

МОДЕЛИРОВАНИЕ СЛУЧАЙНЫХ ЯВЛЕНИЙ И ПРОЦЕССОВ: принципы построения вероятностных моделей

Межфакультетский курс

Лектор: В. Ю. Королев (факультет ВМК)

Часть I.

Введение. 1

Идея данного курса лекций возникла в результате частого общения автора с инженерами, экономистами, медиками, юристами, физиками и специалистами в других областях, которые используют аппарат теории вероятностей и математической статистики для решения своих специфических задач.

Среди математиков за теорией вероятностей прочно закрепился статус прикладной науки. Благодаря этому, к сожалению, несмотря на большое число выдающихся математических результатов, полученных специалистами в области теории вероятностей и совершенных ими действительно фундаментальных математических открытий, ни один из них не удостоен (и, по-видимому, не будет удостоен) премии Филдса, которая вручается за выдающийся вклад в фундаментальную математику. Тем не менее, несмотря на этот статус прикладной математической дисциплины, между “академической” теорией вероятностей и практическим применением ее результатов и методов возникла огромная пропасть.

Введение. 2

В классической теории вероятностей и математической статистике много очень математически красивых и мощных результатов. Однако каждый такой результат начинается со слов “*предположим, что...*” после которых идет перечень условий, которые должны выполняться, чтобы этот результат был справедлив. Очень часто чрезвычайно сложно или просто невозможно проверить, выполнены ли эти условия в конкретной реальной ситуации. Поэтому правильное применение результатов теории вероятностей и математической статистики на практике сопряжено с глубоким осмыслением предпосылок и реальных возможностей применения тех или иных методов.

К сожалению, этот важнейший этап очень часто игнорируется. Вместо такого анализа, во многих случаях исследователи-нематематики просто принимают на веру многие рекомендации математиков, находясь в плену стереотипов о всемогуществе “точнейшей из наук”.

Этим обстоятельством часто вольно или невольно пользуются авторы некоторых книг или статей, в которых приводятся результаты статистической обработки тех или иных данных. Почти все наверняка неоднократно встречали в различных публикациях “заклинания” типа

- “вычисления на компьютере показывают, что...”,
- “применяя методы математической статистики, мы убеждаемся, что...”,
- “согласно закону больших чисел...”
или
- “основываясь на центральной предельной теореме теории вероятностей, мы заключаем, что...”,

где вместо многоточий приводятся удобные авторам выводы. Такой прием рассчитан на читателя, привыкшего слепо доверять печатному слову и/или слепо верящего в непогрешимость компьютера, результатов теории вероятностей или методов статистического анализа.

Для такого читателя, по сути являющегося пользователем моделей и методов теории вероятностей на уровне любителя, перечень источников моделей сводится к следующему:

- “от фонаря” (из справочника);
- “с потолка” (из Википедии);
- “из пальца” (когда-то слышал);
- “кто-то сказал” (где-то читал)

и далее в таком же роде.

Во многом сложность выбора правильной (адекватной) вероятностной обусловлена тем, что в отличие от большинства объектов математического моделирования, вероятность нельзя “пощупать” или измерить непосредственно

(«– Видишь суслика?

– Нет...

– И я нет. А он есть!»).

Однако в действительности компьютер является лишь орудием в руках его пользователя, равно как и подготовка данных, выбор статистического метода для их анализа и интерпретация результатов применения этого метода всегда была, есть и будет прерогативой исследователя.

В данном курсе лекций сделана попытка, не углубляясь в дебри математических выкладок, но сохраняя логическую строгость, ознакомить слушателей с основными принципами построения математических моделей в условиях стохастической неопределенности. При этом особый акцент будет сделан на особенностях применения результатов и методов теории вероятностей и математической статистики к анализу больших массивов данных, накопленных в результате наблюдений за массовыми случайными явлениями. Мы постараемся по возможности разрушить некоторые вредные стереотипы, связанные с практическим применением результатов и методов этих наук.

Почему необходимо знать принципы построения вероятностных моделей при анализе больших данных?

- 1 Значения параметров адекватных математических моделей статистических закономерностей, наблюдаемых в больших данных, можно использовать как дополнительные признаки при применении методов машинного обучения. При этом объем информации, изначально доступной исследователю, не увеличивается, а качество обучения повышается.
- 2 Выбор адекватной вероятностной модели крайне важен при анализе рисков, связанных с редкими событиями (отказами, сбоями, переполнениями, авариями, катастрофами). “Стандартные” модели (например, нормальное распределение) как правило недооценивают вероятности таких событий. Поэтому в таких ситуациях крайне важно уметь делать правильный выбор между “классическими” вероятностными моделями и моделями с “тяжелыми хвостами”.
- 3 Наконец, в большинстве задач выбор адекватной вероятностной модели важен сам по себе для принятия правильного решения.

Можно предложить следующую условную классификацию подходов к формированию вероятностных моделей.

- ① **Исторический (эмпирический).** В рамках такого подхода в качестве вероятностной модели используется ее непараметрическая статистическая оценка (например, эмпирическая функция распределения).
- ② **Асимптотический.** При таком подходе в модель вводится некоторый условный бесконечно малый или бесконечно большой параметр и в качестве модели берется асимптотическая аппроксимация рассматриваемой модели. Пример – нормальное распределение, являющееся предельным для сумм независимых случайных величин при неограниченном увеличении числа слагаемых.
- ③ **Энтропийный (информационный).** В рамках такого подхода в качестве вероятностной модели используется наиболее неблагоприятная модель, обладающая свойствами наибольшей неопределенности при заданных предположениях. По сути этот подход реализует минимаксный принцип выбора наилучшего метода в наихудших условиях.

По ходу курса мы повторим и обсудим базовые положения теории вероятностей, основанной на аксиоматике Колмогорова, и подробно остановимся на каждом из перечисленных выше принципов построения вероятностных моделей.

Данный курс, естественно, не дает готовых рецептов на все случаи жизни. Он предназначена для заинтересованных слушателей, которым он может помочь самим разобраться в том, что теория вероятностей может и чего не может в их конкретных задачах, связанных с теми или иными конкретными ситуациями.

Основные понятия прикладной теории вероятностей. Стохастические ситуации. 1

Окружающая нас действительность постоянно порождает неопределенные ситуации, исходы которых невозможно заранее предсказать с исчерпывающей точностью. Иногда это связано просто с недостатком информации. В таких случаях получение дополнительной информации может существенно уменьшить неопределенность и даже совсем ее устранить. Однако иногда неопределенность принципиально нельзя устранить совсем, например, в лотереях или биржевых играх. Но даже в тех ситуациях, в которых неопределенность принципиально не устранима полностью, ее часто можно существенно уменьшить за счет лучшего понимания, уточнения самих механизмов проявления неопределенности. В частности, для этих целей можно использовать математические методы.

Стохастические ситуации. 2

Математика предоставляет средства описания окружающей действительности, которые являются универсальными в том смысле, что они с одинаковым успехом могут быть использованы в самых разных областях – от физики, техники, биологии и медицины до страхования, финансов и юриспруденции. К разделам математики, изучающим механизмы проявления принципиально неустранимой неопределенности, можно отнести и теорию вероятностей.

Теория вероятностей изучает свойства математических моделей случайных явлений или процессов. Под *случайностью* мы будем понимать принципиально неустранимую неопределенность. С помощью понятий и утверждений теории вероятностей можно описать сами механизмы проявления неопределенности, выявить закономерности в проявлениях случайности.

Стохастические ситуации. 3

Любая математическая теория устроена следующим образом. Фундаментом каждой такой теории является набор аксиом, то есть не противоречащих друг другу утверждений или принципов, заведомо считающихся верными и принимаемых без доказательств. Из этих аксиом с помощью логических переходов конструируются понятия и утверждения соответствующей теории. Разные наборы аксиом ведут к разным математическим теориям, которые могут описывать одни и те же процессы и явления. При этом практическая полезность или эффективность той или иной математической теории определяется удобством ее применения и ее адекватностью, то есть степенью согласованности получаемых с ее помощью выводов со свойствами описываемой ею реальности, наблюдаемыми на практике.

Стохастические ситуации. 4

Имеется довольно много математических теорий, описывающих свойства математических моделей случайных явлений или процессов. В каждой из этих теорий так или иначе присутствует понятие вероятности как числового выражения меры возможности осуществления того или иного события, связанного с неопределенной ситуацией. Другими словами, имеется несколько теорий вероятностей. В нашем курсе мы будем иметь дело с теорией вероятностей, основанной на системе аксиом, которая была предложена в 20-х – 30-х годах XX столетия великим русским математиком Андреем Николаевичем Колмогоровым. Как правило, именно эта теория и называется собственно теорией вероятностей. За другими теориями вероятностей закреплены особые названия, например, теория субъективных вероятностей, интервальная теория вероятностей и т. п.

Не останавливаясь пока на формулировке аксиом колмогоровской теории вероятностей, опишем те свойства, которые должны быть присущи реальной неопределенной ситуации, чтобы ее можно было успешно математически описать на языке теории вероятностей. Другими словами, мы выделим те неопределенные ситуации, описание которых с помощью теории вероятностей ведет к адекватным выводам.

Стохастические ситуации. 6

Назовем *стохастической* такую ситуацию, которая характеризуется следующими свойствами или условиями:

- **непредсказуемость**: исход ситуации невозможно заранее предсказать с абсолютной точностью;
- **воспроизводимость**: имеется по крайней мере теоретическая возможность воспроизвести рассматриваемую ситуацию как угодно много раз в остающихся неизменными условиях;
- **устойчивость частот**: каким бы ни было интересующее нас событие, связанное с рассматриваемой ситуацией, при многократном воспроизведении этой ситуации частота события (то есть отношение количества случаев, в которых наблюдалось рассматриваемое событие, к общему числу воспроизведений ситуации) колеблется возле некоторого числа, приближаясь к нему все ближе и ближе по мере увеличения числа воспроизведений ситуации.

Стохастические ситуации. 7

Свойство непредсказуемости довольно очевидно. Если исход ситуации прогнозируем однозначно, то вообще нет никакой необходимости в привлечении аппарата теории вероятностей.

Свойство воспроизводимости ситуации является ключевым для того, чтобы быть уверенным в успехе применения аппарата теории вероятностей к ее описанию. Именно это свойство имеют в виду, когда говорят, что теория вероятностей и математическая статистика направлены на изучение *массовых* явлений. В связи с условием воспроизводимости следует весьма осторожно относиться к попыткам применения теории вероятностей к анализу *уникальных* явлений или систем. Например, известны многочисленные попытки дать количественный ответ на вопрос о том, какова вероятность существования во Вселенной других планет, населенных разумными существами. Однако пока нет достаточных оснований считать, что наличие других планет и, тем более, существование на них разумной жизни является массовым явлением. Поэтому существующие прогнозы весьма разноречивы и потому неадекватны.

