Некоторый списочек литературы

Базовые для теории.

четкий короткий Логинов М.И. Теория вероятностей. Уч. пособие. Екатеринбург

четкий доходчивый Чернова Н.И. Теория вероятностей. Уч. пособие. Новосибирск

кому летать охота Босс В. Том 4. Вероятность. Информация. Статистика.

олдскул Колмогоров А.Н. Основные понятия теории вероятностей

The Best Феллер В. Введение в теорию вероятностей и ее приложения (в 2 томах)

неотвратимая глубина бытия Ширяев А.Н. Вероятность (в 2 томах)

альпинистам Синай Я.Г. Коралов Л.Б. Теория вероятностей и случайные процессы

Дополнительно

цвет по вкусу Гнеденко Б.В. Курс теории вероятностей

цвет по вкусу Тутубалин В.Н. Теория вероятностей и случайных процессов

цвет по вкусу Боровков А.А. Теория вероятностей

цвет по вкусу Кремер Н.Ш. Теория вероятностей и мат статистика

цвет по вкусу Гихман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика

скорее как задачники

для разгона Мостеллер Ф. Пятьдесят занимательных вероятностных задач с решениями

практики-гробовщики Стохастический анализ в задачах. Ч. I

математики-гробовщики Ширяев А.Н. Задачи по теории вероятностей

легкий британский юмор Кельберт М. Я., Сухов Ю. М. Вероятность и статистика в примерах и задачах (3 тома, нам первого хватит за глаза)

для разшаривания межушного пространства

интуиции ради Секей Г. Парадоксы в теории вероятностей и математической статистике

тем кто все забыл Виленкин Н.Я. Комбинаторика

список не прошел и половины...

...темный лес...

Обо всем и ни о чем...

Цель курса — разобраться в основных понятиях теории вероятностей, случайных процессов и математической статистики.

Цель первого семестра — понятия “вероятностное пространство”, “случайная величина”, “(условное) математическое ожидание”.

Первое из них — вероятностное пространство.

Вероятностное пространство $(\Omega, \mathcal{F}, P)$

О вероятности наивно

Попросту вероятностное пространство $(\Omega, \mathcal{F}, P)$ — это тройка:

Ω — множество элементарных событий (исходов), некоторое непустое множество. Может иметь любую природу, любую мощность;

$\mathcal{F}$ — пока не уточняем, что это такое, наверно события;

P — вероятность, пока некая функция.

1-й наивный способ задания вероятности: Пусть множество Ω конечно, его мощность $|\Omega| = n$. Будем считать, что все исходы равновероятны. Тогда для всякого $\omega \in \Omega$ зададим

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|} = \frac{1}{n}.$$

Примеры

Уточняем: наверное, P — функция из Ω в $\mathbb{R}$.

Пример 1. ("правильная монета") Подбрасываем монетку; $\Omega = \{O, P\}$. Вероятности орла и решки по $1/2$.

Пример 2. ("правильный кубик") Подбрасываем кубик; $\Omega = \{1, 2, 3, 4, 5, 6\}$. Вероятности всех исходов по $1/6$.

Пример 2' ("кубики") Подбрасываем два кубика. Найти вероятность того, что сумма равна 2.

Выпишем различные суммы: $\Omega = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, $|\Omega| = 11$. Вывод: наша наивная вероятность равна $1/11$.

Формально правильно, а по существу — издевательство. Обычный. не квантовый, мир устроен сложнее, воздух и Авогадро тому свидетели.

Пример 2" ("правильные кубики") Подбрасываем два кубика. Найти вероятность того, что сумма равна 2.

Примем $\Omega = \{1, \ldots, 6\}^2$, теперь $|\Omega| = 36$. Поскольку нас интересует вероятность выпадения пары $(1, 1)$ получаем правильный ответ: $1/36$

Мы моделировали (прибивали практику к определению) разными способами. Математика законна в каждом примере, все остальное на совести моделирующего.

Пример 2"' ("правильные кубики") Подбрасываем два кубика. Найти вероятность того, что сумма равна 7.

Поскольку нас интересуют пары $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ получаем ответ: $6/36 = 1/6$

Пример 3 Достаем наудачу одно домино. Найти вероятность того, что число точек на доминошке равно 7?

Все похоже, но $\Omega = \{\text{неупорядоченные пары чисел от 0 до 6}\}$ другое, теперь $n = 28$. Нам годится 3 варианта. Вероятность равна $3/28$.

В двух последних примерах мы приписали вероятность не элементарному исходу. Надо уточнить.

Что хочется?

Наверное, P — функция из множества 2^{Ω} всевозможных подмножеств множества Ω в $\mathbb{R}$. Назовем пока эти подмножества событиями. $\mathbb{P}(A)$ — вероятность того, что произошло событие A.

Объявим хотелки:

Хочу0 $\mathbb{P}(\varnothing) = 0$ (невозможное событие)

Хочу1 $\mathbb{P}(\Omega) = 1$ (достоверное событие)

Хочу2 Из $A \subset B$ следует $\mathbb{P}(A) \le \mathbb{P}(B)$.

На самом деле мы еще доказали, что P - это функция в $[0,1]$. Почему? Потому, что монотонность.

Все хотелки выполнены например в...

1й поправленный способ задания вероятности (классическая вероятность): эксперимент удовлетворяет классическому определению вероятности, если Ω состоит из конечного числа равновозможных исходов. При этом вероятность события A задается отношение числа исходов в событии A к общему числу исходов.

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

А если точек много...

Пример 4. Палку случайно разломили на два куска. Найдите вероятность того, что один из кусков не больше другого.

Наивное определение уже не годится, Ω уже бесконечно. Впрочем понятно, что вероятность равна 0. Более того, это событие вероятности ноль теоретически осуществимо (не невозможно). Не менее невозможно, чем выбрать наугад среди всех точек на отрезке заданную.

Собственно на идее “по одиночке звать никак, а вместе мы сила” основан

2й наивный способ задания вероятности: Пусть Ω — область в плоскости, а A — её непустая подобласть. Назовем вероятностью события A отношение площадей: $\mathbb{P}(A) = \frac{S(A)}{S(\Omega)}$.

Все указанные выше хотелки снова выполнены. Ну если $meas(\Omega)$, конечно, конечно и положительно: не бесконечность, и не ноль.

О случайной хорде

Пример 5. В окружности случайно провели хорду. Найти вероятность того, что ее длина больше стороны вписанного в окружность равностороннего треугольника.

Решение 1. Хорду восстановим по ее середине. Середина — произольная точка на круге. Ω — большой круг. Вписанный в треугольник круг — в точности то, что требуется. Поскольку их радиусы — части одной медианы

точка пересечения медиан делит каждую медиану в отношении 1 к 2, то отношение площадей кругов равно $(1/2)^2$. Ответ: $1/4$.

Решение 2. Вместо хорды рассматриваем её середину. Середину понимаем как случайно взятые точку на окружности S^1 и точку на радиусе, проведенном к первой точке. Теперь $\Omega = [0,1] \times S^1$. Годятся точки, до которых от центра расстояние меньше половины радиуса (точка пересечения медиан делит каждую медиану в отношении 1 к 2). Ответ: $1/2$.

Решение 3. Вместо хорды выбираем ее концы (или точку и центральный угол, по часовой, от нее). Получаем $\Omega = S^1 \times S^1$ или $\Omega = [0, 2\pi] \times S^1$ соответственно. Выбор первой точки вероятность не меняет, второй выбор нас устраивает в трети случаев. Ответ: $1/3$.

Решение 4. Вместо хорды рассматриваем точку и длину хорды. $\Omega = S^1 \times [0,2]$. От точки ничего не зависит. Нужна длина не меньше длины стороны вписанного треугольника. Ответ: $1 - \sqrt{3}/2$.

Какой ответ правильный? Парадокс... (парадокс Бертрана, если что).

Вывод: не важно, что сказали в условии, важно как отображаешь (моделируешь).

Да, в решении 5.3 мы искали площадь у четырехмерного объекта.

Геометрическая вероятность

2й поправленный способ задания вероятности (геометрическая вероятность): эксперимент удовлетворяет геометрическому определению вероятности, если все исходы можно изобразить в виде некоторой области Ω конечномерного пространства, а вероятность подмножества зависит только от его меры (длины, площади, объема, ... n-мерного объема) этого подмножества и задается по формуле $\mathbb{P}(A) = \frac{meas(A)}{meas(\Omega)}$.

Но как определить что такое область? Какой именно взять объем? Что такое изобразить? Что такое не зависит от формы и расположения?... А если Ω — фрактал, например. Или еще проще — прямая.

Подумать: задать какую-нибудь вероятность на прямой (как на Ω), ну или, например, на множестве натуральных чисел. А вот еще бы равномерную выдумать...

Подумать: сколько нулей надо сложить, чтобы получить единицу?

Фишка теории вероятностей — умение делать суждения о вероятностях сложных событий из вероятностей простых событий (поддающихся опыту).

Всякое осуществление определенных условий и действий, при которых наблюдается изучаемое случайное явление будем называть опытом (экспериментом). Любую качественную характеристику опыта назовем событием. Любую количественную характеристику опыта назовем случайной величиной.

О событиях

Попросту говоря, событие — суждение с ответом да/нет. При этом точки ω из Ω можно трактовать как возможные исходы некоторого случайного эксперимента. Каждое подмножество множества Ω теперь связано с некоторым событием, при этом утверждение $\omega \in A$ можно трактовать как "при исходе эксперимента ω произошло событие A". Если всегда когда происходит событие A, то происходит и событие B — про такую пару событий можно писать $A \subset B$.

$\varnothing, \Omega \in \mathcal{F}$ — невозможное и достоверное событие.

Если одновременно события A, B не наступают, то они называются взаимоисключающими (или несовместными)

Если A — некоторое событие - то есть $A \in \mathcal{F}$, то имеется событие не $A = \overline{A} = \Omega \setminus A$, противоположное событие.

Обозначим через $A \setminus B$ событие, заключающееся в том, что " A произошло, а B — нет".

Обозначим через $A \cap B$ событие: "наступило и событие A, и событие B".

Обозначим через $A \cup B$ событие: "произошло хотя бы одно из событий A, B"

Обозначим через $\cap_{i=1}^{k} A_i$ событие: "наступило каждое из событий $A_1, A_2, \ldots, A_k$."

Обозначим через $\cup_{i=1}^{k} A_i$ событие: "наступило хотя бы одно из событий $A_1, A_2, \ldots, A_k$."

Обозначим через $\cap_{i \in \mathbb{N}} A_i$ событие: "наступило каждое из событий $A_1, A_2, \ldots$"

Обозначим через $\cup_{i \in \mathbb{N}} A_i$ событие: "хотя бы для одного $k \in \mathbb{N}$ наступило A_k."

Обозначим через $\cap_{\alpha \in \mathcal{A}} A_\alpha$ событие: "события A_α наступили для каждого $\alpha \in \mathcal{A}$."

Обозначим через $\cup_{\alpha \in \mathcal{A}} A_\alpha$ событие: "хотя бы для одного $\alpha \in \mathcal{A}$ событие A_α наступило".

Подумать: События $\cap_{k \in \mathbb{N}} \cup_{i \geq k} A_i \in \mathcal{F}$ и $\cup_{k \in \mathbb{N}} \cap_{i \geq k} A_i \in \mathcal{F}$ называются верхним и нижним пределом последовательности событий $(A_i)_{i \in \mathbb{N}}$. Первое из них — "произошло бесконечное число этих событий (они никогда не кончатся)", второе — "начиная с некоторого произошли все эти события (не произошло только конечное число событий)".

Что хочется от $\mathcal{F} \neq 2^\Omega$

Писали раньше $P : 2^\Omega \to [0,1]$, но лучше $P : \mathcal{F} \to [0,1]$. Пока понимаем под $\mathcal{F}$ — в точности множество тех событий, на которых удастся задать вероятность. Иногда именно элементы из $\mathcal{F}$ называют событиями, иногда событиями называют произвольные подмножества множества Ω. Совсем корректно, но малоосмысленно, говорить об элементах из $\mathcal{F}$ как о ($\mathcal{F}$-)измеримых событиях.

Итак, пусть Ω — множество элементарных исходов, тогда всякое подмножество множества Ω является событием.

Что при этом хотелось бы:

x0 $\varnothing \in \mathcal{F}$;

x1 $\Omega \in \mathcal{F}$;

x2 $\forall A \in \mathcal{F}\ \Omega \setminus A = \bar{A} \in \mathcal{F}$;

x2п $\forall A \in \mathcal{F}\ \exists k \in \mathbb{N},\ A_1, A_2, \ldots, A_k \in \mathcal{F}\ \Omega \setminus A = \cup_{i=1}^{k} A_i$;

x3 $\forall A, B \in \mathcal{F}\ B \setminus A \in \mathcal{F}$;

x3п $\forall A, B \in \mathcal{F}\ \exists k \in \mathbb{N},\ A_1, A_2, \ldots, A_k \in \mathcal{F}\ B \setminus A = \cup_{i=1}^{k} A_i$;

x4 $\forall A,B\in\mathcal{F}\ A\cap B\in\mathcal{F}$;

x4k $\forall k\in\mathbb{N}\ A_1,A_2,\dots,A_k\in\mathcal{F}\ \cap_{i=1}^{k}A_i\in\mathcal{F}$;

x4 σ $\forall A_1,A_2,\dots,A_k,\dots\in\mathcal{F}\ \cap_{i\in\mathbb{N}}A_i\in\mathcal{F}$;

x4 α $\forall (A_\alpha)_{\alpha\in\mathcal{A}}\in\mathcal{F}^{\mathcal{A}}\ \cap_{\alpha\in\mathcal{A}}A_\alpha\in\mathcal{F}$;

x5 $\forall A,B\in\mathcal{F}\ A\cup B\in\mathcal{F}$;

x5k $\forall k\in\mathbb{N}\ A_1,A_2,\dots,A_k\in\mathcal{F}\ \cup_{i=1}^{k}A_i\in\mathcal{F}$;

x5 σ $\forall A_1,A_2,\dots,A_k,\dots\in\mathcal{F}\ \cup_{i\in\mathbb{N}}A_i\in\mathcal{F}$;

x5 α $\forall (A_\alpha)_{\alpha\in\mathcal{A}}\in\mathcal{F}^{\mathcal{A}}\ \cup_{\alpha\in\mathcal{A}}A_\alpha\in\mathcal{F}$.

Случай $\mathcal{F}=2^\Omega$ обеспечивает все эти хотелки.

Случай $\mathcal{F}=\{\varnothing,\Omega\}$ тоже обеспечивает все эти хотелки.

На самом деле (смотрите задачу за 1 балл) можно начинать не с Ω, а с булевой алгебры $\mathcal{F}$, и кто там более первороден Ω или $\mathcal{F}$ еще вопрос...

Булевы структуры

Совокупность подмножеств некоторого множества Ω называется *алгеброй (булевой алгеброй, полем событий)*, если его элементы замкнуты (относительно конечного числа) теоретико-множественных операций и $\Omega\in\mathcal{F}$
(эквивалентно - выполнены x0,x1,x2,x3,x4,x5,x4k,x5k; достаточно потребовать x1,x2,x4; или x1,x2,x5).

Совокупность подмножеств некоторого множества Ω называется σ *-алгеброй (борелевским полем)*, если его элементы замкнуты относительно счетного числа теоретико-множественных операций и $\Omega\in\mathcal{F}$
(эквивалентно - выполнены x0,x1,x2,x3,x4 σ ,x5 σ ; достаточно потребовать x1,x2,x4 σ ; или x1,x2,x5 σ).

Совокупность подмножеств некоторого множества Ω называется *кольцом (булевым кольцом)*, если его элементы замкнуты (относительно конечного числа) теоретико-множественных операций (эквивалентно - выполнены x0,x2,x3,x4,x5; достаточно потребовать x3,x4; или x3,x5).

Совокупность подмножеств некоторого множества Ω называется *полукольцом (булевым полукольцом)*, если выполнены x3п,x4.

Самостоятельно дайте определения σ-кольца, или полуалгебры.

Пространство Ω вместе с своей σ-алгеброй $\mathcal{F}$ называется *измеримым пространством* и обозначается $(\Omega,\mathcal{F})$.

Название “измеримое” не вполне удачно, полезнее было бы назвать информационное. И вот почему...

Способы создавать σ-алгебры

Способ главный: создать из чего угодно

Принцип в основе теории меры и теории вероятности: не знаешь, что делать дальше — расширяй то, что уже есть.

У Вас есть что-то, некая совокупность множеств. Сделаем из нее σ-алгебру.

Теорема. Пусть $\varnothing\neq\mathcal{G}\subset 2^\Omega$. Тогда имеется минимальное кольцо $\mathcal{R}(\mathcal{G})$, содержащее $\mathcal{G}$, то есть для любого кольца R из $\mathcal{G}\subset R$ следует $\mathcal{R}(\mathcal{G})\subset R$.

Идея доказательства. Рассмотреть пересечение всех таких колец R, для которых $\mathcal{G}\subset R\subset 2^\Omega$.

Теорема. (С-но) Пусть $\varnothing\neq\mathcal{G}\subset 2^\Omega$. Тогда имеется минимальное σ-кольцо $\mathcal{S}(\mathcal{G})$, содержащее $\mathcal{G}$.

Следствие. Пусть $\varnothing\neq\mathcal{G}\subset 2^\Omega$. Тогда имеется минимальная алгебра $\alpha(\mathcal{G})$, содержащая $\mathcal{G}$.

Следствие. Пусть $\varnothing\neq\mathcal{G}\subset 2^\Omega$. Тогда имеется минимальная σ-алгебра $\sigma(\mathcal{G})$, содержащая $\mathcal{G}$.

Подумать про $\alpha(\mathcal{R}(\mathcal{G}))$, $\mathcal{S}(\mathcal{R}(\mathcal{G})\cup\{\Omega\})$.

Пример 4. Дано $A,B\subset\Omega$. Построить $\mathcal{S}(\{A\})$, $\mathcal{S}(\{A,B\})$.

Способ сужать σ-алгебры.

Даны Ω с σ-алгеброй $\mathcal{F}$ и непустое подмножество $A \in \Omega$. Тогда над A можно построить σ-алгебру

$$\mathcal{F}|_A = \{A \cap B \,|\, B \in \mathcal{F}\}.$$

Способ объединять σ-алгебры

Даны Ω_1, Ω_2 с σ-алгебрами $\mathcal{F}_1, \mathcal{F}_2$ и $\Omega_1 \cap \Omega_2 = \varnothing$. Тогда над $\Omega_1 \sqcup \Omega_2$ можно построить σ-алгебру

$$\mathcal{F}_1 \uplus \mathcal{F}_2 = \{A \cup B \,|\, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}.$$

Способ умножать σ-алгебры

Даны Ω_1, Ω_2 и их σ-алгебры $\mathcal{F}_1, \mathcal{F}_2$. Тогда σ-алгеброй над $\Omega_1 \times \Omega_2$ является

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{A \times B \,|\, A \in \mathcal{F}_1, B \in \mathcal{F}_2\}).$$

Подумать как умножать конечное число измеримых пространств, счетное число измеримых пространств, произвольное число измеримых пространств.

