

УДК 519.2:303.732.4

UDC 519.2:303.732.4

**ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ
МЕТОДЫ В РАБОТАХ А.Н. КОЛМОГОРОВА****PROBABILISTIC-STATISTICAL METHODS
IN KOLMOGOROV'S RESEARCHES**

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor
Bauman Moscow State Technical University,
Moscow, Russia

*Московский государственный технический
университет им. Н.Э. Баумана, Россия, 105005,
Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru*

С современной точки зрения рассмотрены работы А.Н. Колмогорова по аксиоматическому подходу к теории вероятностей, критерию согласия эмпирического распределения с теоретическим, свойствам медианы как оценки центра распределения, эффекту «вздувания» коэффициента корреляции, теории средних величин, статистической теории кристаллизации металлов, методу наименьших квадратов, свойствам сумм случайного числа случайных слагаемых, статистическому контролю, несмещенным оценкам, аксиоматическому получению логарифмически нормального закона распределения при дроблении, методам обнаружения различий при экспериментах типа погодных

From a modern point of view we have discussed Kolmogorov's researches in the axiomatic approach to probability theory, the goodness-of-fit test of the empirical distribution with theoretical, properties of the median estimates as a distribution center, the effect of "swelling" of the correlation coefficient, the theory of averages, the statistical theory of crystallization of metals, the least squares method, the properties of sums of a random number of random variables, statistical control, unbiased estimates, axiomatic conclusion of logarithmic normal distribution in crushing, the methods of detecting differences in the weather-type experiments

Ключевые слова: СТАТИСТИЧЕСКИЕ МЕТОДЫ, ИСТОРИЯ СТАТИСТИКИ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, А.Н. КОЛМОГОРОВ, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ПРИКЛАДНАЯ СТАТИСТИКА, СТАТИСТИКА В РОССИИ

Keywords: STATISTICAL METHODS, HISTORY OF STATISTICS, MATHEMATICAL STATISTICS, A.N. KOLMOGOROV, NONPARAMETRIC STATISTICS, APPLIED STATISTICS, STATISTICS IN RUSSIA

1. Введение

В нашей стране развитием и применением статистических методов в XX – XXI вв. занимались десятки тысяч специалистов. Среди них выделяется гигантская фигура Андрея Николаевича Колмогорова. В развитии теории вероятностей и математической статистики четко виден переход от предыстории к современности, и этот переход – работы академика АН СССР А.Н. Колмогорова (1903 – 1987). Именно его работы дали первоначальный толчок дальнейшему развитию ряда направлений, важных для современных статистических методов. Зачастую еще 60 – 80 лет назад А.Н. Колмогоров рассматривал те проблемы, которые только сейчас начинают широко обсуждаться. Поразительно, но длинный ряд

современных публикаций с научной точки зрения являются устаревшими по сравнению с работами А.Н. Колмогорова 1930-х годов.

Предыстории статистических методов посвящена статья [1]. В настоящей публикации с современной точки зрения [2] рассмотрим вероятностно-статистические методы исследования, созданные А.Н. Колмогоровым. Обсудим работы А.Н. Колмогорова по аксиоматическому подходу к теории вероятностей, критерию согласия эмпирического распределения с теоретическим, свойствам медианы как оценки центра распределения, эффекту «вздувания» коэффициента корреляции, теории средних величин, статистической теории кристаллизации металлов, методу наименьших квадратов, свойствам сумм случайного числа случайных слагаемых, статистическому контролю, несмещенным оценкам, аксиоматическому получению логарифмически нормального закона распределения при дроблении, методам обнаружения различий при экспериментах типа погодных.

Факты жизни и творчества А.Н. Колмогорова подробно рассмотрены в сборнике [3]. Его основные работы изданы в трех томах [4 - 6]. Работы отобраны им самим и прокомментированы его учениками. Андрей Николаевич считал, что хорошая математическая работа должна содержать простую идею (желательно геометрического характера) и использовать «тонкую» аналитику, а хорошая и полезная прикладная работа должна опираться на фундаментальные теоретические основы.

2. Аксиоматический подход к теории вероятностей

Он позволил рассматривать теорию вероятностей и математическую статистику как часть математики, проводить рассуждения на математическом уровне строгости. В частности, было введено четкое различие между частотой и вероятностью, случайная величина стала рассматриваться как функция от элементарного исхода, и т.д. За основу

методов статистического анализа данных стало возможным брать вероятностно-статистические модели, сформулированные в математических терминах. В результате удалось четко отделить строгие утверждения от обсуждения философских вопросов случайности, преодолеть подход на основе понятия равновозможности, имеющий ограниченное практическое значение. Наиболее существенно, что после работ А.Н. Колмогорова нет необходимости связывать вероятности тех или иных событий с пределами частот или иными методами оценок вероятностей. В частности, так называемые «субъективные вероятности» имеют смысл экспертных оценок вероятностей.

После выхода (в 1933 г. на немецком языке и в 1936 г. – на русском) основополагающей монографии [7] аксиоматический подход к теории вероятностей стал общепринятым в научных исследованиях в этой области. Во многом перестроилось преподавание. Повысился научный уровень многих прикладных работ. Однако традиционный подход оказался на удивление живучим. С целью повышения строгости формулировок приходится помещать в наших учебниках ([8], [9] и др.) сводки терминов и определений в области вероятностно-статистических методов, опирающаяся на аксиоматику [7].

В послевоенные годы А.Н. Колмогоров формализовал понятие случайности на основе теории информации [6]. Грубо говоря, числовая последовательность является случайной, если ее нельзя заметно сжать (т.е. описать существенно короче) без потери информации. Однако этот подход не был предназначен для использования в прикладных работах и преподавании. Он представляет собой важное методологическое и теоретическое продвижение.

