В.И. Романовский

Элементарный курс математической статистики

Москва
«Книга по Требованию»

УДК 51
ББК 22.1
В11

В.И. Романовский
В11 Элементарный курс математической статистики / В.И. Романовский – М.: Книга по Требованию, 2024. – 359 с.

ISBN 978-5-458-31219-6

Когда Госпланиздат предложил мне приготовить второе издание моего "Элементарного курса математической статистики", вышедшего в 1924 г., я сильно колебался, следует ли это делать. Действительно, незадолго перед эти предложение вышла моя книга "Математическая статистика" (1938 г.), в которой в числе многих других вопросов были рассмотрены и все, входившие в "Элементарный курс математической статистики".Но новая моя книга "Математической статистики" отличалась от курса, вышедшего в 1924 г., гораздо более сложным математическим аппаратом; в старой книге от читателя требовалось лишь знание элементарной математики, в курсе же "Математической статистики" предполагалось, что читатель владеет математической подготовкой, которую дают первые три курса физико-математических факультетов наших университетов.Кроме того "Математическая статистика" отличается от изданной в 1924 г. и целью, с которой она написана: она ставит задачей введение в теорию современной математической статистики и ознакомление читателя с методами исследования в ней, тогда как старая книга была более скромного и практического направления и имела задачей ознакомление читателя в возможно более простой, но достаточно строгой и обоснованной форме с важнейшими результатами математической статистики и их применениями.

ISBN 978-5-458-31219-6

ЧАСТЬ ПЕРВАЯ

ТЕОРИЯ ВЕРОЯТНОСТЕЙ

ГЛАВА I

ВВЕДЕНИЕ

§ 1. Понятие о статистическом коллективе

В различных научных исследованиях и в практических вопросах часто приходится рассматривать совокупности вещей или явлений, которые по некоторым признакам представляют нечто цельное и единое, а по другим признакам разбиваются на группы или классы определенных численностей. Такие совокупности называются статистическими коллективами (или статистическими совокупностями); вещи или явления, их составляющие, именуются их членами; признаки, по которым коллектив разбивается на классы или группы, называются аргументами коллектива, а число членов в коллективе называется объемом коллектива.

Статистические коллективы могут состоять из вещей или явлений, относящихся к самым разнообразным областям знания. Это можно видеть из примеров статистических коллективов, помещенных в таблице 1.

Таблица 1

Примеры статистических совокупностей

Совокупность	Аргументы	Члены
1. Совокупность умерших в определенной стране за определенный период времени.	Возраст умерших; пол; причины смерти; профессия умершего и т. п.	Каждое отдельное умершее лицо
2. Совокупность рабочих определенного района в определенный момент времени.	Категория труда; возраст; образование; национальность и т. п.	Каждый отдельный рабочий
3. Совокупность браков в определенной стране за определенный период времени.	Возраст вступающих в брак; гражданское состояние до брака; национальность; и т. п.	Каждый отдельный заключенный брак

Совокупность	Аргументы	Члены
4. Заводы какого-либо треста по состоянию их за отчетный год.	Число рабочих, показатели производительности труда, снижения себестоимости и т. п.	Каждый отдельный завод
5. Группа детей, подвергнутых педагогическому исследованию.	Возраст, одаренность; умственное развитие и т. п.	Каждый ребенок группы
6. Скважины определенного нефтяного месторождения по их состоянию на 1-е число данного месяца	Оборудование, дебит нефти, срок эксплоатации и т. д.	Каждая отдельная скважина
7. Совокупность молекул газа, заключенного в некотором сосуде	Скорости молекул; направление скоростей и т. п.	Каждая отдельная молекула
8. Совокупность звезд в мировом пространстве	Спектральные типы звезд; скорость собственного движения их; направление их движений и т. п.	Отдельная звезда
9. Совокупность солнечных пятен за определенный период времени	Число пятен за подразделения этого периода; широта пятен и т. п.	Отдельные солнечные пятна
10. Совокупность электрических лампочек определенной партии и определенного вольтажа	Продолжительность горения лампочки	Отдельная лампочка

Число и разнообразие подобных примеров может быть безгранично увеличено. Нет области явлений или вещей, в которой не могли бы быть образованы статистические совокупности, и на это обстоятельство нужно обратить особое внимание: оно показывает чрезвычайную важность теоретического изучения статистических совокупностей. Действительно, отдельные совокупности образуются и изучаются для исследования различных отдельных групп явлений или вещей, а теоретическое изучение статистических совокупностей дает нам закономерности, которым подчинены все частные статистические совокупности, и методы, при помощи которых мы можем изучать любую статистическую совокупность.

