

В. В. НАЛИМОИ

**ПРИМЕНЕНИЕ
МАТЕМАТИЧЕСКОЙ
СТАТИСТИКИ
ПРИ АНАЛИЗЕ
ВЕЩЕСТВА**



ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1960

АННОТАЦИЯ

В книге рассматривается применение аппарата математической статистики в химических и физических методах анализа вещества. Изложение материала иллюстрируется многочисленными примерами, доведенными до численных расчетов. Большое внимание уделяется физической интерпретации результатов статистических исследований. Подробно освещен опыт зарубежных работ в этой области.

Книга предназначена для инженеров-физиков и химиков, работающих в аналитических лабораториях. Она может служить настольным пособием по применению математической статистики при анализе вещества.

ОГЛАВЛЕНИЕ

Предисловие	7
Глава I. Задачи математической статистики	11
Глава II. Классификация аналитических ошибок . .	18
Глава III. Случайная величина и ее характеристики .	34
§ 1. Распределение случайной величины	34
§ 2. Среднее значение случайной величины и дисперсия Среднее значение случайной величины (39). Дисперсия и средняя квадратичная ошибка (42).	38
§ 3. Вычисление дисперсий по текущим измерениям . .	49
§ 4. Закон сложения ошибок	52
§ 5. Ошибки косвенных измерений Абсорбционный спектральный анализ (61). Эмиссионный спектральный анализ (63).	60
Глава IV. Нормальное распределение	68
§ 1. Функция нормального распределения	68
§ 2. Некоторые специальные распределения, связанные с нормальным распределением <i>t</i> -распределение (79). χ^2 -распределение (89). <i>F</i> -распреде- ление (93). <i>g</i> -распределение (96).	78
§ 3. Критерии для оценки степени близости наблюдае- мого распределения к нормальному распределению Оценка с помощью χ^2 -критерия (99). Оценка с помощью λ -критерия (107). Проверка гипотезы нормальности по большому числу малых выборок (112). Метод спрямлен- ных диаграмм (118).	98
§ 4. Отклонения от нормального распределения в анали- тической работе	122
Глава V. Распределение Пуассона и биномиальное распределение	135
§ 1. Распределение Пуассона	135
§ 2. Оценка результатов полуколичественных определе- ний при помощи распределения Пуассона	145
§ 3. Биномиальное распределение	153

Глава VI. Оценка результатов анализа	159
§ 1. Сравнение двух средних с помощью t -критерия	159
§ 2. Сравнение нескольких дисперсий	164
§ 3. Проверка гипотезы однородности результатов измерений. Оценка резко выделяющихся определений	168
§ 4. Секвенциальный (последовательный) анализ	177
§ 5. Непараметрическая статистика	185
Проверка гипотезы о наличии постоянного расхождения в результатах наблюдений (186). Проверка гипотезы о случайном характере флуктуаций (189). Метрологические оценки на основании неравенства Чебышева (193).	
Глава VII. Дисперсионный анализ	197
§ 1. Определение дисперсии, обусловленной действием одного фактора	197
Идея метода и простейшие примеры (197). Дисперсионный анализ при неравных столбцах (205). Условия, при которых можно применять дисперсионный анализ (207). Пример применения дисперсионного анализа при изучении методических ошибок (209).	
§ 2. Многоступенчатая классификация	214
Двухступенчатая группировка (214). Трехступенчатая группировка (221).	
§ 3. Комплексный опыт	227
Двухсторонняя классификация (227). Трехсторонняя классификация (242).	
§ 4. Планирование эксперимента по методу латинского квадрата	253
§ 5. Эффективность дисперсионного анализа	256
Глава VIII. Статистика линейных связей	258
§ 1. Способ наименьших квадратов	258
Определение параметров градуировочного графика (258). Среднее взвешенное неравноточных измерений (267).	
§ 2. Регрессионный анализ	268
Проверка гипотезы линейности (268). Сравнение параметров градуировочных графиков с теоретически ожидаемыми значениями (272). Сравнение двух градуировочных графиков (284). Проверка гипотезы о параллельном смещении градуировочных графиков (290). Оценка результатов анализа, полученных с помощью градуировочного графика (293).	
§ 3. Корреляционный анализ	298
§ 4. Пример комплексного планирования эксперимента с применением различных статистических методов анализа	314

Глава IX. Некоторые приемы работы, связанные со статистическим планированием эксперимента 325