3. Критерии согласия

В работе 1933 г. «Об эмпирическом определении закона распределения» [5, с.134-141] А.Н. Колмогоров предложил и изучил «критерий Колмогорова», предназначенный для проверки согласия эмпирического распределения с полностью известным теоретическим. Пусть элементы выборки (независимые случайные величины) объема n имеют непрерывную функцию распределения $F(x)$. Эмпирической функцией распределения $F_n(x)$ называется доля элементов выборки, не превосходящих x . Критерий Колмогорова предназначен для проверки гипотезы

$$H_0 : F(x) \equiv F_0(x),$$

где $F_0(x)$ – заданная функция распределения. Его статистика имеет вид

$$D_n = \sqrt{n} \sup_x |F_n(x) - F_0(x)|.$$

В [5, с.134-141] показано, что функция распределения статистики D_n имеет предел,

$$\lim_{n \rightarrow \infty} P\{D_n \leq \lambda\} = K(\lambda),$$

и рассчитана первая в истории таблица функции распределения Колмогорова $K(\lambda)$.

Работа [5, с.134-141] породила одно из основных направлений непараметрической статистики. И в настоящее время непараметрические критерии согласия (Колмогорова, Смирнова, омега-квадрат и др.) широко используются. Они были разработаны для проверки согласия с *полностью известным* теоретическим распределением. Основная идея критериев Колмогорова, омега-квадрат и аналогичных им состоит в измерении расстояния между функцией эмпирического распределения и функцией теоретического распределения. Различаются эти критерии видом расстояний в пространстве функций распределения. Расчетные формулы, таблицы распределений и критических значений широко распространены

(см., например, лучший сборник таблиц математической статистики на русском языке [10]).

Часто возникает задача проверки гипотезы согласия эмпирического распределения с параметрическим семейством, например, с семейством нормальных, Вейбулла – Гнеденко или гамма-распределений. Представляется естественным оценить параметры распределения по выборке, а затем формально воспользоваться критериями согласия Колмогорова, Смирнова или омега-квадрат. При этом вместо фиксированной теоретической функции распределения подставляют функцию из параметрического семейства, в которой параметры заменены на их выборочные оценки. В отличие от классических критериев, при этом измеряются расстояния от эмпирической функции распределения до многообразий (в пространстве функций распределения), соответствующих параметрическим семействам. Развита [11] математическая техника проектирования в функциональных пространствах, которая позволяет строить методы проверки рассматриваемых гипотез.

Однако распределения таких критериев (как предельные, так и при конечных объемах выборок) *существенно отличаются* от распределений классических критериев согласия Колмогорова, Смирнова или омега-квадрат (подробнее см. [12]). Такие критерии в отличие от классических обычно называют «критериями согласия с параметрическим семейством типа Колмогорова – Смирнова и типа омега-квадрат». (Как показано в [12, 13] на основе анализа исходных публикаций, корректно употреблять термины «критерий Колмогорова», «критерий Смирнова», «критерий типа Колмогорова – Смирнова», но нельзя говорить о несуществующем «критерии Колмогорова – Смирнова».) В [14] собраны основные факты о критериях согласия с параметрическими семействами типа Колмогорова – Смирнова и типа омега-квадрат и необходимые краткие таблицы. Современное положение дел в этой области отражено в [2, 12]. Наиболее

существенное продвижение в изучении критериев типа Колмогорова-Смирнова достигнуто профессором МГУ им. М.В. Ломоносова Ю.Н. Тюриным [15] и его научной школой.

4. «Вздувание» коэффициента корреляции

Это явление обнаружено А.Н. Колмогоровым в работе 1933 г. «К вопросу о пригодности найденных статистическим путем формул прогноза» [5, с. 161-167]. Предположим, что имеется много наборов предикторов (факторов, признаков, переменных, регрессоров). Для каждого из них строится наилучшее приближение отклика с помощью линейной функции от предикторов. Показателем качества приближения служит коэффициент корреляции между откликом и наилучшей линейной функцией от предикторов (в настоящее время чаще используют его квадрат, называемый коэффициентом детерминации). Эффект «вздувания» коэффициента корреляции состоит в том, что при увеличении числа проанализированных наборов предикторов заметно растет максимальный из соответствующих коэффициентов корреляции - показателей качества приближения. Создается впечатление, что тот набор предикторов, на котором достигается рассматриваемый максимум, дает хорошее приближение для отклика. Однако это впечатление развеивается при попытке использовать соответствующую зависимость для прогноза – по новым данным коэффициент корреляции между откликом и ранее найденной линейной функцией от предикторов оказывается значительно меньшим.

В настоящее время весьма популярны методы поиска «наиболее информативного множества признаков» в регрессионном и дискриминантном анализе. Соответствующие алгоритмы, как правило, основаны на переборе большого числа наборов признаков. Поэтому, как показано в [16], актуальность работы А.Н.Колмогорова [5, с. 161-167] в

настоящее время существенно повысилась. Эффект «вздувания» коэффициента корреляции является одним из проявлений неклассического поведения статистических характеристик в ситуации, когда одна и та же статистическая процедура осуществляется многократно, например, при множественных проверках статистических гипотез (см. [17, разд. 4.3]).

В течение полувека А.Н. Колмогоров интересовался статистическими постановками, в которых число неизвестных параметров растет вместе с объемом данных. К ним относится и работа [5, с. 161-167]. А в 1970-х годах он стимулировал исследования по т.н. «асимптотике Колмогорова» (в современной терминологии)

$$p \rightarrow \infty, n \rightarrow \infty, p/n \rightarrow \lambda > 0,$$

где p - число параметров, n - объем выборки. Эта асимптотика весьма актуальна как для многомерного статистического анализа, так и для статистики нечисловых данных [18], а также для задач статистического приемочного контроля [8, раздел 13.5] и анализа социологических данных (см. [17, гл. 13]).

