

1. ПРОБЛЕМА СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

1.1. Наблюдение и измерение

Одним из основных способов изучения окружающего мира, средством познания было и остаётся наблюдение. Под наблюдением принято понимать целенаправленное восприятие свойств процессов и явлений. Опираясь на результаты наблюдения, исследователь строит физическую модель новых процессов и явлений, выдвигает научные гипотезы, создаёт теории, принимает решения. Наблюдение, как правило, сопровождается определённым преобразованием и регистрацией информации о свойствах наблюдаемого объекта.

Для повышения эффективности процесса наблюдения могут использоваться различные технические средства. Наблюдение присуще любому эксперименту. Фактически, эксперимент – это и есть создание условий для наиболее эффективного наблюдения тех или иных свойств изучаемого объекта. Поэтому каждый экспериментатор является в первую очередь наблюдателем. Учитывая большую, во многих случаях решающую роль эксперимента в современных науках, особенно в технических и экономических, можно утверждать, что понятие эксперимента является фундаментальным.

Принятие любого решения невозможно без наличия необходимой информации. Так, в случае управления нужна информация о состоянии управляемого объекта, окружающей среды и цели управления. Такая информация не может быть получена без наблюдения. В то же время от наблюдения до принятия решения информация претерпевает, как правило, существенные преобразования. Наиболее характерными этапами таких преобразований являются измерения и обработка результатов измерений.

В широком смысле измерение состоит в сравнении наблюдаемой величины с эталоном и получении в результате этого её численного значения. Если результаты наблюдений представлены количественно, то имело место измерение. Нередко, кроме измерения, различают подсчёт. Данную операцию можно квалифицировать как завершающий этап наблюдения – регистрацию величин дискретного типа (число студентов в аудитории, количество проросших зёрен на единице площади посева, доход работника фирмы и т.п.). Тогда измерения можно рассматривать как

регистрацию величин непрерывного типа (вес, расстояние, скорость и т.д.).

Часто понятия наблюдения и измерения отождествляются. Представляется, что в рамках настоящей брошюры это не принципиально и вполне допустимо. Более того, без особой оговорки, используя как синонимы термины «наблюдение» и «измерение», будем включать в их содержание и подсчёт. Действительно, статистические методы обработки применимы к данным, получаемым как путём измерения, так и подсчёта. В дальнейшем, в силу указанных выше причин, все эти данные будут называться экспериментальными данными.

Говоря об экспериментальных данных, будем предполагать, что речь идёт о числовых величинах, векторах или функциях, т.е. о результатах количественных наблюдений, получаемых путём измерения или подсчёта. Если данные не могут быть представлены количественно, а являются качественными характеристиками, высказываниями, утверждениями, то их обработку следует предоставить специалистам по различным логическим методам.

Задачей обработки данных во многих случаях является принятие решения относительно значений определённых параметров (величин), характеризующих изучаемые явления или процессы. Обозначим эти параметры a_1, a_2, \dots, a_m . При этом возможны два случая. В первом из них непосредственно измеряются указанные величины. Говорят, что имеют место прямые измерения.

Во втором случае величины a_1, a_2, \dots, a_m непосредственно измерены быть не могут, а измеряются другие переменные x_1, x_2, \dots, x_n , с которыми функционально связаны величины a_1, a_2, \dots, a_m :

$$x_i = f_i(a_1, a_2, \dots, a_m), \quad i = \overline{1, n}. \quad (1.1.1)$$

Выражение (1.1.1) в векторной форме запишем как

$$X_{<n>} = F_{<n>}(A_{<m>}), \quad (1.1.2)$$

где $X_{<n>} = (x_1, x_2, \dots, x_n)^T$; $F_{<n>} = (f_1, f_2, \dots, f_n)^T$; $A_{<m>} = (a_1, a_2, \dots, a_m)^T$.