- § 1. Отбор проб и рандомизация условий эксперимента.
Применение таблицы случайных чисел 325
- § 2. Выбор числа параллельных определений 328
- § 3. Документация материала 339
- § 4. Контрольные диаграммы 357

Приложение

- Таблица 1. Значения функции $\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{u^2}{2}} du$ 362
- Таблица 2. Значения удвоенной нормированной функции Лапласа $2\Phi(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-\frac{u^2}{2}} du$ 366
- Таблица 3. Значения t для различных уровней значимости 367
- Таблица 4. Значения χ_p^2 в зависимости от вероятности $P(\chi^2 > \chi_p^2)$ и числа степеней свободы χ^2 -распределения 368
- Таблица 5. Вероятность $P(\chi^2 > \chi_p^2)$ в зависимости от значения χ_p^2 и числа степеней свободы χ^2 -распределения 370
- Таблица 6. Значения F для различных уровней значимости 374
- Таблица 7. Значения r для различных уровней значимости 378
- Таблица 7А. Значения $r_{\max}$ (или $r_{\min}$) для различных уровней значимости 378
- Таблица 8. Значения коэффициента корреляции r_{xy} для различных уровней значимости 379
- Таблица 9. Вероятности $P(\lambda)$ для различных значений λ 380
- Таблица 10. Вероятности появления значений r , заключенных в интервале шириной 0,1 380
- Таблица 11. Число испытаний с менее часто встречающимся знаком для различных уровней значимости при различном числе наблюдений n 382
- Таблица 12. Значения верхних и нижних пределов общего числа серий R для различного числа наблюдений n 383
- Таблица 13. Пяти- и однопроцентные пределы для отношения наибольшей эмпирической дисперсии к сумме k эмпирических дисперсий 384

ОГЛАВЛЕНИЕ

Таблица 14. Случайные числа	386
Таблица 15. Квадраты трехзначных чисел	397
Л и т е р а т у р а	401
I. Основные руководства и монографии по математической статистике и теории вероятностей	401
II. Обзоры и библиографические источники по вопросам, связанным с применением математической статистики при анализе вещества	407
III. Статьи по вопросам, связанным с применением математической статистики при анализе вещества, опубликованные в периодической печати и сборниках	408
А л ф а в и т н ы й у к а з а т е л ь	425

ПРЕДИСЛОВИЕ

Целью настоящей книги является изложение методов математической статистики в применении к задачам, связанным с анализом вещества.

В конце XIX и начале XX века на базе теории вероятностей началось создание современной математической статистики в связи с запросами биологии и экономики. За последние десятилетия математическая статистика как метод исследования стала интенсивно применяться в таких областях науки и техники, как агробиология, медицина, машиностроение и приборостроение, химическая промышленность, металлургия и др. Особенно интенсивное развитие статистических методов исследования наблюдается в последние годы. Совсем недавно на основе теории вероятностей создана совершенно новая дисциплина—теория информации, первоначальной задачей которой было изучение вопросов, связанных с передачей сигналов в радиотехнике. На базе теории информации стала развиваться кибернетика—наука об управлении. Совершенно неожиданно теория информации нашла применение в оптике. Весьма перспективным представляется сейчас применение идей теории информации при документации научных и технических материалов. В связи с интенсивным развитием ядерной физики появилась новая область применения теории вероятностей—статистика счета ядерных частиц.

Одной из новых областей применения математической статистики являются исследования, связанные с анализом вещества. Необходимость применения статистических методов при анализе вещества обуславливается рядом факторов. Здесь надо прежде всего указать на то, что внедрение в производство новых сложных по своему

составу сплавов, материалов и непрерывное ускорение процесса производства заставили широко применять новые физические методы анализа, основанные на протекании мало изученных процессов, не поддающихся строгому контролю и точному регулированию. Наличие множества новых аналитических методов наряду со старыми классическими методами анализа остро ставит вопрос об отыскании разумных критериев для сравнения результатов анализа, полученных разными методами.

Развитие и внедрение новых аналитических методов происходит значительно быстрее, чем их стандартизация. Это неизбежно приводит к тому, что в каждой, даже небольшой, аналитической лаборатории приходится постоянно сталкиваться со сложными метрологическими проблемами, рациональное решение которых невозможно без применения методов современной математической статистики. Уже сейчас стало ясно, что аналитик должен так же хорошо владеть методами современной математической статистики, как геодезист владеет методом наименьших квадратов.