Геометрическая вероятность. Плюсы

Пример 5. Алгебра ли $\mathcal{F} = \{[a,b] \subset [0,1]\}$? Нет, надо рассмотреть всевозможные конечные объединения всевозможных промежутков из $[0,1]$.

Догадайтесь как задать на $\alpha(\mathcal{F})$ длину.

Пример 6. Алгебра ли множество Π всех прямоугольников на плоскости? Нет, надо хотя бы всевозможные конечные объединения прямоугольников $\alpha(\Pi)$. Можно взять $\alpha(\diamond) \bigotimes \alpha(\diamond)$.

Догадайтесь как задать на $\alpha(\Pi)$ площадь.

Умеем ли мы считать площади всего разумного? Нет. Например, треугольник, или ромб у нас пока отсутствуют. Их мы можем приблизить, а до σ-алгебры можно дополнить.

Пример 7. Борелевские множества на открытом вправо промежутке. Рассмотрим открытый вправо промежуток $I \subset \mathbb{R}$ и множество $\mathcal{I}$ всех открытых вправо полуинтервалов $[a,b) \subset I$, это полукольцо. Ее можно продолжить до σ-алгебры $\sigma(\mathcal{I})$, σ-алгебры борелевских множеств. А как выглядят $\mathcal{R}(\mathcal{I})$?

Подумать. А что будет, если определить на открытых влево промежутках, на отрезках, на интервалах?

Подумать. А как вероятность определить на таких множествах?

Определение. (очень умное, пользоваться им мы не будем, но иметь ввиду полезно) Для топологического пространства (Ω, τ), где τ - совокупность всех открытых подмножеств множества Ω, под σ-алгеброй борелевских множеств понимается $\sigma(\tau)$.

Пример 8. Борелевские множества на плоскости. Рассмотрим два открытых вправо промежутка $I_1, I_2 \subset \mathbb{R}$ и множества $\mathcal{I}_1, \mathcal{I}_2$ всех открытых вправо полуинтервалов в I_1, I_2. Теперь элементы $\sigma(\mathcal{I}_1) \bigotimes \sigma(\mathcal{I}_2)$ называют борелевскими множествами на плоскости.

Этот пример показывает, почему σ-алгебра, как измеримая структура, победила алгебру.

Определение. (рабочее) Минимальную σ-алгебру содержащую все n-мерные параллелепипеды $\{[a,b] \times \ldots \times [c,d] \subset \mathbb{R}^n\}$ называют борелевской σ-алгеброй, а ее элементы — борелевскими множествами.

Почему счетное число операций

В стандартной булевой алгебре никто не требует, чтобы предел возрастающей или убывающей последовательности множеств оставался в алгебре. Нам это нужно. И не только для того, чтобы считать площади круга или треугольника, изначально зная лишь площади прямоугольников. Нам нужно уметь не только задать вопрос “А верно ли что в идеальной монетке при большом числе испытаний отношение числа решек к числу орлов стремится к единице?”, но и говорить “Да верно, с вероятностью единица это так.” А для этого нужно чтобы предельные объекты (верхние и нижние пределы) имелись в алгебре событий. Почему счетное, но и не более число операций. Ну, например, потому, что если их будет континуум, то считая вероятность каждой точки, равной нулю, складывая их, получаем, что вероятность каждого отрезка — тоже ноль. А это как-то неудобно.

О вероятности

Пусть задано некоторое измеримое пространство $(\Omega, \mathcal{F})$. Что хотелось бы:

$x0$ $\mathbb{P}(\varnothing) = 0$;

$x1$ $\mathbb{P}(\Omega) = 1$;

$x2$ $\mathbb{P}(A) \le \mathbb{P}(B)$ для любых $A, B \in \mathcal{F}$ с $A \subset B$;

$x3$ $\mathbb{P}(\Omega \setminus A) = 1 - \mathbb{P}(A)$ для любых $A \in \mathcal{F}$;

$x4$ $\mathbb{P}(\cup_{i=1}^k A_i) = \sum\limits_{i=1}^{k} \mathbb{P}(A_i)$ для всех $A_i \in \mathcal{F}$ со свойством $A_i \cap A_j = \varnothing$ при $i \ne j$ (для попарно несовместных событий);

$x4^2$ $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ для всех $A, B \in \mathcal{F}$;

$x4^k$

$$\begin{aligned}\mathbb{P}(\cup_{i=1}^k A_i) \quad = \quad & \sum_{i=1}^{k} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<l} \mathbb{P}(A_i \cap A_j \cap A_l) \\ & + \ldots + (-1)^{k+1} \mathbb{P}(A_1 \cap A_2 \cap \ldots A_k);\end{aligned}$$

$x5$ $\mathbb{P}(\cup_{i\in\mathbb{N}} A_i) = \sum\limits_{i\in\mathbb{N}} \mathbb{P}(A_i)$ для всех таких $A_i \in \mathcal{F}, i \in \mathbb{N}$, что $A_i \cap A_j = \varnothing$ при $i \ne j$ (для попарно несовместных событий);

$x5_{\uparrow}$ $\mathbb{P}(\cup_{i\in\mathbb{N}} B_i) = \lim\limits_{i\to\infty} \mathbb{P}(B_i)$ для всех таких $B_i \in \mathcal{F}, i \in \mathbb{N}$, что $B_1 \subset B_2 \subset \ldots \subset B_k \subset \ldots$ (непрерывность сверху);

$x5_{\downarrow}$ $\mathbb{P}(\cap_{i\in\mathbb{N}} C_i) = \lim\limits_{i\to\infty} \mathbb{P}(C_i)$ для всех таких $C_i \in \mathcal{F}, i \in \mathbb{N}$, что $C_1 \supset C_2 \supset \ldots \supset C_k \supset \ldots$ (непрерывность снизу);

$x5_{\varnothing}$ $\lim\limits_{i\to\infty} \mathbb{P}(D_i) = 0$ для всех $D_i \in \mathcal{F}, i \in \mathbb{N}$, если $D_1 \supset D_2 \supset \ldots \supset D_k \supset \ldots$ и $\cap_{i\in\mathbb{N}} D_i = \varnothing$ (непрерывность в $\varnothing$);

Задачка про Шарапову. Вероятность выигрыша турнира Большого шлема Шараповой 0,3. Значит хотя бы один из них (четырех) она выиграет с вероятностью 1,2?

Аддитивность: от заряда к вероятностному пространству

Определение. Функцию из $\mathcal{F}$ в $\mathbb{R} \cup \{+\infty\}$ называют *аддитивной*, если выполнено свойство $x4$, σ*-аддитивной* — если выполнено свойство $x5$.

Зарядом называют произвольную σ-аддитивную функцию из σ-алгебры событий $\mathcal{F}$ (над Ω) в $\mathbb{R} \cup \{+\infty\}$.

Мерой — произвольную неотрицательную σ-аддитивную функцию из σ-алгебры событий $\mathcal{F}$ (над Ω).

Вероятностью *(распределением вероятности, вероятностным распределением)* называют неотрицательную σ-аддитивную функцию из σ-алгебры событий $\mathcal{F}$ (над Ω) со свойством $\mathbb{P}(\Omega) = 1$.

Вероятностным пространством называют совокупность непустого множества, булевой σ-алгебры ее подмножеств, и вероятности, определенной на этой σ-алгебре.

Об эквивалентности

Теорема (с-но). Для всякой неотрицательной аддитивной функции на алгебре событий со свойством x1 выполнено $x0 - x4, x4^2$.

Замечание. Свойство $x4^k$ тоже выполнено, см. задачи.

Теорема Для всякой неотрицательной аддитивной функции на алгебре событий свойства $x5$, $x5_\uparrow$, $x5_\downarrow$, $x5_\varnothing$ эквивалентны.

На самом деле достаточно доказать $x5_\varnothing \Rightarrow x5$. Такое доказательство имеется, например, в [Колмогоров, теорема II.1.1],[Синай, теорема 1.36]. Но проще доказать самим.

Эквивалентность $x5_\downarrow \Leftrightarrow x5_\uparrow$. Достаточно задать $C_k = \Omega \setminus D_k, D_k = \Omega \setminus C_k$, теперь

$$\cup_{i\in\mathbb{N}} D_i = \Omega \setminus (\cap_{i\in\mathbb{N}} C_i)$$

осталось воспользоваться посылкой и свойством $x2$.

Импликация $x5_\downarrow \Leftarrow x5_\varnothing$. Достаточно задать $C = \cap_{i\in\mathbb{N}} C_i$, $D_k = C_k \setminus C$, теперь

$$\cap_{i\in\mathbb{N}} C_i = C \cup (\cap_{i\in\mathbb{N}} D_i)$$

осталось воспользоваться посылкой и свойством $x4$.

Импликация $x5_\uparrow \Leftrightarrow x5$. Достаточно задать $B_k = \cup_{i\le k} A_k$ или $A_k = B_k \setminus B_{k-1}$ и подставить в уже известное; после $\cup_{i\in\mathbb{N}} A_i = \cup_{i\in\mathbb{N}} B_i$, $\mathbb{P}(B_k) = \sum_{i\le k} \mathbb{P}(A_k)$ останется лишь заметить, что сумма ряда это предел его частичных сумм.

О продолжении

Теорема Каратеодори. Всякую σ-аддитивную функцию, заданную на полукольце Π, можно продолжить на минимальное ее содержащее σ-кольцо, на $\alpha(\Pi)$, и притом единственным образом.

Идея доказательства.

Существование: Помимо σ-аддитивной функции $\mu : \Pi \to \mathbb{R}$ рассмотрим отображение

$$\mu^*(A) = \inf\left\{\sum_{k=1}^{\infty} \mu(D_k) \,|\, D_k \in \Pi, A \subset \cup_{k=1}^{\infty} D_k\right\} \qquad \forall A \in 2^\Omega.$$

На Π это отображение совпадает с μ в силу счетно-аддитивности μ. Теперь (это не очень тривиально показать) σ-кольцом является

$$\{A \subset \Omega \,|\, \forall D \subset \Omega\ \mu^*(D \cap A) + \mu^*(D \setminus A) = \mu^*(D)\},$$

оно содержит Π (тоже не просто), значит содержит и $\alpha(\Pi)$. Но тогда $\mu^*|_{\alpha(\Pi)}$ — счетно-аддитивное продолжение μ.

Единственность. Пусть есть две такие функции, тогда те множества, на которых эти функции принимают одни и те же значения, образуют σ-кольцо, и это σ-кольцо содержит Π. Поскольку $\alpha(\Pi)$ — минимальное σ-кольцо, содержащее Π, то на $\alpha(\Pi)$ эти функции совпадают.

Замечание Условие счетно-аддитивности существенно. Контрпример строится уже на множестве натуральных чисел. Пример построить тоже полезно.

Следствие. Всякую счетно-аддитивную неотрицательную функцию, заданную на полуалгебре Π, можно продолжить на минимальную ее содержащую σ-алгебру, на $\sigma(\Pi)$, и притом единственным образом.

Вывод. Вероятность достаточно задать на полуалгебре.

Прикол этой теоремы в том, что на $\sigma(\Pi)$ (или $\alpha(\Pi)$) продолжается любая σ-аддитивная функция. Понятно, что каждую конкретную σ-аддитивную функцию иногда можно продолжить единственным образом и на большую измеримую структуру, смотрите доказательство теоремы, но здесь объект от функции не зависит, про саму σ-аддитивную функцию можно ничего не знать.

В частности, на множество $\mathcal{B}$ всех борелевских множеств соответствующей размерности, длину (площадь, объем, в общем некую $meas$) продолжить с сохранением счетно-аддитивности можно единственным образом.

Геометрическая вероятность. Итог

Определение. (очень умное, пользоваться им мы не будем, но иметь ввиду полезно) Для топологического пространства (Ω, τ), где τ - совокупность всех открытых подмножеств множества Ω, под σ-алгеброй борелевских множеств понимается $\sigma(\tau)$.

Определение. (рабочее) Минимальную σ-алгебру содержащую все n-мерные параллелепипеды $\{[a,b] \times \ldots \times [c,d] \subset \mathbb{R}^n\}$ называют борелевской σ-алгеброй, а ее элементы — борелевскими множествами.

Борелевской мерой каждого параллелепипеда $[a,b] \times \ldots \times [c,d]$ назовем

$$P\Big([a,b] \times \ldots \times [c,d]\Big) = (b-a) \cdot \ldots \cdot (d-c).$$

Эта функция счетно-аддитивна, например, на полукольце всех параллелепипедов. По теореме Каратеодори мы можем ее продолжить на все элементы из $\sigma(\tau)$.

Сие рассуждение сыру подобно, но дыры указывать полезно

Итак, имея меру на всех борелевских подмножествах, мы можем свести геометрическую вероятность к общей схеме.

Полученную нами меру (как и всякую другую меру) можно попробовать распространить и дальше. Введем так множество $\mathcal{L}$ всех лебеговских множеств. Идея: $\mathcal{L} = \sigma(\mathcal{B} \cup \mathcal{N})$, где $\mathcal{N} = \{N \mid \exists A \in \mathcal{B}, N \subset A, meas(A) = 0\}$.

Отметим, что $\mathcal{N}$, а значит и так пополненное σ-кольцо, уже учитывает исходную меру $meas$, в отличие от конструкции Каратеодори. Меры со свойством $meas|_{\mathcal{N}} \equiv 0$ называются полными. До второго семестра это знание нам не пригодится.

Условная вероятность. Наивный взгляд

Пусть дано вероятностное пространство $(\Omega, \mathcal{F}, P)$. Пусть даны также $A, B \in \mathcal{F}, \mathbb{P}(B) > 0.$

Определение. *Условной вероятностью события* A при условии события B называется число

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

Как функция от A это обычная вероятность. Например, это монотонное отображение в $[0,1]$, выполнены

$x0$ $\mathbb{P}(\varnothing|B) = 0$;

$x1$ $\mathbb{P}(\Omega|B) = \mathbb{P}(B|B) = 1$;

$x2$ $\mathbb{P}(A|B) = 1 - \mathbb{P}(\bar{A}|B)$ для любых $A \in \mathcal{F}$;

$x3$ $\mathbb{P}(A_1) \leq \mathbb{P}(A_2)$ для любых $A_1, A_2 \in \mathcal{F}$ с $A_1 \subset A_2$;

$x4$ $\mathbb{P}(\cup_{i=1}^k A_i|B) = \sum\limits_{i=1}^k \mathbb{P}(A_i|B)$ для всех $A_i \in \mathcal{F}$ со свойством $A_i \cap A_j = \varnothing$ при $i \neq j$ (для попарно несовместных событий);

$x5$ $\mathbb{P}(\cup_{i \in \mathbb{N}} A_i|B) = \sum\limits_{i \in \mathbb{N}} \mathbb{P}(A_i|B)$ для всех таких $A_i \in \mathcal{F}, i \in \mathbb{N}$, что $A_i \cap A_j = \varnothing$ при $i \neq j$ (для попарно несовместных событий).

В частности, это способ задавать настоящую вероятность при сужении алгебры, мы получаем вероятностное пространство $(B, \mathcal{F}|_B, \mathbb{P}(\cdot|B))$.

Тождество очевидное, и верное. $\mathbb{P}(A|B) + \mathbb{P}(\bar{A}|B) = 1.$

Тождество очевидное, но неверное. $\mathbb{P}(A|B) + \mathbb{P}(A|\bar{B}) = 1.$

Теорема умножения. Пусть $A, A_1, \dots, A_n, B \in \mathcal{F}$. Если $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(A_1), \dots, \mathbb{P}(A_n) > 0$, то $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)$, более того

$$\mathbb{P}(\cap_{i=1}^n A_i) = \mathbb{P}(A_n | \cap_{i=1}^{n-1} A_i)\mathbb{P}(A_{n-1} | \cap_{i=1}^{n-2} A_i) \dots \mathbb{P}(A_2|A_1)\mathbb{P}(A_1).$$

Заметим, что $\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)$ выполнено вне зависимости от положительности $\mathbb{P}(A), \mathbb{P}(B)$.

Пример задачи. Среди N экзаменационных билетов ровно 15 — на мартингалы. Найдите вероятность того, что а) первый студент возьмет такой билет;
б) второй студент возьмет такой билет;
в) второй студент возьмет такой билет, если первому такой билет достался;
г) второй студент возьмет такой билет, если первому такой билет не достался.

О гипотезах и Байесе

Определение Говорят, что набор событий $H_1, H_2, \dots, H_n \in \mathcal{F}$ образует полную группу событий, если $H_i \cap H_j = \varnothing (i \neq j)$ и $\cup_i H_i = \Omega.$ Удобно называть эти события гипотезами.

Чем это удобно — мы можем разбивать сложную задачу на подзадачи.

Для полной группы событий $H_1, H_2, \dots, H_n \in \mathcal{F}$ и любого события $A \in \mathcal{F}$ выполнены

(формула полной вероятности) $\mathbb{P}(A) = \mathbb{P}(A|H_1)\mathbb{P}(H_1) + \dots + \mathbb{P}(A|H_n)\mathbb{P}(H_n);$

(формула Байеса)

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\mathbb{P}(A|H_1)\mathbb{P}(H_1) + \dots + \mathbb{P}(A|H_n)\mathbb{P}(H_n)}.$$

Для доказательства достаточно воспользоваться в первом случае формулой умножения и аддитивностью, во втором — показанной формулой полной вероятности и определением условной вероятности.

Задача (0,5 баллов). Три карточки AA, BB, AB (на одной стороне левая буква, на другой стороне — правая). Одна из карточек берется наугад. Вы видите A. Что написано с другой стороны, с какой вероятностью?

Зачем нужна формула Байеса...

Предположим мы знаем некое "экспертное" мнение о вероятности $\mathbb{P}(H_i)$. Это априорная вероятность (до опыта). Провели эксперимент, в ходе которого случилось событие A. После этого можно "уточнить" вероятность, посчитав апостериорная вероятность (после опыта).

Вообще говоря, если априорная вероятность высосана из пальца, то и апостериорная тоже, но тут есть один момент... за что все Байеса и любят

Образцовая задачка. Есть две монеты. Одна монета правильная — орел и решка выпадают с вероятностью 0,5. И монета неправильная — решка выпадает с вероятностью 0,75, орел с вероятностью 0,25.
а) Пусть выбор монет равновероятен. Бросили монету 1 раз. С какими шансами выпал орел?
б) Пусть выбор монет равновероятен. Бросили монету 1 раз. Выпал орел. С какими шансами это правильная монета?
в) Пусть выбор монет равновероятен. Бросили монету 10 раз. Выпал орел 2 раза, решка — 8 раз. С какими шансами это правильная монета? $3^8 = 6561$
г) Пусть выбор монет равновероятен. Бросили монету 20 раз. Выпал орел 4 раза, решка — 16 раз. С какими шансами это правильная монета? $3^{16} \approx 43 * 10^6$.
д) Априори монета честная с вероятностью 9/10. Бросили монету 20 раз. Выпал орел 4 раза, решка — 16 раз. С какими шансами это правильная монета?
е) Априори монета честная с вероятностью $9999/10000$. Бросили монету 20 раз. Выпал орел 4 раза, решка — 16 раз. С какими шансами это правильная монета?