Эти измерения называются косвенными. В общем случае функциональные связи (1.1.1) являются нелинейными. Если же схема косвенных измерений линейна, то имеют место соотношения

$$x_i = \sum_{j=1}^m u_{ij} a_j, \quad i = \overline{1, n}$$

или в векторно-матричной форме

$$X_{<n>} = U_{[n;m]} A_{<m>}. \quad (1.1.3)$$

Во многих случаях свойства объекта, над которым ставится эксперимент, изменяются с течением времени. Тогда соответственно вместо соотношений (1.1.2) и (1.1.3) при непрерывных измерениях имеем

$$X_{\langle n \rangle} = F_{\langle n \rangle}(t; A_{\langle m \rangle}),$$

$$X_{\langle n \rangle} = U_{[n; m]}(t) A_{\langle m \rangle},$$

а при дискретных измерениях

$$X_{\langle n \rangle_i} = F_{\langle n \rangle}(t_i; A_{\langle m \rangle}),$$

$$X_{\langle n \rangle_i} = U_{[n; m]}(t_i) A_{\langle m \rangle}.$$

Предполагается, что непрерывные и дискретные измерения проводятся в моменты времени, лежащие внутри интервала $[0, T]$, т.е. $t \in [0, T]$, $t_i \in [0, T]$, $i = \overline{1, N}$.

Результат измерения всегда содержит ошибку (погрешность), представляющую собой отклонение результата измерения от истинного значения. Данное утверждение не относится в полной мере к подсчёту. Подсчитать количество интересующих объектов можно безошибочно. Так, в нормальных условиях можно точно определить количество компьютеров в интернет-классе. Редактор безошибочно может сосчитать число страниц рукописи и т.д. В то же время в некоторых условиях (дефицит времени, несовершенство средства наблюдения и др.) можно допустить ошибки при подсчёте. Например, с вертолёта трудно подсчитать точно количество коров в большом стаде.

Существующая в настоящее время теория ошибок измерений развита в основном применительно к оцениванию непрерывных величин. При этом ошибки принято разделять на систематические и случайные.

Систематические ошибки – это составляющие общей ошибки, вызываемые факторами, действующими одинаковым образом при многократном повторении одних и тех же измерений. Эти ошибки при повторных измерениях остаются неизменными или, если изменяются, то закономерно. Систематические ошибки, как правило, обусловлены погрешностями измерительных приборов (инструментальные ошибки) и несовершенством методов измерений (методические ошибки).

Случайные ошибки – составляющие общей ошибки, изменяющиеся случайным образом при повторных измерениях одной и той же величины. Причиной случайных ошибок являются неконтролируемые факторы, проявление которых неодинаково в каждом измерении и которые заранее не могут быть учтены. Другими словами, случайные ошибки проявляются тогда, когда при измерениях имеют место случайные события. Объективно такие события при наблюдениях и измерениях происходят всегда, как бы ни повышалась «чистота» эксперимента.

Среди случайных ошибок особо следует выделить грубые ошибки (промахи). Их характер и причины существенно отличаются от характера других случайных ошибок измерений. Основная масса случайных ошибок появляется при исправно работающих средствах измерений и пра-

вильных действия экспериментатора. Причиной появления грубых ошибок являются неисправность приборов или неточность в работе экспериментатора.

В данной брошюре рассматриваются методы обработки экспериментальных данных, содержащих случайные ошибки (исключая грубые). Обозначим вектор ошибок через $\hat{Z}(t)$. Тогда при их аддитивном учёте и прямых измерениях выражение вектора экспериментальных данных принимает вид

$$\hat{X}_{<m>} = A_{<m>} + \hat{Z}_{<m>} ; \quad (1.1.4)$$

в случае косвенных измерений

$$\hat{X}_{<n>} = F_{<n>}(A_{<m>}) + \hat{Z}_{<n>} . \quad (1.1.5)$$

При мультипликативном учёте ошибок можно записать

$$\hat{X}_{<m>} = (a_1 \hat{z}_1, a_2 \hat{z}_2, \dots, a_m \hat{z}_m)^T ; \quad (1.1.6)$$

$$\hat{X}_{<n>} = (\hat{z}_1 f_1(A_{<m>}), \hat{z}_2 f_2(A_{<m>}), \dots, \hat{z}_n f_n(A_{<m>}))^T . \quad (1.1.7)$$