Каждая новая область применения математической статистики требует своего особого методического подхода. Опыт, полученный при статистических исследованиях в одной области, нельзя механически переносить на соседние, даже, казалось бы, близкие области. В частности, например, математическая теория ошибок, разработанная, исходя из задач метрологии и геодезии, не может быть без существенного видоизменения перенесена в область аналитической химии. Поэтому наряду с руководствами общего характера по математической статистике появилась необходимость в специализированных руководствах, рассчитанных на работников данной узкой области. Большой опыт в издании специализированных руководств накопился за рубежом, где вместе с выпуском значительного количества пособий общего характера, посвященных применению математической статистики в исследовательских работах, появился ряд специальных руководств по применению статистических методов исследования в химии.

В книге делается попытка систематизировать и обобщить те работы, которые появились за последние годы

в области статистических исследований, связанных с анализом вещества. Мы полагаем, что в ближайшем будущем можно будет построить общую теорию анализа вещества на базе теории вероятностей так, как в свое время удалось создать общую теорию измерений—метрологию. В настоящее время создание такой теории находится еще в зачаточном состоянии. Первым и наиболее трудным шагом в построении этой теории является формулировка на математическом языке всех задач, связанных с анализом вещества.

Книга написана преимущественно в рецептурном плане. Основные положения математической статистики не доказываются, а разъясняются на примерах, заимствованных из работ, связанных с анализом вещества. Теоретические вопросы рассматриваются только в той мере, в какой это необходимо для понимания метрологической стороны рассматриваемой проблемы.

Для чтения книги необходимо знание математического аппарата в объеме втузовского курса и знакомство с основами теории вероятностей. Главы, посвященные общим вопросам теории вероятностей, ставят задачей напомнить читателю основные положения этой дисциплины.

Книга не претендует на систематическое и полное изложение идей современной математической статистики. В ней рассматриваются только те статистические методы, которые уже нашли применение в лабораторной работе при анализе вещества и были освещены в периодической печати.

Расположение материала в книге подчинено развитию основных идей статистического анализа, которые иллюстрируются примерами их применения при анализе вещества. При таком расположении материала к решению задач, связанных с одим и тем же видом аналитических работ, приходится возвращаться несколько раз. Тем не менее такое изложение является с нашей точки зрения вполне оправданным, поскольку при решении задач, связанных с одними и теми же видами аналитической работы, приходится пользоваться как некоторыми очень простыми статистическими методами, так и сложными приемами, понимание которых возможно только на базе изучения предыдущего материала.

В конце книги дается аннотированный указатель основных руководств по математической статистике и журнальных статей, посвященных статистическим исследованиям проблем, связанных с анализом вещества.

Мы надеемся, что издание этой книги будет способствовать дальнейшему развитию и внедрению статистических методов исследования в аналитических лабораториях.

Автор заранее выражает благодарность как за замечания принципиального характера, так и за указания на возможные опечатки.

В. Налимов

ГЛАВА I

ЗАДАЧИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Математическая статистика—наука, опирающаяся на теорию вероятностей и использующая результаты опыта для изучения объективных закономерностей исследуемых явлений.

Это определение математической статистики носит весьма общий характер, обусловливаемый тем, что математическая статистика находит применение в самых разнообразных областях науки и техники. Применение математической статистики в какой-либо одной научной дисциплине всегда связано с преимущественным использованием определенных ее аспектов. В лабораторной работе и, в частности, при анализе вещества математическая статистика используется преимущественно для свертывания (сокращения) и анализа экспериментального материала методами, основанными на теории вероятностей. Объясняется это тем, что в исследовательских работах приходится иметь дело с действием и взаимодействием большого числа факторов, трудно поддающихся учету, поэтому постановка одной серии экспериментов обычно не дает возможности обнаружить действующие здесь физические закономерности. Эти закономерности могут быть выявлены только при сравнении результатов исследований, выполненных над различными объектами в различных условиях и разных лабораториях. Такое сравнение становится возможным только в том случае, если результаты опытов с помощью математической статистики представляются в компактной форме, удобной для хранения, передачи и дальнейшей обработки. Свертка (сокращение) информации, в частности, заключается, например, в том, что с помощью аппарата математической статистики всю

информацию о точности аналитического метода можно представить в виде функции (закона) распределения ошибок этого метода, характеризующегося параметрами распределений: дисперсией или средним квадратичным отклонением и математическим ожиданием*).