§ 2. Задачи математической статистики

Выше было уже указано, что статистические совокупности могут быть взяты из самых разнообразных областей. Отсюда следует, что математическая статистика может найти себе применение во всевозможных исследованиях. Действительность подтверждает этот вывод. Математическая статистика находит применение не только в ряде социально-экономических дисциплин — политической экономии, демографии, социологии и т. д., где она служит надежнейшим орудием исследования, но и в астрономии (исследования по распределению и движению звезд в небесном пространстве, имеющие огромное значение для космогонии), в физике (кинетическая теория газов, термодинамика, статистическая механика), в биологии (учение об изменчивости, законы наследственности, учение о видах), в прикладной ботанике (растениеводство), в прикладной зоологии (животноводство), в метеорологии (связь различных метеорологических факторов между собой), в гидрологии (прогноз режима рек, взаимная зависимость гидрологических факторов), в психологии (современная экспериментальная психология), в агрономии (исследование сортов культурных растений, влияния различных удобрений), в индустрии и технике (контроль производства, контроль качества изделий, исследование свойств материалов) и т. д. И в последнее время применения математической статистики становятся все более разнообразными и важными.

В чем причины роста применений математической статистики?

Возьмем, например, таблицу, представляющую распределение 100 колхозных участков Майкопской и Белореченской МТС (б. Азово-Черноморского края) по урожайности озимой пшеницы в центнерах с гектара (ц/га) и по срокам уборки по трехдневкам июля 1936 г.

Дни / ц/га	Июль 1936 г.										Итоги по строкам
	до 3 —	6 —	9 —	12 —	15 —	18 —	21 —	24 —	27 —	30	
21	—	—	1	—	—	—	—	—	—	—	1
19	1	—	1	—	2	—	1	—	—	—	5
17	1	1	2	1	1	—	—	—	—	—	6
15	—	1	1	3	3	1	—	1	—	—	10
13	—	—	2	1	3	2	2	2	2	4	18
11	—	2	1	—	4	5	2	4	2	1	21
9	—	1	1	1	—	4	7	3	1	1	19
7	—	—	1	1	—	4	4	2	—	1	13
5	—	—	—	—	—	3	2	—	2	—	7
Итоги по столбцам	2	5	10	7	13	19	18	12	7	7	100

(из материалов Н. К. Дружинина)

Эта таблица обнаруживает такие детали в зависимости между урожаем и сроком уборки, которые без нее могут остаться неизвестными и которые, во всяком случае, очень трудно, если только вообще возможно, получить из анализа весьма сложного комплекса условий, влияющих на урожай и изменяющихся со сроком уборки. Связь между сроком уборки и урожаем во всяком данном случае маскируется множеством случайных отклонений, происходящих от влияния различных обстоятельств. Эти случайные отклонения могут быть исключены только путем увеличения числа аналогичных наблюдений, т. е. чисто статистическим путем. В подобном же положении находится бесчисленное множество других вопросов, которые и разрешаются подобным же образом, при помощи методов статистики.

Кроме задач разработки методов разыскания и исследования закономерностей, математическая статистика имеет еще одну цель, которая часто бывает чрезвычайно важной и которая заключается в описании статистических коллективов при помощи возможно более простых, исчерпывающих и немногих характеристик их. Важность описания коллективов ясна уже из того, что исследование закономерностей в коллективах невозможно без предварительного описания их. Но описание коллективов может быть нужным и важным и само по себе, когда требуется, например, установить распределение заработной платы среди членов определенного рабочего коллектива или определить сортовой состав культурных растений в определенной области и т. п. Таким образом и описательные задачи статистики подтверждают важность статистических методов.

§ 3. Математическая статистика и теория вероятностей

Из определения статистической совокупности, данного в § 1, уже ясно, что, как бы сложны методы статистического исследования ни были и каких бы задач они ни касались, основным фактом, с которым статистику приходится иметь дело, остается всегда то число, которое показывает, сколько раз наблюдалось известное свойство, явление или вещь в некоторой статистической совокупности, иначе говоря, ч а с т о с т ь этого явления, свойства, вещи. И научное и практическое значение статистического метода покоятся затем на том факте, проверяющемся почти для всех статистических совокупностей, что заключения, которые строятся на частостях признаков одной совокупности, оказываются имеющими в большей или меньшей степени силу и для других совокупностей, однородных с изученной. Этот факт дает нам возможность находить закономерности, он позволяет нам статистически предсказывать явления.