Наконец, свойство устойчивости частот позволяет связать математическое определение вероятности события с интуитивным представлением о ней как о понимаемом в определенном смысле пределе частоты осуществления события при неограниченном воспроизведении соответствующей ситуации.

Классические вероятностные модели. 1

Рассмотрим следующую классическую задачу.

Задача 1. Предположим, что в некотором сосуде (урне) имеется N шаров, различающихся лишь цветом: из этих шаров M – красные ($M \leq N$) и $N - M$ – белые. Из сосуда случайно извлекаются n шаров так, что каждый вынутый шар не возвращается обратно. Пусть m – некоторое целое число, $0 \leq m \leq n$. Какова вероятность того, что среди n вынутых шаров окажется ровно m красных?

Классические вероятностные модели. 2

По смыслу задачи ясно, что порядок шаров в выборке безразличен. Поэтому общее число разных способов, которыми мы можем извлечь n шаров из совокупности, содержащей N шаров, то есть общее число возможных исходов ситуации, равно C_N^n . При этом все эти исходы равновероятны. Найдем число исходов, в результате каждого из которых наступит интересующее нас событие. Выбрать m шаров из имеющихся M красных можно C_M^m способами. При этом остальные $n - m$ шаров в выборке должны быть белыми. Выбрать $n - m$ шаров из имеющихся $N - M$ белых можно C_{N-M}^{n-m} способами. Ясно, что каждому набору из m красных шаров в выборке может соответствовать любой из C_{N-M}^{n-m} наборов белых шаров. Таким образом, общее число “благоприятных” исходов равно $C_M^m \cdot C_{N-M}^{n-m}$. Поскольку все возможные исходы равновероятны, в качестве искомой вероятности разумно взять отношение числа благоприятных исходов к общему числу исходов, а именно,

$$p_m = \frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n}. \quad (1)$$

Классические вероятностные модели. 3

Казалось бы, данная задача в том виде, как она сформулирована, имеет чисто познавательный интерес. Однако, как только мы зададим конкретные значения для параметров N , M , n и m , мы сразу можем заметить, что эта задача очень важна для практики. Пусть, например, $N = 49$, $M = 6$, $n = 6$. Тогда при $m = 4$, $m = 5$ и $m = 6$ с помощью формулы (1) мы получаем правило вычисления вероятностей угадать соответственно 4, 5 и 6 номеров в лотерее типа Спортлото “6 из 49”.

Набор чисел p_m , определяемых по формуле (1) для $m = 0, 1, \dots, n$, называется *гипергеометрическим распределением* вероятностей.

Задача 1 является примером использования так называемой *классической вероятностной модели*, которая характеризуется тем, что рассматриваемая стохастическая ситуация имеет конечное число равновероятных исходов. В такой модели под вероятностью события понимается отношение количества исходов, каждый из которых влечет осуществление этого события, к общему числу возможных исходов.

Дискретные вероятностные модели. 1

Классические вероятностные модели приводят к самому простому и хорошо согласующемуся с интуитивными представлениями определению вероятности. Однако далеко не каждая стохастическая ситуация имеет *равновероятные* исходы, не говоря уже о том, что исходов может быть бесконечно много. Поэтому нужны более общие и более гибкие подходы к определению вероятности. Рассмотрим еще одну классическую задачу, известную как *задача о справедливом дележе ставки*. Решение этой задачи было найдено в ходе знаменитой переписки великих ученых XVII столетия Блеза Паскаля (1623-1662) и Пьера Ферма (1601-1665), в которой, как считает большинство историков науки, и были заложены основы теории вероятностей.

Дискретные вероятностные модели. 2



Blais Pascal



Pierre Fermat

Задача 2. Два игрока (скажем, A и B), поставив одинаковые суммы денег, играют в игру, состоящую из отдельных последовательных партий. В каждой партии возможны лишь два исхода: выигрыш игрока A или выигрыш игрока B . Вся игра считается выигранной тем из игроков, кто первым выиграет 6 партий (не обязательно подряд). Однако при счете 5:3 в пользу игрока A игра прервана и не может быть продолжена. В какой пропорции следует разделить ставку между A и B ?

Дискретные вероятностные модели. 3

Ясно, что речь идет о наиболее справедливом способе дележа, ведь всего способов довольно много.

Например, поровну – $1:1$. Однако такой способ не устраивает первого игрока, который к моменту прекращения игры уже выиграл больше партий, чем игрок B .

Можно было бы разделить ставку в отношении $1:0$, отдав все игроку A , который выиграл больше партий. Но такой способ, очевидно, не устраивает игрока B , ибо такой дележ предусмотрен лишь для случая, когда весь матч закончен, в то время как матч не закончен, и игрок B все-таки выиграл 3 партии.

Кажется, что дележ ставки в отношении $5:3$ будет справедливым, ведь такой дележ пропорционален количеству выигранных партий. Однако и он не является идеальным, поскольку на самом деле ничем не лучше дележа в отношении $3:1$, обратно пропорциональном числу партий, которые осталось выиграть игрокам до окончательной победы.

Дискретные вероятностные модели. 4

Паскаль и Ферма пришли к обоснованному выводу, что делить ставку следует пропорционально шансам игроков на окончательную победу, то есть пропорционально вероятностям того, что матч закончится победой соответствующего игрока, если бы он был продолжен. Найдем эти вероятности. Сначала опишем все возможные исходы данной стохастической ситуации. Победу игрока A в отдельной партии условно обозначим a , победу игрока B в отдельной партии обозначим b . Если первую после возобновления игры партию выигрывает игрок A , уже выигравший до этого пять партий, то весь матч (играющийся до шестой победы одного из игроков) заканчивается победой игрока A , выигравшего нужные шесть партий. Если эту партию выигрывает игрок B , то матч продолжается. Если следующую партию выигрывает игрок A , то весь матч заканчивается победой игрока A . Если же следующую партию выигрывает B , то в матче должна быть сыграна еще одна, решающая, партия. Если эту партию выигрывает A , то он побеждает в матче. Но если же эту партию выигрывает B , то весь матч заканчивается победой игрока B . Таким образом, возможные пути развития событий в матче, то есть возможные исходы рассматриваемой стохастической ситуации имеют вид a , ba , bba и bbb .

Дискретные вероятностные модели. 5

Термин “исход” здесь используется не для *результата* партии или матча, но для описания того *процесса*, который привел к тому или иному результату матча. Будем считать, что игроки одинаково сильны в этой игре и, отложив математическую конкретизацию сказанного до следующих разделов, предположим, что результат каждой последующей партии не зависит от того, как закончились предыдущие. Тогда, очевидно,

- вероятность исхода a равна $\frac{1}{2}$,
- вероятность исхода ba равна $\frac{1}{4}$, а
- вероятности каждого из исходов bba и bbb равны $\frac{1}{8}$.

Событие “игра выиграна игроком A ” состоит из исходов a , ba и bba . В качестве вероятности этого события следует взять сумму вероятностей элементарных исходов, в результате каждого из которых наступает рассматриваемое событие. Таким образом, вероятность события “игра выиграна игроком A ” равна $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8}$.

В то же время, вероятность события “игра выиграна игроком B ” равна $\frac{1}{8}$. Следовательно, оптимальная пропорция, в которой следует поделить ставку, как это ни странно, составляет **7:1**.

Дискретные вероятностные модели. 6

Задача 2 приводит нас к следующей конструкции *дискретной вероятностной модели*. Пусть Ω – совокупность всех исходов рассматриваемой стохастической ситуации, элементарных в том смысле, что каждый из них нельзя разбить на более мелкие исходы. Под *событием* мы будем понимать любое подмножество A элементов совокупности Ω . При этом вероятность события A определим как суммарную вероятность элементарных исходов, его составляющих. Более формально, пусть $\Omega = \{\omega_1, \dots, \omega_n\}$ и вероятность исхода ω_i равна p_i , $i = 1, \dots, n$. Пусть A – некоторое событие. Тогда вероятность события A , обозначаемая $P(A)$, определяется как

$$P(A) = \sum_{i: \omega_i \in A} p_i. \quad (2)$$

В (2) суммирование идет по всем номерам i элементарных исходов, составляющих событие A .

Дискретные вероятностные модели. 7

При этом должно выполняться естественное условие нормировки

$$\sum_{i=1}^n p_i = 1.$$

Заметим, что в этой модели допускается возможность бесконечного числа элементарных исходов стохастической ситуации ($n = \infty$), а вероятности p_i совсем не обязаны быть равны между собой.

Классическая вероятностная модель является частным случаем дискретной модели, в котором $p_i = \frac{1}{n}$, $i = 1, \dots, n$.

Геометрические вероятностные модели. 1

Обе вероятностные модели, рассмотренные выше, характеризуются тем, что *все* элементарные исходы соответствующих стохастических ситуаций можно занумеровать, поставив каждому исходу в соответствие натуральное число так, что разным исходам будут соответствовать разные числа. Именно это свойство позволяет определить вероятность события по формуле (2). Однако зачастую исходы стохастической ситуации не могут быть перенумерованы разумным образом. В качестве примера рассмотрим следующую задачу.

Задача 3. Два человека, назовем их, скажем, X и Y , договорились встретиться в условленном месте между 12⁰⁰ и 13⁰⁰. Согласно обоюдной договоренности, пришедший первым ждет второго лишь в течение 20 минут (например, чтобы не привлекать излишнее внимание окружающих). Считая, что каждый из X и Y может прийти на место встречи в любой момент указанного интервала времени с одинаковой возможностью, найти вероятность того, что встреча состоится.

Геометрические вероятностные модели. 2

Чтобы решить эту задачу, сначала построим модель множества возможных элементарных исходов данной стохастической ситуации. Совокупность возможных моментов появления X на месте встречи – это отрезок длиной 1 час. Аналогично, другой точно такой же отрезок является совокупностью моментов появления Y на месте встречи. Отложим эти отрезки на соответствующих перпендикулярных координатных осях, резервируя горизонтальную ось за временем прихода X , а вертикальную – за временем прихода Y . Пусть x – момент появления X , y – момент появления Y . Тогда каждый исход стохастической ситуации характеризуется парой чисел (x, y) . Если мы отождествим каждую такую пару с точкой, имеющей координаты (x, y) на плоскости с указанными координатными осями, то станет ясно, что совокупность исходов данной стохастической ситуации можно представить в виде квадрата с вершинами $(0, 0)$, $(0, 1)$, $(1, 0)$ и $(1, 1)$.

