить в виде суммы дискретной, абсолютно непрерывной и сингулярной случайной величины. То есть

$$F(x) = \alpha_d F_d(x) + \alpha_c F_c(x) + \alpha_s F_s(x), \quad \alpha_d + \alpha_c + \alpha_s = 1.$$

Доказательство. Не вошло в содержание и не вышло с публикацией. Ищите в других учебниках. ∎

Список литературы

[1] Королев В. Ю. Теория вероятностей и математическая статистика: учеб. М.: ТК Вебли, Изд-во Проспект, 2006. с. 160.

[2] Гусев Н. А. Заметки к курсу: мера и интеграл Лебега. 2019. https://www.ngusev.ru/mt/mt-lectures.pdf.