В настоящее время в асимптотике Колмогорова вместо от третьего предельного перехода обычно отказываются, но требуют отделенности дроби p/λ от 0 и ∞ , т.е. требуют существования числа $\varepsilon > 0$ такого, что

$$p \rightarrow \infty, n \rightarrow \infty, p/n > \varepsilon, p/n < 1/\varepsilon.$$

5. Метод медианы в теории оценивания

Пусть X_1, X_2, \dots, X_n - независимые одинаково распределенные случайные величины с функцией распределения F и непрерывной плотностью f . Пусть μ и σ^2 - соответственно математическое ожидание и дисперсия, а m - медиана распределения F (т.е. $P\{X_1 \geq m\} \geq 1/2$ и $P\{X_1 \leq m\} \geq 1/2$). Медиана всегда существует, но не всегда определяется

однозначно. Обычно в качестве оценки для μ используют (в случае нормального закона, прежде всего) выборочное среднее арифметическое

$$\bar{X}_n = n^{-1} \sum_{k=1}^n X_k,$$

обладающее при условии нормальности F оптимальными свойствами. Что делать, если распределение F отлично от нормального? В работе 1931 г. «Метод медианы в теории ошибок» [5, с.111-114] А.Н. Колмогоров предлагает в этом случае оценивать по выборке другую среднюю характеристику распределения – медиану m (для симметричных распределений эти две характеристики совпадают). Пусть $X_n(k)$ – k -ая порядковая статистика, построенная по рассматриваемой выборке. Если n четно, то в качестве оценки m_n медианы m возьмем $X_n(n/2)$; если же $n = 2k+1$, то в качестве оценки m возьмем $X_n(k)$. С целью сравнения оценок \bar{X}_n и m_n рассмотрим преобразованные величины

$$\alpha_n = \sqrt{n}(\bar{X}_n - \mu), \quad \beta_n = \sqrt{n}(m_n - m).$$

Согласно центральной предельной теореме предельное (при $n \rightarrow \infty$) распределение величины α_n является асимптотически нормальным с нулевым средним и дисперсией σ^2 . Можно показать [5, с.111-114], что распределение величины β_n является асимптотически нормальным с нулевым средним и дисперсией $\sigma_m = (1/2)/f(m)$, если $f(m)$ отлично от 0. Мерой сравнительной точности обоих методов является отношение $\lambda = \sigma_m/\sigma = (1/2)/[\sigma f(m)]$. В случае нормальной плотности f имеем $\lambda = (\pi/2)^{1/2} \approx 5/4$. Как показал А.Н. Колмогоров [5, с.111-114], для унимодальных распределений отношение λ может принимать любое значение из интервала $(0; \sqrt{3})$, но не может превосходить $\sqrt{3}$.

Дискуссия о том, какую выборочную характеристику использовать для оценивания среднего значения совокупности (понимаемого как типичное значение), продолжаются. В силу закона больших чисел выборочное среднее арифметическое при росте объема выборки

приближается к математическому ожиданию (если оно существует). Но оно весьма неустойчиво по отношению к отклонениям «хвостов» распределения [9, разд. 4.7]. Как говорят, эта оценка не является робастной, в отличие от медианы. Медиана хороша тем, что является допустимым средним при измерениях в порядковой шкале, в то время как выборочное среднее арифметическое не обладает этим свойством [19, разд. 3.1], зато в ряде случаев имеет меньшую дисперсию (как для выборки из нормального распределения) и хорошо известно среди неспециалистов. Заслуга А.Н. Колмогорова в том, что он выявил рассматриваемую спорную точку в статистической теории и исчерпывающим образом сравнил асимптотические распределения двух видов средних величин.

6. Средние по Колмогорову

Естественная система аксиом приводит к так называемым ассоциативным средним. Их общий вид нашел в 1930 г. А.Н. Колмогоров [4, с.136-138]. Теперь их называют «средними по Колмогорову» (или «средними Колмогорова»). Для чисел X_1, X_2, \dots, X_n среднее по Колмогорову вычисляется как

$$G\{(F(X_1)+F(X_2)+\dots+F(X_n))/n\},$$

где F - строго монотонная функция (т.е. строго возрастающая или строго убывающая), G - функция, обратная к F . Среди средних по Колмогорову - много хорошо известных средних величин. Так, если $F(x) = x$, то среднее по Колмогорову - это среднее арифметическое. Для положительных X_1, X_2, \dots, X_n : если $F(x) = \ln x$, то среднее по Колмогорову - это среднее геометрическое, если $F(x) = 1/x$, то среднее гармоническое, если $F(x) = x^2$, то среднее квадратическое, и т.д. Однако такие популярные средние, как медиана и мода, нельзя представить в виде средних по Колмогорову. В прикладной статистике к средним по Колмогорову обращаются в связи с задачей выбора алгоритмов для анализа данных, измеренных в той или

иной шкале (см. [17, разд. 10.3]). Так, для алгоритмов усреднения установлено, что в шкале интервалов из всех средних по Колмогорову допустимым является только среднее арифметическое, а в шкале отношений - только степенные средние с $F(x) = x^c$, (при c , отличном от 0) и среднее геометрическое. Таким образом, среднее геометрическое или среднее квадратическое температур (в шкале Цельсия) или расстояний не имеют смысла. В качестве среднего в шкале интервалов надо применять среднее арифметическое. А также можно использовать медиану или моду.

В теории средних величин рассматриваемая работа А.Н. Колмогорова – самое значительное продвижение за последние 200 лет, после того, как О. Коши ввел общее понятие среднего.

7. Статистическая теория кристаллизации металлов.

В работе 1937 г. [5, с. 178-182] А.Н. Колмогоров разработал модель возникновения центров кристаллизации и нарастания закристаллизованной массы. При широких допущениях им была найдена точная формула для вероятности $p(t)$, с которой наудачу выбранная точка P из объема, заполненного подлежащим кристаллизации веществом, попадет в течение промежутка кристаллизации t внутрь уже закристаллизованной массы. С достаточным приближением можно считать, что доля вещества, закристаллизовавшегося за время t , также равно $p(t)$. Рассчитано число центров кристаллизации, образующихся в течение всего процесса кристаллизации. Полученные в работе [5, с. 178-182] результаты до сих пор представляют интерес для всех специалистов, связанных с изучением и использованием процессов кристаллизации металлов и иных веществ.