Геометрические вероятностные модели. 3

Теперь найдем совокупность тех исходов, в результате каждого из которых наступает встреча X и Y . Если x отстоит от y более чем на $\frac{1}{3}$ (20 минут составляют $\frac{1}{3}$ часа), то встреча X с Y не может состояться и наоборот. Таким образом, совокупность исходов, “благоприятствующих” встрече X с Y , содержит те и только те точки с координатами (x, y) , которые, во-первых, попадают внутрь указанного квадрата и, во-вторых, удовлетворяют условию $|x - y| \leq \frac{1}{3}$. С учетом того, что по условию все исходы равновозможны, искомая вероятность будет равна отношению площади шестиугольника, задаваемого условиями $0 \leq x \leq 1$, $0 \leq y \leq 1$, $|x - y| \leq \frac{1}{3}$, и равной, как несложно подсчитать, $\frac{5}{9}$, к площади всего квадрата, равной, очевидно, единице. Таким образом, искомая вероятность равна $\frac{5}{9}$.

Геометрические вероятностные модели. 4

Подобные рассуждения вполне применимы и в других задачах, где оказывается возможным отождествить совокупность всех элементарных исходов стохастической ситуации с каким-либо одно-, дву- или трехмерным множеством и все исходы считаются *равновозможными*. Тогда вероятность события, отождествляемого с каким-либо подмножеством совокупности всех элементарных исходов, будет определяться как отношение длин в первом случае, площадей во втором и объемов в третьем случае. Такие вероятностные модели принято называть *геометрическими*. Эти модели являются частными случаями так называемых непрерывных вероятностных моделей, связанных с описанием поведения непрерывных случайных величин.

Структура вероятностных моделей стохастических ситуаций

В каждом из примеров, рассмотренных выше, каждая *вероятностная модель* той или иной стохастической ситуации имеет вид тройки объектов $(\Omega, \mathcal{A}, P)$, где Ω – совокупность всех *элементарных исходов* стохастической ситуации, каждое возможное событие является некоторым множеством элементарных исходов, то есть подмножеством множества Ω , так что $\mathcal{A}$ – совокупность всех возможных *событий* (которая в первых двух примерах совпадала с совокупностью всех подмножеств совокупности Ω), P – неотрицательная функция, областью определения которой является множество $\mathcal{A}$, а областью значений – отрезок $[0, 1]$. Каждому событию $A \in \mathcal{A}$ функция P ставит в соответствие число $P(A)$, понимаемое как вероятность события A . При этом $P(\Omega) = 1$.

Каждая такая тройка является математическим описанием (математической моделью) соответствующей стохастической ситуации. Условимся в дальнейшем отождествлять термин *вероятностная модель* с конкретной тройкой $(\Omega, \mathcal{A}, P)$.

Парадокс Бертрана. 1

Выбор правильной, адекватной вероятностной модели в конкретном случае очень важен и иногда очень не прост. Дело в том, что иногда стохастическая ситуация, которая, казалось бы, вполне конкретно, однозначно описывается словами, допускает несколько разных вероятностных моделей, как показывает следующий пример.

Парадокс Бертрана. 2



Joseph Louis François
Bertrand (1822-1900)

Пример 1 (парадокс Бертрана, см. J. Bertrand. *Calcul des Probabilités*. Paris, 1899). Рассмотрим следующую задачу. Пусть имеется окружность, в которую вписан равносторонний треугольник с вершинами, скажем, A , B и C . Спрашивается: какова вероятность того, что наугад (случайно) выбранная хорда данной окружности окажется длиннее стороны вписанного равностороннего треугольника?

Парадокс Бертрана. 3

Первое решение. Так как хорда выбирается случайно, то каждый из ее концов может попасть в любую точку окружности. Предположим, что один из концов хорды попал в точку A , являющуюся одной из вершин треугольника. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его вершина A совместилась с одним из концов хорды (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы другой ее конец (напомним, что один из ее концов совмещен с точкой A) попал на дугу BC , соединяющую две другие вершины треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то каждый ее конец может с равными возможностями попасть в любую точку окружности. Следовательно, искомая вероятность будет равна отношению длины дуги BC , каждая точка которой “благоприятствует” наступлению интересующего нас события, к длине всей окружности. А так как треугольник ABC – равносторонний, то это отношение равно $\frac{1}{3}$.

Парадокс Бертрана. 4

Второе решение. Так как хорда выбирается случайно, то она может быть сориентирована произвольным образом по отношению к стороне AC треугольника. Предположим, что хорда оказалась параллельна этой стороне. Это предположение никак не ограничивает общность наших рассуждений, поскольку в противном случае мы всегда можем так повернуть треугольник, чтобы его сторона AC оказалась параллельна хорде (вопрос, сформулированный в задаче, никак не ограничивает взаимное расположение хорды и треугольника). Середина всех хорд, параллельных стороне AC , обязательно попадает на диаметр окружности, перпендикулярный стороне AC . Более того, так как треугольник ABC – правильный, то точка пересечения этого диаметра со стороной AC , скажем, точка D , делит радиус пополам. Пусть D' – точка, симметричная точке D относительно центра окружности. Теперь ясно, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь отрезка DD' . Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку диаметра, перпендикулярного стороне AC . Следовательно, искомая вероятность будет равна отношению длины отрезка DD' , каждая точка которого “благоприятствует” наступлению интересующего нас события, к длине всего диаметра. Но из сказанного выше вытекает, что длина отрезка DD' составляет ровно половину длины диаметра окружности. Таким образом, искомая вероятность равна $\frac{1}{2}$.

Парадокс Бертрана. 5

Третье решение. Так как хорда выбирается случайно, то ее середина может попасть в любую точку круга, ограниченного рассматриваемой окружностью. Используя построения, аналогичные проведенным в ходе второго решения, легко убедиться, что для того чтобы хорда была длиннее стороны треугольника, необходимо и достаточно, чтобы ее середина попала внутрь окружности, вписанной в треугольник ABC , радиус которой, как несложно видеть, равен половине радиуса окружности, описанной около этого же треугольника. Вспоминая геометрическое определение вероятности, мы заключаем: так как хорда выбирается случайно, наугад, то ее середина может с равными возможностями попасть в любую точку круга, ограничиваемого описанной окружностью. Следовательно, искомая вероятность будет равна отношению площади круга, ограниченного вписанной окружностью, каждой точке которого “благоприятствует” наступлению интересующего нас события, к площади круга, ограничиваемого описанной окружностью. А так как это отношение равно отношению квадратов радиусов окружностей, то искомая вероятность равна $\frac{1}{4}$.

Парадокс Бертрана. 6

Налицо явный парадокс: одна задача имеет сразу три решения, и каждое из этих решений, будучи полученным логически непротиворечивым путем, является верным! Однако с математической точки зрения здесь нет парадокса. Дело в том, что исходная задача была сформулирована довольно нестрого, неоднозначно, так сказать, некорректно. Это проявилось в том, что понятие "случайного выбора хорды" допускает несколько математических конкретизаций.

Данная стохастическая ситуация может быть описана несколькими разными вероятностными моделями.

При первом решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с окружностью, совокупность событий $\mathcal{A}$ включает в себя все дуги, а функция P , определенная на $\mathcal{A}$, каждой дуге приписывает число, равное отношению длины этой дуги к длине всей окружности.

Парадокс Бертрана. 7

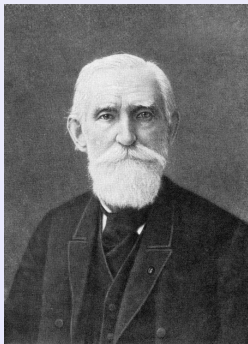
При втором решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с диаметром окружности, совокупность событий $\mathcal{A}$ включает в себя все отрезки, лежащие на этом диаметре, а функция P , определенная на $\mathcal{A}$, каждому отрезку приписывает число, равное отношению длины этого отрезка к диаметру.

При третьем же решении была использована модель $(\Omega, \mathcal{A}, P)$, в которой множество элементарных исходов Ω отождествлено с кругом, ограниченным описанной окружностью, совокупность событий $\mathcal{A}$ включает в себя все подмножества этого круга, площадь которых можно измерить, а функция P , определенная на $\mathcal{A}$, каждому подмножеству указанного круга ставит в соответствие число, равное отношению площади этого подмножества к площади круга.

В данной задаче самым сложным является адекватный выбор конкретной вероятностной модели. Такая ситуация типична для практики.

Большинство задач, выдвигаемых жизнью, являются некорректными, допускают неоднозначную математическую интерпретацию. Ниже мы еще столкнемся с примерами подобных задач.

Парадокс Бертрана. 8



P. L. Chebyshev

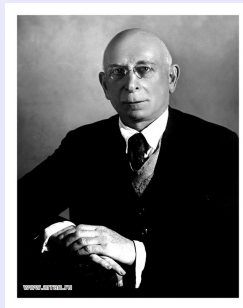
Можно сказать, что настоящий парадокс заключается в том, что выбор адекватной *математической*, вероятностной модели часто связан с *нематематическими* обстоятельствами. Еще в XIX веке великий русский математик П. Л. Чебышев говорил, что *правильно поставить математическую задачу – значит наполовину ее решить*.

Формальное определение вероятностной модели. 1

Задачи типа парадокса Бертрана стимулировали поиск удобных аксиоматизаций теории вероятностей. Наиболее известными стали те попытки, которые связаны с именами Р. Фон Мизеса и С. Н. Бернштейна, которые в качестве отправного пункта рассматривали свойство устойчивости частот.



R. Von Mises

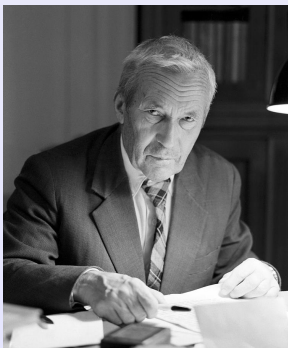


S. N. Bernstein

Однако такие системы аксиом оказались неудобными и сложно интерпретируемыми.

Формальное определение вероятностной модели. 2

Наиболее удачной и удобной стала система аксиом, предложенная А. Н. Колмогоровым, который заметил, что вероятность можно трактовать как некий аналог объема, и предложил использовать систему аксиом, принятую в хорошо к тому времени разработанной теории меры, дополнив ее требованием нормированности вероятности как меры множества (события).



A. N. Kolmogorov

Формальное определение вероятностной модели. 3

Пусть Ω – непустое множество, элементы которого будут обозначаться ω . Мы будем отождествлять элементы Ω с возможными элементарными, то есть неделимыми исходами некоторой стохастической ситуации. В связи с этим множество Ω будет называться *множеством элементарных исходов*.

Пусть $\mathcal{A}$ – множество подмножеств множества Ω элементарных исходов, обладающее свойствами

- $\Omega \in \mathcal{A}$;
- если $B \in \mathcal{A}$, то $B^c \in \mathcal{A}$;
- если $B_i \in \mathcal{A}$, $i = 1, 2, \dots$, то

$$\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}, \quad \bigcap_{i=1}^{\infty} B_i \in \mathcal{A}.$$

Множество $\mathcal{A}$ называется *σ -алгеброй событий*, а его элементы (являющиеся подмножествами множества Ω) называются *событиями*.

Множество Ω вместе с σ -алгеброй его подмножеств образуют *измеримое пространство* $(\Omega, \mathcal{A})$.