1.2. Обобщённая модель обработки экспериментальных данных

Цель обработки данных эксперимента заключается в получении из них сведений о свойствах изучаемого объекта или процесса. На заключительном этапе экспериментатор принимает решение относительно этих свойств. Данное решение может быть связано, например, с оцениванием конкретных значений величин, характеризующих свойства объекта, а также с проверкой предположений о нахождении этих величин в определённых пределах, предположений о возможных законах распределения и т.д. Следует отметить, что процедура оценивания связана с получением оценок характеристик объекта. Под **оценкой** будем понимать приближённое значение оцениваемой величины, которое целесообразно принимать за её истинное значение. На рис.1.1 приведена обобщённая модель обработки экспериментальных данных.

Состояние (свойства) исследуемого объекта или процесса абстрагируется пространством ситуаций $\{E\}$. При этом под $E \in \{E\}$ в общем случае понимается абстрактный параметр, характеризующий ситуацию, т.е. состояние объекта (фазовые координаты, структуру и т.д.). Абстрактный параметр может выражаться числом, функцией, функционалом, оператором, отношением, событием и т.п. Пространство ситуаций при оценивании вектора $A_{<m>}$ представляет множество всех возможных его значений. Экспериментатора интересует истинное значение вектора $A_{<m>}$.

В процессе наблюдений регистрируется случайный вектор \hat{X} , который связан с вектором A выражениями (1.1.4) – (1.1.7). В общем виде истинную связь обозначим соотношением $X = N(E)$. Случайные воздействия (помехи) \hat{Z} , образующие пространство воздействий $\{\hat{Z}\}$, искажают эту связь и она принимает вид $\hat{X} = N(E; \hat{Z})$.

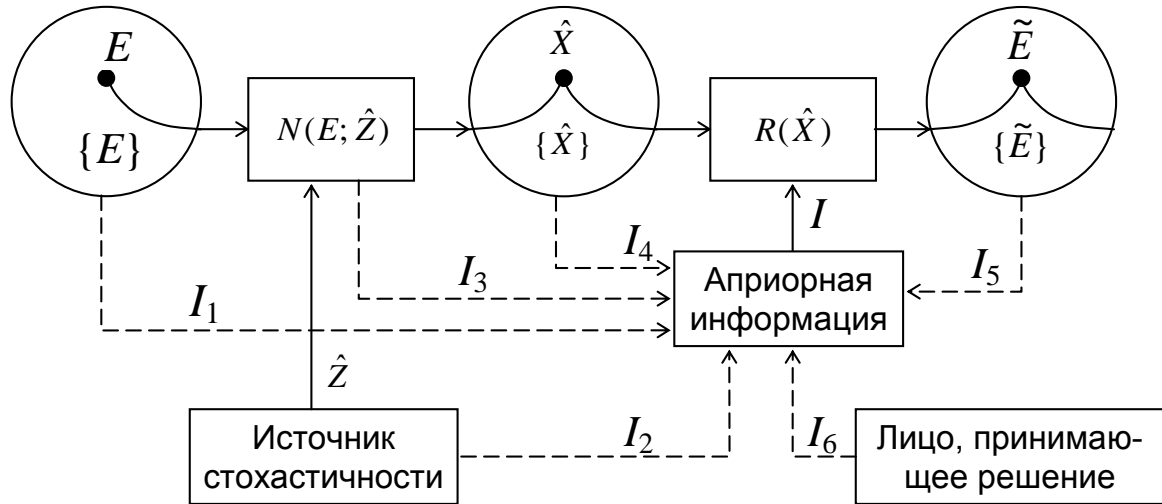


Рис.1.1. Обобщённая модель обработки экспериментальных данных

Помехи \hat{Z} генерируются источником стохастичности. Операция над параметром E и случайной величиной (процессом) \hat{Z} , в результате которой формируется вектор \hat{X} , обозначена на схеме $N(E; \hat{Z})$.