8. Метод наименьших квадратов

В двух работах А.Н. Колмогорова [5, с.267-283, с. 283-288] опубликованных в 1946-1947 гг., построена геометрическая теория метода

наименьших квадратов, выявляющая роль ортогонального проектирования на подпространства конечномерного евклидова пространства с целью получения оценок параметров. Эта идея затем широко использовалась как в научных исследованиях, так и при преподавании, в частности, в работах В.Н. Тутубалина, Ю.Н. Тюрина.

Другая составляющая этих работ А.Н. Колмогорова - построение алгоритмов доверительного оценивания и проверки гипотез на основе предположения о нормальности распределения погрешностей измерения. К настоящему времени эти результаты ушли в прошлое, поскольку установлено, что в подавляющем большинстве случаев распределение погрешностей заметно отличается от нормального (см. [20] и [17, разд. 2.1]). Поэтому современный подход (см., например, [21, 22] и [17, гл. 6]) к методу наименьших квадратов является непараметрическим, т.е. в определенном смысле наблюдается возврат к доколмогоровским взглядам.

9. Суммы случайного числа случайных слагаемых

Они глубоко изучены в работе 1949 г. [5, с.308-313], выполненной совместно с Ю.В. Прохоровым (1929 - 2013), в дальнейшем академиком АН СССР. Эта статья стимулировала исследования по важному для приложений виду предельных теорем (см. [23, с.300-312], [24, с.223-228]). Речь идет прежде всего о статистическом последовательном анализе [25], в частности, об изучении времени наблюдения в задаче последовательного различения двух простых гипотез. Предельные теоремы [26, 27] о суммах случайного числа случайных слагаемых находят применения в задачах статистического контроля качества и надежности по Вальду, в моделях управления запасами в логистике и других прикладных областях. Нами получен ряд результатов о суммах случайного числа случайных слагаемых при изучении двухуровневой модели управления запасами и асимптотики

квантования в связи с выбором числа градаций в социологических анкетах (см. [17, разд. 8.4 и 12.3] и монографию [28]).

10. Статистический контроль

А.Н. Колмогоров – основоположник современной теории статистического приемочного контроля в нашей стране. Около 150 лет статистические методы применяются в России для проверки соответствия продукции установленным требованиям, т.е. для сертификации. Так, еще в 1846 г. действительный член Петербургской академии наук М.В. Остроградский рассматривал задачу статистического контроля партий мешков муки или штук сукна армейскими поставщиками [29]. Однако современный этап начался в 1951 г. с брошюры А.Н. Колмогорова [30]. С тех пор в России в статистическом контроле качества было сделано многое, особенно в области теории [31 - 33]. (К сожалению, до сих пор нет доступной информации об отечественных разработках и стандартах в области статистического контроля качества в оборонных отраслях промышленности до, во время и после Великой Отечественной Войны.) Вопросы статистического контроля постоянно рассматриваются на страницах журнала «Заводская лаборатория» (с 1994 г. – «Заводская лаборатория. Диагностика материалов») – основного места публикации отечественных работ по статистическим методам [34, 35].

Большое значение для развития статистических методов управления качеством имеют статья А.Н.Колмогорова 1933 г. [5, с.134-141] о критерии согласия эмпирического распределения с теоретическим и статья 1950 г. о несмещенных оценках [5, с. 340-363]. Актуальность первой из них определяется недостатками в используемых до сих пор статистических методах управления качеством. Широко распространенные ошибки состоят в том, что для критериев согласия с параметрическими семействами используют критические значения классических критериев.

При этом, например, гипотеза нормальности принимается гораздо чаще, чем следует [12]. Поскольку в действующей нормативно-технической документации дальнейшие этапы анализа данных часто зависят от того, принимается нормальность или нет, то ошибки при такой проверке могут иметь далеко идущие последствия. Так, при анализе характеристик эластомерных материалов при ошибочном подходе из 30 выборок нормальность была отвергнута лишь для 2, а при правильном - для 26, т.е. в подавляющем большинстве случаев. Указанные ошибки встречаются в массе публикаций (хотя специалистам суть дела хорошо известна уже почти 50 лет [36]). Наиболее известным примером является полностью ошибочный ГОСТ 11.006-74 (СТ СЭВ 1190-78) «Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим» (хотя он формально отменен в 1987 г., но продолжает использоваться неспециалистами как методический материал; об ошибочных стандартах по статистическим методам см. [35]).

Популярны и другие ошибки при применении рассматриваемых критериев согласия. Некоторые пытаются их использовать для сгруппированных данных, что приводит к излишне частому принятию гипотез [37]. Другие вместо эмпирической функции распределения рассматривают иные оценки теоретической функции распределения. Например, при использовании вероятностной бумаги удобно ординату точки, соответствующей i -ой порядковой статистике, установить равной $(i-0.5)/n$, а не i/n , как в классической эмпирической функции распределения. Возникает искушение построенную таким методом оценку использовать в критериях согласия вместо эмпирической функции распределения. Увы, распределение изменится (впрочем, в данном случае при росте объема выборки различие будет исчезать). Ряд ошибок рассмотрен в пояснительной части таблиц [10].

11. Несмещенные оценки

При оценивании по выборке параметров распределений (либо функций от них) рекомендуют использовать метод максимального правдоподобия, дающий при выполнении условий регулярности асимптотически оптимальные оценки. Однако часто возникают трудности с решением уравнений правдоподобия. Поэтому вместо оценок максимального правдоподобия применяют асимптотически им эквивалентные одношаговые оценки (см. [17, разд. 3.2]) или оценки иных видов. Среди последних популярными (см., например, [38, гл.2]) являются несмещенные оценки. При конечном объеме выборки оценки максимального правдоподобия в ряде случаев хуже несмещенных оценок, как показал Я.П. Лумельский в статье [39].