Формальное определение вероятностной модели. 4

Мера, то есть σ -аддитивная функция множеств, P , определенная на $\mathfrak{A}$ и нормированная условием $P(\Omega) = 1$, называется *вероятностной мерой* или *вероятностью*. Напомним, что свойство σ -аддитивности, также называемое счетной аддитивностью, заключается в следующем: если $A_1, A_2, \dots$ – события ($A_i \in \mathfrak{A}$), причем $A_i \cap A_j = \emptyset$ при $i \neq j$, то

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

Для $B \in \mathfrak{A}$ значение $P(B)$ называется *вероятностью события B* . Тройка $(\Omega, \mathfrak{A}, P)$ называется *вероятностным пространством* или *вероятностной моделью*.

Независимые события

События A и B , связанные с некоторой вероятностной моделью $(\Omega, \mathfrak{A}, P)$, называются *независимыми*, если вероятность их одновременного осуществления равна произведению вероятностей каждого из них:

$$P(A \cap B) = P(A) \cdot P(B).$$

События $A_1, A_2, \dots, A_n$ в вероятностной модели $(\Omega, \mathfrak{A}, P)$ ($n \geq 2$) называются *независимыми в совокупности*, если для любого $k \leq n$ и любых индексов $i_1, \dots, i_k$ ($i_p \neq i_q$ при $p \neq q$ и $1 \leq i_p \leq n$, $p = 1, \dots, k$)

$$P\left(\bigcap_{p=1}^k A_{i_p}\right) = \prod_{p=1}^k P(A_{i_p}).$$

Условная вероятность

Пусть A и B – события, причем $P(B) \neq 0$. *Условной вероятностью A при условии B* называется величина

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

Если события A и B независимы, то из определения условной вероятности следует, что $P(A|B) = P(A)$.

Из определения условной вероятности также вытекает формула

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

Эту формулу иногда называют *законом умножения вероятностей*.

Формула полной вероятности

В некоторой вероятностной модели $(\Omega, \mathfrak{A}, P)$ рассмотрим события $A_1, A_2, \dots, A_n$ ($n \geq 2$), которые обладают следующими свойствами:

- a) события $A_1, A_2, \dots, A_n$ несовместны, то есть никакие два из них не могут произойти одновременно;
- b) одно из событий $A_1, A_2, \dots, A_n$ обязательно произойдет, то есть $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$, причем $P(A_i) > 0$, $i = 1, \dots, n$.

Если события $A_1, A_2, \dots, A_n$ обладают свойствами a) и b), то говорят, что они образуют *полную группу*.

Пусть B – некоторое событие, а события $A_1, A_2, \dots, A_n$ образуют полную группу. В таком случае справедлива формула

$$P(B) = \sum_{k=1}^n P(B|A_k)P(A_k),$$

называемое *формулой полной вероятности*.

Формула Байеса. 1

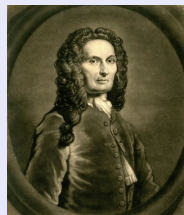
Пусть B – некоторое событие, имеющее положительную вероятность ($P(B) > 0$), а события $A_1, A_2, \dots, A_n$ образуют полную группу. В таком случае справедлива формула

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{k=1}^n P(B|A_k)P(A_k)}, \quad k = 1, \dots, n,$$

называемая *формулой Байеса*. Формула Байеса позволяет уточнить представление о вероятности любого из событий, составляющих полную группу, с учетом информации об осуществлении некоторого события.



Thomas Bayes
(1702-1761)



Abraham de
Moivre
(1667-1756)

Формула Байеса. 2

Рассмотрим весьма примечательную задачу.

Задача 4. При контроле правдивости показаний подозреваемого на “детекторе лжи” вероятность признать ложью ответ, не соответствующий действительности, равна 0.99, вероятность ошибочно признать ложью правдивый ответ равна 0.01. Известно, что ответы, не соответствующие действительности, составляют 1% всех ответов подозреваемого. Какова вероятность того, что ответ, признанный ложью, и в самом деле не соответствует действительности?

Решение. Пусть α – вероятность признать ложью ответ, не соответствующий действительности, β – вероятность ошибочно признать ложью правдивый ответ, γ – доля ответов, не соответствующих действительности, среди всех ответов подозреваемого. Тогда по формуле Байеса искомая условная вероятность p равна

$$p = \frac{\alpha\gamma}{\alpha\gamma + \beta(1 - \gamma)} = \frac{0.99 \cdot 0.01}{0.99 \cdot 0.01 + 0.01 \cdot 0.99} = \frac{1}{2}.$$

Другими словами, на самом деле лишь **половина** признанных ложью ответов не соответствуют действительности, хотя, казалось бы, выбранные значения параметров контроля α и β являются довольно жесткими.

Случайные величины и их распределения. 1

Любая вещественная функция $X = X(\omega)$, определенная на Ω , отображает множество элементарных исходов во множество вещественных чисел $\mathbb{R}$. Пусть A – произвольное множество вещественных чисел: $A \subseteq \mathbb{R}$. Определим множество

$$X^{-1}(A) = \{\omega : X(\omega) \in A\},$$

являющееся подмножеством множества Ω , называемое *прообразом множества A* (при преобразовании X).

Определим *борелевскую σ -алгебру $\mathfrak{B}$* подмножеств вещественной прямой $\mathbb{R}$ как наименьшую σ -алгебру, содержащую все интервалы. Любой элемент борелевской σ -алгебры называется *борелевским множеством*. Таким образом, любое борелевское множество можно построить из интервалов с помощью операций дополнения, счетного объединения или счетного пересечения.

Случайные величины и их распределения. 2

Если $X^{-1}(A) \in \mathfrak{A}$ для любого борелевского множества $A \in \mathfrak{B}$, то функция $X(\omega)$ называется *измеримой*.

Конечная вещественная измеримая функция называется *случайной величиной*.

Простейшим примером нетривиальной случайной величины является индикатор $\mathbf{1}_B(\omega)$ множества $B \in \mathfrak{A}$,

$$\mathbf{1}_B(\omega) = \begin{cases} 1, & \text{если } \omega \in B, \\ 0, & \text{если } \omega \notin B. \end{cases}$$

Случайные величины и их распределения. 3

Рассмотрим события $B_1, B_2, \dots$ такие, что $B_i \cap B_j = \emptyset$ при $i \neq j$ и $\bigcup_i B_i = \Omega$. Пусть $\{x_1, x_2, \dots\}$ – вещественные числа. Случайная величина

$$X(\omega) = \sum_j x_j \mathbf{1}_{B_j}(\omega)$$

называется *дискретной*. Заметим, что свойства событий B_j гарантируют, что для каждого ω в последней сумме один и только один индикатор отличен от нуля. При этом $B_i = \{\Omega : X(\Omega) = x_i\}$.

Случайные величины и их распределения. 4

Обозначим

$$p_i = P(B_i) \quad (= P(\{\Omega : X(\Omega) = x_i\})).$$

Набор $\{(x_i, p_i)\}_{i \geq 1}$ называется *распределением вероятностей* или просто *распределением* дискретной случайной величины X .

В дальнейшем для краткости вместо $P(\{\Omega : X \in A\})$ мы будем писать $P(X \in A)$.

Распределение дискретной случайной величины X полностью определяет вероятности попадания случайной величины X в любое борелевское множество: если $A \in \mathfrak{B}$, то

$$P(X \in A) = \sum_{i: x_i \in A} p_i.$$

Случайные величины и их распределения. 5

Требование измеримости случайной величины как функции элементарного исхода гарантирует возможность рассмотрения множеств вида $\{\omega : X(\omega) \in A\}$ в качестве событий, какими бы ни были борелевские множества A , и, следовательно, определены вероятности попадания случайной величины в любые борелевские множества.

Таким образом, мы можем рассмотреть вероятностную меру P_X , определенную на σ -алгебре $\mathfrak{B}$ с помощью соотношения

$$P_X(A) = P(\{\omega : X(\omega) \in A\}), \quad A \in \mathfrak{B}.$$

Эта вероятностная мера называется *распределением* случайной величины X . Итак, любая случайная величина X порождает новое вероятностное пространство $(\mathbb{R}, \mathfrak{B}, P_X)$.

Случайные величины и их распределения. 6

Пусть X – случайная величина. Рассмотрим вероятность $P(X \in A)$ в том случае, когда множество A является бесконечным интервалом вида $(-\infty, x)$, $x \in \mathbb{R}$. В таком случае положим

$$F_X(x) \equiv F(x) = P(X < x).$$

Функция $F(x)$ определена для каждого вещественного x и называется *функцией распределения* случайной величины X . Если X – дискретная случайная величина, для которой $P(X = x_i) = p_i$, то

$$F_X(x) = \sum_{i: x_i < x} p_i. \quad (3)$$

Функция распределения вида (3) называется *дискретной*.

Случайные величины и их распределения. 7

Функция распределения $F(x)$ любой случайной величины обладает следующими свойствами:

- $F(x)$ не убывает и непрерывна слева;
- $\lim_{x \rightarrow -\infty} F(x) = 0$,
- $\lim_{x \rightarrow +\infty} F(x) = 1$.

Обратное утверждение также верно: для любой функции $F(x)$, удовлетворяющей этим трем условиям, существуют вероятностное пространство и заданная на нем случайная величина такая, что $F(x)$ является ее функцией распределения.

Случайные величины и их распределения. 8

Каждой случайной величине соответствует одна и только одна функция распределения.

Обратное неверно. Например, на вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, в котором $\Omega = [0, 1]$, $\mathfrak{A}$ – совокупность всех борелевских подмножеств интервала $[0, 1]$, а P – мера, каждому интервалу приписывающая его длину, случайные величины $X(\omega) \equiv \omega$ и $X(\omega) \equiv 1 - \omega$ имеют одну и ту же функцию распределения

$$F(x) = \begin{cases} 0 & \text{при } x < 0, \\ x & \text{при } 0 \leq x \leq 1, \\ 1 & \text{при } x > 1. \end{cases}$$

Число x называется *точкой роста* функции распределения $F(x)$, если $F(x + \delta) - F(x - \delta) > 0$ для любого $\delta > 0$.

Случайные величины и их распределения. 9

Среди всех мер, заданных на $(\mathbb{R}, \mathfrak{B})$, особую роль играет *мера Лебега*, которая каждому интервалу (a, b) приписывает его длину (очевидно, равную $b - a$), так что любое одноточечное множество имеет нулевую меру Лебега, что вполне естественно для непрерывных моделей. Кстати, практически ни в одном учебнике по теории вероятностей не обсуждается один важный вопрос, а именно, вопрос о том, почему для того, чтобы определить вероятностную модель (вероятностное пространство), непременно нужно рассматривать специальную σ -алгебру событий. Казалось бы, было бы намного проще, если бы в качестве совокупности событий *всегда* рассматривалось множество всех подмножеств множества Ω (или $\mathbb{R}$, если вероятностная мера задается на подмножествах вещественной прямой). Однако, вообще говоря, такой подход оказывается невозможным. Еще в 1930 выдающийся польский математик С. М. Улам доказал теорему, устанавливающую, что конечная мера, заданная на множестве всех подмножеств некоторого множества мощности континуума и приписывающая нулевую меру каждому множеству, содержащему ровно один элемент (одну точку), обязана приписывать нулевую меру любым другим множествам, то есть является тривиальной (краткое доказательство теоремы Улама приведено, например, в книге (Окстоби, 1971), глава 5).

