

## §1. Проблема статистического вывода

### Лекция 1

Теория вероятностей создает базу для построения моделей реальных явлений, в основе которых лежат соотношения между частотами появления определенных событий. Располагая вероятностной моделью, мы можем рассчитать вероятности (относительные частоты) этих событий и тем самым оптимизировать свое поведение в условиях неопределенности. Математическая статистика строит модели индуктивного поведения в этих условиях на основе имеющихся вероятностных моделей. Основная проблема состоит в том, чтобы по наблюдениям элементарных исходов (обычно это – значения наблюдаемых случайных величин) дать метод выбора действий, при которых частота ошибок была бы наименьшей. Естественно, эта проблема сопряжена с решением сложных задач на экстремум, но даже в том случае, когда эти задачи не удается решить, теория вероятностей дает метод для расчета средней величины потерь, которые мы будем нести, используя конкретное, выбранное нами правило индуктивного поведения. Таким образом, *математическая статистика есть теория принятия оптимальных решений, когда последствия от действий, предпринимаемых на основе этих решений, носят случайный характер*. Математическая статистика использует методы теории вероятностей для расчета частоты “неправильных” решений или, более общо, для величины средних потерь, которые неизбежно возникают в условиях случайности, как бы мы ни пытались оптимизировать свое поведение в этих условиях.

Приведем два примера, иллюстрирующих задачи математической статистики и, отчасти, методы их решения, с тем чтобы в последующем формализовать общую проблему статистического вывода.

**Пример 1.1.** *Определение общего содержания серы в дизельном топливе.* Мы снова обращаемся к примеру 7.2 из курса теории вероятностей, где речь шла о важной в экологическом отношении характеристике дизельного топлива – процентном содержании элементарной серы, которая при сжигании и последующем соединении с водой дает серную кислоту. Необходимость использования методов теории вероятностей при аттестации дизельного топлива по этой характе-

ристике была вызвана значительными расхождениями между результатами  $x_1, \dots, x_n$  параллельных и независимых испытаний  $n$  проб из партии дизельного топлива. Если даже исключить ошибки эксперимента, связанные с неправильным определением веса пробы и титрованием, то все равно разброс в параллельных испытаниях будет значительным в силу случайного характера процесса сжигания пробы топлива и выпадения части элементарной серы в золу. Но в таком случае возникает естественный вопрос, что же мы измеряем и что же это за характеристика дизельного топлива, которую мы назвали “общим содержанием серы”? В практике лабораторных испытаний обычно говорят о *среднем* значении этой характеристики, и дизельное топливо аттестуется величиной  $\bar{x} = n^{-1} \sum_1^n x_k$  – арифметическим средним результатов параллельных испытаний. Это и есть то “индуктивное поведение” статистика в условиях случайности, о котором мы говорили в начале лекции, и оправдание разумности такого поведения естественно искать в рамках закона больших чисел.

Действительно, в примере 7.2 мы интерпретировали результат  $x$  определения общего содержания серы в одной пробе как результат наблюдения случайной величины  $X$ , распределенной по нормальному закону со средним  $\mu$  и дисперсией  $\sigma^2$ , причем значение (неизвестное экспериментатору) параметра  $\mu$  являлось математическим выражением той, не совсем понятной для нас характеристики испытуемого топлива, которая называлась “общим содержанием серы”. В рамках этой вероятностной модели естественно трактовать результаты  $x_1, \dots, x_n$  параллельных испытаний  $n$  проб дизельного топлива как наблюдения  $n$  независимых копий  $X_1, \dots, X_n$  случайной величины  $X$ . Термин “копия” в данном случае употребляется для обозначения того факта, что каждая из наблюдаемых случайных величин имеет то же распределение, что и  $X$ . Таким образом, постулируется, что  $X_1, \dots, X_n$  независимы и одинаково распределены  $\mathcal{N}(\mu, \sigma^2)$ , так что в силу закона больших чисел при неограниченном возрастании объема испытаний  $n$

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu.$$

Итак, закон больших чисел гарантирует нам, что при достаточно большом объеме испытаний мы будем близки к истинному значению исследуемой характеристики топлива. Однако на практике в завод-