Основная идея использования несмещенных оценок состоит по А.Н. Колмогорову [5, с.340-363] в следующем. Во многих важных случаях эти оценки существуют. С другой стороны, чрезмерное разнообразие несмещенных оценок может быть значительно сокращено, если воспользоваться несмещенными оценками, которые выражаются через надлежащим образом выбранные достаточные статистики. Надо употреблять только несмещенные оценки, выражающиеся через достаточные статистики: оказывается, что при этом мы не суживаем круг задач, в котором несмещенные оценки существуют, и при переходе от произвольной (даже плохой) несмещенной оценки к осредненной оценке, выражающейся через достаточную статистику, мы можем только уменьшить дисперсию оценки. Имеет место [38, гл.2] теорема Рао – Блекуэлла – Колмогорова: оптимальная оценка, если она существует, является функцией от достаточной статистики.

А.Н. Колмогоров первым ([5, с.340-363], [30]) применил несмещенные оценки в задачах статистического контроля. Он впервые использовал несмещенные оценки для определения эффективности

реально используемых планов выборочного контроля по альтернативному признаку. На основе идей А.Н. Колмогорова рядом авторов были построены несмещенные оценки для предъявленного и пропущенного брака, для априорного распределения числа дефектных изделий в контролируемых партиях, а также получены несмещенные оценки при контроле по альтернативному и количественному признакам (см. [33], а также комментарии Ю.К. Беляева и Я.П. Лумельского в [5, с.522-523]). Несмещенные оценки основных показателей контроля включены в некоторые государственные стандарты (ГОСТ 24660-81 «Статистический приемочный контроль по альтернативному признаку на основе экономических показателей», например).

Полученная А.Н.Колмогоровым несмещенная оценка плотности нормального распределения нашла широкое применение в задачах контроля по количественному признаку. В дальнейшем этот результат был перенесен на многомерное нормальное распределение, а также применен для задач статистической классификации. Метод проверки гипотез по совокупности малых выборок, разработанный нами в [28], также основан на использовании несмещенных оценок. Этот метод применяется при статистическом приемочном контроле по нескольким альтернативным признакам [8, раздел 13.5]. Отметим, что в этом случае, как и в теории люсианов [19], оказывается нецелесообразным переход к осредненной оценке, выражающейся через достаточную статистику.

Введенные А.Н.Колмогоровым верхние и нижние оценки могут быть использованы и в тех случаях, когда несмещенные оценки не существуют. Именно так обстоит дело при оценивании пропущенного брака при биномиальном распределении и плане одноступенчатого контроля. Рядом авторов были получены верхние и нижние оценки функций неизвестных параметров, а также оценки с минимальным смещением.

12. О логнормальном законе распределения

В 1940 г. Н.К. Разумовский привел много случаев, в которых логарифмы размеров частиц (золотин в золотоносных россыпях, частиц горных пород при их дроблении и т.п.) приближенно подчиняются нормальному закону распределения. В 1941 г. А.Н. Колмогоров указал общую схему случайного процесса последовательного дробления частиц, при которой в пределе, при неограниченном продолжении дробления, нормальный закон для логарифмов размеров частиц может быть установлен теоретически [5, с.264-266]. (Напомним, что положительная случайная величина X имеет логнормальный закон распределения, если логарифм величины X имеет нормальный закон распределения; условия, при которых вероятностная модель приводит к нормальному закону, хорошо известны.)

Идея А.Н. Колмогорова о выводе вида распределения случайной величины в конкретной прикладной задаче из некоторой системы аксиом нашла свое развитие, в частности, в экономике. Установлено, что распределения различных видов доходов (индивидуальных, подушевых, заработной платы и т.п.) хорошо приближаются с помощью логарифмически нормальных функций распределения. Исходя из этого факта, Росстат для сверхбольших и сверхмалых доходов вместо наблюдаемых значений доходов использует расчетные на основе логарифмически нормальных функций распределения. Это объясняется большими трудностями в замере величин сверхбольших и сверхмалых доходов, необходимостью привлечения для таких замеров правоохранительных органов. Сам же логарифмически нормальный закон выводят аксиоматически, полагая, что доход того или иного вида можно считать полученным как произведение независимых случайных величин.

Если же принять, что практически значимый результат определяется с помощью максимального значения из ряда независимых одинаково

распределенных случайных величин, как это делают в теории рекордов [40], то приходим к возможности описать распределение максимума одним из трех законов, исходя из предельных теорем.

13. Обнаружение различий

В семидесятых – восьмидесятых годах XX в. под научным руководством А.Н. Колмогорова на механико-математическом факультете МГУ им. М.В. Ломоносова работала группа исследователей, занимавшаяся статистическим анализом эффективности экспериментальных методов управления погодой. Речь идет об изменении количеств выпавших осадков, борьбе с градом и рассеянии туманов. Среди прочих [41] вероятностных моделей использовалась и следующая.