В аналитической работе часто приходится ограничиваться сравнительно небольшим числом определений. Это небольшое количество наблюдаемых величин можно рассматривать как случайную выборку из некоторого гипотетического бесконечного множества—генеральной совокупности, которая является математической моделью реально наблюдаемых величин. Задача свертывания информации с математической точки зрения сводится в этом случае к тому, что по выборке определяют некоторые величины (выборочную дисперсию и среднее арифметическое значение случайной величины), которые являются оценкой неизвестных параметров (соответственно дисперсии и математического ожидания) функции распределения этой генеральной совокупности. При оценке (определении) параметров генеральной совокупности по выборке, естественно, вносится известный элемент неопределенности, который можно учесть методами математической статистики. Среди экспериментаторов распространено совершенно неправильное мнение о том, что математическая статистика применима только к большому цифровому материалу. Современная математическая статистика дает возможность оценивать параметры генеральных совокупностей и устанавливать для них доверительные пределы даже по весьма малым выборкам,—в некоторых случаях всего по двум измерениям. Но при этом, естественно, что чем меньше экспериментальный материал, тем менее точно может быть произведена оценка параметров генеральной совокупности по их выборочным значениям. Таким образом, математическая статистика, с одной стороны, дает возможность компактным образом представить результаты эксперимента, а с другой стороны, позволяет количественно оценить тот элемент сомнения, который сопутствует каждому эксперименту при малом числе опытов.

*) Определение этих понятий см. в гл. III.

Вопрос о разработке и стандартизации достаточно надежных и удобных способов свертывания информации становится особенно актуальным в последние годы в связи со все увеличивающимся числом исследовательских работ во всех областях науки. На рис. 1 приведена кривая,

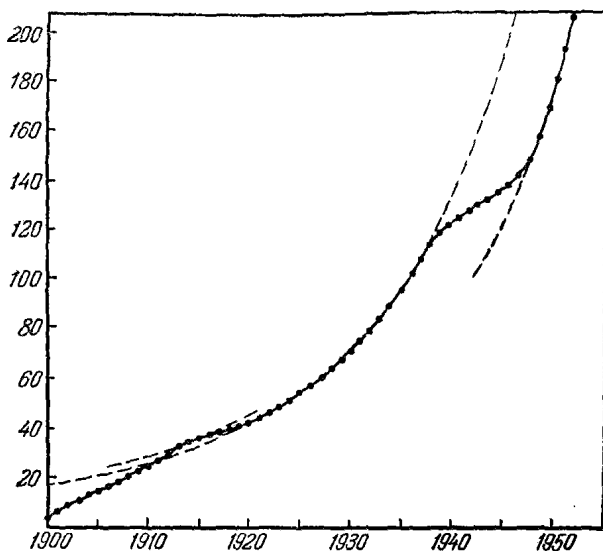


Рис. 1. Рост суммарного числа публикаций в реферативном журнале «Physics Abstracts», начиная с 1900 г. [162].

По оси ординат — накопленные суммы в тысячах, по оси абсцисс — годы.

показывающая рост числа публикаций в реферативном журнале «Physics Abstracts», начиная с 1900 г. [162]. На протяжении почти 50 лет удается проследить экспоненциальный ход кривой*), который нарушился только

*) Экспоненциальный характер роста наблюдается для числа научных журналов, количества работников, занятых на исследовательских работах, и для ассигнований на эти работы. Общее число журналов в настоящее время достигает 100 000, количество реферативных журналов близко к 300.

Во всех случаях постоянная экспонент такова, что показатели развития науки удваиваются за каждые 10—15 лет. Экстра-

в годы войны. Каждые 10—15 лет число публикаций удваивается. Увеличение числа публикаций неизбежно приводит к сокращению среднего объема публикации. Некоторые журналы за последние годы сократили объем публикуемых статей почти в два раза, в дальнейшем надо предвидеть еще большее сокращение объема публикуемых работ. Произвольное сокращение статей редакторами журналов приводит к потере значительной части информации, содержащейся в экспериментальном материале. Если раньше читатель имел возможность оценить степень достоверности результатов на основании пространственных описаний как условий эксперимента, так и способов его обработки, то теперь эта возможность исключается, и с особой остротой возникает необходимость в компактной количественной оценке того элемента неопределенности, который связан с экспериментом.