Вывод первый. Мнение любого эксперта проверяемо, если эксперимент достаточно длинный.
Вывод второй. Хороший эксперт экономит.
Вывод третий. Никакой эксперимент не убедит всех экспертов.

А если нечестность игрока лишь стечение обстоятельств?
Что делать, если про монету изначально неизвестно ничего? Сколько гипотез потребуется?

Данетка про “застрелился из-за музыки”' дает прекрасный пример, когда даже мудрый эксперт ошибается.

Независимость

Определение. События $A, B \in \mathcal{F}$ называют независимыми, если $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A \cap B)$.

Интуитивно не очевидно, но факт: 1) $\varnothing$ и Ω независимы с любым событием A.
2) A, B независимы и $A \cap B = \varnothing \;\Leftrightarrow\; \mathbb{P}(A)$ или $\mathbb{P}(B)$ равно нулю.

Более удобно думать про независимость так: Знание того, произошло B или не произошло, не влияет на вероятность события A.

Действительно, в случае $\mathbb{P}(B) > 0$ имеется эквивалентное определение независимости: A и B независимы $\Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A)$.

История про Колмогорова и материализм как пример независимых событий.

Определение События $A_1, \ldots, A_m, \ldots$ называют попарно независимыми, если для любых различных i, j события A_i, A_j независимы. События $A_1, \ldots, A_m, \ldots$ называют независимыми в совокупности, если для всякого конечного набора $k_1, k_2, \ldots, k_n$ выполнено $\mathbb{P}(\cap_{i=1}^n A_{k_i}) = \prod_{i=1}^n \mathbb{P}(A_{k_i})$.

А выполнено ли это для любого счетного набора?

Пример Бернштейна Четырехгранная пирамидка, ее грани раскрашены в серый, бурый, малиновый цвет, а последняя, в серо-буро-малиновую полосочку. Все грани считаем равновероятными. События С,Б,М — на выпавшей грани имеются соответственно серый, бурый, малиновый цвет. Тогда все эти события попарно независимы, но не независимы в совокупности.

Пара свойств

Если A, B независимы, то $\bar{A}, B$ независимы, $A, \bar{B}$ независимы, $\bar{A}, \bar{B}$ независимы.

Если A, B_1 независимы, A, B_2 независимы и $B_1 \cap B_2 = \varnothing$, то $A, B_1 \cup B_2$ также независимы.

Доказательства этих фактов тривиальны, есть, например, у Синая и у Скорохода.

Снова о σ-алгебрах

Определение. σ-подалгебры $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n$ независимы, если любой конечный набор множеств A_i $(A_i \in \mathcal{F}_i)$ независим; σ-подалгебры $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_n, \ldots$ независимы, если любой их конечный набор независим.

Например, тривиальная σ-алгебра $\{\varnothing, \Omega\}$ независима с любой другой подалгеброй; σ-подалгебра независимой σ-подалгебры также независима.

Пример Рассмотрим σ-алгебру $\mathcal{B}$ всех борелевских подмножеств на $[0,1]^2$ и геометрическую вероятность P на ней. Зададим две σ-подалгебры $\mathcal{B}_x, \mathcal{B}_y$:

$$\mathcal{B}_x = \sigma\{[a,b] \times [0,1] \,|\, 0 \le a \le b \le 1\},$$
$$\mathcal{B}_y = \sigma\{[0,1] \times [a,b] \,|\, 0 \le a \le b \le 1\}.$$

Эти две σ-подалгебры $\mathcal{B}_x, \mathcal{B}_y$ независимы. В частности $\mathbb{P}([a,b]\times[c,d]) = \mathbb{P}([a,b]\times[0,1])\times\mathbb{P}([0,1]\times[c,d])$.

Рассмотрим полную систему гипотез $H_1, \ldots, H_n$. Можно образовать минимальную σ-подалгебру $\mathcal{H}$, их содержащую.

Теперь можно задать условную вероятность относительно σ-подалгебры (информации, содержащейся в) $\mathcal{H}$. Формально $\mathbb{P}(A|\mathcal{H}) : \Omega \to [0,1]$ задается правилом: $\mathbb{P}(A|\mathcal{H})(\omega) = \mathbb{P}(A|H_i)$ если $\omega \in H_i$. Фокус в том, что теперь(!) перед нами отображение не из алгебры, а из множества элементарных событий.

В частности, если при этом A не зависит от σ-алгебры $\mathcal{H}$, то $\mathbb{P}(A|\mathcal{H}) \equiv \mathbb{P}(A)$.

СЛУЧАЙНЫЕ ВЕЛИЧИНЫ.

Дискретный случай

Если Ω не более чем счетно, и $\mathcal{F} \subset 2^{\Omega}$, то найдется некоторое не более чем счетное разбиение $H_1, \ldots, H_i, \ldots$ множества Ω, что $\mathcal{F} = \{\cup_{i \in I} H_i \,|\, I \subset I_0\}$ (простая задачка есть в задачах, ее решили чуть ли не раньше все остальных). Теперь, приписывая каждому H_i вероятность $\mathbb{P}(H_i) = p_i$ мы можем (почему?) гарантировать $\mathbb{P}(\cup_{i \in I} H_i) = \sum_{i \in I} \mathbb{P}(H_i) = \sum_{i \in I} p_i$.

А как проверить, что при этом получится действительно вероятность? Когда $\sum_{i \in \mathbb{N}} p_i = 1$.

Итак, классическая вероятность прибита к общему случаю с сигма-алгебрами. Для задания вероятности (вероятностного распределения) в дискретном случае достаточно задать табличку вида:

H_1	H_2	H_3	H_4	$\ldots$
$\mathbb{P}(H_1)$	$\mathbb{P}(H_2)$	$\mathbb{P}(H_3)$	$\mathbb{P}(H_4)$	$\ldots$

Так задают дискретные распределения. Некоторые считают, что ничего другого и не существует... Не будем их пока разочаровывать, дадим определение дискретной случайной величины.

Дискретные случайные величины. Определения

Определение. В заданном измеримом пространстве $(\Omega, \mathcal{F})$ отображение $\xi : \Omega \to \mathbb{R}$ называется дискретной случайной величиной, если найдутся не более чем счетное разбиение $H_1, \ldots, H_i, \ldots \in \mathcal{F}$ множества Ω и не более чем счетный набор чисел x_i, такие, что $\xi(\omega_i) = x_i$ для всех $\omega_i \in H_i$.

Эквивалентное определение. В заданном измеримом пространстве $(\Omega, \mathcal{F})$ отображение $\xi : \Omega \to \mathbb{R}$ называется дискретной случайной величиной, если оно принимает не более чем счетное число значений $x_1, \ldots, x_i, \ldots$ и при этом для всякого числа x_i выполнено

$$H_i = \{\omega \in \Omega \,|\, \xi(\omega) = x_i\} \in \mathcal{F}.$$

Если при этом задана вероятность, то мы, принимая $p_i = \mathbb{P}(H_i)$, также получаем распределение этой случайной величины.

x_1	x_2	x_3	x_4	$\ldots$
p_1	p_2	p_3	p_4	$\ldots$

Рабочее, но не вполне аккуратное определение: дискретным распределением называют отображение из счетного подмножества множества $\mathbb{R}$ в $[0,1]$, сумма элементов образа которого равна единице.

Значительная часть стандартного курса теории вероятности посвящена исследованию распределений случайных величин. В чем прелесть этого понятия — мы абстрагируемся от вероятностного пространства, от исходной модели, оставляя лишь числа и их вероятности. За счет чего это получилось? Просто мы переместили вероятности с исходного Ω на множество действительных чисел (точнее, на не более чем счетное его подмножество). [Тут надо бы нарисовать красивую диаграммку, но сайт не справился].

Пример 1. Самая простая случайная величина — константы, им соответствуют вырожденные распределения.

Пример 2. Дискретное равномерное распределение. Принимает конечный набор целых значений из заданного диапазона, все эти значения равновероятны. Например, $U\{1,n\}$ принимает значения от 1 до n с вероятностью $1/n$ каждое: $\mathbb{P}(\xi_{U\{1,n\}} = k) = 1/n \cdot 1_{\{1,2,\ldots,n\}}(k)$

Пример 3. Более интересное распределение — распределение Бернулли. Обозначение $B(1,p)$, где $p \in [0,1]$. Параметр p задает вероятность единицы (“успеха”). Тогда $1-p$ — вероятность нуля (“неудачи”).

$$\mathbb{P}(\xi_{B(1,p)} = 1) = p, \mathbb{P}(\xi_{B(1,p)} = 0) = 1 - p = q.$$

Заметьте, что в этих примерах мы не указали вероятностное пространство. Это не случайно, достаточно часто про него можно не вспоминать. В простых задачах нам оно и не нужно.

Да, задачек будет много. Вот одна на затравку

Задача 5.4 Даны две дискретные случайные величины. Докажите, что сумма этих дискретных случайных величин также является дискретной случайной величиной. А как найти сумму дискретных распределений?

Биномиальная схема независимых испытаний Бернулли

Пусть элементарное событие — последовательность из нулей и единиц, результат последовательности испытаний, в каждом из которых вероятности успеха и неуспеха не меняются: вероятность успеха равна $p \in (0,1)$, неуспеха — $q = 1 - p$.

Пусть пока число испытаний конечно, n. Тогда

$$\Omega_n = \{\text{конечные строки из нулей и единиц длиной } n\} = \{0,1\}^n \cong 2^n, \qquad \mathcal{F}_n = 2^{\Omega_n}.$$

Например, $\mathcal{F}_1 = \{\varnothing, \{0\}, \{1\}, \{0,1\}\}$. Тогда, $P_1\{0\} = q, P_1\{1\} = p$. Мы получили распределение Бернулли $B(1,p)$.

Теперь, поскольку σ-алгебру $\mathcal{F}_n$ можно задать как произведение $\mathcal{F}_n = \otimes_{k=1}^n \mathcal{F}_1$, то и вероятность можно задать также. Нам достаточно задать ее на элементах Ω_n (почему?).

Например, при $p = q = 1/2$ все исходы равновероятны, вероятность каждой конечной строки равна 2^{-n}. А вероятность конечной строки $\mathbb{P}(\text{“}100111110001\text{''})$ равна видимо $p^7 q^5$.

В общем случае, в схеме Бернулли вероятность задается следующим образом:

$$\mathbb{P}(\text{“конечная строка”}) = p^{\text{число успехов}} q^{\text{число неуспехов}}.$$

Почему именно так?

Можем ли мы сузить нашу алгебру, рассмотрев меньшее множество исходов. Например, только первый? Или только второй? Или первый и второй вместе? А объединить? А чему это соответствует?

О счетных произведениях

Интересно, что полученная формула не зависит от n, значит есть шанс продолжить вероятность на бесконечные строки. Зачем это нужно?

Задачка. Найдите вероятность того, что успех наступит хоть в каком-то испытании. Пусть k — номер первого успеха в схеме испытаний Бернулли. Найдите распределение этого k.

Если число испытаний конечно — мы все умеем. Но обычно никакие ограничения сверху явно не заданы, и значит ни в какое Ω_n задача не влезет.

Нужно ввести $\Omega_\infty = \{\text{бесконечные (вправо) строки из нулей и единиц}\}$ и соответствующую сигма-алгебру. Как?

На самом деле все просто, сначала надо заметить, что $\Omega_n \not\subseteq \Omega_\infty$, но $\Omega_n \times \Omega_\infty \subset \Omega_\infty$, тогда $\mathcal{F}_\infty = \sigma\{A_n \times \Omega_\infty \,|\, n \in \mathbb{N}, A \in \mathcal{F}_n\}$. При этом $A_n \times \Omega_\infty$ фактически играет роль: “до момента n знаем, что произошло A_n, а что будет потом — не знаем ничего”.

Вот собственно тривиальная алгебра $\{\varnothing, \Omega\}$ и пригодилась...

Геометрическое распределение

Решая задачку в общем случае получаем геометрическое распределение $Geom(p)$.

Геометрическое распределение — дискретное распределение, зависящее от параметра p. Моделирует число испытаний до первого успеха в схеме Бернулли с вероятностью p успеха в каждом испытании.

Так называют в разных книжках разные распределения:

Пример 4_0. “Число промахов до первого попадания”. Тогда r_i принимают целые значения от 0 включительно до бесконечности, $\mathbb{P}(\xi_{Geom^0(p)} = k) = (1-p)^k p = pq^k$.

Пример 4_1. “Число выстрелов до первого попадания”. Тогда r_i принимают целые значение от 1 включительно до бесконечности, $\mathbb{P}(\xi_{Geom^1(p)} = k) = (1-p)^{k-1} p = pq^{k-1}$.

А сумма чисел точно 1? Сумму геометрической прогрессии можно посчитать.

Распределение Бернулли

Пример 5. Пусть число испытаний равно n. Пусть все, что нас интересует — число попаданий (успехов), случившихся в ходе этих n испытаний. Мы получили очередную случайную величину. Ее распределение — распределение Бернулли $B(n,p)$.

Это распределение принимает значения — целые числа от 0 до n. Вероятность числа k равна $\mathbb{P}(\xi_{B(n,p)} = k) = C_n^k p^k q^{n-k}$. Здесь $C_n^m = \frac{n!}{m!(n-m)!}$, $0! = 1$.

[как это доказать? вероятность каждой строки длиной n с ровно k единицами в каждой равна $p^k q^{n-k}$, число таких строк равно C_n^k, а аддитивность никто не отменял.]

Биномиальное распределение — дискретное распределение, зависящее от параметров n, p. Моделирует число успехов k при n испытаниях в схеме Бернулли с вероятностью p успеха в каждом испытании.

$$\mathbb{P}(\text{число успехов в } n \text{ испытаниях} = k) = C_n^k p^k (1-p)^{n-k} \quad k \in \{0, 1, 2, \ldots, n\}.$$

А сумма чисел точно 1?

Отметим, что мы сделали очень сильное предположение, когда именно так задали вероятность. От порядка нулей и единиц в строке ничего не зависит, конечная подстрока ничего не знает про то, что будет потом. Более того, каждый следующий символ будто ничего не знает о предыдущих. Ничего не помнящий, ничего не предсказывающий. Фактически мы получили простейшую модельку, чтобы потом упражняться с такими словами как марковость, независимость, однородность. (найдите, где в схеме Бернулли марковость, независимость и однородность...)

Да, но в реальных задачах как-то надо находить p... Как? Идеи в студию...

О наивности и поисках p.

Можно проверить, что m — наиболее вероятное число успехов в распределении Бернулли (мода этого распределения) в точности тогда, когда $m \in [np - q, np + p]$, то есть $m \approx np$. Ну отсюда вроде ясно, что надо взять $p = m/n$. Однако, например при $p = 1/2$, $n = 2m$, с помощью формулы Валлиса можно доказать

$$\mathbb{P}(\text{число успехов} = m) = C_n^m 2^{-n} \approx \sqrt{\frac{2}{\pi n}}.$$

Вывод : в точности $m/n = p$ бывает редко. А как же наивное понимание вероятности "типа примерно равно числу p"?

Что-то говорит о том, как искать вероятность при известных k, n

Локальная теорема Муавра–Лапласа. Пусть k зависит от n так, что $\left|\frac{k(n)-np}{\sqrt{npq}}\right| \le C$ для некоторого C (равномерно ограничено). Тогда имеет место формула Муавра-Лапласа

$$P_n(\text{число успехов} = k(n)) \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k(n)-np)^2}{2npq}} (1 + o(n^{-1/2})).$$

Основная идея доказательства — использование формулы Стирлинга:

$$n! = \sqrt{2\pi n} \cdot n^n e^{-n} (1 + \frac{1}{12n} + \frac{1}{288n^2} + \ldots)$$

Подробнее см., например wiki.

Считается, что теорему имеет смысл использовать при $n > 100, npq > 20$. Точные оценки погрешности формулы имеются: неравенство Бернштейна, неравенство Берри-Эссеена. В принципе, всю зиму мы будем крутиться вокруг этого и схожих с ним результатов...

А еще пару месяцев перед этим мы будем разбираться, что такое абсолютно непрерывные распределения, как появившееся в этой теореме нормальное распределение.

Из схемы независимых испытаний Бернулли можно также вывести отрицательное биномиальное распределение (число неуспехов до r-го успеха), распределение Пуассона.

Теорема о редких событиях. Распределение Пуассона

Теорема Пуассона. (Теорема о редких событиях) Пусть $p = p_n$ зависит от n так, что $\lambda_n = np_n \to \lambda > 0$ при $n \to \infty$. Тогда для любого $k \in \{0, 1, 2, \dots\}$

$$P_n(\text{число успехов} = k) = C_n^k p_n^k (1-p_n)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}.$$

Доказательство не найдет только ленивый, но если коротко, то так:

$$\frac{n(n-1)\dots(n-k+1)}{k!}\frac{(p_n n)^k}{n^k}(1-p_n)^{n-k} \approx \frac{n}{n}\frac{n-1}{n}\dots\frac{n-k+1}{n}\frac{\lambda^k}{k!}\Big((1-p_n)^{\frac{1}{p_n}}\Big)^{p_n(n-k)} \to \frac{\lambda^k}{k!}e^{-\lambda}.$$

Соответствующее распределение называется распределением Пуассона. Можно показать, что ошибка в теореме не превосходит $2\lambda_n^2/n$.

Распределение Пуассона — дискретное распределение, зависящее от положительного параметра λ. Моделирует число успехов m за единицу времени, если в среднем за эту единицу времени происходит λ успехов. Каждому целому неотрицательному числу k сопоставляется вероятность

$$\mathbb{P}(\xi_{Pois(\lambda)} = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

А сумма чисел точно 1?

Классический пример: потери от удара копытом в прусской кавалерии.

Полиномиальная схема независимых испытаний Бернулли

Элементарное событие — последовательность элементов множества $\{1, 2, \dots, k\}$. Вероятность каждого элемента p_i $(i \in \{1, 2, \dots, k\})$ Вероятность конечной строки с количеством n_i каждого символа $\prod_{i=1}^k p_i^{n_i})$.

Каждому набору $(n_1, n_2, \dots, n_k)$ со свойством $(n_i \geq 0, n_1 + \ldots + n_k = n)$ соответствует вероятность

$$n! \prod_{i=1}^k \frac{p_i^{n_i}}{n_i!}.$$

Мы получили полиномиальное распределение. А оно точно дискретное?