Имеется n объектов U_1, U_2, \dots, U_n и с каждым объектом U_k связана пара чисел a_k и b_k , $k = 1, 2, \dots, n$. Пусть $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ – последовательность независимых случайных величин, причем величина ε_k принимает значение 1 (считаем, что имеет место воздействие) с вероятностью p_k и значение 0 (воздействие отсутствует) с вероятностью $q_k = 1 - p_k$, $k = 1, 2, \dots, n$. В результате наблюдений над объектами нам известны значения случайных пар (ε_k, X_k) , $k = 1, 2, \dots, n$, где $X_k = a_k$ при $\varepsilon_k = 0$ и $X_k = b_k$ при $\varepsilon_k = 1$. Задача состоит в сравнении двух последовательностей $a^{(n)} = (a_1, a_2, \dots, a_n)$ и $b^{(n)} = (b_1, b_2, \dots, b_n)$. Тем самым в этой модели (могущей быть использованной и в других случаях, когда необходимо установить наличие или отсутствие эффекта воздействия) предполагается, что числа a_k и b_k неслучайны и вся случайность связана с процессом рандомизации. С помощью оценок Горвица – Томпсона и их обобщений [42] можно построить [41] ряд статистических критериев для проверки гипотезы

$$H(n): A(n) = B(n),$$

где

$$A(n) = \sum_{k=1}^{k=n} a_k, \quad B(n) = \sum_{k=1}^{k=n} b_k.$$

А.Н. Колмогоров заметил, что дисперсии оценок в критериях могут быть заметно уменьшены, если имеются хорошие методы прогноза, позволяющие до начала наблюдений указывать оценки a_k^* и b_k^* для a_k и b_k соответственно. Полагая $a_k = a_k^* + \Delta a_k$ и $b_k = b_k^* + \Delta b_k$, мы можем упомянутые выше процедуры применить не к a_k и b_k , а к Δa_k и Δb_k . При этом получаются оценки, правильные независимо от качества прогноза, но они будут лучше оценок без обращения к прогнозам лишь в случае хороших прогнозов, когда величины $|\Delta a_k|$ и $|\Delta b_k|$ значительно меньше $|a_k|$ и $|b_k|$ соответственно.

Эти модели и методы нашли ряд применений, в частности, при разработке рандомизированной процедуры оценки айсберговой угрозы, связанной с добычей нефти и газа на морском арктическом шельфе [43].

Упомянем также работы А.Н. Колмогорова по теории стрельбы, выполненные в военные годы, по генетике и биологии, по лингвистике. Особенно велик вклад, сделанный А.Н. Колмогоровым и его учениками (М.Д. Миллионщиков, А.С. Монин, А.М. Обухов, А.М. Яглом и др.) в теорию турбулентности. Здесь прежде всего следует упомянуть знаменитый Колмогоровский «закон двух третей» о распределении энергии в спектре турбулентности, полученный из простых соображений размерности (подробнее см. [3, с.445, 475], [4]).

Бесспорно, что многие работы А.Н.Колмогорова [4 - 6] представляют несомненный интерес для всех, кто разрабатывает или применяет статистические методы. Его мысли еще долго будут приносить нашей стране и всему миру всем практическую пользу. Отечественная вероятностно-статистическая научная школа порождена идеями А.Н. Колмогорова. Это хорошо видно на примере работ его ученика Б.В.

Гнеденко, которым будет посвящена следующая статья по истории статистических методов в нашей стране.

Литература

1. Орлов А.И. Основные этапы становления статистических методов // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 1209 – 1233. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/86.pdf>.
2. Кудлаев Э.М., Орлов А.И. Вероятностно-статистические методы исследования в работах А.Н.Колмогорова // Заводская лаборатория. Диагностика материалов. 2003. Т.69. № 5. С.55-61.
3. Колмогоров в воспоминаниях / Под ред. А.Н.Ширяева. – М.: Физматлит, 1993. - 736 с.
4. Колмогоров А.Н. Избранные труды: Математика и механика. - М.: Наука, 1985. - 470 с.
5. Колмогоров А.Н. Теория вероятностей и математическая статистика. – М.: Наука, 1986. - 535 с.
6. Колмогоров А.Н. Теория информации и теория алгоритмов. – М.: Наука, 1987. - 304 с.
7. Колмогоров А.Н. Основные понятия теории вероятностей. – М.-Л.: ОНТИ, 1936. - 80 с. (3-е издание. – М.: Фазис, 1998. - 144 с. – Серия «Стохастика», вып.1.)
8. Орлов А.И. Эконометрика. Учебник для вузов. Изд. 3-е, исправленное и дополненное. - М.: Экзамен, 2004. – 576 с.
9. Орлов А.И. Прикладная статистика. Учебник. - М.: Экзамен, 2006. - 672 с.
10. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. - М.: Наука, 1965 (1-е изд.), 1968 (2-е изд.), 1983 (3-е изд.).
11. Тюрин Ю.Н. Линейная модель в многомерной непараметрической статистике. - В сб.: Многомерный статистический анализ в социально-экономических исследованиях. - М.: Наука, 1974. С.7-24.
12. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – Режим доступа: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
13. Орлов А.И. О критериях Колмогорова и Смирнова // Заводская лаборатория. 1995. Т.61. № 7. С.59-61.
14. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория.1985. Т.51. №1. С.60-62.
15. Тюрин Ю.Н. Исследования по непараметрической статистике (непараметрические методы и линейная модель). Автореф. дисс. докт. физ.-мат. наук. - М.: МГУ, 1985. - 33 с.
16. Орлов А.И. Методы поиска наиболее информативных множеств признаков в регрессионном анализе // Заводская лаборатория. 1995. Т.61. №1. С.56-58.
17. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. - М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. - 624 с.
18. Сердобольский В.И., Орлов А.И. Статистический анализ при большом числе параметров. – В сб.: Программно-алгоритмическое обеспечение прикладного

многомерного статистического анализа / Тезисы докладов III Всесоюзной школы-семинара. - М.: ЦЭМИ АН СССР, 1987. С.151-160.

19. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. — 541 с.
20. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. 1991. Т.57. №.7. С.64-66.
21. Орлов А.И. Восстановление зависимости методом наименьших квадратов на основе непараметрической модели с периодической составляющей // Научный журнал КубГАУ [Электронный ресурс]. – Краснодар: КубГАУ, 2013. – №07(091). С. 189 – 218. – Режим доступа: <http://ej.kubagro.ru/2013/07/pdf/13.pdf>
22. Орлов А.И. Непараметрический метод наименьших квадратов с периодической составляющей // Заводская лаборатория. Диагностика материалов. 2014. Т.80. №1. С.65-75.
23. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т.1: Пер. с англ. – М.: Мир, 1984. - 528 с.
24. Гнеденко Б.В. Курс теории вероятностей: Учебник. 7-е изд., исправл. - М.: Эдиториал УРСС, 2001. - 320 с.
25. Ширяев А.Н. Статистический последовательный анализ. – М.: Наука, 1976. - 240 с.
26. Круглов В.М., Королев В.Ю. Предельные теоремы для случайных сумм. – М.: МГУ, 1990. - 188 с.
27. Gnedenko B.V., Korolev V.Yu. Random summation: limit theorems and applications. – CRC Press, Boca Raton, Fl., 1996. - 268 pp.
28. Орлов А.И. Устойчивость в социально-экономических моделях. – М.: Наука, 1979. - 296 с.
29. Остроградский М.В. Об одном вопросе, касающемся вероятностей / Полное собрание трудов. Т.3. – Киев: Издательство Академии наук УССР, 1961. – С.215-237.
30. Колмогоров А.Н. Статистический приемочный контроль при допустимом числе дефектных изделий, равном нулю. – Л.: Знание, 1951. - 24 с.
31. Гнеденко Б.В. Математика и контроль качества продукции. – М.: Знание, 1978. - 64 с.
32. Беляев Ю.К. Вероятностные методы выборочного контроля. – М.: Наука, 1975. - 407 с.
33. Лумельский Я.П. Статистические оценки результатов контроля качества. – М.: Изд-во стандартов, 1979. - 200 с.
34. Орлов А.И. О современных проблемах внедрения прикладной статистики и других статистических методов // Заводская лаборатория. 1992. Т.58. № 1. С.67-74.
35. Орлов А.И. Сертификация и статистические методы // Заводская лаборатория. 1997. Т.63. № 3. С.55-62.
36. Кас М., Kiefer J., Wolfowitz J. On test of normality and other tests of goodness of fit based on distance methods / Ann. Math. Statist. 1955. V.26. No.2. P.189-211.
37. Рунион Р. Справочник по непараметрической статистике. - М.: Финансы и статистика, 1982. - 198 с.
38. Ивченко Г.И., Медведев Ю.И. Математическая статистика. – М.: Высшая школа, 1984. - 248 с.
39. Лумельский Я.П. К вопросу сравнения несмещенных и других оценок. – В сб.: Прикладная статистика. – М.: Наука, 1983. С.316-319.
40. Невзоров В.Б. Рекорды. Математическая теория. - М.: Фазис, 2000. - XII+244 с.

41. Журбенко И.Г., Кудлаев Э.М. О выявлении эффекта воздействия в рандомизированных экспериментах // Успехи математических наук. 1984. Т.39. Вып.1. С.3-38.
42. Булинский А.В., Колмогоров А.Н. Линейные выборочные оценки сумм / Теория вероятностей и ее применения. 1979. Т.24. № 2. С. 241-251.
43. Кудлаев Э.М. Рандомизированная процедура оценки айсберговой угрозы, связанная с добычей нефти и газа на морском арктическом шельфе // Заводская лаборатория. Диагностика материалов. 2009. Т.75. № 1. С.75-77.

References

1. Orlov A.I. Osnovnye jetapy stanovlenija statisticheskikh metodov // Nauchnyj zhurnal KubGAU [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №03(097). S.1209 – 1233. – Rezhim dostupa: <http://ej.kubagro.ru/2014/03/pdf/86.pdf>.
2. Kudlaev Je.M., Orlov A.I. Verojatnostno-statisticheskie metody issledovanija v rabotah A.N.Kolmogorova // Zavodskaja laboratorija. Diagnostika materialov. 2003. T.69. № 5. S.55-61.
3. Kolmogorov v vospominanijah / Pod red. A.N.Shirjaeva. – M.: Fizmatlit, 1993. - 736 s.
4. Kolmogorov A.N. Izbrannye trudy: Matematika i mehanika. - M.: Nauka, 1985. - 470 s.
5. Kolmogorov A.N. Teorija verojatnostej i matematicheskaja statistika. – M.: Nauka, 1986. - 535 s.
6. Kolmogorov A.N. Teorija informacii i teorija algoritmov. – M.: Nauka, 1987. - 304 s.
7. Kolmogorov A.N. Osnovnye ponjatija teorii verojatnostej. – M.-L.: ONTI, 1936. - 80 s. (3-e izdanie. – M.: Fazis, 1998. - 144 s. – Serija «Stohastika», vyp.1.)
8. Orlov A.I. Jekonometrika. Uchebnik dlja vuzov. Izd. 3-e, ispravlennoe i dopolnennoe. - M.: Jekzamen, 2004. – 576 s.
9. Orlov A.I. Prikladnaja statistika. Uchebnik. - M.: Jekzamen, 2006. - 672 s.
10. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoi statistiki. - M.: Nauka, 1965 (1-e izd.), 1968 (2-e izd.), 1983 (3-e izd.).
11. Tjurin Ju.N. Linejnaja model' v mnogomernoj neparametricheskoi statistike. - V sb.: Mnogomernyj statisticheskij analiz v social'no-jekonomicheskix issledovanijah. - M.: Nauka, 1974. S.7-24.
12. Orlov A.I. Neparametricheskie kriterii soglasija Kolmogorova, Smirnova, omega-kvadrat i oshibki pri ih primenenii // Nauchnyj zhurnal KubGAU [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №03(097). S. 647 – 675. – Rezhim dostupa: <http://ej.kubagro.ru/2014/03/pdf/47.pdf>
13. Orlov A.I. O kriterijah Kolmogorova i Smirnova // Zavodskaja laboratorija. 1995. T.61. № 7. S.59-61.
14. Orlov A.I. Rasprostranennaja oshibka pri ispol'zovanii kriteriev Kolmogorova i omega-kvadrat // Zavodskaja laboratorija. 1985. T.51. №1. S.60-62.
15. Tjurin Ju.N. Issledovanija po neparametricheskoi statistike (neparametricheskie metody i linejnaja model'). Avtoref. diss. dokt. fiz.-mat. nauk. - M.: MGU, 1985. - 33 s.
16. Orlov A.I. Metody poiska naibolee informativnyh mnozhestv priznakov v regressionnom analize // Zavodskaja laboratorija. 1995. T.61. №1. S.56-58.
17. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch.3. Statisticheskie metody analiza dannyh. - M.: Izd-vo MGTU im. N.Je. Bauman, 2012. - 624 s.