ских лабораториях обычно сжигаются всего две пробы топлива, и только в исключительных случаях при поверке приборов или тестировании лаборантов делается четыре испытания. Естественно, при  $n = 2$  говорить о законе “больших” чисел просто смешно, — следует искать некоторую количественную характеристику последствий от неточной аттестации партии дизельного топлива. Легко понять, что в основу такой характеристики следует положить ошибку  $|\bar{X} - \mu|$  в оценке параметра  $\mu$ , но, к сожалению, значение  $\mu$  нам неизвестно, а  $\bar{X}$  есть случайная величина, что окончательно делает проблему прогноза ожидаемых ошибок при аттестации конкретной партии топлива неразрешимой. Здесь наблюдается та же ситуация, что и при попытке предсказать сторону монеты, которая выпадет при ее подбрасывании. Точный прогноз невозможен, но методы теории вероятностей позволяют нам рассчитать, как часто мы будем ошибаться в прогнозе при достаточно длительной игре в орлянку. Следовательно, мы должны решить задачу о вычислении вероятности того, что ошибка в оценке  $\mu$  будет слишком большой — превосходить некоторую предписанную величину  $\Delta$ . Эта вероятность  $P(|\bar{X} - \mu| > \Delta)$  обычно называется *риском* оценки  $\bar{X}$ , а вероятность  $P(|\bar{X} - \mu| \leq \Delta)$  противоположного события — *надежностью* этой оценки.

Таким образом, риск оценки указывает частоту тех партий дизельного топлива, в паспорте которых общее содержание серы указано с недопустимо большой ошибкой. Зная риск оценки, мы можем вычислить средние затраты на выплату рекламаций по искам потребителей дизельного топлива. Вывести формулу для вычисления риска не представляет особого труда, если обратиться к теореме сложения для нормального распределения (предложение 12.2 курса ТВ). Выборочное среднее  $\bar{X}$  есть нормированная на  $n$  сумма независимых одинаково распределенных  $\mathcal{N}(\mu, \sigma^2)$  случайных величин. В силу теоремы сложения эта сумма имеет также нормальное распределение, среднее значение которого равно сумме средних  $n\mu$ , а дисперсия равна сумме дисперсий  $n\sigma^2$ . При умножении на  $1/n$  среднее умножается на ту же величину, а дисперсия умножается на ее квадрат. Таким образом,  $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$ , надежность оценки

$$P(-\Delta \leq \bar{X} - \mu \leq \Delta) = \Phi(\Delta\sqrt{n}/\sigma) - \Phi(-\Delta\sqrt{n}/\sigma) = 2\Phi(\Delta\sqrt{n}/\sigma) - 1$$

(напомним,  $\Phi(-x) = 1 - \Phi(x)$ ), а ее риск

$$P(|\bar{X} - \mu| > \Delta) = 2(1 - \Phi(\Delta\sqrt{n}/\sigma)).$$

При вычислении риска оценки необходимо знать величину стандартного отклонения  $\sigma$ . Но значение  $\sigma$ , очевидно, остается постоянным при аттестации различных партий – это параметр, характеризующий точность метода химического анализа топлива, и не имеет отношения к его химическому составу. Естественно, за достаточно короткий срок в лабораториях накапливается большой архивный материал данных испытаний различных партий топлива, что позволяет оценить значение  $\sigma$  с достаточно высокой точностью. С тем, как это делается, мы познакомимся в одной из ближайших лекций.

Используя формулу риска, мы можем определить минимальный объем испытаний  $n$ , гарантирующий предписанную, достаточно малую величину риска. Действительно, если  $\alpha$  – заданное ограничение на риск оценки, то разрешая неравенство  $2(\Phi(\Delta\sqrt{n}/\sigma) - 1) \leq \alpha$  относительно переменной  $n$ , получаем, что требуемый объем испытаний определяется неравенством

$$n \geq \left( \frac{\Phi^{-1}(1 - \alpha/2)\sigma}{\Delta} \right)^2.$$

**Пример 1.2. Выявление эффекта лечения.** Группа пациентов в количестве 10 человек, обладающих схожими антропометрическими и антропологическими характеристиками, подвергается лечению по некоторой новой методике, подтверждение или опровержение эффективности которой составляет предмет статистического исследования. После лечения дается только качественное заключение о состоянии здоровья каждого пациента, так что результат испытания новой методики можно представить в виде последовательности  $x_1, \dots, x_{10}$ , компоненты которой принимают значения 1 (положительный исход лечения) или 0 (отрицательный исход).