Множество $\{\hat{X}\}$ всех значений \hat{X} , которые могут быть реализованы при всех возможных значениях E и \hat{Z} , составляет пространство наблюдений. Информация о параметре E из вектора \hat{X} получается при обработке последнего. Статистическую процедуру преобразования над вектором \hat{X} обозначим через $R(\hat{X})$. Величина I характеризует всю совокупность априорной информации, которая используется в статистической процедуре совместно с результатами наблюдений \hat{X} . Следует подчеркнуть, что эта информация известна до начала эксперимента. Результатом преобразования величин \hat{X} и I является принятие решения о параметре E . Поэтому функцию R можно назвать решающей функцией.

Как видно из рис.1.1 (пунктирные линии), априорная информация $I_1, I_2, I_3, I_4, I_5, I_6$ может отражать определённые свойства пространства ситуаций, источника стохастичности, оператора N , пространства наблюдений, пространства решений и требований лица, принимающего решение.

Множество всех решений $\tilde{E} = R(\hat{X})$ образует пространство решений $\{\tilde{E}\}$. Идеальный исход обработки экспериментальных данных сво-

дится к тождеству $\tilde{E} \equiv E$. Однако на практике это тождество оказывается нереализуемым. Решение \tilde{E} является случайной величиной (вектором, функцией), с помощью которой экспериментатор оценивает интересующую его величину E . При этом степень близости данной оценки к истинному значению E определяется как характером априорной информации об условиях эксперимента и о задаче обработки данных, так и особенностями процедуры принятия решений.

Выше указывалось, что помехи и регистрируемые в процессе наблюдения сигналы представляют собой случайные объекты (величины, векторы, функции). Из теории вероятностей известно, что самой полной характеристикой случайного объекта является закон распределения. Наиболее употребительны формы закона распределения в виде функции распределения $F_{\hat{x}}(x)$ и плотности распределения $\varphi_{\hat{x}}(x)$. В рассматриваемом случае принципиальное влияние на выбор способов обработки и качества обработки имеют законы распределения на пространствах $\{\hat{Z}\}$ и $\{\hat{X}\}$, а также оператор $N(E; \hat{Z})$. Обычно $N(E; \hat{Z})$ предполагается известным. Наиболее распространён случай, когда по отношению к помехе \hat{Z} оператор N является оператором суммирования (помеха аддитивная).

В зависимости от того, известны законы распределений на пространствах $\{\hat{Z}\}$ и $\{\hat{X}\}$ или нет, статистические методы обработки экспериментальных данных подразделяются на параметрические (классические) и непараметрические.

Параметрические методы используются тогда, когда законы распределений на указанных пространствах известны. Однако, на практике они не всегда априори известны. При постановке экспериментов для исследования новых явлений и процессов истинные законы распределений могут быть неизвестны вообще. В других случаях условия эксперимента и характер помех настолько сложны и нестабильны, что трудно говорить о конкретных законах непосредственно в период проведения экспериментальных работ. Так, например, в задачах приёма и обработки телеметрической информации о техническом состоянии бортовых устройств космических аппаратов характер помех меняется в зависимости от времени суток, взаимного расположения пункта приёма и аппарата на орбите, наличия целенаправленных возмущающих воздействий. В настоящее время проблема обработки экспериментальных данных при неизвестных законах распределений решается двумя путями.

Во-первых, используется минимаксный подход. При этом решаемая задача фактически сводится к параметрической, так как данный подход приводит к нахождению закона распределения на $\{\hat{Z}\}$, в некотором

смысле «наилучшего среди плохих» (максимин) или «наихудшего среди хороших» (минимакс).

Во-вторых, применяются методы непараметрической статистики.

Во многих научных дисциплинах (исследование операций, теория принятия решений, теория игр и т.д.) по информированности специалиста, решающего задачу, различают три уровня – детерминированный, статистический и неопределённый.

Наиболее простым является первый уровень – детерминированный, когда условия решения задачи известны полностью и случайные факторы отсутствуют. Применительно к обобщённой модели на рис.1.1 это означает, что операторы N и R точно известны, Z – неслучайная величина, все измерения проводятся абсолютно точно. Крайний случай этой ситуации – достоверно известно значение параметра E . Говорят, что решение задачи на детерминированном уровне производится в условиях определённости.