При общем — теоретическом — изучении совокупностей весьма важно иметь совокупности, в которых связи между частостями их признаков обусловливались бы простыми, легко обозримыми и прослеживаемыми причинами, а не сложными, запутанными, часто совершенно недоступными анализу комплексами причин, как это бывает для действительно встречающихся совокупностей. Ибо только в таких совокупностях мы можем вполне разобраться, проследить в них влияние причин на частость признаков и добыть, следовательно, надежные общие результаты. Далее, изучая многие такие совокупности, различные по объему, но однородные, мы можем также обнаружить, как и почему констатируется тот факт, о котором мы выше говорили, дающий возможность результаты, полученные для одной совокупности, переносить на другую совокупность, однородную с первой.

Теория вероятностей дает нам возможность построить абстрактные совокупности, прозрачные по своей структуре и представляющие собой отображения реальных совокупностей. В ней изучаются частости, построенные на основании простых и определенных предположений, которые вносят своего рода прозрачность в совокупность и делают ее доступной точному математическому анализу. Только эти теоретически и упрощенно сконструированные частости носят уже название вероятностей.

Из сказанного ясно, что теория вероятностей для математической статистики должна играть основную роль. И действительно, сколько-нибудь сознательное и глубокое изучение современной математической статистики совершенно невозможно без помощи теории вероятностей. Все в математической статистике становится стройным и ясным, когда привлекается теория вероятностей, и остается запутанным и обоснованным лишь эмпирически — без нее. Поэтому мы начнем изложение основ математической статистики с необходимых для нашего элементарного курса сведений из теории вероятностей. Читатели, знакомые с теорией вероятностей, могут пропустить две следующие главы и сразу после § 4 приступить к § 38 „Устойчивость статистических рядов и статистические вероятности", стр. 79.

§ 4. Развитие математической статистики

Математическая статистика занимается теоретическим изучением при помощи математических методов статистических коллективов, рассматриваемых в общем виде.

Первые задачи математической статистики нужно отнести к тому времени, когда нарождалось и развивалось исчисление вероятностей усилиями Паскаля, Якова Бернулли и Лапласа. Ими были заложены те основы этого исчисления, без которых невозможно более или менее сознательное и глубокое изучение вопросов статистики. В этом отношении особенно много сделано Я. Бернулли и Лапласом, работы которых по исчислению вероятностей появились: первого — в начале XVIII века, второго — в конце его и в начале XIX века.

В этих работах мы находим и первые применения теории вероятностей к статистическим вопросам, главным образом к статистике населения: к изучению смертности, брачности, рождаемости и т. п.

Весьма важны для развития математической статистики работы по теории вероятностей русских ученых: академика П. Л. Чебышева (1821—1894), академика А. М. Ляпунова (1857—

1918), академика А. А. Маркова (1856—1922) и наших современников: академика С. Н. Бернштейна, академика А. Н. Колмогорова, профессора Е. Е. Слуцкого, профессора А. Я. Хинчина и других.

К первым работам по математической статистике могут быть причислены работы по теории ошибок Лапласа, Лежандра и Гаусса. В теории ошибок совокупность измерений с их случайными погрешностями в каждом отдельном измерении исследуется при помощи теории вероятностей. Лучший из учеников Лапласа Пуассон, много работавший в области теории вероятностей, ввел в науку термин „Закон больших чисел" и дал математические основы стрельбы (внешняя баллистика в артиллерии). Дальнейшие применения теории вероятностей к изучению массовых явлений были сделаны другим учеником Лапласа—Адольфом Кетле (1796—1874).

Кетле применял теорию вероятностей к изучению разнообразных социальных явлений и дал математические основы теории статистики.

Среди первых крупнейших работ по математической статистике надо отметить работы английского ученого Френсиса Гальтона (1822—1911). Его важнейшим вкладом в математическую статистику являются первые основания теории корреляции.