Многомерные распределения

Иногда, например полиномиальное распределение, удобнее не перенумеровывать некоторым линейным порядком, создавая строку вероятностей p_i и конечную строку чисел x_i, а сопоставить все возможные x в некоторую вообще говоря многомерную таблицу, получая вероятности каждой строки чисел уже в виде многомерного (возможно бесконечного) массива $p_{ij\dots}$.

	x_1	x_2	x_3	$\dots$
y_1	p_{11}	p_{12}	p_{13}	$\dots$
y_2	p_{21}	p_{22}	p_{23}	$\dots$
y_1	$\dots$	$\dots$	$\dots$	$\dots$

Пример. В случае полиномиального распределения с тремя опытами и тремя вариантами исходов X, Y, Z в каждой с вероятностями $1/2, 1/3, 1/6$ получаем такую табличку

$\xi_{(X,Y)}$	0	1	2	3
0	$\frac{1}{6^3}$	$\frac{3}{2\cdot 6^2}$	$\frac{3}{2^2\cdot 6}$	$\frac{1}{2^3}$
1	$\frac{3}{3\cdot 6^2}$	$\frac{3\cdot 2}{2\cdot 3\cdot 6}$	$\frac{3}{2^2\cdot 3}$	0
2	$\frac{3}{3^2\cdot 6}$	$\frac{3}{2\cdot 3^2}$	0	0
3	$\frac{1}{3^3}$	0	0	0

Из этой таблички можно извлечь еще по крайней 10 дискретных распределений (на самом деле существенно больше).

Во-первых маргинальные распределения по каждой переменной (в общем случае по набору переменных). Маржин — поле. Поэтому на полях таблицы можно написать

$\xi_{(X,Y)}$	0	1	2	3	ξ_Y
0	$\frac{1}{6^3}$	$\frac{3}{2\cdot 6^2}$	$\frac{3}{2^2\cdot 6}$	$\frac{1}{2^3}$	$\frac{2^3}{3^3}$
1	$\frac{3}{3\cdot 6^2}$	$\frac{3\cdot 2}{2\cdot 3\cdot 6}$	$\frac{3}{2^2\cdot 3}$	0	$\frac{3\cdot 2^2}{3^3}$
2	$\frac{3}{3^2\cdot 6}$	$\frac{3}{2\cdot 3^2}$	0	0	$\frac{3\cdot 2}{3^3}$
3	$\frac{1}{3^3}$	0	0	0	$\frac{1}{3^3}$
ξ_X	$\frac{1}{2^3}$	$\frac{3}{2^3}$	$\frac{3}{2^3}$	$\frac{1}{2^3}$	

Складывая по строкам и столбцам соответственно, мы получили распределения числа успехов типа X и Y соответственно, распределения случайных величин ξ_X, ξ_Y. Это маргинальные для $\xi_{X,Y}$ распределения, распределения, возникающие при забывании информации об остальных переменных.

Можно было сделать по другому. Зная информацию о числе успехов типа X, получить распределение для числа успехов типа Y.

Например,

$\xi_(Y\|X=0)$	0	1	2	3
0	$1/64$	$9/64$	$27/64$	$27/64$

То, что мы получили, называется условным распределением. Распределением Y при условии, что $X = 0$.

А куда Z делся? А почему? Ну хорошо, пока забывали про Z совсем было 10 распределений, а сколько будет, если вспомнить?

Независимость случайных величин

Пусть у нас имеется дискретное рапределение случайной величины ξ:

x_1	x_2	x_3	x_4	$\ldots$
p_1	p_2	p_3	p_4	$\ldots$

Она порождает σ-алгебру $\sigma(\xi) = \sigma(\{\omega|\xi(\omega) = x_1\}, \{\omega|\xi(\omega) = x_2\}, \ldots)$ (фактически это σ-алгебра, порожденная разбиением на гипотезы H_i).

Пусть у нас имеется еще одно дискретное рапределение случайной величины η:

y_1	y_2	y_3	y_4	$\ldots$
q_1	q_2	q_3	q_4	$\ldots$

Определение Случайные величины ξ, η независимы, если независимы порожденные ими σ-алгебры.

Раскроем это общее определение в более понятный вид.

Построим распределение дискретной случайной величины (ξ, η). Тогда

(ξ,η)	y_1	y_2	$\ldots$
x_1	p_1q_1	p_1q_2	$\ldots$
x_2	p_2q_1	p_2q_2	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$

Как и положено, $\mathbb{P}(\xi = x_i, \eta = y_i) = \mathbb{P}(\xi = x_i)\mathbb{P}(\eta = y_i)$. Никакая информация о y_i никоим образом не поколеблет вероятности чисел x_i.

Скалярные характеристики случайных величин

Модой дискретной случайной величины называют те значения x_i, вероятности которых не меньше всех остальных p_i. У дискретной случайной величины мода всегда есть (доказать!), но она не обязана быть единственной.

Медианой дискретной случайной величины ξ называют любое число x, для которого $\mathbb{P}(\xi \le x) \ge 0,5$ и $\mathbb{P}(\xi \ge x) \ge 0,5$. У дискретной случайной величины медиана всегда есть и она или одна, или их континуум (доказать!).

Самой известной характеристикой, возможно не настолько уж и заслуженно, является

Математическое ожидание

Определение. Для всякой дискретной случайной величины ξ выражение

$$\mathbb{E}\xi = \sum_{i\in\mathbb{N}} p_i x_i$$

назовем математическим ожиданием случайной величины ξ.

А ряд всегда сходится? Приведите пример дискретной случайной величины, не имеющей математического ожидания.

Будем говорить. что математическое ожидание существует, если это выражение корректно определено и конечно.

Пример 1 (Распределение Пуассона)

$$\mathbb{E}\xi = e^{-\lambda}\sum_{k\in\mathbb{N}}\frac{k\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda.$$

Пример 2 (Геометрическое распределение)

$$\begin{aligned}\mathbb{E}\xi &= \sum_{k\in\mathbb{N}} kq^k p = pq\sum_{k=1}^{\infty} kq^{k-1} = pq\sum_{k=1}^{\infty}\frac{d}{dq}(q^k)\\ &= pq\frac{d}{dq}\sum_{k=1}^{\infty} q^k = pq\frac{d}{dq}\frac{q}{1-q} = pq\frac{d}{dq}(\frac{1}{1-q}-1) = pq\frac{1}{(1-q)^2} = q/p.\end{aligned}$$

Пример 3 (Равномерное дискретное распределение) $\xi \in U\{a,b\}$, $\mathbb{E}\xi = \frac{b+a}{2}$.

Свойства математического ожидания

Они верны не только в дискретном случае, но и в обшем, впрочем доказывать пока будем только в дискретном случае.

1. $\mathbb{E}\xi \ge 0$, если $\xi \ge 0$ для всех $\omega \in \Omega$ (и $\mathbb{E}\xi$ существует);

Верно ли, что $\mathbb{E}\xi > 0$, если $\xi > 0$ для всех $\omega \in \Omega$ (и $\mathbb{E}\xi$ существует)?

2. $\mathbb{E}\xi_1 \ge \mathbb{E}\xi_2$, если $\xi_1(\omega) \ge \xi_2(\omega)$ для всех $\omega \in \Omega$ (и все матожидания существуют)
3. $\mathbb{E}c = c$ для случайной величины равной константе c при любых $\omega \in \Omega$;
4. $\mathbb{E}(c\xi) = c\mathbb{E}\xi$ для любого $c \in \mathbb{R}$ (при $c \ne 0$ если существует одно, то имеется и другое)
5. $\mathbb{E}\xi_1 + \mathbb{E}\xi_2 = \mathbb{E}(\xi_1 + \xi_2)$; (если существуют два из них, то существует и третье)
6. $a\mathbb{E}\xi_1 + b\mathbb{E}\xi_2 = \mathbb{E}(a\xi_1 + b\xi_2)$ при любых $a, b \in \mathbb{R}$ (если существуют два из них, то существует и третье $ab \ne 0$);
7. $\mathbb{E}(\sum_{i=1}^{\infty}\xi_i) = \sum_{i=1}^{\infty}\mathbb{E}\xi_i$ если сходится ряд $\sum_{i=1}^{\infty}\mathbb{E}|\xi_i|$; (если в этом предположении ряд конечен, то матожидания существуют заведомо)
8. $|\mathbb{E}\xi| \le \mathbb{E}|\xi|$; (если существует матожидание справа, то и слева тоже существует)
9. $g(\mathbb{E}\xi) \le \mathbb{E}g(\xi)$ для любой выпуклой (вниз) скалярной функции g (если существуют оба);

Здесь доказательство сводится к неравенству Йенсена или замечанию, что надграфик выпуклой функции тоже выпуклый.

10. Если ξ_i независимы, то $\mathbb{E}\xi_1\xi_2 = \mathbb{E}\xi_1\mathbb{E}\xi_2$. (если существуют матожидания справа, то и слева существует)

11. Если ξ_i равномерно (или монотонно) сходится к ξ, то $\mathbb{E}\xi_i \to \mathbb{E}\xi$. (в случае монотонности приходится требовать и существование $\mathbb{E}\xi$)

Единственное свойство "на вырост" — просто определять равномерную и монотонную сходимость в дискретном случае замаемся. Но вот как задачка это интересно...

12. Если ξ ограничено, то $\mathbb{E}\xi$ существует.

13. Если $\mathbb{E}\xi^2$ существует, то и $\mathbb{E}|\xi|, \mathbb{E}\xi$ существуют.

14. Если $\mathbb{E}\xi^2$ и $\mathbb{E}\eta^2$ существуют, то и $\mathbb{E}|\xi\eta|$ существует.

Дисперсия

Определения. *Моментом* k*-го порядка* ($k \in \mathbb{N}$) случайной величины ξ называют выражение $\mathbb{E}\xi^k$. Для любого r *абсолютным моментом* r-го порядка ($k \in \mathbb{N}$) случайной величины ξ называют выражение $\mathbb{E}|\xi|^r$, а *центральным моментом* r-го порядка ($k \in \mathbb{N}$) случайной величины ξ называют выражение $\mathbb{E}|\xi - \mathbb{E}\xi|^r$.

Дисперсией называют центральный момент второго порядка

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

Корень из нее называют *среднеквадратичным отклонением* $\sigma = \sqrt{\mathbb{D}\xi}, \sigma^2 = \mathbb{D}\xi$.

Дисперсия описывает меру отклонения, размах вокруг матожидания, его разбросанность. Итак,

$$\mathbb{D}\xi = \mathbb{E}\Big((\xi - \mathbb{E}\xi)^2\Big) = \sum_{i \in \mathbb{N}} p_i(x_i - \mathbb{E}\xi)^2.$$

А ряд точно сходится?

Эквивалентное определение (что там насчет существования всех частей?)

$$\begin{aligned}
\mathbb{D}\xi &= \mathbb{E}\big((\xi - \mathbb{E}\xi)^2\big) = \mathbb{E}\big(\xi^2\big) - \mathbb{E}\big(2\xi\mathbb{E}\xi\big) + \mathbb{E}\big((\mathbb{E}\xi)^2\big) \\
&= \mathbb{E}\big(\xi^2\big) + 2\mathbb{E}\xi\mathbb{E}\big(\xi\big) + (\mathbb{E}\xi)^2 \\
&= \mathbb{E}\big(\xi^2\big) - (\mathbb{E}\xi)^2.
\end{aligned}$$

Примеры на посчитать

Пример 1 (Равномерное дискретное распределение) $\xi \in U\{a, b\}$, $\mathbb{E}\xi = \frac{b+a}{2}$.

Пример 2 (Распределение Пуассона)

$$\begin{aligned}
\mathbb{E}\xi &= e^{-\lambda}\sum_{k\in\mathbb{N}} \frac{k\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda, \\
\mathbb{E}\xi^2 &= e^{-\lambda}\sum_{k\in\mathbb{N}} \frac{k^2\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda k\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda\lambda^{k-1}}{(k-1)!} + e^{-\lambda}\sum_{k=2}^{\infty}\frac{\lambda(k-1)\lambda^{k-1}}{(k-1)!} \\
&= \lambda + e^{-\lambda}\lambda^2\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda + \lambda^2, \\
\mathbb{D}\xi &= \mathbb{E}\xi^2 - (\mathbb{E}\xi)^2 = \lambda.
\end{aligned}$$

Производящие функции

Для любой случайной величины ξ, принимающей целочисленные неотрицательные значения, можно рассматривать производящую функцию (pgf: probability generating function)

$$\phi_\xi(t) = \mathbb{E}t^\xi = \sum_{k=0} \mathbb{P}\{\omega \,|\, \xi(\omega) = k\} t^k \qquad \forall t.$$

А откуда, из какого множества, здесь t?

Несколько тривиальных свойств:

0. $\phi_\xi(1) = 1$

1. $\mathbb{P}\{\omega \,|\, \xi(\omega) = k\} = \frac{1}{k!}\phi_\xi^{(k)}(0)$. А производная точно существует?

Кстати, это свойство гарантирует, что не только по дискретному распределению можно получить производящую функцию, но и по функции восстановить распреление можно.

2. $\phi'_\xi(1) = \mathbb{E}\xi, \phi_\xi^{(k)}(1) = \mathbb{E}\xi(\xi-1)\dots(\xi-k+1)$ если соответствующие матожидания существуют.

Для доказательства достаточно продифференцировать.

3. Для независимых целочисленных неотрицательных случайных величин $\xi_1, \xi_2, \dots, \xi_k$ выполнено $\psi_{\xi_1+\xi_2+\dots+\xi_k} = \psi_{\xi_1}\psi_{\xi_2}\dots\psi_{\xi_k}$.

Достаточно просто посчитать “в лоб” для $k = 2$.

4. Для независимых целочисленных неотрицательных одинаково распределенных случайных величин $\xi_1, \xi_2, \dots, \xi_k$ и независящей от них случайной величины η, принимающей натуральные значения, выполнено $\phi_{\xi_1+\xi_2+\dots+\xi_\eta} = \phi_\eta(\phi_{\xi_1})$, если все они есть.

Как задачка сойдет.

Выразить производящую функцию разности двух независимых случайных величин, принимающих лишь целочисленные неотрицательные значения.

5. $\phi_{a\xi+b}(t) = t^b\psi_\xi(t^a)$ для целочисленных неотрицательных a, b.

Совсем тривиально.

Примеры на посчитать еще проще

Пример 3 (Геометрическое распределение) Для $\xi \in Geom_1(p)$ получаем

$$\phi_\xi(t) = \sum_{k=1}^{\infty} q^{k-1}pt^k = pt\sum_{k=0}^{\infty}(qt)^k = \frac{pt}{1-qt},$$

$$\phi'_\xi(t) = \frac{d}{dt}\left(\frac{pt}{1-qt}\right) = \frac{\mathbb{P}(1-qt)+pqt}{(1-qt)^2} = \frac{p}{(1-qt)^2}, \qquad \mathbb{E}\xi = \phi'_\xi(1) = \frac{1}{p},$$

$$\phi''_\xi(t) = \frac{d}{dt}\left(\frac{p}{(1-qt)^2}\right) = \frac{2pq}{(1-qt)^3}, \quad \frac{2q}{p^2} = \phi''_\xi(1) = \mathbb{E}\xi(\xi-1) = \mathbb{E}(\xi^2) - \frac{1}{p},$$

$$\mathbb{D}\xi = \frac{2q}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} = \frac{q}{p^2}.$$

Пример 4 (Биномиальное распределение) Тривиально но факт, что для $\xi \in B(1,p)$ выполнено

$$\phi_\xi(t) = q + pt,\ \mathbb{E}\xi = \phi'_\xi(1) = p,\ \mathbb{D}\xi = \phi''_\xi(1) + \phi'_\xi(1) - (\phi'_\xi(1))^2 = pq.$$

Тогда по свойствам выше для $\eta \in B(n,p)$ выполнено

$$\phi_\eta(t) = (q+pt)^n,\ \mathbb{E}\eta = np,\ \mathbb{D}\eta = npq.$$

Ковариация и корреляция

Ковариацией двух случайных величин назовем выражение

$$cov(\xi_1, \xi_2) = \mathbb{E}(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2).$$

Более того,

$$cov(\xi_1,\xi_2)=\mathbb{E}(\xi_1-\mathbb{E}\xi_1)(\xi_2-\mathbb{E}\xi_2)=\mathbb{E}(\xi_1\xi_2)-\mathbb{E}\xi_1\mathbb{E}\xi_2.$$

Если $cov(\xi_1,\xi_2)=0$, то случайные величины называют *некоррелированными*, в частности независимые случайные величины всегда некоррелированы (функционально зависимы).

Обратное неверно. Например, $\sin\frac{2\pi k}{n},\cos\frac{2\pi k}{n}$ для $k\in U_{\{1,n\}}$.

Свойства: для любых случайных величин ξ_1,ξ_2 для любых c_1,c_2:

1. $\mathbb{D}(\xi_1+\xi_2)=\mathbb{D}\xi_1+\mathbb{D}\xi_2+2cov(\xi_1,\xi_2)$;
2. $cov(\xi_1,\xi_2)=0$ в случае независимых ξ_1,ξ_2;
3. $cov(\xi_1+c_1,\xi_2+c_2)=cov(\xi_1,\xi_2)$;
4. $cov(c_1\xi_1,c_2\xi_2)=c_1c_2cov(\xi_1,\xi_2)$;
5. $cov(\xi_1,\xi_2+a\xi_1)=cov(\xi_1,\xi_2)+a\mathbb{D}\xi_1$ для любых ненулевых a;
6. $cov(\xi_1,\xi_1)=\mathbb{D}\xi_1$.

Корреляцией двух случайных величин ξ_1,ξ_2 называют выражение

$$\rho(\xi_1,\xi_2)=\frac{cov(\xi_1,\xi_2)}{\sqrt{\mathbb{D}\xi_1\cdot\mathbb{D}\xi_2}}.$$

$\rho(\xi_1,\xi_2)=0$ выполнено в точности для некоррелированных случайных величин, в частности это так для независимых случайных величин.

Банальное свойство: $\rho(a_1\xi_1+b_1,a_2\xi_2+b_2)=\rho(\xi_1,\xi_2)$, то есть Корреляция инвариантна относительно сдвига и растяжения (на число) каждого аргумента.

Теорема. 1) Если коэффициент корреляции существует, то $\rho(\xi_1,\xi_2)\in[-1,1]$.

2) Если при этом $|\rho(\xi_1,\xi_2)|=1$ и $\mathbb{D}\xi_1\cdot\mathbb{D}\xi_2\neq 0$, то для некоторых не равных нулю одновременно чисел b,c выполнено $\xi_2=c+b\xi_1$ почти для всех ω, и b имеет знак $\rho(\xi_1,\xi_2)$.

Доказательство раз.