На втором, стохастическом уровне известны множество возможных ситуаций и априорное вероятностное распределение случайных факторов и этих ситуаций. В обобщённой модели обработки экспериментальных данных (см. рис.1.1) это касается законов распределений величин \hat{Z} и \hat{X} . На стохастическом уровне информированности специалист находится в «условиях риска».

Наконец, на уровне неопределённости известно множество ситуаций, но отсутствует априорная информация о распределениях. Принято говорить, что задача решается в условиях неопределённости.

Последние два уровня обобщаются термином «условия неполной информации». Статистические методы обработки экспериментальных данных используются именно в условиях неполной информации.

При этом параметрические методы в основном используются на стохастическом уровне, а непараметрические – на уровне неопределённости. Утверждение «в основном» подчёркивает, что в принципе и параметрические и непараметрические методы при необходимости могут быть использованы на обоих рассматриваемых уровнях неполной информации.

Многие задачи обработки данных решаются в условиях, когда относительно исходной информации (характеристик исследуемого объекта, ошибок измерений и т.д.) выдвигаются определённые гипотезы. В действительности эти гипотезы могут отличаться от фактических, реально существующих характеристик. Иначе говоря, модель может оказаться неадекватной реальным процессам. В связи с этим необходимо отметить, что различные статистические процедуры обладают различной чувствительностью к такой неадекватности. Чувствительность означает степень

влияния экспериментальных данных на результаты их обработки. Поскольку непараметрические методы требуют наименьшего объёма априорной информации, то очевидно, что они и менее чувствительны к искажениям исходных данных.

Выше отмечалось, что экспериментальная информация может моделироваться с помощью случайных величин, векторов или функций. Случайная функция представляет собой наиболее общую модель наблюдаемого сигнала. Если аргументом случайной функции является время, то для неё используется термин «случайный процесс». Значительная часть результатов экспериментов описывается случайными процессами.

При наиболее распространённой аддитивной ошибке модель непрерывного сигнала имеет вид

$$\hat{X}(t) = \hat{F}(t; A) + \hat{Z}(t),$$

где $\hat{Y} = \hat{F}(t; A)$ – полезная составляющая наблюдаемого сигнала; $A_{\langle m \rangle}$ – вектор оцениваемых параметров.

Характер полезной составляющей может быть использован как один из признаков для классификации моделей наблюдения. При этом можно выделить следующие типы полезных сигналов: детерминированный сигнал с неизвестными параметрами, случайный сигнал с известной функцией распределения (с точностью до параметров) и случайный сигнал с неизвестной функцией распределения (функция распределения может быть задана только классом распределений).

1.3. Сущность выборочного метода

При рассмотрении обобщённой модели обработки экспериментальных данных отмечалось, что пространство наблюдений $\{\hat{X}\}$ представляет собой множество реализаций вектора \hat{X} , или выборочное множество. Остановимся подробнее на понятии выборки, так как оно играет фундаментальную роль в статистических методах обработки информации как параметрических, так и непараметрических.

При экспериментальных исследованиях закономерностей в массовых случайных явлениях предполагается, что опыты могут быть повторены большое число раз при одинаковых условиях. В каждом опыте регистрируется определённый признак изучаемого объекта. Различают общий и основной признаки. **Общим признаком** называется свойство, по которому объекты объединяются в однородные совокупности, а **основным признаком** – свойство объектов, исследуемое в данном эксперименте. Под однородной будем понимать совокупность, все элементы которой

являются реализациями одной и той же случайной величины (функции) с одним и тем же законом распределения.

Если производится исследование веса совокупности однотипных деталей (например, гаек), то тип деталей (гайки) характеризует их общий признак, а вес деталей – основной признак.