Современная математическая статистика развивалась, главным образом, в Англии. Среди других стран, где также шло развитие математической статистики, можно назвать лишь скандинавские страны и СССР. Среди скандинавских ученых следует особенно отметить шведского астронома Шарлье, занимавшегося применениями математической статистики к звездной астрономии и опубликовавшего важные работы в области теоретической статистики. В США, Франции и в других странах в последнее время также замечается интерес к математической статистике и вышел ряд важных работ по математической статистике. Но ведущая роль в математической статистике принадлежит Англии и СССР.

ГЛАВА II

ОСНОВНЫЕ ПОНЯТИЯ И ТЕОРЕМЫ ТЕОРИИ ВЕРОЯТНОСТЕЙ

§ 5. Определения

Предположим, что рассматривается совокупность некоторых определенных явлений, признаков или событий

$$a_1, a_2, \ldots, a_n, \tag{5.1}$$

число которых конечно и каждое из которых может осуществиться при выполнении некоторых условий C, характеризуемых определенными общими признаками и могущих повторяться неограниченное число раз. Каждое отдельное осуществление условий C, приводящих к тому или иному из событий a_i ряда (5.1), мы будем называть испытанием или, точнее, испытанием класса C, подразумевая под классом C совокупность всех возможных условий C.

Например, бросание в воздух игральной кубической кости и падение ее на горизонтальную плоскость, неограниченно повторяемые с одной и той же костью, можно считать условиями C, и тогда номера 1, 2, 3, 4, 5, 6, нанесенные на гранях кости и выпадающие в результате каждого бросания, будут представлять события a_i.

Испытания класса C мы будем называть независимыми стохастическими испытаниями или просто стохастическими (т. е. вероятностными), если они удовлетворяют по отношению к событиям a_i следующим условиям:

1. События a_i для испытаний класса C единственно возможны, т. е. таковы, что всякое испытание класса C непременно приводит к осуществлению одного из событий a_i.

2. События a_i для испытаний класса C несовместимы, т. е. в каждом отдельном испытании класса C появление одного из событий a_i исключает возможность появления любого другого из них.

3. События a_i в условиях C представляют случайные события, т. е. испытания C таковы, что для любого из них мы не можем точно предсказать результат его, пока оно не будет осуществлено.

4. События a_i равновозможны для испытаний класса C, т. е. испытания C таковы, что при достаточно большом числе повторений их обеспечено приблизительно одинаково частое появление событий a_i.

О равновозможности событий a_i вообще мы можем судить только по числу их появлений в достаточно большом числе испытаний. Но иногда о равновозможности можно судить и не прибегая к опыту, исходя лишь из общих свойств условий C. Например, если бросается геометрически правильная и материально однородная игральная кость и бросается так, что она делает большое непредвычислимое число оборотов в воздухе, то выпадение тех или иных из ее граней мы можем считать равновозможными случаями и до опыта, априори. И действительно, если читатель попробует при таких условиях бросить кость, скажем, 600 раз, то он убедится, что каждая грань выпадет приблизительно 100 раз.

Важно заметить, что равновозможность случаев a_i представляет объективное свойство рассматриваемого класса испытаний, вытекающее из их реальных свойств, а не из наших субъективных представлений о них. Следует также заметить, что равновозможность, как всякое реальное свойство вещей или явлений, может быть установлена только с известной степенью точности, а не абсолютно точно.

Предположим теперь, что при осуществлении одного из событий или случаев:

$$a_1, \ a_2, \ldots, \ a_m, \tag{5.2}$$

составляющих часть ряда событий (5.1), принадлежащих к стохастическим испытаниям класса C, всегда появляется некоторое определенное событие A, а при осуществлении одного из прочих событий или случаев:

$$a_{m+1}, \ a_{m+2}, \ldots, \ a_n \tag{5.3}$$

ряда (5.1) событие A не появляется. Тогда мы будем говорить, что случаи (5.2) благоприятствуют событию A, а случаи (5.3) ему не благоприятствуют и назовем вероятностью события A число $\frac{m}{n}$.

Таким образом мы приходим к следующему определению.

В е р о я т н о с т ь ю н е к о т о р о г о с о б ы т и я A в н е к о т о р о м к л а с с е с т о х а с т и ч е с к и х и с п ы т а н и й м ы н а з о в е м о т н о ш е н и е ч и с л а с л у ч а е в, б л а г о п р и я т с т в у ю щ и х A, к ч и с л у в с е х с л у ч а е в, п р и н а д л е ж а щ и х к э т о м у к л а с с у и с п ы т а н и й.