Случайные величины и их распределения. 10

Пусть μ – мера, заданная на измеримом пространстве $(\mathbb{R}, \mathfrak{B})$.

Распределение P_X называется *абсолютно непрерывным относительно μ* , если существует неотрицательная функция $p(x)$ такая, что

$$P_X(A) = \int_A p(y) \mu(dy) \quad (4)$$

для любого $A \in \mathfrak{B}$ (здесь интеграл понимается в смысле Лебега). При этом функция $p(x)$ называется *плотностью распределения* или *плотностью вероятностей* или просто *плотностью* случайной величины X . Дискретное распределение $\{(x_i, p_i)\}_{i \geq 1}$ не является абсолютно непрерывным относительно меры Лебега, но является абсолютно непрерывным относительно *считающей меры*, которая каждому множеству $A \in \mathfrak{B}$ приписывает число, равное количеству тех точек $\{x_i\}$, которые попадают в A . Более того, в последнем случае

$$p(x) = P(X = x) = \begin{cases} p_i, & x = x_i, \\ 0, & x \neq x_i \text{ для любого } i. \end{cases}$$

Случайные величины и их распределения. 11

Распределение, абсолютно непрерывное относительно меры Лебега, мы будем ниже называть просто *абсолютно непрерывным*. Можно показать, что распределение случайной величины X абсолютно непрерывно, если $P(X \in A) = 0$ для любого множества $A \in \mathfrak{B}$ нулевой меры Лебега. Для абсолютно непрерывного распределения $p(x) = F'(x)$. Случайная величина, распределение которой абсолютно непрерывно, и соответствующая функция распределения также называются абсолютно непрерывными.

Дискретные и абсолютно непрерывные распределения не исчерпывают все возможные виды распределений. Существуют также функции распределения, множество точек роста которых имеет лебегову меру нуль. Такие функции распределения и соответствующие им случайные величины и распределения вероятностей называются *сингулярными*. Сингулярные распределения представляют собой довольно экзотические объекты и практически не используются в прикладной теории вероятностей, теории риска или актуарной математике.

Случайные величины и их распределения. 12

Приведем несколько общих свойств функций распределения, доказательство которых можно найти в стандартных курсах теории вероятностей или теории функций вещественной переменной.

Теорема 1. Для любого $\delta > 0$ функция распределения $F(x)$ имеет не более чем счетное число точек скачков, в которых скачок превышает δ , и, следовательно, не более чем счетное число точек разрыва. Производная $F'(x)$ функции распределения $F(x)$ существует почти во всех точках x .

Теорема 2. Любая функция распределения $F(x)$ может быть однозначно представлена в виде взвешенной суммы двух компонент:

$$F(x) = a_1 F_1(x) + a_2 F_2(x),$$

где a_1, a_2 – неотрицательные числа, сумма которых равна единице, $F_1(x)$ – непрерывная функция распределения, $F_2(x)$ – дискретная функция распределения, в точке x равная сумме всех скачков функции $F(x)$ в точках разрыва, не превосходящих x .

Теорема 3. Любая функция распределения $F(x)$ может быть однозначно представлена в виде взвешенной суммы трех компонент:

$$F(x) = a_1 F_1(x) + a_2 F_2(x) + a_3 F_3(x),$$

где a_1, a_2, a_3 – неотрицательные числа, сумма которых равна единице, $F_1(x)$ – абсолютно непрерывная функция распределения, $F_2(x)$ – дискретная функция распределения, в точке x равная сумме всех скачков функции $F(x)$ в точках разрыва, не превосходящих x , $F_3(x)$ – сингулярная функция распределения.

Случайные величины и их распределения. 14

Если X – дискретная случайная величина и $P(X = x) > 0$, то x называется *возможным значением* случайной величины X .

Случайная величина X имеет *решетчатое распределение*, если все ее возможные значения имеют вид $\{b + nh, n = 0, \pm 1, \pm 2, \dots\}$, где b и $h > 0$ – фиксированные числа. Для решетчатого распределения максимальное из чисел h называется *шагом распределения*.

Говорят, что случайная величина X имеет *симметричное распределение*, если $P(X < x) = P(-X < x)$ для любого $x \in \mathbb{R}$ или, что эквивалентно, $F_X(x) \equiv 1 - F_X(x + 0)$. Если при этом случайная величина X абсолютно непрерывна, то ее плотность $p_X(x)$ четна: $p_X(x) = p_X(-x)$, $x \in \mathbb{R}$.

Случайные величины и их распределения. 15

Случайные величины $X_1, X_2, \dots, X_n$ называются *независимыми в совокупности*, если для любых борелевских множеств $B_1, B_2, \dots, B_n$ события $\{\omega : X_1(\omega) \in B_1\}, \{\omega : X_2(\omega) \in B_2\}, \dots, \{\omega : X_n(\omega) \in B_n\}$ независимы в совокупности.

Говорят, что случайные величины $\{X_j\}_{j \geq 1}$ образуют *последовательность независимых случайных величин*, если для любого $n \geq 1$ случайные величины $X_1, X_2, \dots, X_n$ независимы в совокупности.

Случайные величины и их распределения. 16

Если $X_1 = X_1(\Omega), \dots, X_n = X_n(\Omega)$ – случайные величины, определенные на одном и том же вероятностном пространстве $(\Omega, \mathfrak{A}, P)$, то вектор $\mathbf{X} = (X_1, \dots, X_n)$ называется случайным вектором, или n -мерной случайной величиной.

Областью значений случайного вектора $\mathbf{X}$ является n -мерное евклидово пространство $\mathbb{R}^n$.

Борелевской σ -алгеброй $\mathfrak{B}_n$ подмножеств $\mathbb{R}^n$ называется минимальная σ -алгебра, содержащая все n -мерные параллелепипеды. Элементы борелевской σ -алгебры, как и ранее, будем называть борелевскими множествами.

Для каждого борелевского множества B пространства $\mathbb{R}^n$ определена вероятность $P(\mathbf{X} \in B)$. Набор $\{P(\mathbf{X} \in B) : B \in \mathfrak{B}_n\}$ называется *распределением случайного вектора $\mathbf{X}$* . В частности, для любых действительных чисел $x_1, \dots, x_n$ определена функция

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n),$$

которая называется *функцией распределения случайного вектора $\mathbf{X}$* .

Случайные величины и их распределения. 17

Случайные величины $X_1, \dots, X_n$ независимы тогда и только тогда, когда

$$F(x_1, \dots, x_n) = \prod_{k=1}^n F_k(x_k)$$

для любых действительных $x_1, \dots, x_n$. Здесь

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n) \quad \text{и} \quad F_k(x) = P(X_k < x).$$

Для любой последовательности функций распределения $F_1(x), F_2(x), \dots$ существует вероятностное пространство $(\Omega, \mathfrak{A}, P)$ и определенная на нем последовательность независимых случайных величин $X_1, X_2, \dots$ такая, что для любого n функция распределения случайной величины X_n есть $F_n(x)$.

Числовые характеристики случайных величин.

Характеристики центра. 1

Часто при анализе той или иной стохастической ситуации основной интерес представляет даже не распределение вероятностей наблюдаемой случайной величины, а лишь такое число, вокруг которого концентрируются возможные значения этой случайной величины при многократном воспроизведении стохастической ситуации. Примером может служить непосредственное измерение некоторого параметра, когда вследствие случайных погрешностей каждое новое измерение отличается от предыдущего, но вместе с тем ясно, что все результаты измерения группируются вокруг некоторого “центрального” значения. Возможными синонимами понятия “центрального” значения случайной величины являются “среднее”, “ожидаемое”, “наиболее вероятное” значение. Несмотря на кажущуюся простоту, понятие “центрального” значения случайной величины является довольно тонким, и каждому из словесных описаний, приведенных выше, можно придать свое строгое математическое определение. Эти определения, вообще говоря, приводят к *разным* понятиям и, соответственно, к *разным* числам.

Характеристики центра. 2

Определение. Пусть X – дискретная случайная величина, принимающая значения $x_1, x_2, \dots$ с вероятностями соответственно $p_1, p_2, \dots$. Математическим ожиданием дискретной случайной величины называется число

$$EX = x_1 p_1 + x_2 p_2 + \dots \quad (5)$$

Если ряд (5) расходится, то говорят, что математическое ожидание не существует

Пример дискретной случайной величины, у которой математическое ожидание не существует:

$$P(X = \pm k) = \frac{3}{\pi^2 k^2}, \quad k = 1, 2, \dots$$

$$\left(\frac{6}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} = 1, \text{ но ряд } \frac{3}{\pi^2} \sum_{k=-\infty}^{\infty} \frac{1}{k} \text{ расходится.} \right)$$

Характеристики центра. 3

Смысл математического ожидания становится ясным, если мы предположим, что случайная величина может с равными вероятностями принять одно из конечного числа значений $x_1, \dots, x_n$. Тогда $p_k = \frac{1}{n}$, $k = 1, \dots, n$, и математическое ожидание такой случайной величины совпадает со средним арифметическим чисел $x_1, \dots, x_n$:

$$EX = \frac{1}{n} \sum_{k=1}^n x_k.$$

Таким образом, мы можем заключить, что в общем случае математическое ожидание дискретной случайной величины, определенное формулой (5), является обобщением понятия среднего арифметического на случай (возможно, бесконечного числа) неравновероятных возможных значений.

Характеристики центра. 4

Определение. Пусть X – абсолютно непрерывная случайная величина, имеющая плотность $p_X(x)$. Математическим ожиданием непрерывной случайной величины называется число

$$EX = \int_{-\infty}^{\infty} xp_X(x)dx. \quad (6)$$

Если в (6) интеграл расходится, то говорят, что математическое ожидание не существует.

Пример абсолютно непрерывной случайной величины, у которой математическое ожидание не существует: пусть $a \in \mathbb{R}$,

$$p_X(x) = \frac{1}{\pi(1 + (x - a)^2)}, \quad x \in \mathbb{R}.$$

Тогда плотность $p_X(x)$ имеет максимум в точке a , но интеграл

$$\int_{-\infty}^{\infty} \frac{x dx}{\pi(1 + (x - a)^2)}$$

расходится.