Для всех $t\in\mathbb{R}$ выполнено $0\le\mathbb{E}[(t(\xi_1-\mathbb{E}\xi_1)+(\xi_2-\mathbb{E}\xi_2))^2]\le t^2\mathbb{E}(\xi_1-\mathbb{E}\xi_1)^2+2t\,cov(\xi_1,\xi_2)+\mathbb{E}(\xi_2-\mathbb{E}\xi_2)^2]=t^2\mathbb{D}\xi_1+2tcov(\xi_1,\xi_2)+\mathbb{D}\xi_2$. Тогда детерминант неотрицателен, то есть $|cov(\xi_1,\xi_2)|^2\le\mathbb{D}\xi_1\cdot\mathbb{D}\xi_2$, что и требовалось доказать.

При этом, если $|cov(\xi_1,\xi_2)|=1$, то многочлен имеет корень, то есть при некотором t выполнено $\mathbb{E}[(t(\xi_1-\mathbb{E}\xi_1)+(\xi_2-\mathbb{E}\xi_2))^2]=0$, откуда $t(\xi_1-\mathbb{E}\xi_1)+(\xi_2-\mathbb{E}\xi_2)=0$ почти всегда, то есть $t\xi_1+\xi_2$ константа почти всегда.

Доказательство два.

Можно считать, что $\mathbb{E}\xi_1=\mathbb{E}\xi_2=0,\mathbb{D}\xi_1=\mathbb{D}\xi_2=1$, Теперь $2|cov(\xi_1,\xi_2)|=|\mathbb{E}(2\xi_1\xi_2)|\le\mathbb{E}\xi_1^2+\mathbb{E}\xi_2^2=2$. В случае равенства имеем $\xi_1\pm\xi_2=0$.

Верно ли, что в формулировке теоремы можно написать $\xi_1=c+b\xi_2$?

Случайные величины.

Измеримые отображения.

Даны измеримые пространства $(\Omega,\mathcal{F}),(\Omega',\mathcal{G})$.

Определение. Отображение $\xi:\Omega\to\Omega'$ измеримо, если для любого измеримого множества $B\in\mathcal{G}$ выполнено

$$\xi^{-1}(B)=\{\omega\,|\,\xi(\omega)\in B\}\in\mathcal{F}.$$

В частности, если $B\in\mathcal{G}$ — событие, то $\{\omega|\xi(\omega)\in B\}$ принадлежит $\mathcal{F}$, то есть также событие. В частности на вероятностном пространстве (у нас пока лишь измеримое) законен запрос $\mathbb{P}\{\omega\,|\,\xi(\omega)\in B\}$.

Примеры.

1. Если $\mathcal{F}=2^\Omega$, то любое отображение $\xi\colon\Omega\to\Omega'$ измеримо.
2. Если $\mathcal{F}=\{\varnothing,\Omega\}$, то отображение $\xi\colon\Omega\to\Omega'$ измеримо тогда и только тогда, когда равно константе.
3. Если $\Omega=\{w,b,r\}$, $\mathcal{F}=\{\varnothing,\{w,b,r\},\{w\},\{b,r\}\}$, то отображение $\xi\colon\Omega\to\Omega'$ измеримо тогда и только тогда, когда $\xi(b)\in B\Leftrightarrow\xi(r)\in B$ для всех $B\in\mathcal{G}$. Таким образом элементами $B\in\mathcal{G}$ нельзя отделить b и r друг от друга.
4. Пусть Ω — некоторое множество непрерывных на отрезке $[0,1]$ функций. Можно так определить неубывающее (параметризованное параметром t) семейство σ-алгебр $\mathcal{F}^t\subset\mathcal{F}^1$ (фильтрация), что для

каждого $t \in [0,1]$ отображение $\xi\colon \Omega \to \Omega'$ измеримо относительно $\mathcal{F}^t$ в точности тогда, когда $\xi(f) = \xi(g)$ при $f|_{[0,t]} = g|_{[0,t]}$ (произошедшее после момента времени t не влияет на значение отображения ξ). Таким образом $\mathcal{F}^t$ описывает информацию о процессе, имеющуюся к моменту t.
5. Всякое отображение ξ задает σ-алгебру $\mathcal{F}(\xi) = \sigma\{\xi^{-1}(B) \,|\, B \in \mathcal{G}\}$ — по сути вся информация, которая может быть восстановлена по значению отображения ξ.
6. Всякий набор Ξ отображений задает σ-алгебру $\mathcal{F}(\Xi) = \sigma\{\xi^{-1}(B) \,|\, B \in \mathcal{G}, \xi \in \Xi\}$ — по сути вся информация, которая может быть восстановлена по значениям отображений $\xi \in \Xi$.

Далее пока будем считать, что $(\Omega', \mathcal{F}') = (\mathbb{R}, \mathcal{B})$. Таким образом иы рассматриваем скалярные случайные величины. Здесь $\mathcal{B}$ — совокупность борелевских множеств.

Борелевские множества, напоминание

Определение 0. Для топологического пространства (Ω, τ), где τ - совокупность всех открытых подмножеств множества Ω, под σ-алгеброй борелевских множеств понимается $\mathcal{B} = \sigma(\tau)$.

Определение 1. Минимальную σ-алгебру, содержащую все n-мерные параллелепипеды $\{[a,b] \times \ldots \times [c,d] \subset \mathbb{R}^n\}$, называют борелевской σ-алгеброй над $\mathbb{R}^n$, а ее элементы — борелевскими множествами.

Определение 2. Минимальную σ-алгебру, содержащую все промежутки

$$\{[a,b], [a,b), (a,b], (a,b) \,|\, a, b \in \mathbb{R}, a < b\},$$

называют борелевской σ-алгеброй над $\mathbb{R}$, а ее элементы — борелевскими множествами.

Отметим, что любая точка метрического пространства является борелевским множеством (почему?). В частности, борелевские множества над $\mathbb{R}$ можно строить взяв за основу, например, лишь все полуинтервалы вида $(-\infty, a]$.

Случайные величины

Определение 1. Случайной (скалярной) величиной ξ называется такая функция $\xi : \Omega \to \mathbb{R}$, что для любого борелевского множества $B \in \mathcal{B}$ выполнено

$$\xi^{-1}(B) = \{\omega \,|\, \xi(\omega) \in B\} \in \mathcal{F}.$$

Определение 2 Случайной (скалярной) величиной ξ называется такая функция $\xi : \Omega \to \mathbb{R}$, что для всех $a, b \in \mathbb{R}(a < b)$ выполнено

$$\xi^{-1}((a,b]) = \{\omega \,|\, a < \xi(\omega) \le b\} \in \mathcal{F}.$$

Определение 3 Случайной (скалярной) величиной ξ называется такая функция $\xi : \Omega \to \mathbb{R}$, что для всех $a, b \in \mathbb{R}(a < b)$ выполнено

$$\xi^{-1}((-\infty,b]) = \{\omega \,|\, \xi(\omega) \le b\} \in \mathcal{F}.$$

Теорема. Определения 1-3 эквивалентны.
Очевидно, что из 1 следует 2, а из 2 следует 3. Покажем, что из 3 следует 1.

В частности для случайной величины ξ событиями являются $\{\omega \,|\, \xi(\omega) = x\}$, $\{\omega \,|\, \xi(\omega) < x\}$, $\{\omega \,|\, \xi(\omega) > x\}$, $\{\omega \,|\, \xi(\omega) \le x\}$, $\{\omega \,|\, \xi(\omega) \ge x\}$, $\{\omega \,|\, y \le \xi(\omega) \le x\}$, $\{\omega \,|\, y < \xi(\omega) \le x\}$, $\{\omega \,|\, y \le \xi(\omega) < x\}$, $\{\omega \,|\, y < \xi(\omega) < x\}$.

Задачи на баллы.

Докажите, что всякая непрерывная функция из $(\mathbb{R}, \mathcal{B})$ в $(\mathbb{R}, \mathcal{B})$ измерима.

Докажите, что функция, являющаяся поточечным пределом последовательности случайных величин, сама является случайной величиной. А вот если слово “последовательность” убрать, то факт будет неверен, приведите соответствующий контрпример.

Докажите, что всякую случайную величину можно представить в виде предела последовательности случайных величин, принимающих не более чем счетное число значений. При этом такой предел можно считать и равномерным, и монотонным одновременно.

Как из предыдущего легко доказать, что сумма, произведение, максимум случайных величин также являются случайными величинами?

О случайных величинах на вырост

Даны измеримые пространства $(\Omega, \mathcal{F})$, $(\Omega', \mathcal{G})$. Пусть имеется некоторое измеримое отображение $f : \Omega \to \Omega'$. Оно естественно порождает некоторую, вообще говоря возможно меньшую, σ-подалгебру $\sigma(f) \subset \mathcal{F}$. Когда некоторая случайная величина $g : \Omega \to \mathbb{R}$ является $\sigma(f)$-измеримой?

Теорема Дуба. Функция $g : \Omega \to \mathbb{R}$ является $\sigma(f)$-измеримой тогда и только тогда, когда для некоторой случайной величины $h : \Omega' \to \mathbb{R}$ выполнено $g(\omega) = h(f(\omega))$ для всех $\omega \in \Omega$.

Доказательство сами. 1,5 балла.

Собственно факт очень полезный и без всякой вероятности. Вы имеете точный прибор, дающий показание f, и еще один прибор, дающий показание g. Когда один из этих приборов можно выкинуть как лишний?

В статистике на подобные вещи завязано понятие “достаточная статистика”, когда из большого массива данных мы вычисляем одно-два числа (например среднее значение) и уже только по нему восстанавливаем нужную нам характеристику.

Функции распределения

Пусть с нами снова вероятностное пространство $(\Omega, \mathcal{F}, P)$.

Определение Функцией распределения случайной величины ξ называют отображение $F_\xi : \mathbb{R} \to [0, 1]$, заданное по правилу:

$$F_\xi(x) = \mathbb{P}\{\omega \,|\, \xi(\omega) \le x\}.$$

Можно порисовать картинки для дискретных распределений: вырожденное, Бернулли, геометрическое.

Теперь при любых $x, y \in \mathbb{R}, (x < y)$

$$\begin{aligned}
\mathbb{P}\{\omega \,|\, \xi(\omega) = x\} &= F_\xi(x) - F_\xi(x-0), & \mathbb{P}\{\omega \,|\, \xi(\omega) < x\} &= F_\xi(x-0),\\
\mathbb{P}\{\omega \,|\, \xi(\omega) > x\} &= 1 - F_\xi(x), & \mathbb{P}\{\omega \,|\, \xi(\omega) \le x\} &= F_\xi(x),\\
\mathbb{P}\{\omega \,|\, \xi(\omega) \ge x\} &= 1 - F_\xi(x-0), & \mathbb{P}\{\omega \,|\, y \le \xi(\omega) \le x\} &= F_\xi(x) - F_\xi(y-0),\\
\mathbb{P}\{\omega \,|\, y < \xi(\omega) \le x\} &= F_\xi(x) - F_\xi(y), & \mathbb{P}\{\omega \,|\, y \le \xi(\omega) < x\} &= F_\xi(x-0) - F_\xi(y-0),\\
& & \mathbb{P}\{\omega \,|\, y < \xi(\omega) < x\} &= F_\xi(x-0) - F_\xi(y).
\end{aligned}$$

Свойства:

1. F_ξ не убывает,
2. $\lim_{x\to+\infty} F_\xi(x) = 1$,
3. $\lim_{x\to-\infty} F_\xi(x) = 0$,
4. для любого $x_0 \in \mathbb{R}$ существуют односторонние пределы

$$F_\xi(x_0+0) = \lim_{x\to x_0+0} F_\xi(x), \quad F_\xi(x_0-0) = \lim_{x\to x_0-0} F_\xi(x),$$

5. функция F_ξ непрерывна справа: $F_\xi(x_0+0) = F_\xi(x_0)$.

Свойства сии доказаны где угодно, в любом учебнике выше уровня ПТУ.

Теорема Колмогорова. Всякой функции $F : \mathbb{R} \to \mathbb{R}$, удовлетворяющей свойствам 1,2,3,5, найдутся такие измеримое пространство Ω и случайная величина $\xi \colon \Omega \to \mathbb{R}$, что $F = F_\xi$.

Эскиз доказательства. Принимается $\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}$. Для всякого непустого интервала $(a, b]$ принимается $\mathbb{P}((a, b]) = F(b) - F(a)$. Остается доказать, что это отображение имеет счетно-аддитивное продолжение на $\mathcal{B}$, то есть задает вероятность. В силу теоремы Каратеодори для этого достаточно доказать, что оно имеет счетно-аддитивное продолжение на алгебре $\mathcal{A}$ всевозможных не более чем конечных объединений промежутков вида $(a, b]$, поскольку аддитивное продолжение оно заведомо имеет, то нужно лишь показать, его счетно-адитивность, то есть показать

$$\mathbb{P}(\cup_{i\in\mathbb{N}}(a_i, b_i]) = \sum_{i\in\mathbb{N}} \mathbb{P}((a_i, b_i])$$

в случае, если $\cup_{i\in\mathbb{N}}(a_i, b_i] = (a, b]$, и все промежутки $(a_i, b_i]$ попарно не пересекаются.

Для доказательства "$\leq$", достаточно перейти в пределу в очевидном

$$\mathbb{P}((a,b]) = \mathbb{P}(\cup_{i\in\mathbb{N}}(a_i,b_i]) \geq \sum_{i=1}^{n} \mathbb{P}((a_i,b_i]).$$

Для доказательства "$\geq$" заметим, что можно найти конечные $a^\varepsilon > a$ и $b_i^\varepsilon > b_i$ так, что $\mathbb{P}((a^\varepsilon,b]) \geq \mathbb{P}((a,b]) - \varepsilon$, $\mathbb{P}((a_i,b_i^\varepsilon)) \leq \mathbb{P}((a_i,b_i]) + \varepsilon 2^{-i}$. Теперь

$$(a^\varepsilon, b] \subset \cup_{i\in\mathbb{N}}(a_i, b_i^\varepsilon),$$

и из этого открытого покрытия найдется конечное подпокрытие $(a^\varepsilon,b] \subset \cup_{k=1}^{n}(a_{i_k}, b_{i_k}^\varepsilon)$. Тогда

$$\mathbb{P}((a,b]) - \varepsilon \leq \mathbb{P}((a^\varepsilon,b]) \leq \sum_{k=1}^{n} \mathbb{P}((a_{i_k}, b_{i_k}^\varepsilon)) < \varepsilon + \sum_{k=1}^{n} \mathbb{P}((a_{i_k}, b_{i_k})) < \varepsilon + \sum_{i\in\mathbb{N}} \mathbb{P}((a_i,b_i)).$$

Переходя к пределу по ε все показано

Вопрос попроще: где использовалась монотонность? Посложнее: а где использовались свойства 2,3? Ну и наконец, пока не разберетесь, где здесь применяется понятие "компакт", на зачет с этой теоремой лучше не появляться...

Образ меры

Пусть имеется измеримое пространство $(\Omega,\mathcal{F})$ с некоторой мерой μ и измеримое пространство $(\tilde{\Omega},\tilde{\mathcal{F}})$. Пусть $\xi : \Omega \to \tilde{\Omega}$ измеримо.

Введем индуцированную меру (образ меры, push-forward) $\tilde{\mu} = \mu \circ \xi^{-1} = \xi\#\mu$:

$$\tilde{\mu}(\tilde{A}) = \mu(\xi^{-1}(\tilde{A})) \qquad \forall \tilde{A} \in \tilde{\mathcal{F}}.$$

Таким образом мы перетащили меру с одного измеримого пространства на другое.

Функция распределения как образ меры. Пусть $\tilde{\Omega} = \mathbb{R}$. Тогда произвольная случайная величина задает на $\mathbb{R}$ вероятность μ правилом $\mu((\infty, a]) = F_\xi(a)$. Таким образом, F_ξ задает вероятностную меру на борелевских подмножествах $\mathbb{R}$.

Абсолютно непрерывные случайные величины

Напомним, что действительная случайная величина ξ дискретна, если имеется не более чем счетное число различных значений a_i которые она принимает, и все остальные значения она принимает с вероятностью 0, то есть $\mathbb{P}\{\omega \,|\, \forall i \in \mathbb{N}\ \xi(\omega) \neq a_i\} = 0$. При этом $F_\xi(x) = \sum_{x_i \leq x} p_i$.

В этом случае понятно, насколько каждое значение вероятнее другого.

Пример 1. Рассмотрим $\xi \in U[0,1]$, она равномерно распределена на $[0,1]$, и $F_\xi(x) = x$ для всех $x \in [0,1]$, $F_\xi(x) = 0$ для всех $x < 0$, $F_\xi(x) = 1$ для всех $x > 1$.

Пример 2. (Экспоненциальное распределение) Для всякого $\lambda > 0$ можно ввести действительную случайную величину $\xi \in Ex\mathbb{P}(\lambda)$, задав распределение $F_\xi(x) = 1 - e^{-\lambda x}$ для всех $x \geq 0$, и $F_\xi(x) = 0$ в противном случае.

Понятно, что во втором примере ноль вроде более вероятен, чем один, но вероятность выпадения и нуля и единицы везде ноль... В том числе и поэтому введем еще одно определение.

Определение. Действительная случайная величина ξ *абсолютно непрерывна*, если найдется такая измеримая борелевская функция $f_\xi : \mathbb{R} \to \mathbb{R}$ (называемая *плотностью*), что $F_\xi(b) - F_\xi(a) = \int_a^b f_\xi(x)\,dx$ для всех $a < b$.

Легко проверить, что при этом $F'_\xi(x) = f_\xi(x)$, $\qquad F_\xi(x) = \int_{-\infty}^{x} f_\xi(t)\,dt$.

Отсюда для равномерного распределения $f_\xi(x)$ равна 1 на интервале $(0,1)$, 0 вне отрезка $[0,1]$. А вот экспоненциальное распределение имеет плотность $f_\xi(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x)$. А чему они равны в точке 0?

Вывод: плотность определяется неоднозначно, а насколько неоднозначно? Да, dx это фактически тоже мера...

Задача. Докажите, что непрерывная плотность обязана быть неотрицательной, а не являющаяся таковой — не обязана.

Мера Лебега

Определение. Мерой Лебега λ над $\mathbb{R}$ называют определенную на всех борелевских подмножествах прямой счетно-аддитивную функцию, равную на каждом конечном промежутке его длине: $\lambda([a,b]) = b-a, \lambda([a,b)) = b-a, \lambda((a,b]) = b-a, \lambda((a,b)) = b-a$ для всех $a,b \in \mathbb{R}, a<b.$

По теореме Каратеодори такая мера единственна.

Теперь можно определить понятие “почти всюду.”