Сведения о точности и правильности анализа, приводимые в разного рода рабочих инструкциях и аналитических

полюция в историческое прошлое приводит к значению, равному единице в 1700 г., т. е. в эпоху Ньютона, которая может считаться началом современного периода развития науки. Таким образом, экспоненциальный характер развития науки, по-видимому, выполнялся в течение последних 200—250 лет.

Экспоненциальный характер развития науки можно вывести из весьма вероятного предположения достаточно общего характера. Аналитическое выражение для экспоненты

$$y = ae^{kt}, \quad k > 0,$$

можно рассматривать как решение дифференциального уравнения

$$\frac{dy}{dt} = ky,$$

где производная $\frac{dy}{dt}$ означает скорость роста интересующих нас показателей, т. е. увеличение их за единицу времени. Таким образом, экспоненциальный характер развития является следствием весьма вероятного предположения о том, что относительная скорость роста

$$\frac{dy}{y dt} = \text{const}$$

является постоянной величиной. Нетрудно видеть, что удвоению всех показателей за 10—15 лет соответствует относительная скорость роста в 5—7% в год [172].

ГОСТ'ах, и данные, характеризующие эталоны и стандартные образцы (однородность эталонов, достоверность, с которой установлено среднее содержание вещества в них и пр.), также надо рассматривать как свертывание информации, содержащейся в тех больших экспериментальных работах, которые проводились при изучении данного вопроса.

Наконец, результаты текущих анализов также можно рассматривать как свертывание той информации, которая была получена в результате нескольких параллельных определений пробы неизвестного состава, а в ряде случаев еще и контрольных анализов эталонов или стандартных образцов.

Таким образом, можно утверждать, что свертывание информации при помощи математической статистики должно явиться составной частью любого аналитического процесса. Анализ вещества может считаться законченным только тогда, когда материал представлен компактным образом и произведена оценка надежности полученных данных.

Идея применения математической статистики для свертывания информации была выдвинута Р. Фишером [28а]. Основатели англо-американской статистической школы Р. Фишер и К. Пирсон, как известно, стояли на махистских позициях, и в их работах свертка информации часто превращалась в самоцель. Такая интерпретация одной из задач математической статистики, конечно, не может быть признана правильной, она противоречит позиции советской статистической школы, которая считает главной задачей математической статистики выявление объективных закономерностей.

Свертывание информации в действительности представляет собой не самоцель, а одну из составных частей сложного процесса познания объективно существующих закономерностей. Поясним это следующим примером: такой важный для эмиссионного спектрального анализа вопрос, как влияние «третьих элементов», освещен в той или иной степени в сотнях работ, причем каждая из работ рассматривает этот вопрос для какого-то частного случая, используя ограниченный экспериментальный материал. Ни одна из этих работ, взятая в отдельности, не содержит достаточно данных для того, чтобы можно было проникнуть

в природу этого сложного явления. Если бы результаты всех работ были свернуты при помощи математической статистики и представлены некоторым стандартным образом, удобным для сопоставления и дальнейшей обработки, то, по-видимому, можно было бы, обобщая весь материал в целом, приблизиться к пониманию природы этого сложного явления, с одной стороны, а с другой стороны, сделать практически важные выводы для разработки аналитических методов.

Таким образом если мы воспользуемся математической статистикой для свертки информации, то это даст возможность рассматривать отдельные работы как части одного большого коллективного эксперимента, а сам процесс свертки рассматривать как некоторую составную часть процесса познания, которая становится особенно актуальной сейчас, в связи со все увеличивающимся числом отдельных исследований частного значения, нуждающихся в дальнейшем обобщении. При такой постановке вопроса интересующие нас аспекты математической статистики могут рассматриваться и как задачи кибернетики—дисциплины, которая занимается прежде всего проблемами свертки и переработки информации.

Более сложной является вторая из интересующих нас задач математической статистики—анализ экспериментального материала. Обе задачи математической статистики—анализ экспериментального материала и компактное представление полученных при этом результатов—оказываются органически связанными между собой.

Целью статистического анализа является, с одной стороны, получение максимальной информации при минимальной затрате труда на проведение экспериментальных работ, с другой стороны—оценка достоверности полученных результатов.

К статистическому анализу исследователю приходится обращаться каждый раз, когда он, пользуясь ограниченным экспериментальным материалом, хочет трезво оценить роль того или иного изучаемого фактора. Статистический анализ значительно усложняется, если приходится изучать одновременное действие нескольких факторов и их взаимодействие. Благодаря развитию статистических методов анализа в ряде случаев оказалось возможным