Характеристики центра. 5

В общем случае математическое ожидание произвольной случайной величины X с функцией распределения $F_X(x)$ можно записать в виде *интеграла Лебега–Стильтьеса*

$$EX = \int_{-\infty}^{\infty} x dF_X(x).$$

Основным свойством математического ожидания можно считать его линейность: если a и b – два числа, а X и Y – две случайные величины, то

$$E(aX + bY) = aEX + bEY$$

при условии, что любые два из участвующих математических ожиданий существуют.

Отсюда, в частности, вытекает, что

- ★ математическое ожидание суммы случайных величин равно сумме математических ожиданий этих случайных величин;
- ★ неслучайный коэффициент при случайной величине можно выносить за знак математического ожидания.

Характеристики центра. 6

Другие свойства математического ожидания.

- Пусть X – случайная величина, A – некоторое подмножество ее значений. Пусть $1_A(x)$ – индикаторная функция множества A . Тогда $1_A(X)$ – случайная величина, принимающая значение 1, если $X \in A$, и значение 0, если $X \notin A$, причем

$$P(1_A(X) = 1) = P(X \in A) = 1 - P(1_A(X) = 0).$$

Тогда в соответствии с определением математического ожидания дискретной случайной величины мы имеем

$$E1_A(X) = 1 \cdot P(X \in A) + 0 \cdot P(X \notin A) = P(X \in A).$$

- Если случайные величины X и Y независимы, то

$$EXY = EX \cdot EY.$$

Характеристики центра. 7

Смысл математического ожидания как центрального значения случайной величины поясняет следующая задача.

Задача 5. Предположим, что некий человек на перемещение из точки a в точку b затрачивает усилия, равные $(b - a)^2$. Это означает, что он легко перемещается на малые расстояния, но большие перемещения даются ему с большим трудом. Скажем, мы рассматриваем пожилого немощного человека. Пусть X – случайная величина, характеризующая координату точки, в которую необходимо переместиться этому пожилому человеку (скажем, он может пойти или поехать в поликлинику, в собес, в магазин, к внукам и т. п.). Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?

Характеристики центра. 8

Решение. Интуитивно ясно, что если этот человек будет жить в “средней”, “центральной” точке между возможными конечными пунктами своих перемещений, то его усилия в среднем будут невелики. Подкрепим эту догадку строгими рассуждениями. Средние усилия, затрачиваемые этим человеком на перемещение из неслучайной точки x в случайную точку X , имеют вид

$$E(X - x)^2 = E(X^2 - 2xX + x^2) = x^2 - 2xE(X) + E(X^2).$$

Правая часть этого выражения как функция от x представляет собой хорошо знакомую квадратичную параболу. Ее минимум находится в точке $x_0 = E(X)$. Таким образом, математическое ожидание характеризует центр распределения случайной величины X (или, в терминах данной задачи, определяет “среднюю” или “центральную” точку населенного пункта) в смысле квадратичной функции потерь.

Характеристики центра. 9

Определение. Пусть q – некоторое число из интервала $(0, 1)$. *Квантилью порядка q (или q -квантилью)* случайной величины X называется число x_q такое, что одновременно

$$P(X \leq x_q) \geq q \text{ и } P(X \geq x_q) \geq 1 - q.$$

Определение. *Медианой* случайной величины X называется ее квантиль порядка $\frac{1}{2}$.

Медиана случайной величины X обозначается $\text{med}X$. По определению медианы, одновременно выполняются неравенства

$$P(X \leq \text{med}X) \geq \frac{1}{2} \quad \text{и} \quad P(X \geq \text{med}X) \geq \frac{1}{2}.$$

Другими словами, медиана случайной величины X характеризует “центр” ее распределения в том смысле, что случайная величина X может с *равными возможностями* быть как меньше медианы, так и больше нее.

Характеристики центра. 10

Для $q \in (0, 1)$ квантиль x_q *непрерывной* случайной величины можно вычислить как корень уравнения

$$F_X(x) = q,$$

где $F_X(x)$ – функция распределения случайной величины X .

Таким образом, медиана *непрерывной* случайной величины является корнем уравнения

$$F_X(x) = \frac{1}{2}.$$

Смысл понятия медианы поясняет следующая задача, являющаяся в некотором смысле дополнением к Задаче 5.

Задача 6. Предположим, что некий человек на перемещение из точки a в точку b затрачивает усилия, равные $|b - a|$. Это означает, что его затраты на перемещение пропорциональны длине пути. Скажем, мы рассматриваем делового человека, для которого фактор времени имеет решающее значение и который может перемещаться в любую точку с постоянной скоростью. Пусть X – случайная величина, характеризующая координату точки, куда необходимо переместиться этому человеку. Спрашивается: где должен жить этот человек, чтобы минимизировать свои ожидаемые усилия, связанные с перемещениями?

Характеристики центра. 12

Сведем **решение** этой задачи к доказательству неравенства

$$E|X - \text{med}X| \leq E|X - d|, \quad (7)$$

справедливого для любого числа d . Это будет означать, что решением задачи является медиана случайной величины X .

Сначала рассмотрим случай $d > \text{med}X$. Тогда

$$|X - d| - |X - \text{med}X| = \begin{cases} \text{med}X - d & \text{при } X \in [d, \infty), \\ d + \text{med}X - 2X & \text{при } X \in (\text{med}X, d), \\ d - \text{med}X & \text{при } X \in (-\infty, \text{med}X]. \end{cases}$$

Характеристики центра. 13

Так как $d + \text{med}X - 2X > \text{med}X - d$ при $\text{med}X < X < d$, то

$$\begin{aligned} E|X - d| - E|X - \text{med}X| &= E[|X - d| - |X - \text{med}X|] \geqslant \\ &\geqslant (\text{med}X - d)E\mathbf{1}_{[d, \infty)}(X) + \\ &\quad + (\text{med}X - d)E\mathbf{1}_{(\text{med}X, d)}(X) + \\ &\quad + (d - \text{med}X)E\mathbf{1}_{(-\infty, \text{med}X]}(X) = \\ &= (\text{med}X - d)P(X \geqslant d) + \\ &\quad + (\text{med}X - d)P(\text{med}X < X < d) + \\ &\quad + (d - \text{med}X)P(X \leqslant \text{med}X) = \\ &= (d - \text{med}X)[P(X \leqslant \text{med}X) - P(X > \text{med}X)] \geqslant 0. \end{aligned}$$

Последнее неравенство вытекает из определения медианы. Но это и означает справедливость (7) для случая $d > \text{med}X$. Случай $d < \text{med}X$ рассматривается аналогично.

Характеристики центра. 14

Таким образом, медиана характеризует центр распределения случайной величины X (определяет центр населенного пункта в терминах Задачи б) в смысле функции потерь, имеющей вид абсолютной величины.

В отличие от математического ожидания, медиана всегда определена, но не всегда единственна. Поясним сказанное на примере случайной величины X , принимающей значения 0 и 1 с равными вероятностями. Тогда, как легко видеть, определению медианы удовлетворяет любая точка отрезка $[0, 1]$ (включая концы). В то же время, очевидно, $EX = \frac{1}{2}$.

Характеристики центра. 15

Определение. Модой абсолютно непрерывной случайной величины X , имеющей плотность распределения $p_X(x)$, называется число $\text{mod}X$, удовлетворяющее условию

$$p_X(x) \leq p_X(\text{mod}X)$$

для любого x .

Другими словами, мода непрерывной случайной величины X – это такая точка, в которой плотность распределения случайной величины X принимает наибольшее значение.

Определение. Модой дискретной случайной величины X , принимающей значения $x_1, x_2, \dots$, называется такое ее значение x_{k_0} , которое удовлетворяет условию

$$P(X = x_k) \leq P(X = x_{k_0})$$

для любого $k = 1, 2, \dots$

В общем случае распределение P_X случайной величины X называется *унимодальным*, если существует значение $x = a$ такое, что при $x < a$ функция распределения $F_X(x)$ выпукла, а при $x > a$ – вогнута. При этом число a называется *модой* распределения.

Из этих определений вытекает, что моду случайной величины X можно интерпретировать как наиболее вероятное ее значение.

Характеристики центра. 15

Каждая из введенных выше числовых характеристик – математическое ожидание, медиана и мода – описывает “центральную” тенденцию в поведении случайной величины. Однако эти характеристики совсем не обязаны совпадать. Чтобы в этом убедиться, рассмотрим следующую весьма поучительную задачу.

Задача 7. Предположим, что вероятность того, что среднегодовой доход наугад выбранного жителя некоторой страны превосходит уровень x , равна $e^{-\lambda x}$, где $\lambda > 0$. Другими словами, доля индивидуумов, чей среднегодовой доход превосходит уровень x , во всем населении страны, составляет $100 \cdot e^{-\lambda x} \%$. Каков средний доход жителей этой страны?

Эта кажущаяся очень простой задача весьма знакома. Однако на самом деле эта задача не является тривиальной, но все затруднения не являются математическими. Главная особенность связана с интерпретацией понятия среднего дохода.

Характеристики центра. 16

Решение. Пусть X – случайная величина, равная среднегодовому доходу наугад выбранного жителя. Тогда $F_X(x) = P(X < x) = 1 - e^{-\lambda x}$. Такая функция распределения соответствует абсолютно непрерывной случайной величине с плотностью вероятностей $p_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Параметр λ влияет лишь на выбор единиц измерения. Поэтому, не ограничивая общности, выберем такую “условную единицу”, которая соответствует значению $\lambda = 1$. Тогда

$$EX = \int_0^{\infty} x e^{-x} dx = 1.$$

Медиана случайной величины X находится как решение уравнения $1 - e^{-x} = \frac{1}{2}$ и оказывается равной

$$\text{med}X = \ln 2 \approx 0.693.$$

В то же время, мода случайной величины X , то есть та точка, в которой плотность $p_X(x) = e^{-x}$ максимальна, равна нулю, то есть

$$\text{mod}X = 0.$$

Характеристики центра. 17

Таким образом, если мы по какой-либо причине желаем приукрасить ситуацию (скажем, готовя отчет правительства перед парламентом), но *не хотим подтасовывать данные*, то, конечно же, мы будем характеризовать средний доход как математическое ожидание, которое в данном случае почти в полтора раза (весьма существенно!) превосходит медиану. Если же мы, к примеру, представляем ультра-радикальную оппозицию, то на митингах мы будем, опять-таки, *не подтасовывая данные*, говорить о среднем доходе, имея в виду моду случайной величины X .

В определении математического ожидания в равной степени учитываются как вероятности значений случайной величины, так и *сами значения*.

Поэтому наличие одного-двух мультимиллиардеров довольно заметно повышает средний доход граждан страны, характеризуемый математическим ожиданием. В определении же медианы, грубо говоря, величина больших значений рассматриваемой характеристики не играет большой роли. Важно лишь их количество. Поэтому наличие нескольких мультимиллиардеров практически не влияет на средний доход граждан страны, характеризуемый медианой.