Определение. Говорят, что две функции f,g действительного переменного совпадают почти всюду, если для некоторого борелевского множества $B, \lambda(B)=0$ при всех $x \in \mathbb{R}$ из $f(x) \neq g(x)$ следует $x \in B.$

Аналогично вводится понятие “почти всюду относительно вероятности $P.$”

Абсолютная непрерывность относительно меры

Определение. Заряд S назовем абсолютно непрерывным относительно меры μ, если для всякого $A \in \mathcal{F}$ из $\mu(A)=0$ следует $S(A)=0.$

Эквивалентное определение. Заряд S назовем абсолютно непрерывным относительно вероятности μ, если для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что из $\mu(A) < \delta$ следует $|S(A)| < \varepsilon.$

Что такое “случайная величина ξ абсолютно непрерывна относительно меры Лебега”?

Формула замены переменных

Для абсолютно непрерывной д.с.в ξ и строго возрастающей дифференцируемой функции $g:\mathbb{R} \to \mathbb{R}$ рассмотрим д.с.в $\eta = g(\xi)$. Она также абсолютно непрерывна: действительно, для всех a', b' найдутся $a = g^{-1}(a') < b = g^{-1}(b')$ (g^{-1} неубывающая), и из $a' < b'$ следует $a < b$.

Теперь

$$F_\eta(b') - F_\eta(a') = F_\eta(g(b)) - F_\eta(g(a)) = F_\xi(b) - F_\xi(a)$$
$$= \int_a^b f_\xi(y)\,dy = \int_{g^{-1}(a')}^{g^{-1}(b')} f_\xi(g^{-1}(x))\,dg^{-1}(x) = \int_{a'}^{b'} \frac{f_\xi(g^{-1}(x))}{g'(g^{-1}(x))}\,dx.$$

Итак, для строго монотонных функций g

$$f_\eta(x) = \frac{f_\xi(g^{-1}(x))}{|g'(g^{-1}(x))|},\ f_{g(\xi)}(g(y)) = \frac{f_\xi(y)}{|g'(y)|}.$$

Откуда взялся модуль? А зачем требование строгой монотонности?

В частности,

$$f_{a\xi+b}(x) = \frac{f_\xi(\frac{x-b}{a})}{|a|}.$$

Характеристики распределений

Мода — значение во множестве наблюдений, которое встречается наиболее часто. Модой абсолютно непрерывного распределения называют любую точку глобального максимума плотности распределения, иногда вводят понятие локальной моды. Распределение называется унимодальным, если мода, даже с учетом локальных, одна. Для дискретных распределений модой считают любое значение a_i, вероятность которого p_i больше, чем вероятности соседних значений.

Медиана распределения F_ξ (распределения случайной величины ξ)— такое x, что $F_\xi(x) = 0,5$. Верно ли, что для абсолютно непрерывных величин она определяется однозначно?

Первая (нижняя) квартиль — такое x, что $F_\xi(x) = 0,25$; третья (верхняя) квартиль — такое x, что $F_\xi(x) = 0,75$.

Вам дали абсолютно непрерывную случайную величину ξ. Как найти ее математическое ожидание?

Примеры абсолютно непрерывных распределений

Пример 3. (Стандартное нормальное распределение $N(0,1)$) Задается плотностью

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad \forall x \in \mathbb{R}.$$

При этом $F_\xi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \qquad \forall x \in \mathbb{R}.$

Для стандартного нормального распределения $N(0,1)$

Уровень	99,99	99,90	99,00	97,72	97,50	95,00	90,00	84,13	50,00
Квантиль	3,715	3,090	2,326	2,000	1,960	1,645	1,282	1,000	0,000

Пример 4. (Нормальное распределение $N(m,\sigma)$) Для каждого $m \in \mathbb{R}, \sigma > 0$ задается плотностью

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-m)^2/2\sigma^2} \qquad \forall x \in \mathbb{R}.$$

При этом $F_\xi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(t-m)^2/2\sigma^2}\, dt \qquad \forall x \in \mathbb{R}.$

Легко видеть, что это семейство выдерживает сдвиги и растяжения.

Пример 5. (Распределение Коши) Задается плотностью

$$f_\xi(x) = \frac{1}{\pi} \frac{1}{1+x^2} \qquad \forall x \in \mathbb{R}.$$

Тогда $F_\xi(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg} x \qquad \forall x \in \mathbb{R}.$

Распределением Коши характеризуется длина отрезка, отсекаемого на оси абсцисс прямой, закреплённой в точке на оси ординат на высоте 1, если угол между прямой и осью ординат имеет равномерное распределение на $[-\pi,\pi]$ (т.е. направление прямой изотропно на плоскости).

(Если η распределена по $U(-\pi,\pi)$, то $\mathrm{tg}\,\#\eta$ распределено по Коши).

Выборочное среднее выборки из какого-нибудь распределения Коши само имеет такое же распределение Коши.

Характеристики распределений

Мода — значение во множестве наблюдений, которое встречается наиболее часто. Модой абсолютно непрерывного распределения называют любую точку глобального максимума плотности распределения, иногда вводят понятие локальной моды. Распределение называется унимодальным, если мода, даже с учетом локальных, одна. Для дискретных распределений модой считают любое значение a_i, вероятность которого p_i больше, чем вероятности соседних значений.

Медиана распределения F_ξ (распределения случайной величины ξ)— такое x, что $F_\xi(x) = 0,5$. Верно ли, что для абсолютно непрерывных ыеличин она определяется однозначно?

Первая (нижняя) квартиль - такое x, что $F_\xi(x) = 0,25$; третья (верхняя) квартиль - такое x, что $F_\xi(x) = 0,75$.

Вам дали абсолютно непрерывную случайную величину ξ. Как найти ее математическое ожидание?

ИНТЕГРАЛ

Интеграл. Хотелки

Пока $\int$ некий абстрактный значок, а f, g — не менее абстрактные функции.

Что хочется

Х1. $\int f + \int g = \int (f+g)$;
Х2. $\int f - \int g = \int (f-g)$;
Х3. $c \int f = \int (cf)$ для всех $c \in \mathbb{R}$;
Х4. $\int f \le \int g$ в случае $f \le g$.

Что забыли — по-какому множеству ищем:

Х5. $\int_A f + \int_B f = \int_{A \cup B} f$ для непересекающихся A, B;

Это некритично, если можно интегрировать функцию 1_A и умножать на нее. Теперь $\int_A f + \int_B g = \int f 1_A + \int f 1_B = \int f(1_A + 1_B) = \int_{A\cup B} f$. Но тогда…

X6. $\int 1_A$ должен быть определен.

Однако множество A может сколь угодно плохим множеством (парадокс Банаха-Тарского и прочая...)

Первое ограничение: не все придется интегрировать.

Вывод — есть отображения которые можно интегрировать, а есть которые нельзя.

Рецепт (один из возможных, есть и другие): задать на $(\Omega, \mathcal{F})$ какую-нибудь меру μ и воспользоваться измеримостью.

Говорят, что функция f μ**-интегрируема**, если интеграл $\int f\,d\mu$ определен.

Да, даже так не все измеримые функции будут μ-интегрируемы.

Алгебра индикаторных функций

Каждому множеству $A \subset \Omega$ сопоставим его индикаторную функцию $1_A : \Omega \to \{0,1\}$ правилом $1_A(\omega) = 1$ при $\omega \in A$, и 0 иначе. Множество всех индикаторных функций $1_A, A \in \mathcal{F}$ алгебры $\mathcal{F}$ обозначим через $\mathcal{I}$. Можно заметить, что это, как и $\mathcal{F}$, булева алгебра и полное частично упорядоченное пространство с единицей. Значит можно попробовать обеспечить $X4$, $X5$, $X6$.

Пусть на измеримом пространстве $(\Omega, \mathcal{F})$ задана некоторая мера μ. Тогда примем

$$\int_A d\mu(\omega) = \int_\Omega 1_A(\omega)\,d\mu(\omega) = \mu(A).$$

По возможности лишние значки ω, Ω будем опускать, например

$$\int 1_A\,d\mu = \mu(A).$$

Но сильно к хорошему не привыкайте, в учебниках могут быть и другие варианты обозначений, например

$$\int_\Omega 1_A(\omega)\,\mu(d\omega).$$

Итак, **все индикаторные функции** μ**-интегрируемы**.

В случае меры Лебега λ получаем

$$\int_{[0,1]} d\lambda = \lambda([0,1]) = 1, \ \int_{[a,b]} d\lambda = \lambda([a,b]) = b - a, \ \int_{\mathbb{R}} d\lambda = \lambda(\mathbb{R}) = +\infty.$$

Второе ограничение: при интегрировании могут возникнуть бесконечности.

Рецепт (один из возможных, есть и другие): допустить $+\infty$ как значение интеграла. Такие интегралы мы допустим, но будем их называть расходящимися.

Говорят, что функция f μ**-суммируема**, если интеграл от нее определен и конечен.

Для вероятностной меры μ **все индикаторные функции** μ**-суммируемы**.

Да, $X1$-$X3$ не выполнены на $\mathcal{I}$. Ради них придется создать линейное пространство...

Ступенчатые действительные функции

Определение. Отображение $\xi : \Omega \to \mathbb{R}$ называется ступенчатой (простой) случайной величиной, если найдутся конечное разбиение $H_1, \ldots, H_i, \ldots \in \mathcal{F}$ множества Ω и конечный набор чисел x_i, такие, что $\xi(\omega_i) = x_i$ для всех $\omega_i \in H_i$. Множество всех ступенчатых функций назовем $\mathcal{S}$.

Свойства: С1. это линейное пространство, более того, это линейное подпространство всех функций $f : \Omega \to \mathbb{R}$, натянутое на $\mathcal{I}$;

С2. оно замкнуто относительно произведения;

С3. на нем можно ввести частичный порядок, продолжающий порядок на $\mathcal{I}$ (поточечный порядок), при этом нижняя и верхняя грань достигаются, то есть $\inf(\xi_1, \xi_2), \sup(\xi_1, \xi_2) \in \mathcal{S}$ для всех $\xi_1, \xi_2 \in \mathcal{S}$.

Легко видеть, что оно это множество не выдерживает счетного числа операций $\sup$, $\inf$, $+$, $-$, $*$. Также оно не замкнуто относительно взятия предела.

Но определить интеграл можно.

Интеграл для ступенчатых функций

Всякое ступенчатое отображение f можно разложить в сумму вида

$$f = \sum_{k=1}^{m} a_k 1_{\{\omega | f(\omega) = a_k\}} = \sum_{k \in \mathbb{N}} a_k 1_{\{\omega | f(\omega) = a_k)\}}.$$

Для всякой ступенчатой функции f определим $\int f d\mu = \sum_{k=1}^{m} p_k \int 1_{\{\omega | f(\omega) = a_k)\}} \, d\mu$.

Докажите, что для всякой вероятностной меры μ все ступенчатые функции f μ-суммируемы.

Докажите, что $X1$-$X6$ выполнены на $\mathcal{S}$ для всякой вероятностной меры μ.

Чему равен $\int sign(x) \, d\lambda(x)$?

Третье ограничение: при интегрировании нужно избежать неопределенности типа $\infty - \infty$.

При этом, **все неотрицательные ступенчатые функции μ-интегрируемы** для всякой, не обязательно вероятностной, меры μ.

Докажите, что $X1$, $X3$-$X6$ выполнены для всякой меры μ если f, g неотрицательны.

Докажите, для всякой меры μ, если $\int f d\mu$, $\int g d\mu$ конечны, то $X2$ выполнено.

Теорема. Для всякого отображения $\xi : \Omega \to \mathbb{R}$ следующие условия эквивалентны:
1. ξ является действительной случайной величиной;
2. ξ является пределом последовательности ступенчатых случайных величин;

Более того, для всякого неотрицательного отображения эти условия эквивалентны также условию:
3. ξ является пределом неубывающей последовательности случайных величин, равномерной на всяком множестве, на котором ξ ограничена снизу.

Доказательство: $2 \Rightarrow 1$. Достаточно проверить измеримость $\xi^{-1}((-\infty, a]))$.

Пусть ξ является пределом монотонно невозрастающей последовательности д.с.в. ξ_k. Тогда имеем $\xi^{-1}((-\infty, a])) = \cap_{k \in \mathbb{N}} \xi_k^{-1}((-\infty, a]))$, и все показано. Аналогично для монотонно неубывающей последовательности.

В общем случае, пусть ξ является пределом последовательности д.с.в. ξ_k. Тогда последовательности $\eta_k^j = \sup_{j \le i \le i+k} \xi_i$ также неубывающие, и их пределы $\eta^j = \sup_{j \le i} \xi_i$ также невозрастающие последовательности д.с.в., и все показано.

$1 \Rightarrow 2, 1 \Rightarrow 3$. Достаточно, начиная с достаточно большого n, задать

$$\xi_n = n 1_{\{\omega | \xi(\omega) \ge n\}} - n 1_{\{\omega | \xi(\omega) < n\}} + \sum_{q=-n2^n+1}^{n2^n} \frac{q-1}{2^n} 1_{\{\omega | 2^n \xi(\omega) \in [q-1, q)\}}.$$

Интеграл как предел

Пусть ξ — неотрицательная действительная случайная величина.

Определение. Назовем число R интегралом д.с.в. ξ по мере μ (и обозначим его через $\int \xi d\mu$), если для некоторой монотонно неубывающей сходящейся к ξ последовательности ступенчатых д.с.в. ξ_n выполнено $R = \lim_{n \to \infty} \int \xi_n d\mu$.

Все неотрицательные действительные случайные величины μ-интегрируемы для всякой, не обязательно вероятностной, меры μ.

Всякая функция ξ может быть представлена в виде

$$\xi 1_{\{\omega | \xi(\omega) > 0\}} + \xi 1_{\{\omega | \xi(\omega) < 0\}} = \xi_+ - \xi_-.$$

При этом ξ_+, ξ_- неотрицательные д.с.в.

Положим в этом случае

$$\int \xi d\mu = \int \xi_+ d\mu - \int \xi_- d\mu.$$

Итак, д.с.в. ξ **μ-интегрируема $\Leftrightarrow$ μ-суммируема хотя бы одна** из д.с.в ξ_+, ξ_-;
и д.с.в. ξ **μ-суммируема $\Leftrightarrow$ μ-суммируема каждая** из д.с.в ξ_+, ξ_-.

Сделанное здесь определение не вполне корректно, поскольку не гарантирует единственность интеграла для каждой функции, чуть позже мы докажем независимость от выбора последовательности ступенчатых функций в лоб, но сначала взгляд чуть со стороны.

Интеграл как абстракция

Спрашивать ни формулировку, ни тем более доказательство не буду, также постараемся обойтись без ее использования, но в качестве ориентира сойдет.

Теорема Даниэля. Пусть имеется векторное пространство $\mathcal{S}$ функций $\xi : \Omega \to \mathbb{R}$, содержащее константы (в частности содержащее 1) и замкнутое относительно операции $\sup$. Пусть имеется неотрицательный линейный функционал I над $\mathcal{S}$, для которого $I\,1 = 1$. Тогда для существования на $(\Omega, \sigma(\mathcal{S}))$ вероятностной меры μ со свойством: каждая $\xi \in \mathcal{S}$ μ-интегрируема и

$$I\,\xi = \int_\Omega \xi(\omega)\, d\mu(\omega),$$

необходимо и достаточно, чтобы для всякой убывающей к нулю последовательности $\xi_n \in \mathcal{S}$ было выполнено $\lim_{n\to\infty} I\,\xi_n = 0$.

На тех же условиях, в качестве еще одной путеводной звезды:

Теорема Рисса. Пусть Ω — метрический компакт, $\mathcal{B}$ — его борелевская σ-алгебра, $C(\Omega)$ — пространство всех непрерывных функций $f : \Omega \to \mathbb{R}$. Для всякого непрерывного линейного функционала I над $C(\Omega)$ со свойством $f \geq 0 \Rightarrow I\,f \geq 0$ найдется единственная мера μ над $(\Omega, \mathcal{B})$ со свойством

$$I\,f = \int f\, d\mu \forall f \in C(\Omega).$$

Если свойство неотрицательности ($f \geq 0 \Rightarrow I\,f \geq 0$) убрать, то слово мера надо заменить на слово заряд.

Единственность

Как можно заметить, например из формулировки теоремы Даниэля, просто так единственность интеграла не гарантируется, нам придется ее доказать. Самой теоремой Даниэля при этом пользоваться мы не будем.

Лемма 1. Пусть f_k — монотонно убывающая последовательность ступенчатых функций, сходящаяся к нулю. Тогда $\lim_{n\to\infty} \int f_n d\mu = 0$.

Доказательство. Зададим $M = \sup_{\omega\in\Omega, k\in\mathbb{N}} f_k(\omega) = \max_{\omega\in\Omega} f_1(\omega)$. Теперь определим $A_n = \{\omega | f_n(\omega) > 1$ Поскольку $\cap_{n\in\mathbb{N}} A_n = \emptyset$, то в силу счетно-аддитивности $\mu(A_n) \to 0$. Примем $g_n = \frac{1}{n} 1_{\Omega\setminus A_n} + M 1_{A_n}$. В силу $g_n > f_n \geq 0$ имеем

$$0 \leq \lim_{n\to\infty} \int f_n d\mu \leq \lim_{n\to\infty} \int g_n d\mu \leq \lim_{n\to\infty} \Big(\frac{1}{n}\mu(\Omega \setminus A_n) + M\mu(A_n)\Big) = 0.$$

Лемму 1 доказали.

Лемма 2. Пусть g — ступенчатая функция, а последовательность ступенчатых измеримых функций f_n сходится к некоторой f. Тогда из $0 \leq g \leq f$ следует

$$\int g d\mu \leq \lim_{n\to\infty} \int f_n d\mu.$$

Доказательство. Фиксируем $\alpha \in (0,1)$. Для каждого n введем $A_n = \{\omega | \alpha g(\omega) \leq f_n(\omega)\}$. Теперь $\cup_{n\in\mathbb{N}} A_n = \Omega$ и $1_{\{\Omega\setminus A_n\}} \to 0$. Но тогда, по Лемме 1, $\int \alpha g 1_{A_n} d\mu \to \int \alpha g d\mu$.

При этом

$$\alpha \int g 1_{A_n} d\mu = \int \alpha g 1_{A_n} d\mu \leq \int f_n d\mu.$$

Переходя к пределу, имеем $\alpha \int g d\mu \leq \lim_{n\to\infty} \int f_n d\mu$ для всех $\alpha \in (0,1)$, а значит и для $\alpha = 1$. Лемму 2 доказали.

Где в доказательстве Леммы 2 использовали ступенчатость f_n?

Теорема об единственности. Интеграл от измеримой неотрицательной функции не зависит от выбора аппроксимирующей последовательности.

Доказательство. Рассмотрим некоторую измеримую неотрицательную функцию f, пусть даны две монотонно неубывающие последовательности неотрицательных ступенчатых функций f_n, g_n. По Лемме 2, $\int g_n d\mu \le \lim_{n\to\infty}\int f_n d\mu$, то есть

$$\lim_{n\to\infty}\int g_n d\mu \le \lim_{n\to\infty}\int f_n d\mu.$$

В силу симметричности теорема доказана.

Замечание 1. В силу теоремы два определения — абстрактное и через функции — эквивалентны.

