

Предисловие

Гораздо лучше предвидеть даже без уверенности, чем не предвидеть вообще.

Анри Пуанкаре (1854—1912), французский математик

За последние годы развитие статистики характеризовалось дальнейшим обособлением этой дисциплины от собственно математики, хотя, разумеется, математические методы, в первую очередь вероятностные (и в растущей мере геометрические), составляют ее сердцевину. Можно сказать, что современная статистика концентрируется на изучении потоков данных, часто в режиме их поступления. Говоря наглядно (и используя прозрачные исторические параллели), статистика изучает объективную реальность, данную в виде информационных потоков. С развитием и распространением компьютеров объем данных быстро растет и уже достиг астрономических уровней. Соответственно, традиционный взгляд на статистику как область исследований, а также на ее преподавание нуждается в определенной ревизии и корректировке. Возникла ситуация, когда теоретическая база явно отстает от практических решений, находимых на основе учета специфики того или иного случая.

В последние годы быстро развивается многомерная статистика. В качестве примера рассмотрим задачу обработки изображений. Предположим, что нам нужно обработать несколько сотен фотографий, каждая из которых содержит 320×243 пикселей с различными уровнями затемнения. Размерность этой задачи $d = 77760 \gg n$ много больше размера выборки $n \approx 10^2$. Методы классической статистики не применимы в этой ситуации: для того чтобы проявились типичные вероятностные закономерности, например закон больших чисел и центральная предельная теорема, размер выборки n должен экспоненциально расти с ростом размерности d . Это означает, что число фотографий должно быть много больше числа песчинок в пустыне Сахара. Для введения в специальные методы, развитые

для анализа подобных ситуаций, потребуется написать отдельную книгу¹⁾.

Такая ситуация породила достаточно серьезный вызов, с которым даже ведущие исследовательские и образовательные центры не могут успешно справиться. К сожалению, даже на этом фоне положение статистической науки в России нельзя оценить иначе как плачевное. Сказалось отсутствие лидеров в исследовательской работе, сопоставимых с А. Н. Колмогоровым и Ю. В. Линником, а несколько ранее Н. В. Смирновым и Е. Е. Слуцким. Существенно упал уровень преподавания теоретических основ статистики, не говоря уже о специализированных курсах, ориентированных на приложения. На кафедре статистики и теории случайных процессов мехмата МГУ нет ни одного признанного специалиста по статистике. Статистика отсутствует как направление в Математическом институте имени Стеклова, где она была раньше серьезно представлена. Не лучше положение и в Петербурге, не говоря уже о других городах России. Последний раз, когда яркое достижение в статистических исследованиях было связано с именами российских ученых, — это, по-видимому, теория В. Н. Вапника — А. Я. Червоненкиса. В то же время в американских университетах департаменты статистики имеются практически везде и играют активную роль в экономической деятельности штатов и на федеральном уровне, а также в экологии и медицине, не говоря об электронных коммуникациях и взаимодействии с вооруженными силами.

Тем не менее определенные основания для оптимизма есть. В России сохранился высокий уровень образования в математике и физике, а также имеется развитая культура компьютерных вычислений. Для подъема статистической культуры нужно не так уж и много: небольшой группы энтузиастов может быть достаточно. Но конечно, современные курсы статистики, в первую очередь базовые, совершенно необходимы. Мы надеемся, что данная книга внесет свою лепту в этот процесс.

Как и первая часть, посвященная основам теории вероятностей, эта книга основана на методике преподавания статистики

¹⁾Мы рекомендуем недавнюю монографию [141], в которой сделан акцент на статистических задачах в высоких размерностях. Вероятностная техника, использованная в высоких размерностях, обсуждается в книге [140] и курсе лекций [139]. Необходимо также отметить книгу [93], посвященную специальному разделу высокоразмерной теории вероятностей: теории случайных матриц.

в Кембриджском университете. Важной особенностью учебного процесса в Кембридже всегда было сочетание серьезной математической компоненты с практическими применениями, а также наличие большого количества разнообразных задач, тщательно подобранных и отредактированных. В книге эти задачи приведены с решениями, и мы настоятельно рекомендуем читателю не злоупотреблять возможностью сравнительно легкого чтения, а пытаться самостоятельно решить хотя бы часть из них.

* * *

Книга издана при поддержке Российского научного фонда, грант № 23-21-00052. М. Я. Кельберт является профессором-исследователем департамента статистики и анализа данных факультета экономики Национального исследовательского университета «Высшая школа экономики».