18. Serdobol'skij V.I., Orlov A.I. Statisticheskij analiz pri bol'shom chisle parametrov. – V sb.: Programmno-algoritmicheskoe obespechenie prikladnogo mnogomernogo statisticheskogo analiza / Tezisy dokladov III Vsesojuznoj shkoly-seminara. – M.: CJeMI AN SSSR, 1987. S.151-160.
19. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch. 1. Nechislovaja statistika. – M.: Izd-vo MGTU im. N.Je. Bauman, 2009. — 541 s.
20. Orlov A.I. Chasto li raspredelenie rezul'tatov nabljudenij javljaetsja normal'nym? // Zavodskaja laboratorija. 1991. T.57. No.7. S.64-66.
21. Orlov A.I. Vosstanovlenie zavisimosti metodom naimen'shix kvadratov na osnove neparametricheskoj modeli s periodicheskoj sostavljajushhej // Nauchnyj zhurnal KubGAU [Elektronnyj resurs]. – Krasnodar: KubGAU, 2013. – №07(091). S. 189 – 218. – Rezhim dostupa: <http://ej.kubagro.ru/2013/07/pdf/13.pdf>
22. Orlov A.I. Neparametricheskij metod naimen'shix kvadratov s periodicheskoj sostavljajushhej // Zavodskaja laboratorija. Diagnostika materialov. 2014. T.80. №1. S.65-75.
23. Feller V. Vvedenie v teoriju verojatnostej i ee prilozhenija. V 2-h tomah. T.1: Per. s angl. – M.: Mir, 1984. - 528 s.
24. Gnedenko B.V. Kurs teorii verojatnostej: Uchebnik. 7-e izd., ispravl. - M.: Jeditorial URSS, 2001. - 320 s.
25. Shirjaev A.N. Statisticheskij posledovatel'nyj analiz. – M.: Nauka, 1976. - 240 s.
26. Kruglov V.M., Korolev V.Ju. Predel'nye teoremy dlja sluchajnyh summ. – M.: MGU, 1990. - 188 s.
27. Gnedenko B.V., Korolev V.Yu. Random summation: limit theorems and applications. – CRC Press, Boca Raton, Fl., 1996. - 268 pp.
28. Orlov A.I. Ustojchivost' v social'no-jekonomicheskix modeljah. – M.: Nauka, 1979. - 296 s.
29. Ostrogradskij M.V. Ob odnom voprose, kasajushhemsja verojatnostej / Polnoe sobranie trudov. T.3. – Kiev: Izdatel'stvo Akademii nauk USSR, 1961. – S.215-237.
30. Kolmogorov A.N. Statisticheskij priemochnyj kontrol' pri dopustimom chisle defektnyx izdelij, ravnom nulju. – L.: Znanie, 1951. - 24 s.
31. Gnedenko B.V. Matematika i kontrol' kachestva produkcii. – M.: Znanie, 1978. - 64 s.
32. Beljaev Ju.K. Verojatnostnye metody vyborochnogo kontrolja. – M.: Nauka, 1975. - 407 s.
33. Lumel'skij Ja.P. Statisticheskie ocenki rezul'tatov kontrolja kachestva. – M.: Izd-vo standartov, 1979. - 200 s.
34. Orlov A.I. O sovremennyh problemah vnedrenija prikladnoj statistiki i drugih statisticheskix metodov // Zavodskaja laboratorija. 1992. T.58. № 1. S.67-74.
35. Orlov A.I. Sertifikacija i statisticheskie metody // Zavodskaja laboratorija. 1997. T.63. № 3. S.55-62.
36. Kac M., Kiefer J., Wolfowitz J. On test of normality and other tests of goodness of fit based on distance methods / Ann. Math. Statist. 1955. V.26. No.2. P.189-211.
37. Runion R. Spravochnik po neparametricheskoj statistike. - M.: Finansy i statistika, 1982. - 198 s.
38. Ivchenko G.I., Medvedev Ju.I. Matematicheskaja statistika. – M.: Vysshaja shkola, 1984. - 248 s.
39. Lumel'skij Ja.P. K voprosu sravnenija nesmeshhennyh i drugih ocenok. – V sb.: Prikladnaja statistika. – M.: Nauka, 1983. S.316-319.
40. Hevzopov V.B. Rekordy. Matematicheskaja teorija. - M.: Fazis, 2000. - XII+244 s.

41. Zhurbenko I.G., Kudlaev Je.M. O vyjavlenii jeffekta vozdejstvija v randomizirovannyh jeksperimentah // Uspehi matematicheskikh nauk. 1984. T.39. Vyp.1. S.3-38.

42. Bulinskij A.V., Kolmogorov A.N. Linejnye vyborochnye ocenki summ / Teorija verojatnostej i ee primenenija. 1979. T.24. № 2. S. 241-251.

43. Kudlaev Je.M. Randomizirovannaja procedura ocenki ajsbergovoj ugrozy, svjazannaja s dobychej nefi i gaza na morskome arkticheskom shel'fe // Zavodskaja laboratorija. Diagnostika materialov. 2009. T.75. № 1. S.75-77.