Замечание 2. В силу теоремы в формулировке Леммы 2 предположение о ступенчатости функции f_n можно убрать.

Итак, для каждой меры μ мы определили вообще говоря свой интеграл $\int f d\mu$, заданный на μ-интегрируемых (μ-суммируемых) функциях. Иногда эти функции надо чуть поправить.

Пространство $L^1(\Omega,\mathcal{F},\mu)$.

Заметим, что, если испортить функции (и/или) их сходимость на множестве N меры ноль $\mu(N)=0$, то значение интеграла не изменится. Назовем функции $f,g:\Omega\to[-\infty,+\infty]$ μ-эквивалентными, если $\mu\{\omega|f(\omega)=g(\omega)\}=1$ (почти всюду совпадают, совпадают почти наверняка, совпадают с вероятностью 1, отличаются лишь на множестве меры 0). Интегралы эквивалентных функций совпадают. Классы эквивалентности для всех суммирумых функций образуют множество, его обозначают $L^1(\Omega,\mathcal{F},\mu)$.

Определение. f_n сходится к f почти всюду (относительно меры μ), если для некоторого множества $N\in\mathcal{F},\mu(N)=0$ выполнено $f_n(\omega)\to f(\omega)$ для всех $\omega\in\Omega\setminus N$.

В пространстве $L^1(\Omega,\mathcal{F},\mu)$ задан интеграл $\int f d\mu$, он выдерживает $+$, $-$, умножение на скаляр, сохраняет порядок, в нем возможен переход к пределу. Слово “возможен” уточняет ряд теорем: Леви, Фату, Лебега.

Пределы интегралов. Теорема Лебега

Теорема Лебега о мажорируемой сходимости. Если последовательность измеримых функций f_n почти всюду сходится к некоторой функции f, и для некоторой суммируемой функции ϕ выполнено $|f_n|\le\phi$ при всех натуральных n, то их интегралы ограничены, f суммируема и

$$\int f d\mu = \lim_{n\to\infty}\int f_n d\mu.$$

Набросок д-ва теоремы Лебега для случая вероятностной меры.

Теорема Егорова. (без д-ва) Пусть последовательность измеримых функций f_k почти всюду сходится к почти всюду конечной измеримой функции f. Тогда для любого $\varepsilon>0$ найдется такое событие $B_\varepsilon\in\mathcal{F},\mathbb{P}(B_\varepsilon)>1-\varepsilon$ что f_k сходится к f равномерно на B_ε.

Отметим, что f измерима, поскольку для всех $a\in\mathbb{R}$ множество $\{\omega|f\le a\}=\cap_{n\in\mathbb{N}}\cup_{k>n}\{\omega|f_k\le a\}$ является событием, более того, в силу $|f_n|\le\phi$ выполнено $|f_-|,|f_+|\le\phi$, отсюда f почти всюду конечна.

Фиксируем $\varepsilon>0$. Найдется такое M, что для $A=\{\omega\,|\,|\phi(\omega)|\ge M\}$ имеем $\int|\phi|1_A d\mu<\varepsilon$. Теперь найдется такое $B\in\mathcal{F}$, $\mu(B)<\varepsilon/M$, что f_n сходится к f равномерно на $\Omega\setminus(A\cup B)$. Тогда

$$\Big|\int f_n 1_{\Omega\setminus(A\cup B)}d\mu-\int f1_{\Omega\setminus(A\cup B)}d\mu\Big|<\varepsilon$$

при достаточно больших n. Отсюда, наконец, получим

$$\begin{aligned}\Big|\int f_n d\mu-\int f d\mu\Big| &< \varepsilon+\Big|\int f_n 1_B d\mu-\int f1_B d\mu\Big|+\int|f_n|1_A d\mu+\int|f|1_A d\mu\\ &\le \varepsilon+2M\varepsilon/M+2\int|\phi|1_A d\mu\le 5\varepsilon.\end{aligned}$$

Доказательство завершено.

Как доказывать теорему Лебега для меры Лебега, например?

Пределы интегралов. Теоремы Леви и Фату

Теорема Леви. Если последовательность измеримых функций f_k не убывает, а их интегралы $\int f_k d\mu$ ограничены, то поточечный предел $f = \lim\limits_{k\to\infty} f_k$ этих функций μ-почти всюду существует и является μ-суммируемой функцией, при этом

$$\int f d\mu = \lim_{k\to\infty} \int f_k d\mu.$$

Д-во теоремы Леви.

Обозначим через $S = \lim\limits_{n\to\infty} \int f_n d\mu.$

Воспользовавшись при необходимости заменой $f_k \Rrightarrow f_k - f_1$ можно считать, что все функции неотрицательны. Предел f (возможно равный $+\infty$) существует для всех ω в силу монотонности. Пусть A — в точности множество тех $\omega \in \Omega$, для которых $f(\omega)$ не является конечным. Повторив начало д-ва теоремы Лебега показываем, что все множества $\{\omega | f(\omega) \leq a\}$ измеримы. Тогда $A \in \mathcal{F}$ измеримо, $f 1_{\Omega\setminus A}$ измеримо. Для функции $c 1_A$ для всех $c \geq 0$, по лемме 2, имеем $c\mu(A) \leq S$. Значит $\mu(A) = 0$. Итак, исправив f_n на нулевом множестве можно считать, что f измерима и конечна всюду.

Отметим, что $\int f d\mu$ можно представить как возможно бесконечный предел $\int s_k d\mu$ для некоторой последовательности неотрицательных ступенчатых функций $s_k \leq f$. Из Леммы 2 имеем $\int f d\mu \leq S = \lim\limits_{n\to\infty} \int f_n d\mu.$ В частности предел конечен, а f суммируема. С другой стороны, в силу $f \geq f_n$, имеем и

$$\int f d\mu \geq \lim_{n\to\infty} \int f_n d\mu.$$

Лемма Фату. Для последовательности неотрицательных измеримых функций f_k их нижний предел также измерим, при этом

$$\int \liminf_{k\to\infty} f_k \, d\mu \leq \liminf_{k\to\infty} \int f_k \, d\mu,$$

если правая часть конечна.

Д-во леммы Фату.

Пусть g_k — такая подпоследовательность f_k, что $\liminf\limits_{k\to\infty} \int f_k d\mu = \lim\limits_{k\to\infty} \int g_k d\mu.$ При этом автоматически $\inf_{k\geq n} f_k \leq \inf_{k\geq n} g_k$, поскольку справа нижняя грань лишь по части того, отчего нижняя грань слева.

Для всех натуральных n, $\inf\limits_{k\geq n} f_k = \lim\limits_{k\to\infty} \inf\{f_n, \dots, f_{n+k}\}$ существует как монотонно невозрастающий предел ограниченных снизу функций, следовательно он измерим и интегрируем. Аналогично для $\inf_{k\geq n} g_k$. Теперь в

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} \inf_{k\geq n} f_k \leq \lim_{n\to\infty} \inf_{k\geq n} g_k \qquad (*),$$

пределы существуют как монотонно неубывающие пределы, следовательно они измеримы.

Если $\liminf\limits_{n\to\infty} \int f_n d\mu$, то, начиная с некоторого, g_k суммируемы. Тогда правая часть у $(*)$ суммируема по теореме Леви. Применяя уже теорему Лебега для правой части, получаем требуемое.

Интеграл как новая мера

Пусть дано некоторое вероятностное пространство $(\Omega, \mathcal{F}, \mu)$.

Пусть имеется некоторая μ-суммируемая функция $f : \Omega \to \mathbb{R}$. Определим $S : \mathcal{F} \to \mathbb{R}$ $S(A) = \int 1_A f \, d\mu$. Теперь S счетно-аддитивна, $\int hf \, d\mu = \int h \, dS$, и $S \ll \mu$ (то есть S абсолютно непрерывна относительно μ). Если f была неотрицательна, мы получили меру, в общем случае — заряд.

Напомним, что заряд S абсолютно непрерывен относительно вероятности μ, если для всякого $A \in \mathcal{F}$ из $\mathbb{P}(A) = 0$ следует $S(A) = 0$ (Эквивалентное определение: если для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что из $\mu(A) < \delta$ следует $|S(A)| < \varepsilon$.) Напомним, что счетно-аддитивные функции $S : \mathcal{F} \to (-\infty, +\infty)$ называют зарядами, знакопеременными мерами.

От зарядов, как и от мер, тоже можно считать интегралы. Свойства те же. Нам достаточно будет определения данного выше. Полноты ради можете следующую часть прочитать, а можете пропустить.

Интегрирование по заряду

Теорема Жордана-Хана (без доказательства). Для всякого заряда $S : \mathcal{F} \to \mathbb{R}$ имеется разложение Хана $\Omega = \Omega_+ \cup \Omega_-$, $\Omega_+ \cap \Omega_- = \varnothing$, то есть правила $S_+(A) = S(\Omega_+ \cap A) \geq 0, S_-(A) = -S(\Omega_- \cap A) \geq 0$ задают неотрицательные счетно-аддитивные функции S_+, S_-, причем $S_-(\Omega)$ конечно, и $S = S_+ - S_-$ (разложение Жордана).

Неотрицательную счетно-аддитивную функцию $|S| = S_+ + S_-$ называют полной вариацией заряда S (иногда так называют число $|S|(\Omega)$).

Законно ввести $\int h\,dS = \int h\,dS_+ - \int h\,dS_-$, если h интегрируема относительно меры $|S|$.

Производная меры

Теорема Радона-Никодима (без д-ва). Если заряд S абсолютно непрерывен относительно вероятности μ, то найдется такая измеримая функция $f : \Omega \to \mathbb{R}$, что функция f_- μ-суммируема и

$$S(A) = \int_A f\,d\mu = \int 1_A f\,d\mu.$$

При этом любые такие функции μ-эквивалентны (почти всюду совпадают, отличаются лишь на множестве меры 0). Кроме того, f суммируема, если $|S|(\Omega)$ конечно.

Такая функция f называется *плотностью (производной)* заряда S относительно вероятности μ (есть даже не самое распространенное обозначение $f = \frac{dS}{d\mu}$).

Следствие. Для вероятности η (на измеримом пространстве $(\Omega, \mathcal{F})$) следующие условия эквивалентны:

1) вероятность η абсолютно непрерывна относительно вероятности μ;
2) для всякого $\varepsilon > 0$ найдется такое $\delta > 0$, что из $\mu(A) < \delta$ следует $\eta(A) < \varepsilon$;
3) для некоторой μ-суммируемой f при всех $A \in \mathcal{F}$ выполнено $\eta(A) = \int 1_A f\,d\mu$.
4) для некоторой μ-суммируемой f при всех η-интегрируемых h выполнено $\int hf\,d\mu = \int h\,d\eta$.

При этом так полученная функция f почти всюду неотрицательна и $\int f\,d\mu = 1$.

Зачем сказано "почти всюду"?

Теорема Радона-Никодима достаточно тонкая вещь, редко где имеет место (настолько хорошие банаховы пространства даже называют “smooth Banach spaces”). Например, нельзя заменить $\mathbb{R}$ на $L^1([0,1], \mathcal{B}, \lambda)$.

Интеграл от суперпозиции

Пусть имеется еще одно измеримое пространство $(\tilde{\Omega}, \mathcal{H})$, но мера на нем не задана. Пусть есть измеримое отображение $\xi : \Omega \to \tilde{\Omega}$.

Ранее мы вводили индуцированную меру (образ меры, push-forward) $\tilde{\mu} = \mu \circ \xi^{-1} = \xi \# \mu$:

$$\tilde{\mu}(H) = \mu(\xi^{-1}(H)) \qquad \forall H \in \mathcal{H}.$$

Таким образом мы перетащили меру с одного измеримого пространства на другое (“по стрелке” ξ).

Заметим, что, если некоторое отображение $h : \tilde{\Omega} \to \mathbb{R}$ $\mathcal{H}$-измеримо, то $f = h \circ \xi : \Omega \to \mathbb{R}$ уже $\mathcal{F}$-измеримо. Тут мы перетащили отображение (но “против стрелки” ξ).

Докажем, что интегрирование инвариантно при таком двойном переходе.

Теорема. Если $\xi : \Omega \to \tilde{\Omega}$ измеримо, $h : \tilde{\Omega} \to \mathbb{R}$ $\xi\#\mu$-интегрируемо, то $f = h \circ \xi : \Omega \to \mathbb{R}$ уже μ-интегрируемо и

$$\int_\Omega h \circ \xi\,d\mu = \int_{\xi(\Omega)} h\,d(\mu \circ \xi^{-1}) = \int_{\tilde{\Omega}} h\,d(\xi\#\mu),$$

или, что эквивалентно,

$$\int_\Omega f\,d\mu = \int_\Omega h \circ \xi\,d\mu = \int_\Omega h(\xi(\omega))d\mu(\omega) = \int_{\tilde{\Omega}} h(\tilde{\omega})d\tilde{\mu}(\tilde{\omega}) = \int_{\tilde{\Omega}} h\,d\tilde{\mu}.$$

При этом для $H \in \mathcal{H}$ имеем

$$\int_{\xi^{-1}(H)} f\, d\mu = \int_\Omega 1_H(\xi(\omega)) h(\xi(\omega)) d\mu(\omega) = \int_H h\, d\tilde{\mu} = \int_H h\, d(\xi\#\mu).$$

Д-во для ступенчатых f следует из определения индуцированной меры, дальше через монотонный предел.

Частный случай этой формулы ранее был получен как формула замены переменных для интеграла Лебега. Пока нам хватит вроде простейшего случая этой формулы, когда или g, или ξ, тождественны.

Математическое ожидание

Определение. Для всякой μ-суммируемой функции $\xi : \Omega \to \mathbb{R}$ назовем ее математическим ожиданием число $\mathbb{E}\xi = \int_\Omega \xi(\omega)\, d\omega$.

Для дискретной случайной величины получаем в точности $\mathbb{E}\xi = \sum_{i\in\mathbb{N}} p_i x_i$.

В общем случае напомним, что случайная величина ξ задает на $(\mathbb{R}, \mathcal{B})$ вероятность μ правилом $\mu((-\infty, a]) = F_\xi(a)$. Таким образом, F_ξ задает меру на борелевских подмножествах $\mathbb{R}$. Обозначим ее тем же значком, хотя это не вполне корректно. Теперь пусть $(\mathbb{R}, \mathcal{B}) = (\tilde{\Omega}, \mathcal{H})$. Тогда $F_\xi = \xi\#\mu$.

Возьмем в качестве h — тождественное отображение. Теперь $f = h \circ \xi = \xi$. Отсюда

$$\int_\Omega \xi\, d\mu = \int_\Omega h \circ \xi\, d\mu = \int_\mathbb{R} h(x)\, d(\xi\#\mu)(x) = \int_\mathbb{R} x\, dF_\xi(x).$$

В самом приятном случае, для абсолютно непрерывных случайных величин, в силу равенства $dF_\xi(x) = f_\xi(x)\, dx$ имеем

$$\int_\Omega \xi\, d\mu = \int_\mathbb{R} x f_\xi(x)\, dx.$$

Как и прежде, работают выражения через матожидание для k-х моментов и дисперсии. Основные свойства математического ожидания были перечислены уже для дискретных д.с.в., на общий случай д.с.в. почти все они переносятся переходом к монотонной последовательности ступенчатых. Но доказывать мы их все равно будем в еще более общей ситуации.

Все тот же интеграл от суперпозиции да не совсем...

Пусть $\Omega = \tilde{\Omega}$. Пусть ξ — тождественное отображение. Тогда в силу измеримости ξ $\mathcal{H}$ — подалгебра алгебры $\mathcal{F}$. Ранее мы перетащили отображение (но “против стрелки” тождественного отображения ξ). Сейчас попробуем “по стрелке”. Пусть задано μ-измеримое отображение $f : \Omega \to \mathbb{R}$.

Определение. Назовем действительную случайную величину $h : \Omega \to \mathbb{R}$ условным математическим ожиданием д.с.в f относительно $\mathcal{H}$, если h — $\mathcal{H}$-измерима и

$$\int_H f\, d\mu = \int_H h\, d\mu \qquad \forall H \in \mathcal{H}.$$

Введем для него следующее обозначение: $\mathbb{E}(f|\mathcal{H}) = h$.

Пример. $\mathcal{H} = \mathcal{F}$. Тогда $\xi\#\mu = \mu$, $\mathbb{E}(f|\mathcal{H}) = h$.

Тривиальный факт: $\mathbb{E}(h|\mathcal{H}) = h$, если h $\mathcal{H}$-интегрируемо. А если нет, если алгебра $\mathcal{H}$ слишком бедная?

Пример. Пусть $\mathcal{H} = \{\varnothing, \Omega\}$. Тогда $\xi\#\mu = \mu|_\mathcal{H}$ и $\mathbb{E}(f|\mathcal{H}) = \int_\Omega f\, d\omega/\mu(\Omega) = \mathbb{E}f$ — константная действительная случайная величина. Константа в общем.

Условное матожидание относительно тривиальной σ-подалгебры — это действительная случайная величина, равная константе, а именно “обычному” математическому ожиданию.

Пример. Пусть $H_1, \ldots, H_n, \ldots$ — полная система гипотез, а $\mathcal{H}$ — соответствующая ей σ-подалгебра. Рассмотрим $f = 1_A$ для некоторого $A \in \mathcal{F}$. Тогда $\mathbb{E}(1_A|\mathcal{H})(\omega) = \mu(A \cap H_i)/\mu(H_i)$ для всех $\omega \in H_i$.

Условное матожидание индикаторных функций называется условной вероятностью. Ранее вводилась условная вероятность как набор чисел, теперь как действительная случайная величина. Принципиальной разницы нет, но обозначение все-таки поставим другое: $\mathbb{P}(A|\mathcal{H}) = \mathbb{E}(1_A|\mathcal{H})$.

Существование условного математического ожидания

Теорема. Пусть существует матожидание $\mathbb{E}|\xi|$. Тогда существует и условное математическое ожидание случайной величины ξ относительно σ-подалгебры $\mathcal{H}$, некоторая $\mathcal{H}$-измеримая μ-интегрируемая случайная величина $\mathbb{E}(\xi|\mathcal{H})$. Более того, она единственна с точностью до почти всюду.

Доказательство. Отметим, что отображение $\mathcal{H} \ni H \mapsto \int_H \xi(\omega)d\mu(\omega)$ является зарядом на $\mathcal{H}$, и этот заряд равен нулю на тех H, которых $\mu(H) = 0$. Вывод — это абсолютно непрерывный относительно μ заряд. По теореме Радона-Никодима, найдется $\mathcal{H}$-интегрируемая случайная величина $\mathbb{E}(\xi|\mathcal{H})$ с нужным равенством.