Характеристики центра. 18

К примеру, если в некотором населенном пункте живет 10001 человек, из которых 10000 имеет доход 1000 условных единиц (говоря так мы не имеем в виду непременно доллары), а один житель имеет доход 10^9 условных единиц. Тогда медиана дохода наугад выбранного жителя этого населенного пункта, очевидно, равна 1000 условных единиц, тогда как математическое ожидание дохода наугад выбранного жителя этого городка равно

$$\frac{10000}{10001} \cdot 10^3 + \frac{1}{10001} \cdot 10^9 \approx 10^3 + 10^5$$

условных единиц, то есть практически в 100 раз больше дохода наугад выбранного жителя этого городка, характеризующего его медианой. Поэтому если мы хотим дать более взвешенный и объективный ответ (скажем, готовя аналитическую записку для администрации президента), то мы будем характеризовать средний доход медианой.

Характеристики разброса случайных величин. 1

Наряду с характеристикой “центра” распределения вероятностей случайной величины, определенную информацию об этом распределении несет параметр, характеризующий степень разброса возможных значений случайной величины вокруг “центрального” значения. Как и “центр”, разброс могут описывать различные характеристики.

Определение. *Дисперсией* случайной величины X называется число DX , определяемое как

$$DX = E(X - EX)^2.$$

Дисперсия определена только для тех случайных величин X , для которых EX^2 существует.

Дисперсия характеризует “центр” распределения квадратов отклонения случайной величины X от ее “центра” при многократном воспроизведении стохастической ситуации, с которой связана случайная величина X .

Дисперсию иногда бывает удобно вычислять по формуле

$$DX = EX^2 - (EX)^2.$$

Характеристики разброса случайных величин. 2

Перечислим основные свойства дисперсии.

- Неслучайный коэффициент можно выносить из-под знака дисперсии, возведя его в квадрат:

$$D(cX) = c^2 \cdot DX.$$

- Дисперсия любой случайной величины неотрицательна.
- Дисперсия случайной величины равна нулю тогда и только тогда, когда эта случайная величина всегда (с вероятностью единица) принимает одно и то же значение (естественно, равное ее математическому ожиданию).

Для унификации единиц измерения степени разброса с единицами измерения самой случайной величины используют понятие *среднеквадратичного отклонения*, определяемого как квадратный корень из дисперсии: $\sigma = \sqrt{DX}$.

Характеристики разброса случайных величин. 3

Несложно видеть, что

$$DX = E(X - EX)^2 = \min_a E(X - a)^2$$

(см. Задачу 5). Поэтому, вспоминая задачу 6, мы заключаем, что наряду с дисперсией степень разброса значений случайной величины может характеризовать также такой параметр как

$$E|X - \text{med}X| = \min_a E|X - a|.$$

Однако такая характеристика используется редко из-за отсутствия у нее удобных аналитических свойств.

Однако в описательной статистике довольно часто используется еще одна характеристика степени разбросанности возможных значений случайной величины.

Характеристики разброса случайных величин. 4

Определение. Пусть $x_{1/4}$ и $x_{3/4}$ – квантили случайной величины X порядков $\frac{1}{4}$ и $\frac{3}{4}$ соответственно. Разность $R_X = x_{3/4} - x_{1/4}$ называется *интерквартильным размахом* случайной величины X .

Можно сказать, что интерквартильный размах в той же мере характеризует степень разброса возможных значений случайной величины, в какой медиана характеризует ее центральное значение: внутри интервала $[x_{1/4}, x_{3/4}]$ случайная величина X попадает с вероятностью $\frac{1}{2}$. При этом среди всех возможных интервалов с таким свойством (легко видеть, что все такие интервалы имеют своими концами квантили случайной величины порядков $q - \frac{1}{4}$ и $q + \frac{1}{4}$ при $\frac{1}{4} \leq q \leq \frac{3}{4}$) интервал $[x_{1/4}, x_{3/4}]$ занимает “срединное” положение, поскольку вероятность того, что случайная величина X попадет левее этого интервала равна вероятности того, что случайная величина X попадет правее этого интервала (при этом обе эти вероятности равны $\frac{1}{4}$). Поэтому длина этого интервала вполне может быть принята в качестве меры разброса случайной величины.

Неравенства для характеристик центра и разброса случайных величин. Неравенство Маркова

Пусть X – неотрицательная случайная величина. Тогда для произвольного положительного числа x мы можем записать:

$$\begin{aligned} E \min\{x, X\} &= E[X\mathbf{1}_{[0,x)}(X) + x\mathbf{1}_{[x,\infty)}(X)] = \\ &= E[X\mathbf{1}_{[0,x)}(X)] + E[x\mathbf{1}_{[x,\infty)}(X)] = \\ &= xE\mathbf{1}_{[x,\infty)}(X) + E[X\mathbf{1}_{[0,x)}(X)] \geq xE\mathbf{1}_{[x,\infty)}(X) = xP(X \geq x), \end{aligned}$$

откуда, сравнивая левую и правую части этого соотношения, мы получаем неравенство

$$P(X \geq x) \leq \frac{E \min\{x, X\}}{x}, \quad (8)$$

справедливое при любом $x > 0$.

Вытекающее из (8) неравенство

$$P(X \geq x) \leq \frac{EX}{x}$$

называется *неравенством Маркова*.

Неравенство Чебышева

Пусть X – произвольная случайная величина, у которой существует EX^2 . Применим неравенство Маркова к случайной величине $Z = (X - EX)^2$ и получим

$$P(|X - EX| \geq x) = P((X - EX)^2 \geq x^2) \leq \frac{E(X - EX)^2}{x^2} = \frac{DX}{x^2}.$$

Соотношение между левой и правой частями этого неравенства называют *неравенством Чебышева*:

$$P(|X - EX| \geq x) \leq \frac{DX}{x^2}. \quad (9)$$

С помощью неравенств Маркова и Чебышева можно оценивать вероятности больших значений неотрицательных случайных величин или вероятности больших отклонений случайных величин от своего среднего значения.

Неравенство Гаусса

При дополнительных условиях неравенство Чебышева можно уточнить. А именно, рассмотрим абсолютно непрерывную случайную величину X с унимодальным распределением. В таком случае существует такая точка $\nu \in \mathbb{R}$, что плотность $p_X(x)$ случайной величины X не убывает при $x \leq \nu$ и не возрастает при $x \geq \nu$. При этом число ν является *модой* случайной величины X . Предположим, что $EX^2 < \infty$. Обозначим

$$\tau^2 = E(X - \nu)^2.$$

В 1821 году К. Гаусс доказал следующий результат.

Теорема 4 (Gauss, 1821). Пусть X – абсолютно непрерывная случайная величина с унимодальным распределением и модой ν . Тогда для любого $r > 0$ справедливо неравенство

$$P(|X - \nu| \geq r) \leq \frac{4\tau^2}{9r^2}.$$

Правило «трех сигм»

Вообще говоря, такая характеристика как τ^2 нетрадиционна для практических задач. Намного привычнее иметь дело с дисперсией, то есть со средним квадратом отклонения случайной величины X не от моды (которая может быть не единственной), но от математического ожидания. Если распределение случайной величины X унимодально и симметрично, то математическое ожидание (равное нулю) совпадает с модой (по крайней мере с одной из мод). В таком случае неравенство Гаусса позволяет убедиться в справедливости следующего «правила трех сигм» для случайных величин с симметричным унимодальным распределением. Обозначим $\sigma^2 = DX$.

Следствие 1 («правило трех сигм» для симметричных унимодальных распределений). *Если распределение случайной величины X симметрично и унимодально, то*

$$P(|X| \leq 3\sigma) \geq \frac{77}{81} > 0.95.$$

Неравенство Иенсена и его частные случаи. 1

Определение. Функция $g(x)$ называется *выпуклой*, если ее график лежит *выше* касательной к нему, проведенной в любой его точке. Функция $h(x)$ называется *вогнутой*, если ее график лежит *ниже* касательной к нему, проведенной в любой его точке.

Примерами выпуклых функций служат функции $g(x) = x^2$ ($-\infty < x < \infty$), $g(x) = 1/x$ ($x > 0$).

Примерами вогнутых функций служат функции $h(x) = \sqrt{x}$ ($x > 0$), $h(x) = \log_a x$ ($a > 1, x > 0$).

Пусть X – случайная величина, $g(x)$ – произвольная выпуклая функция, а $h(x)$ – произвольная вогнутая функция. *Неравенство Иенсена* устанавливает, что

$$Eg(X) \geq g(EX), \quad Eh(X) \leq h(EX).$$

В частности, применяя неравенство Иенсена для выпуклых функций к произвольной случайной величине Y и функции $g(x) = |x|$, мы получаем

$$|EY| \leq E|Y| \tag{10}$$

Неравенство Иенсена и его частные случаи. 2

С помощью неравенства Иенсена мы можем получить следующие полезные соотношения между характеристиками “центра” случайной величины и характеристиками ее разброса.

Пусть X – произвольная случайная величина. В свою очередь, применим неравенство (10) к случайной величине $Y = X - \text{med}X$. Вследствие линейности математического получаем

$$|EX - \text{med}X| \leq E|X - \text{med}X|. \quad (11)$$

Пусть X – произвольная случайная величина. Применим теперь неравенство Иенсена для выпуклых функций к случайной величине $Y = |X - EX|$ и функции $g(x) = x^2$. Получим

$$(E|X - EX|)^2 \leq E(X - EX)^2$$

или, что то же самое,

$$E|X - EX| \leq \sqrt{DX}. \quad (12)$$

Соотношение (12) принято называть *неравенством Ляпунова*.

Неравенство Иенсена и его частные случаи. 3

Вспомним Задачу 6, согласно которой $E|X - \text{med}X| \leq E|X - a|$ для любого a , в частности, при $a = EX$ мы получаем $E|X - \text{med}X| \leq E|X - EX|$. Используя последнее неравенство как связующее звено между (11) и (12), мы получаем неравенство

$$|EX - \text{med}X| \leq \sqrt{DX}.$$

Неравенства для дисперсии и интерквартильного размаха. 1

Пусть $\mathcal{F}$ – класс всех распределений с конечным вторым моментом. Пусть $\mathcal{F}_0$ – подкласс класса $\mathcal{F}$, содержащий все унимодальные абсолютно непрерывные распределения с конечным вторым моментом и только их. Для простоты, без принципиального ограничения общности последующих рассуждений будем считать, что у распределений класса $\mathcal{F}$ математическое ожидание равно нулю, а у распределений класса $\mathcal{F}_0$ мода совпадает с математическим ожиданием и также равна нулю. Символ $\mathcal{L}(X)$ обозначает распределение случайной величины X . С помощью неравенств Чебышева и Гаусса можно получить следующее утверждение.