Предлагается следующее статистическое правило: новая методика объявляется эффективной, если  $x_i = 1$  для всех  $i = 1, \dots, 10$ , то есть все пациенты выздоровели. Если же лечение хотя бы одного пациента не привело к положительному исходу, новая методика не рекомендуется к дальнейшему клиническому использованию. Что можно ска-

зять о надежности или, как говорят медики, “достоверности” такого правила индуктивного поведения?

Чтобы ответить на этот вопрос, мы должны построить вероятностную модель проводимых наблюдений. Естественно предполагать, что в силу “однородности” группы пациентов они обладают одинаковой вероятностью  $p$  положительного исхода лечения, и если в процессе лечения они не имели возможности излишне тесного общения, то исходы лечений можно представить в виде реализации десяти независимых бинарных случайных величин  $X_1, \dots, X_{10}$ , каждая из которых принимает значение 1 с вероятностью  $p$  и значение 0 с вероятностью  $1-p$ . Таким образом, мы пришли к модели испытаний в схеме Бернулли с вероятностью  $p$  успешного исхода. Вероятность того, что все 10 исходов были успешными равна  $p^{10}$ , и задавая различные значения  $p$  мы можем судить о том, как часто возможны различные результаты апробации нового метода лечения.

Предположим сначала, что новая методика неэффективна. При таком предположении значение  $p$  не должно превосходить величины  $1/2$ , и максимальное значение вероятности события  $X_1 = 1, \dots, X_{10} = 1$  равно  $2^{-10} = 1/1024 < 0,001$ . Это очень редкое событие, и поэтому предположение о неэффективности новой методики должно быть отвергнуто. При этом вероятность  $2^{-10}$  можно интерпретировать как риск внедрения в медицинскую практику неэффективного метода лечения: *используя предложенное правило выбора между двумя действиями (внедрение или отклонение методики) при испытаниях последующих методик, мы рискуем в среднем не более чем один раз из тысячи внедрить неэффективный метод лечения.*

Интересно заметить, что в предположении “нейтральности” нового метода ( $p = 1/2$ ) вероятность любого исхода  $X_1 = x_1, \dots, X_{10} = x_{10}$  одинакова и равна  $2^{-10}$ , но исход  $X_1 = 1, \dots, X_{10} = 1$  обладает наибольшей вероятностью принятия действительно эффективной методики, ибо

$$p \sum_1^{10} x_k (1-p)^{n - \sum_1^{10} x_k} \leq p^{10},$$

если  $p > 1/2$ . Столь же просто проверить, что результаты испытаний, в которых лечение только одного пациента окончилось неудачей, имеют вероятность  $p^9(1-p)$ , и такие 10 результатов  $x_1, \dots, x_{10}$  с одним  $x_i = 1$  и другими  $x_j = 0$  обладают большей вероятностью, чем

исходы с двумя и более количеством неудач, если в действительности  $p > 1/2$ . Это замечание позволяет нам определить статистическое правило, обладающее наибольшей вероятностью принятия в действительности эффективной методики, но не с таким малым риском, как  $2^{-10}$ .

Дело в том, что в медицинской практике установилась определенная граница риска, равная 0.05, и все события, обладающие меньшей вероятностью, объявляются “редкими” – ими можно пренебречь. В связи с этим позволим себе включить в область принятия новой методики дополнительные исходы с ровно одним неуспехом, и вычислим риск такого статистического правила при  $p = 1/2$ . Используя известные нам формулы биномиальных вероятностей, находим, что

$$P\left(\sum_1^{10} X_k \geq 9\right) = p^{10} + C_{10}^1 p^9(1-p),$$

и при  $p = 1/2$  эта вероятность равна  $2^{-10}(1+10) = 11/1024 \approx 0,01$ , что по-прежнему достаточно мало по сравнению с 0.05. Следовательно, мы можем включить в область принятия новой методики еще  $C_{10}^2$  результатов испытаний, в которых присутствуют ровно две неудачи. Риск такого статистического правила становится равным

$$P\left(\sum_2^{10} X_k \geq 8\right) = p^{10} + C_{10}^1 p^9(1-p) + C_{10}^2 p^8(1-p)^2,$$

и при  $p = 1/2$  эта вероятность равна  $2^{-10}(1+10+45) = 56/1024 \approx 0.05$ .