Отдельное конкретное значение наблюдаемого основного признака называется его реализацией или вариантом. При статистическом исследовании вероятностных свойств совокупности объектов нет возможности производить опыты над каждым из них. Так, при изучении роста ежемесячных доходов граждан РФ невозможно установить доход каждого гражданина за ограниченное время. Для определения всхожести зерновых перед посевной бессмысленно пытаться обследовать каждое отдельное зерно. Всё же, несмотря на это, существует метод, который позволяет изучить интересующие свойства всей совокупности объектов. Речь идёт о выборочном методе, согласно которому основные признаки совокупности объектов изучаются по некоторой её части, называемой выборкой. Более строго в математической статистике выборкой называют совокупность наблюдаемых реализаций основного признака.

Совокупность всех возможных объектов (вариантов), из которых производится выборка, называется генеральной совокупностью. Следует отметить, что выборка является однородной, если все её элементы извлечены из одной генеральной совокупности.

Обычно предполагается, что выборки формируются при многократной реализации случайного эксперимента, результат которого нельзя заранее точно предсказать. Поэтому такие выборки называются случайными.

Пусть количественно исследуемый основной признак описывается случайной величиной \hat{x} . Допустим, что в процессе эксперимента получена последовательность из n значений x_1, x_2, \dots, x_n случайной величины \hat{x} . До проведения эксперимента эта последовательность является случайной выборкой и обозначается $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$.

Если основой признак наблюдаемого объекта описывается случайной функцией $\hat{x}(t)$, то в процессе эксперимента получаем конечную совокупность реализаций $x_1(t), x_2(t), \dots, x_n(t)$, $t \in [0; T]$ случайной функции. Случайная выборка представляет в этом случае последовательность случайных функций $\hat{x}_1(t), \hat{x}_2(t), \dots, \hat{x}_n(t)$. Часто выборку целесообразно описывать с помощью n -мерного случайного вектора $\hat{X}_{<n>}$, компонентами которого являются элементы упорядоченной последовательности $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n$.

Случайный характер выборки выражается в том, что нельзя заранее предсказать возможные значения её элементов, и любые две последовательности из n наблюдаемых значений в общем случае будут различными. В конкретных прикладных задачах элементы выборок представляют собой реализации случайных величин, т.е. детерминированные величины. Таким образом, априорно (до проведения эксперимента) выборка будет случайной, а апостериорно (после проведения опыта) – неслучайной.

Число n элементов выборки является конечным и называется **объёмом выборки**. Число элементов генеральной совокупности может быть конечным или бесконечным.

В статистике различают повторные и бесповторные выборки. Выборка называется **повторной**, если отобранный объект после испытания перед отбором следующего объекта снова возвращается в генеральную совокупность. Выборка называется **бесповторной**, если отобранный объект после испытания не возвращается в генеральную совокупность.

С позиций теории вероятностей элементы случайной выборки рассматриваются как независимые случайные величины с одной и той же функцией распределения $F_{\hat{x}}(x)$ и плотностью распределения $\varphi_{\hat{x}}(x)$. Последнее означает, что для плотности распределения случайного вектора $\hat{X}_{<n>}$ имеет место равенство

$$\varphi_{\hat{X}}(X_{<n>}) = \prod_{i=1}^n \varphi_{\hat{x}_i}(x).$$

Далеко не всякая выборка адекватно отражает свойства генеральной совокупности. Убедительный пример: требуется оценить средний рост жителей некоторого города, а в качестве выборки исследователю предлагают городскую баскетбольную команду. Нетрудно понять, насколько будет искажён результат.

Говорят, что выборка должна быть **представительной** или **репрезентативной**. Выборка представительна, если все элементы генеральной совокупности имеют одинаковую вероятность быть выбранными. Чтобы обеспечить это, не имея никаких сведений о генеральной совокупности, можно полагаться только на случайность отбора объектов в выборку. Все прочие способы будут необъективными, носящими следы влияния посторонних факторов. И семена для проверки всхожести, и жителей для оценивания среднего роста – всё нужно отбирать совершенно случайным образом. Иное дело, если экспериментатор заранее знает, что генеральная совокупность состоит из нескольких классов, различных по своим характеристикам. При таких условиях выборку лучше делать из каждого класса в отдельности. Например, изучая рост жителей, целесо-

образно делать отдельную выборку мужчин, отдельную женщин; при этом можно учесть возраст, профессию.