Например, при бросании игральной кости, как было описано выше, вероятность выпадения одной из граней с четным номером равна $\frac{3}{6}$, так как из 6 случаев различных возможных номеров три номера — 2, 4, 6 — четны.

Согласно данному определению каждый из случаев (5.1) имеет вероятность $\frac{1}{n}$. Следовательно, равновозможные случаи, принадлежащие к стохастическому классу испытаний, равновероятны. Не следует думать, что равновозможность и равновероятность случаев (5.1) представляет одно и то же. Равновозможность есть реальное свойство случаев (5.1) в действительных испытаниях, проверяемое, как и всякое реальное свойство, только более или менее точно, тогда как равновероятность их есть абстракция равновозможности и определяется совершенно точно: одним и тем же числом $\frac{1}{n}$.

Вероятность события A обозначается знаком $P(A)$, так что

$$P(A)=\frac{m}{n}.$$

§ 6. Вероятности, равные нулю и единице. Физический закон больших чисел

Если все из возможных для данного класса стохастических испытаний случаев благоприятствуют событию A, то оно будет непременно появляться в каждом испытании и, следовательно, будет достоверным. Вместе с тем его вероятность равна $\frac{n}{n}=1$. Таким образом достоверное событие имеет вероятность, равную единице.

Обратно, если вероятность события A равна 1, то оно достоверно при тех предположениях, которые рассматривались в предыдущем параграфе, так как тогда все случаи (5.1) должны ему благоприятствовать.

Если из всех единственно возможных случаев ни один не благоприятствует событию A, то событие A в рассматриваемых испытаниях не может появиться, т. е. будет невозможным. Вместе с тем его вероятность равна $\frac{0}{n}=0$. Следовательно, невозможное событие имеет вероятностью нуль.

Обратно, если событие A имеет вероятность 0, то оно невозможно, так как тогда среди случаев (5.1) нет ему благоприятствующих.

Отметим теперь следующий чрезвычайно важный факт. Когда производится возрастающее число испытаний стохастического класса C, тогда, в силу равновозможности случаев (5.1), эти случаи должны появляться приблизительно одинаково часто. Отсюда тотчас вытекает, что в большем числе N испытаний событие A появится приблизительно

$$M\approx N\cdot\frac{m}{n}$$

раз. Число $\frac{M}{N}$ называется относительной частостью события A или просто частостью его и представляет отношение числа появлений A к общему числу испытаний. Стало быть, мы можем также утверждать, что при большом числе испытаний относительная частость события A будет приближенно равна его вероятности:

$$\frac{M}{N} \approx \frac{m}{n}.$$

Это равенство вытекает из написанного выше.

Оно подвергалось многочисленным и разнообразным проверкам, о некоторых из которых будет рассказано далее. Выражаемое им соотношение между частостью и вероятностью мы назовем физическим законом больших чисел. На нем, в конечном счете, покоятся все применения теории вероятностей к реальному миру.

В частности, из него следует, что осуществление события с вероятностью, очень близкой к единице, можно ожидать почти во всех испытаниях, так что оно может быть рассматриваемо как практически достоверное. Наоборот, событие с очень малой вероятностью, т. е. очень близкой к нулю, осуществляется очень редко и потому его практически можно считать невозможным. Этими положениями часто приходится руководствоваться в статистических исследованиях.

§ 7. Теорема сложения вероятностей

Пусть при некотором испытании возможно появление одного из ряда несовместимых событий. Тогда можно высказать следующую теорему.

Теорема. Вероятность появления одного из несовместимых событий без указания, какого именно, равна сумме вероятностей этих событий.

Будем, для простоты, рассматривать три несовместимых события A, B, C. Пусть из всех n единственно возможных, несовместимых и равновозможных случаев случаи

$a_1, a_2, \ldots, a_k$ благоприятствуют A,
$b_1, b_2, \ldots, b_l$ » B,
$c_1, c_2, \ldots, c_m$ » C,

так что

$$P(A) = \frac{k}{n}, \quad P(B) = \frac{l}{n}, \quad P(C) = \frac{m}{n}.$$

Ясно, что, соединив все случаи, благоприятствующие A, с случаями, благоприятствующими B, и с случаями, благоприятствующими C, мы получим случаи, благоприятствующие или A,