Это как раз соответствует принятой в медицине норме риска статистического правила. Итак, мы рекомендуем новую методику к дальнейшему использованию в клинике, если лечение не более чем двух пациентов из десяти оказалось неудачным, и применение такого правила в испытаниях дальнейших методик может привести к принятию неэффективного метода лечения в среднем в пяти случаях из 100.

Мы рассмотрели две типичных задачи математической статистики – оценка параметров и проверка гипотез. Естественно, круг проблем математической статистики намного шире, но при надлежащей трактовке проблем большинство из них сводится или к задаче оценки параметров, или к задаче выбора одного из нескольких альтернативных высказываний об исследуемом объекте. Опираясь на рассмот-

ренные примеры, мы можем теперь представить достаточно общую схему статистического вывода.

Любое статистическое исследование, проводимое в рамках математической статистики, начинается с описания объекта исследования и формализации *пространства*  $\mathcal{D}$  решений  $d$ , одно из которых статистик принимает на основе наблюдений независимых копий случайной, возможно векторной, величины  $X$ , характеризующей состояние объекта в момент проведения наблюдений. Так, в примере с аттестацией партии дизельного топлива (объект исследования)  $\mathcal{D}$  есть интервал  $(0; 100)$  (напомним, общее содержание серы измеряется в процентах к весу пробы), а в примере с определением эффективности нового метода лечения (объект исследования) пространство  $\mathcal{D}$  состоит из двух точек:  $d_0$  – решение о неэффективности метода (принятие “нулевой” гипотезы) и  $d_1$  – решение о внедрении нового метода в лечебную практику (принятие альтернативной гипотезы).

Наиболее важной и, по-видимому, наиболее сложной частью статистического исследования является этап построения *вероятностной модели*, который состоит в спецификации семейства  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$  возможных распределений наблюдаемой случайной величины  $X$ . Этот этап связан с достаточно глубоким проникновением в природу исследуемого объекта и метода наблюдений  $X$ , – одной математикой здесь, как правило, не обойдешься. Семейство  $\mathcal{P}$  индексируется абстрактным *параметром*  $\theta$ , совокупность значений которого  $\Theta$  называется *параметрическим пространством*.

В первом примере мы выяснили, что семейство возможных распределений  $X$  есть семейство нормальных распределений  $\mathcal{N}(\mu, \sigma^2)$  с двумерным параметром  $\theta = (\mu, \sigma)$  и параметрическим пространством  $\Theta = \mathbb{R} \times \mathbb{R}^+$ . В дальнейшем мы предположили, что значение  $\sigma$  известно, и свели наше параметрическое пространство к евклидовой прямой:  $\Theta = \mathbb{R}$  с  $\theta = \mu$ . Наконец, поскольку общее содержание серы измеряется в процентах, мы должны окончательно положить  $\Theta = (0; 100.)$ .

Во втором примере мы имели дело с бинарной случайной величиной  $X$ , принимающей значение 1 с вероятностью  $p$  и значение 0 с вероятностью  $1 - p$ . Таким образом, вероятностная модель представлялась семейством двухточечных распределений  $B(1, p)$  с  $\theta = p$  и параметрическим пространством  $\Theta = (0; 1)$ .

Следующий этап статистического исследования состоит в интерпретации решений  $d$  в терминах высказываний о соответствующих этому решению значениях параметра  $\theta$ . Это необходимо сделать, если мы поставили себе задачу количественного измерения последствий от принятия неверных решений, – в наших примерах риск используемых правил представлял собой функцию от  $\theta$ . Нетрудно понять, что в первом примере  $\mathcal{D} = \Theta$ , а во втором примере решению  $d_0$  о неэффективности метода соответствует подмножество параметрического пространства  $(0; 1/2]$ , а альтернативному решению  $d_1$  об использовании новой методики соответствует интервал  $(1/2; 1)$  возможных значений параметра  $\theta = p$ . Именно таким образом мы сводим конкретные задачи по аттестации партии дизельного топлива и выявлению эффективности нового метода лечения к абстрактным задачам математической статистики – оценке параметра (среднего значения)  $\theta$  нормального  $(\theta, \sigma^2)$  распределения и, соответственно, различению двух гипотез  $H_0 : \theta \in (0; 1/2]$  и  $H_1 : \theta \in (1/2; 1)$  о величине вероятности  $\theta$  успешного испытания в схеме Бернулли.