Регрессии. Условное математическое ожидание относительно случайной величины

Определение Условным математическим ожиданием $\mathbb{E}(\xi|\eta)$ д.с.в ξ относительно д.с.в. η (при $\mathbb{E}|f| < +\infty$) называют случайную величину вида $g \circ \eta$ (для некоторой борелевской $g : \mathbb{R} \to \mathbb{R}$), для которой $\mathbb{E}((h \circ \eta)\xi) = \mathbb{E}(h \circ \eta)(g \circ \eta))$ при любых борелевских $h : \mathbb{R} \to \mathbb{R}$. При этом:

$$\mathbb{E}((g \circ \eta)(\xi - \mathbb{E}(\xi|\eta))) = 0.$$

Скалярную функцию g со свойством выше называют регрессией ξ на η.

По теореме Дуба, действительная случайная величина ξ измерима относительно $\sigma(\eta)$ тогда и только тогда, когда найдется такая случайная величина g, что $\xi = g \circ \eta$. Отсюда

$$\mathbb{E}(\xi|\eta) = \mathbb{E}(\xi|\sigma(\eta)),$$

то есть условное математическое ожидание ξ относительно другой случайной величины η в точности условное математическое ожидание ξ относительно минимальной σ-алгебры, в которой измеримо η.

Проще всего условное математическое ожидание относительно д.с.в. вводить через условную плотность. Для этого нам потребуются...

Векторные случайные величины

Пусть $\vec{\xi} = (\xi_1, \xi_1, \ldots, \xi_m)$ — измеримое отображение из Ω в $\mathbb{R}^m$, то есть векторная с.в.

Понятно, что при этом возникает вероятностная мера на $\mathbb{R}^m$, а значит формально можно задать, например, матожидание

$$\mathbb{E}\vec{\xi} = \int_\Omega \vec{\xi}(\omega)\, d\mathbb{P}(\omega),$$

которое может существовать, а может и не существовать. Впрочем, хоть в данном случае матожидание и вектор (почему?), как и в скалярном случае для нахождения матожидания достаточно знать функцию распределения векторной случайной величины, чтобы не путать мы будем ее назвать совместной функцией распределения, но слово “совместной” обычно регулярно опускают.

Определение Совместной функцией распределения в.с.в $\vec{\xi}$ назовем отображение $F_{\vec{\xi}} : \mathbb{R}^m \to [0, 1]$, заданное правилом:

$$F_{\vec{\xi}}(x_1, \ldots, x_m) = \mathbb{P}\{\omega \,|\, \xi_1(\omega) \le x_1, \ldots, \xi_m(\omega) \le x_m\}.$$

Свойства: $F_{\vec{\xi}}$ не убывает по каждой переменной, непрерывна справа по каждой переменной, имеет пределы слева по каждой переменной, в том числе при стремлении x_i к бесконечности.

$$\lim_{x_1,\ldots,x_m \to +\infty} F_{\vec{\xi}}(x_1, \ldots, x_m) = 1, \quad \lim_{x_i \to -\infty} F_{\vec{\xi}}(x_1, \ldots, x_m) = 0.$$

Доказательства такие же, как и для обычной функции распределения. Но полезно разобраться почему в этих пределах сходимость x_i прописана по разному.

Пример. Найти вероятность $\mathbb{P}\{\omega \,|\, \xi_1(\omega) \in (a', a''], \xi_2(\omega) \in (b', b'']\}$.
Ответ $F_{(\xi_1,\xi_2)}(a'', b'') - F_{(\xi_1,\xi_2)}(a'', b') - F_{(\xi_1,\xi_2)}(a', b'') + F_{(\xi_1,\xi_2)}(a', b')$.

Зная совместную функцию распределения можно интеграл считать уже как в матане:

$$\mathbb{E}\vec{\xi} = \int_{\mathbb{R}^m} \vec{x}\, dF_{\vec{\xi}}(\vec{x}).$$

Маргинальные распределения

Для каждой ξ_i можно рассмотреть её маргинальное распределение и соответствующую функцию распределения:

$$F_{\xi_i}(x_i) = \lim_{x_j \to +\infty, i \neq j} F_{\vec{\xi}}(x_1, \ldots, x_m).$$

Пример. Пусть $\vec{\xi}$ равномерно распределена на единичной окружности. Найти плотность маргинального распределения F_{ξ_1}. При $x \in [-1, 1]$

$$1 - F_{\xi_1}(x) = \frac{1}{\pi} \arccos x, \; F_{\xi_1}(x) = 1 - \frac{\arccos x}{\pi}, \; f_{\xi_1}(x) = F'_{\xi_1}(x) = \frac{1}{\pi\sqrt{1-x^2}}.$$

Пример. Пусть $\vec{\xi} = (\xi_1, \xi_2)$ равномерно распределена в единичном круге. Найти плотность маргинального распределения F_{ξ_1}. Теперь

$$F_{\xi_1}(x_1) = \frac{1}{\pi} \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} 1_{\{(x,y) \,|\, x^2+y^2 \leq 1\}}(x,y) dy\, dx = \frac{1}{\pi} \int_{-\infty}^{x_1} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy\, dx = \int_{-\infty}^{x_1} \frac{2}{\pi} \sqrt{1-x^2}\, dx.$$

Итак, $f_{\xi_1}(x_1) = \frac{2}{\pi}\sqrt{1-x^2}$. Отметим, что при этом мы нашли и плотность исходного распределения: $\frac{1}{\pi} 1_{\{(x,y) \,|\, x^2+y^2 \leq 1\}}(x,y)$.

Можно убедиться, что и в том, и в другом примере $\mathbb{E}\xi = (0,0)$, $\mathbb{E}(\xi_1\xi_2) = 0$, тогда $cov(\xi_1, \xi_2) = \mathbb{E}(\xi_1 - \mathbb{E}\xi_1)(\xi_2 - \mathbb{E}\xi_2) = 0$, то есть эти случайные величины некоррелированны. Независимы ли ξ_1, ξ_2 в этих примерах?

Теорема Фубини

Пусть задано несколько измеримых пространств $(\Omega_1, \mathcal{F}_1), \ldots, (\Omega_n, \mathcal{F}_n)$. На каждом из них задана мера μ_i, или вероятностная мера, или мера Лебега (достаточно потребовать σ-конечности меры).

Тогда задано измеримое пространство $(\Omega, \mathcal{F})$, где $\Omega = \Omega_1 \times \ldots \times \Omega_n$.

$$\mathcal{F} = \mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n = \sigma(\{A_1 \times \ldots \times A_n \,|\, A_i \in \mathcal{F}_i \, \forall i\}).$$

По теореме Каратеодори однозначно востанавливается мера $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ правилом $\mu(A_1 \times \ldots \times A_n) = \mu_1(A_1) \cdot \ldots \cdot \mu_n(A_n)$.

Порядок интегрирования можно менять в том случае, если в каждом выражении каждый внутренний интеграл также суммируемая функция. Простоты ради сформулируем для случая $n = 2$.

Теорема Фубини (без д-ва). Пусть ξ_1, ξ_2 — μ_1- и μ_2-измеримые д.с.в, $f : \mathbb{R}^2 \to \mathbb{R}$. Если $f(\xi_1, \xi_2)$ — μ-суммируемая функция, то отображения $x \mapsto \int_{\Omega_2} f(x,y)\, d\mu_2(y), y \mapsto \int_{\Omega_1} f(x,y)\, d\mu_1(x)$ μ_1- и μ_2-суммируемы, и

$$\int_{\Omega_1} \int_{\Omega_2} f(x,y)\, d\mu_2(y)\, d\mu_1(x) = \int_{\Omega_2} \int_{\Omega_1} f(x,y)\, d\mu_1(x)\, d\mu_2(y) = \int_{\Omega_1 \times \Omega_2} f(x,y) d\mu(x,y).$$

В частности, теперь можно всегда при интегрировании многократных интегралов даже по мере Лебега (из матана в общем) менять порядок интегрирования.

Свойства абсолютно непрерывной совместной функции распределения

Определение Назовем совместную функцию распределения абсолютно непрерывной, если ее можно представить в виде

$$F_{\vec{\xi}}(x_1, \ldots, x_m) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \ldots \int_{-\infty}^{x_m} f_{\vec{\xi}}(y_1, \ldots, y_m) dy_m\, dy_{m-1} \ldots\, dy_1$$

для некоторой борелевской неотрицательной функции $f_{\vec{\xi}} : \mathbb{R}^m \to \mathbb{R}$ называемой ее плотностью.

Проблема определения - а если порядок переменных (а значит и интегрирования) будет другой? Смотрим теорему Фубини.

Почему каждый следующий внутренний интеграл законен (предыдущий измерим или интегрируем)? Почему вся функция заведома интегрируема? Потому что ограничена в силу

1^o

$$F_{\vec{\xi}}(+\infty,\ldots,+\infty)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}f_{\vec{\xi}}(y_1,\ldots,y_m)\,dy_m\,dy_{m-1}\ldots\,dy_1=1.$$

Заметим, что это не означает, что интеграл внутри существует (или конечен) всегда...

2^o Порядок интегрирования не важен. В частности можно говорить о $\int_{A_1\times A_2\times\ldots\times A_m}f_{\vec{\xi}}\,d\vec{x}$

3^o Рассматривая равенство сначала для прямоугольников, потом для вмещающей их алгебры, и σ-алгебры, получаем:

$$\mathbb{P}\{\omega\,|\,(\xi_1,\ldots,\xi_m)\in A\}=\int_A f_{\vec{\xi}}\,d\vec{x}.$$

4^o Благодаря теореме Радона-Никодима плотность единственна с точностью до меры ноль; в частности, среди непрерывных если такая есть, то ровно одна.

5^o Если $f_{\vec{\xi}}$ достаточно гладкая ($m+1$ раз непрерывно дифференцируема), то $f_{\vec{\xi}}=\frac{\partial^m F_{\vec{\xi}}}{\partial x_1\ldots\partial x_m}$. Для доказательства достаточно проинтегрировать обе части по каждой переменной и воспользоваться формулой Ньютона-Лейбница.

6^o Маргинальные распределения абсолютно непрерывны, например для ξ_1 имеем

$$f_{\xi_1}(x_1)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}f_{\vec{\xi}}(x_1,y_2,\ldots,y_m)\,dy_2\ldots dy_m.$$

Верно ли обратное, если маргинальные распределения абсолютно непрерывны, будет ли таковым их совместное распределение?

Распределение суммы случайных величин

Плотность суммы двух д.с.величин ξ,η с плотностью совместного распределения $f_{(\xi,\eta)}$:

$$\begin{aligned}F_{\xi+\eta}(z) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z-x}f_{(\xi,\eta)}(y,x)\,dy\,dx\\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{z}f_{(\xi,\eta)}(y-x,x)\,dy\,dx\\ &= \int_{-\infty}^{z}\int_{-\infty}^{+\infty}f_{(\xi,\eta)}(y-x,x)dx\,dy.\end{aligned}$$

Формула свертки Плотность суммы двух независимых абсолютно непрерывных д.с.величин ξ,η :

$$f_{\xi+\eta}(y)=\int_{-\infty}^{+\infty}f_\xi(x)f_\eta(y-x)\,dx.$$

Пример. Найти плотность суммы двух независимых д.с.в. ξ и η, равномерно распределенных на $[0,1]$. Получим треугольное распределение: $f_{\xi+\eta}(y)=\int_0^1 f_\eta(y-x)\,dx=\int_0^1 1_{[0,1]}(y-x)\,dx=(1-|y-1|)1_{[0,2]}(y)$.

Условная плотность

В случае абсолютно непрерывного совместного распределения даже считать математическое ожидание проще:

$$\mathbb{E}\vec{\xi}=\int_{\mathbb{R}^m}\vec{x}f_{\vec{\xi}}(\vec{x})\,d\vec{x}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}(y_1,\ldots,y_m)f_{\vec{\xi}}(y_1,\ldots,y_m)dy_m\,dy_{m-1}\ldots\,dy_1.$$

1. Пусть случайная величина η — суперпозиция некоторой (для простоты скалярной) функции g и векторной случайной величины $\vec{\xi}$, то есть $\eta(\omega)=g(\vec{\xi}(\omega))$. Тогда

$$\mathbb{E}\eta=\int_{\mathbb{R}^m}g(\vec{x})f_{\vec{\xi}}(\vec{x})\,d\vec{x}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}g(y_1,\ldots,y_m)f_{\vec{\xi}}(y_1,\ldots,y_m)dy_m\,dy_{m-1}\ldots\,dy_1.$$

2. Пусть (ξ,η) имеют совместное абсолютно непрерывное распределение с известной плотностью $f_{(\xi,\eta)}$. Эта случайная величина отображает Ω в $\mathbb{R}^2$. Напомним, что

$$\mathbb{E}\xi = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{(\xi,\eta)}(x,y)\,dy\,dx,$$

а $\mathbb{E}(\xi|\eta)$ — это такая действительная случайная величина, что является функцией от η (то есть $g\circ\eta$ для некоторой борелевской $g:\mathbb{R}\to\mathbb{R}$), и при любых борелевских $h:\mathbb{R}\to\mathbb{R}$ выполнено $\mathbb{E}((h\circ\eta)\xi) = \mathbb{E}((h\circ\eta)(g\circ\eta))$.

В нашем случае для всех борелевских $h:\mathbb{R}\to\mathbb{R}$ выражение

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(y) x f_{(\xi,\eta)}(x,y)\,dx\,dy = \int_{-\infty}^{+\infty} h(y)\int_{-\infty}^{+\infty} x f_{(\xi,\eta)}(x,y)\,dx\,dy$$

должно равняться

$$\int_{-\infty}^{+\infty} h(y)g(y)\int_{-\infty}^{+\infty} f_{(\xi,\eta)}(x,y)\,dy\,dx.$$

Тогда

$$\mathbb{E}(\xi|\eta)(\omega) = \frac{\int_{-\infty}^{+\infty} x f_{(\xi,\eta)}(x,\eta(\omega))dx}{\int_{-\infty}^{+\infty} f_{(\xi,\eta)}(x,\eta(\omega))dx} \qquad \forall\omega\in\Omega,$$

или попросту

$$g(y) = \mathbb{E}(\xi|\eta=y) = \frac{\int_{-\infty}^{+\infty} x f_{(\xi,\eta)}(x,y)dx}{\int_{-\infty}^{+\infty} f_{(\xi,\eta)}(x,y)dx} \qquad \forall y\in\mathbb{R}.$$

Определение Отображение $f_{\xi|\eta}(\bar{x}|y) = \frac{f_{(\xi,\eta)}(\bar{x},y)}{\int_{-\infty}^{+\infty} f_{(\xi,\eta)}(x,y)dx}$ также называют условной плотностью, плотность ξ в точке $\bar{x}$ при известном значении y.

Отметим, что вообще говоря все условные математические ожидания (плотности и вероятности) заданы с точностью до меры ноль. Потому определять их лишь для отдельных исходов ω, значений η и прочая, вещь бесмысленная и чреватая ошибками.

Еще проще убедиться, что если ξ,η — д.с.в. с дискретным совместным распределением, то

$$\mathbb{E}(\xi|\eta)(\bar{\omega}) = \frac{\sum_i x_i\mathbb{P}(\xi=x_i|\eta(\omega)=y)}{\sum_i \mathbb{P}(\xi=x_i|\eta(\omega)=y)}, \text{где } y=\eta(\bar{\omega}).$$

Свойства условного математического ожидания

Сначала восстановим написанные ранее для $\mathbb{E}\xi$ свойства,но уже для $\mathbb{E}(\xi|\mathcal{H})$. Будем для простоты предполагать в 1-10 , что все ξ, ξ_i суммируемы. Отметим, что у всех указанных ниже свойств есть корректные трактовки для случая векторных (или, например, комплексных) случайных величин.

1. $\mathbb{E}(\xi|\mathcal{H})\ge 0$, если $\xi(\omega)\ge 0$ для всех $\omega\in\Omega$.

Верно ли, что $\mathbb{E}\xi > 0$ если $\xi > 0$ для всех $\omega\in\Omega$?

2. $\mathbb{E}(\xi_1|\mathcal{H})\ge\mathbb{E}(\xi_2|\mathcal{H})$, если $\xi_1(\omega)\ge\xi_2(\omega)$ для всех $\omega\in\Omega$.

3. $\mathbb{E}(\xi|\mathcal{H}) = c$ для случайной величины $\xi = c$ при любых $\omega\in\Omega$;

4. $\mathbb{E}(c\xi|\mathcal{H}) = c\mathbb{E}(\xi|\mathcal{H})$ для любого $c\in\mathbb{R}$;

5. $\mathbb{E}(\xi_1|\mathcal{H})+\mathbb{E}(\xi_2|\mathcal{H}) = \mathbb{E}(\xi_1+\xi_2|\mathcal{H})$;

6. $a\mathbb{E}(\xi_1|\mathcal{H})+b\mathbb{E}(\xi_2|\mathcal{H}) = \mathbb{E}(a\xi_1+b\xi_2|\mathcal{H})$ при любых $a,b\in\mathbb{R}$;

7. $\mathbb{E}(\sum_{i=1}^{\infty}\xi_i|\mathcal{H}) = \sum_{i=1}^{\infty}\mathbb{E}(\xi_i|\mathcal{H})$, если сходится ряд $\sum_{i=1}^{\infty}\mathbb{E}(|\xi_i|\,|\mathcal{H})$;

8. $|\mathbb{E}(\xi|\mathcal{H})|\le\mathbb{E}(|\xi|\,|\mathcal{H})$;

9. $g(\mathbb{E}(\xi|\mathcal{H}))\le\mathbb{E}(g(\xi)|\mathcal{H})$ для любой выпуклой (вниз) функции g (если существуют оба).

10. Если ξ_i независимы, то $\mathbb{E}(\xi_1\xi_2|\mathcal{H}) = \mathbb{E}(\xi_1|\mathcal{H})\mathbb{E}(\xi_2|\mathcal{H})$.

11.1. Если ξ_i почти всюду сходится к ξ, и для некоторой суммируемой ϕ выполнено $|\xi|,|\xi_i|\le\phi$, то

$$\mathbb{E}(\xi_i|\mathcal{H})\to\mathbb{E}(\xi|\mathcal{H})$$

и все эти математические ожидания существуют и конечны.

12. Если ξ ограничено, то $\mathbb{E}(\xi|\mathcal{H})$ существует.

13. Если $\mathbb{E}(\xi^2|\mathcal{H})$ существует, то и $\mathbb{E}(|\xi|\,|\mathcal{H}), \mathbb{E}(\xi|\mathcal{H})$ существуют.

14. Если $\mathbb{E}(\xi^2|\mathcal{H})$ и $\mathbb{E}(\eta^2|\mathcal{H})$ существуют, то и $\mathbb{E}(|\xi\eta|\,|\mathcal{H})$ существует.

Доказательства 1-6,8-10 сводятся к проверке на ступенчатых функциях, а затем переход к пределу в силу определения матожидания как интеграла по мере. Доказательства 7,11-12 — переформулировки теоремы Лебега. Доказательства 13-14 — комбинация неравенств о среднем и 11.1.