Теорема 5. Пусть $q \in [\frac{1}{2}, 1)$. Справедливы соотношения

$$\sup_{\mathcal{L}(X) \in \mathcal{F}} \frac{\sqrt{1-q}(x_q - x_{1-q})}{\sqrt{DX}} = \sqrt{2} < 1.4142,$$

$$\sup_{\mathcal{L}(X) \in \mathcal{F}_0} \frac{\sqrt{1-q}(x_q - x_{1-q})}{\sqrt{DX}} \leq \frac{2\sqrt{2}}{3} < 0.9429.$$

Неравенства для дисперсии и интерквартильного размаха. 2

Следствие 2. *Справедливы соотношения*

$$\sup_{\mathcal{L}(X) \in \mathcal{F}} \frac{R_X}{\sqrt{DX}} = 2\sqrt{2} < 2.8285,$$

$$\sup_{\mathcal{L}(X) \in \mathcal{F}_0} \frac{R_X}{\sqrt{DX}} \leq \frac{4\sqrt{2}}{3} < 1.8857.$$

Q-средние значения случайных величин. 1

Помимо введенных выше характеристик “центра” распределения случайной величины (математическое ожидание, медиана, мода), которые мы будем условно называть *структурными* средними, в эконометрических задачах традиционно используют так называемые *Q-средние значения*, примерами которых являются среднее арифметическое, среднее гармоническое и среднее геометрическое. Все Q-средние значения определяются единообразно.

А именно, пусть X – положительная случайная величина, а $Q(x)$ – некоторая функция, имеющая обратную функцию $Q^{-1}(x)$ (если $y = Q(x)$, то $x = Q^{-1}(y)$). С помощью функции $Q(x)$ и операции вычисления математического ожидания Q-среднее значение случайной величины X , обозначаемое $M_Q X$, определяется как

$$M_Q X = Q^{-1}(EQ(X)).$$

Q-средние значения случайных величин. 2

В частности, если $Q(x) \equiv x$, то $Q^{-1}(x) \equiv x$ и мы получаем “обобщенное среднее арифметическое”

$$M_Q X \equiv m_A(X) = EX.$$

Если $Q(x) \equiv 1/x$, то $Q^{-1}(x) \equiv 1/x$ и мы получаем “обобщенное среднее гармоническое”

$$M_Q X \equiv m_H(X) = \frac{1}{E[1/X]}.$$

Если $Q(x) \equiv \log_a x$, то $Q^{-1}(x) \equiv a^x$ и мы получаем “обобщенное среднее геометрическое”

$$M_Q X \equiv m_G(X) = a^{E \log_a X}.$$

Q-средние значения случайных величин. 2

Несложно убедиться, что, если случайная величина X принимает положительные значения $x_1, \dots, x_n$ с равными вероятностями $P(X = x_k) = \frac{1}{n}$, $k = 1, \dots, n$, то мы соответственно получаем традиционные средние арифметическое, гармоническое и геометрическое:

$$m_A(X) = \frac{1}{n} \sum_{k=1}^n x_k, \quad m_H(X) = \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{x_k}}, \quad m_G(X) = \sqrt[n]{x_1 \cdot \dots \cdot x_n}.$$

С помощью неравенства Иенсена мы получаем следующее соотношение между средним арифметическим, средним гармоническим и средним геометрическим:

$$m_H(X) \leq m_G(X) \leq m_A(X).$$

Числовые характеристики формы распределения случайных величин. 1

Пусть X – случайная величина и EX и DX – соответственно ее математическое ожидание и дисперсия. Рассмотрим случайную величину

$$\tilde{X} = \frac{X - EX}{\sqrt{DX}}.$$

Несложно убедиться, что

$$E\tilde{X} = 0, \quad D\tilde{X} = E\tilde{X}^2 = 1.$$

Эти соотношения выполняются для любой случайной величины X , у которой существует (конечна) дисперсия. Однако математические ожидания более высоких степеней случайной величины $\tilde{X}$ уже будут разными для разных случайных величин X и несут в себе некоторую информацию о форме распределения случайной величины X .

Числовые характеристики формы распределения случайных величин. 2

Определение. Предположим, что существует $E|X|^3$. Число

$$\kappa_3 = E\tilde{X}^3 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^3$$

называется *коэффициентом асимметрии* случайной величины X .

Если $\kappa_3 < 0$, то левый “хвост” распределения случайной величины X тяжелее правого. Если при этом случайная величина X непрерывна, то в этом случае левый “склон” ее плотности более пологий, нежели правый. Если $\kappa_3 > 0$, то левый “хвост” распределения случайной величины X легче правого. Если при этом случайная величина X непрерывна, то в этом случае левый “склон” ее плотности более крут, нежели правый. Если случайная величина X симметрична, то есть распределения величин X и $-X$ совпадают, то $\kappa_3 = 0$.

Числовые характеристики формы распределения случайных величин. 3

Определение. Предположим, что существует EX^4 . Число

$$\kappa_4 = E\tilde{X}^4 = E\left(\frac{X - EX}{\sqrt{DX}}\right)^4 = \frac{E(X - EX)^4}{(DX)^2}$$

называется *коэффициентом эксцесса* или *коэффициентом островершинности* случайной величины X .

Чем больше коэффициент островершинности, тем острее вершина плотности соответствующей случайной величины и тяжелее “хвосты” распределения. Наоборот, если коэффициент островершинности мал, то график плотности соответствующей случайной величины имеет тупую вершину, а “хвосты” – легкие.

В качестве эталона “островершинности” традиционно принимают так называемое нормальное распределение, задаваемое плотностью

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty.$$

Коэффициент эксцесса случайной величины, имеющей “эталонную” стандартную нормальную плотность $\varphi(x)$, равен 3.

Ковариация. Коэффициент корреляции. 1

Наряду с числовыми характеристиками центра и разброса можно ввести числовые характеристики взаимозависимости случайных величин.

Определение. Пусть X и Y – случайные величины с конечными вторыми моментами. Число

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)]$$

называется *ковариацией* случайных величин X и Y .

Ковариацию удобно вычислять по формуле

$$\text{cov}(X, Y) = EXY - EX \cdot EY.$$

Ковариация. Коэффициент корреляции. 2

Если ковариация случайных величин X и Y положительна, то в среднем, при многократном воспроизведении стохастической ситуации, отклонение случайной величины X от своего среднего значения в большую сторону влечет одновременное отклонение случайной величины Y от своего среднего значения также в большую сторону и отклонение случайной величины Y от своего среднего значения в большую сторону влечет одновременное отклонение случайной величины X от своего среднего значения также в большую сторону. В то же время отклонение случайной величины X от своего среднего значения в меньшую сторону влечет одновременное отклонение случайной величины Y от своего среднего значения также в меньшую сторону и отклонение случайной величины Y от своего среднего значения в меньшую сторону влечет одновременное отклонение случайной величины X от своего среднего значения также в меньшую сторону. В таком случае говорят, что случайные величины X и Y *положительно коррелированы*.

Если же ковариация случайных величин X и Y отрицательна, то эти случайные величины в среднем, при многократном воспроизведении стохастической ситуации, отклоняются от своих средних значений в разные стороны. В таком случае говорят, что случайные величины X и Y *отрицательно коррелированы*.

Ковариация. Коэффициент корреляции. 3

Если $\text{cov}(X, Y) = 0$, то говорят, что случайные величины X и Y *некоррелированы*.

Некоррелированность является более слабым свойством случайных величин, чем их независимость. Действительно, легко видеть, что, если случайные величины X и Y независимы, то $\text{cov}(X, Y) = 0$.

Однако обратное утверждение неверно: из равенства нулю ковариации случайных величин не следует их независимость. В этом нас убеждает следующий пример.

Пример 2. Рассмотрим случайные величины X и Y , совместное распределение вероятностей которых задается следующим соотношением

$$\begin{aligned} P(X = 0; Y = 1) &= P(X = 0; Y = -1) = \\ &= P(X = -1; Y = 0) = P(X = 1; Y = 0) = \\ &= P(X = \sqrt{2}/2; Y = \sqrt{2}/2) = P(X = -\sqrt{2}/2; Y = \sqrt{2}/2) = \\ &= P(X = \sqrt{2}/2; Y = -\sqrt{2}/2) = P(X = -\sqrt{2}/2; Y = -\sqrt{2}/2) = \frac{1}{8}. \end{aligned}$$

Ковариация. Коэффициент корреляции. 4

Несложно видеть, что при этом

$$\text{cov}(X, Y) = 0,$$

то есть случайные величины X и Y некоррелированы.
Однако, например,

$$0 = P(X = 0; Y = 0) \neq P(X = 0) \cdot P(Y = 0) = \frac{1}{8},$$

то есть случайные величины X и Y не являются независимыми.
Более того, случайные величины X и Y связаны жесткой функциональной зависимостью

$$X^2 + Y^2 = 1.$$

Дисперсия и ковариация

Для дисперсии суммы двух случайных величин справедлива формула

$$D(X \pm Y) = DX + DY \pm 2\text{cov}(X, Y).$$

Отсюда вытекает, что дисперсия суммы *или разности* независимых случайных величин равна *сумме* дисперсий слагаемых.

Ковариация. Коэффициент корреляции. 5

Ковариация не является безразмерной и ее величина зависит от того, какими единицами измеряются рассматриваемые случайные величины. Другими словами, по величине ковариации нельзя судить о том, насколько тесна взаимосвязь (взаимозависимость) случайных величин. Чтобы обойти это препятствие, рассмотрим еще одну характеристику.

Определение. Число

$$\rho = \rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{DX \cdot DY}}$$

называется *коэффициентом корреляции* случайных величин X и Y .

Можно показать, что всегда $|\rho(X, Y)| \leq 1$. Более того, если $|\rho(X, Y)| = 1$, то случайные величины X и Y связаны линейной зависимостью: существуют числа a и b такие, что $P(Y = aX + b) = 1$. При этом, если $\rho = 1$, то $a > 0$, а если $\rho = -1$, то $a < 0$. Если случайные величины X и Y независимы, то, очевидно, $\rho = 0$. Таким образом, можно сказать, что величина коэффициента корреляции характеризует степень *линейной* зависимости.

Ковариация. Коэффициент корреляции. 6

Близость коэффициента корреляции к единице означает лишь наличие *статистической* взаимосвязи, но никак не может быть использована для выявления направления *причинно-следственной* взаимосвязи.

Например, коэффициент корреляции между числом пожаров в наугад выбранном районе большого города за год и числом пожарных частей, расположенном в этом районе, как правило, довольно велик (скажем, для Нью-Йорка он приблизительно равен 0.8). Но это совсем не означает, что чем больше пожарных частей, тем больше пожаров.

Другой **пример**. В Голландии значение коэффициента корреляции между числом новорожденных за год и числом гнезд аистов в соответствующем районе довольно близко к единице (≈ 0.8). Однако мы не можем использовать это обстоятельство для обоснования теории о том, что детей приносят аисты. Объяснение здесь, увы, другое. Дело в том, что, в Голландии вообще рождаемость низка и, как правило, дети рождаются в молодых семьях, а молодые семьи поселяются в отдельных новых домах, на крышах которых и вьют свои гнезда аисты.