Параметрическая интерпретация решений позволяет статистику задать *потери*  $L(\theta, d)$ , которые он несет от принятия решения  $d$ , когда  $\theta$  представляет истинное значение параметра. Среднее значение этих потерь в длинном ряду однотипных статистических исследований с одним и тем же правилом принятия решения определяет величину риска, связанную с принятием неправильных решений. Так, в наших примерах риск определялся вероятностью принятия решения, отстоящего достаточно далеко от того решения, которое соответствовало истинному значению параметра, и, следовательно, *функция потерь* определялась индикатором некоторого подмножества в  $\Theta \times \mathcal{D}$ . Это так называемые функции потерь типа 0–1. В первом примере  $L(\theta, d) = 1$ , если  $|d - \theta| > \Delta$ , и  $L(\theta, d) = 0$  в противном случае. Во втором примере  $L(\theta, d) = 1$ , если принималось решение  $d_1$ , а  $\theta \in (0; 1/2]$ , или принималось  $d_0$ , а  $\theta \in (1/2; 1)$ , в остальных точках произведения пространств  $\Theta \times \mathcal{D}$  потери  $L(\theta, d)$  полагались равными нулю. Отметим, что в задаче оценки параметров довольно часто используется квадратичная функция потерь  $L(\theta, d) = |d - \theta|^2$ .

Каждое из решений  $d$  статистик принимает на основе результата  $x^{(n)} = x_1, \dots, x_n$  наблюдений над независимыми копиями  $X^{(n)} = (X_1, \dots, X_n)$  случайной величины  $X$ . Строится измеримое отображе-

ние  $\delta = \delta(\cdot)$  пространства возможных значений  $X^{(n)}$  в пространство решений  $\mathcal{D}$ , с помощью которого принимается решение  $d = \delta(x^{(n)})$ . Это отображение называется *решающей функцией* или *статистическим правилом*. Так, в первом примере  $\delta(X^{(n)}) = \bar{X}$ , а во втором

$$\delta(X^{(n)}) = \begin{cases} d_0, & \text{если } \sum_1^n X_k > 2, \\ d_1, & \text{если } \sum_1^n X_k \leq 2. \end{cases}$$

Последствия от использования конкретной решающей функции в длинном ряду однотипных статистических исследований определяются величиной средних потерь  $R(\theta; \delta) = \mathbf{E}_\theta L(\theta, \delta(X^{(n)}))$ , которая зависит от  $\theta$ ; функция  $R(\theta, \delta)$ ,  $\theta \in \Theta$ , называется *функцией риска*.

*Основная проблема математической статистики состоит в построении решающих функций  $\delta$ , минимизирующих равномерно по  $\theta \in \Theta$  функцию риска  $R(\theta; \delta)$ .* Мы будем решать эту проблему для задач оценки параметров и проверки гипотез. Естественно, будут также изучаться традиционные, возможно не обладающие оптимальными свойствами, статистические правила, и в этом случае нашей основной задачей будет вычисление их функций риска.

Представленная выше схема статистического вывода весьма далека от общности. Большинство статистических задач имеет дело с наблюдениями одновременно за несколькими объектами, например, новый метод лечения применяется к одной группе пациентов, в то время как другая подвергается лечению традиционным методом, и по данным наблюдений копий двух случайных величин делается вывод о предпочтительности нового метода. Если мы хотим сократить число наблюдений, необходимое для достижения заданной (малой) величины риска, то целесообразно не фиксировать заранее  $n$ , а планировать прекращение испытаний после наблюдения каждой копии в зависимости от полученных ранее результатов. Существует большой класс задач управления наблюдениями – оптимального выбора случайной величины, наблюдаемой на каждом шаге статистического эксперимента, а также правила прекращения наблюдений. Все это далеко выходит за рамки тех “кратких начатков” теории статистических выводов, которые будут представлены в нашем семестровом курсе.

Мы завершим этот параграф набором простейших определений и понятий, которые постоянно используются в математической статистике.

Итак, с исследуемым объектом, относительно которого мы должны принять некоторое решение  $d \in \mathcal{D}$ , соотносится наблюдаемая случайная величина  $X$ , распределение которой  $P_\theta$  известно с точностью до значения параметра  $\theta$ . Семейство распределений  $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ , как обычно, называется *вероятностной моделью*. Пусть  $(X, \mathcal{A})$  – измеримое пространство значений  $X$ . В дальнейшем будет всегда предполагаться, что на сигма-алгебре  $\mathcal{A}$  существует такая сигма-конечная мера  $\mu$ , что при любом  $\theta \in \Theta$  распределение  $X$  можно представить в виде интеграла

$$P_\theta(A) = \mathbf{P}(X \in A) = \int_A f(x | \theta) d\mu(x), \quad A \in \mathcal{A},$$

от плотности  $f(x | \theta)$  распределения  $X$  по мере  $\mu$ . В таком случае распределение независимых копий  $X^{(n)} = (X_1, \dots, X_n)$  случайной величины  $X$  на произведении  $(X^n, \mathcal{A}^n)$  измеримых пространств  $(X, \mathcal{A})$  определяется функцией плотности  $f_n(x^{(n)} | \theta)$  по мере  $\mu_n = \underbrace{\mu \times \dots \times \mu}_n$ , то есть

$$P_{\theta,n}(A_n) = \mathbf{P}(X^{(n)} \in A_n) = \int_{A_n} f_n(x^{(n)} | \theta) d\mu_n(x^{(n)}), \quad A_n \in \mathcal{A}^n.$$

**Определение 1.1.** Вектор  $X^{(n)} = (X_1, \dots, X_n)$  независимых, одинаково распределенных по тому же закону, что и наблюдаемая случайная величина  $X$ , случайных величин называется *случайной выборкой объема  $n$* . Измеримое пространство  $(X^n, \mathcal{A}^n)$  значений  $X^{(n)}$  называется *выборочным пространством*, а семейство распределений  $\mathcal{P}_n = \{P_{\theta,n}, \theta \in \Theta\}$  на этом пространстве – *статистической структурой* или *статистическим экспериментом*. Вектор  $x^{(n)} = (x_1, \dots, x_n)$  результатов наблюдения случайной выборки  $X^{(n)}$  называется *вектором (или совокупностью) выборочных данных*.

Зная распределение выборки, мы можем вычислять риск любого статистического правила  $\delta$  с помощью  $n$ -кратного интеграла

$$R(\theta; \delta) = \int_{\mathcal{X}} \dots \int_{\mathcal{X}} L(\theta, \delta(x^{(n)})) f_n(x^{(n)} | \theta) d\mu_n(x^{(n)}).$$

Конечно, если удастся найти распределение  $G_\theta$  решающей функции  $\delta$  на измеримом пространстве решений  $(\mathcal{D}, \mathcal{D})$ , то вычисление риска упрощается:

$$R(\theta; \delta) = \int_{\mathcal{D}} L(\theta, a) dG(a).$$

Так, в первом примере с выборкой из нормального  $(\mu, \sigma^2)$  распределения решающей функцией служило выборочное среднее  $\bar{X}$ . Было показано, что  $\bar{X}$  имеет нормальное  $(\mu, \sigma^2)$  распределение, и именно это обстоятельство позволило нам найти простое выражение риска статистического правила через функцию распределения стандартного нормального закона. Точно так же во втором примере с выбором из двухточечного распределения  $B(1, p)$  решающая функция была основана на случайной величине  $\sum_1^n X_k$ , которая имеет распределение Бернулли  $B(n, p)$ . Риск нашего решающего правила по выявлению эффективности метода лечения выражался через функцию распределения  $B(n, p)$ .

Заметим, что функции от выборочного вектор  $X^{(n)}$  играют важную, можно даже сказать самостоятельную, роль в математической статистике.

**Определение 1.2.** Любое измеримое отображение  $T = T(X^{(n)})$  выборочного пространства  $(\mathcal{X}^n, \mathcal{A}^n)$  в некоторое измеримое пространство  $(\mathcal{T}, \mathcal{B})$  называется *статистикой*.

Существует довольно устоявшийся универсальный набор статистик, постоянно используемых в теории и практике статистического вывода; распределения этих статистик интенсивно изучались на протяжении последних двух столетий. В следующем параграфе мы познакомимся с набором статистик, которые являются выборочными аналогами стандартных характеристик распределения наблюдаемой случайной величины, а также рассмотрим статистики, редуцирующие размерность выборочного вектора до размерности параметрического пространства без потери информации.