Из случайного характера выборок неопровержимо следует, что **любое суждение о генеральной совокупности по выборке само является случайным**. Имеется в виду суждение, затрагивающее хотя бы один элемент генеральной совокупности, не попавший в выборку.

А какова же связь между наблюдениями, отбираемыми в состав экспериментальных данных, и выборочным методом? Имеется случайная величина \hat{x} и в результате n независимых испытаний получают n её допустимых значений. Если все допустимые значения случайной величины \hat{x} считать генеральной совокупностью, то полученные при наблюдениях n значений образуют выборку. По этой выборке изучаются свойства случайной величины \hat{x} . Итак, **производство наблюдений является частным случаем выборочного метода, когда в качестве генеральной совокупности берутся все допустимые значения некоторой случайной величины и исследуются свойства этой величины**.

Если требуется исследовать несколько основных признаков, например рост и вес жителей города, три размера прямоугольных деталей, то экспериментатор наблюдает векторную случайную величину $\hat{X}_{<q>}$. При этом выборка $\hat{X}_{<q>1}, \hat{X}_{<q>2}, \dots, \hat{X}_{<q>n}$ описывается nq -мерным случайным вектором $\hat{X}_{<nq>}$.

Как уже отмечалось, случайные выборки используются для изучения свойств генеральной совокупности. При этом обычно элементы выборки используются для образования по соответствующим правилам новых случайных величин вида

$$\hat{s}_j = f_j(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n), \quad j = 1, 2, \dots$$

которые называются **статистиками**. Примерами статистик являются

выборочная сумма $\hat{s}_1 = \sum_{i=1}^n \hat{x}_i$, выборочное среднее $\hat{s}_2 = \frac{1}{n} \left(\sum_{i=1}^n \hat{x}_i \right)$ и т.д.

Все свойства (характеристики) генеральной совокупности, получаемые на основе выборки, называются **выборочными** или **статистическими** (в отличие от теоретических, изучаемых в теории вероятностей). В дальнейшем все выборочные характеристики будут отмечаться индексом *. Так, $F_{\hat{x}}^*(x)$ обозначает статистическую функцию распределения случайной величины \hat{x} ; $M_{\hat{x}}^* = \bar{x}^*$ – статистическое математическое ожидание; $D_{\hat{x}}^*$ – статистическую дисперсию и т.д. Все выборочные характеристики представляют собой функции элементов выборки, т.е. статистики.

В выборочном методе большое внимание уделяется изучению законов распределения различных статистик, статистических рядов. Теоретической основой выборочного метода являются предельные теоремы теории вероятностей, которые приведены в виде, адаптированном к рассматриваемой тематике, в § 2.2.

1.4. Задачи и методы обработки экспериментальных данных

Как уже отмечалось, основная цель обработки экспериментальных данных состоит в получении определённых сведений об исследуемом объекте (процессе, явлении). Особенности процесса обработки определяются характером решаемых задач и объёмом информации, получаемой в ходе эксперимента. По указанным признакам обработка может быть первичной и вторичной.

Задачами первичной обработки являются выделение полезного сигнала на фоне помех (шумов), сжатие данных и приведение их к системе измерений, пригодной для дальнейшей обработки или отображения. Наиболее характерной операцией выделения полезного сигнала является устранение грубых ошибок.

Задачи вторичной обработки могут быть распределены в две группы. Первая группа задач сводится к построению математических моделей реальных процессов и явлений, вторая группа – к анализу таких моделей. Математическая модель – это абстрактное информационное отражение реального процесса или явления на языке математики.

В кибернетике (как технической, так и экономической) и теории управления процедура построения математических моделей объектов управления по результатам наблюдения их входных и выходных процессов называется идентификацией. Параллельно с теорией идентификации в кибернетике развивается теория распознавания образов, которая также связана с проблематикой построения математических моделей по результатам обработки экспериментальных данных. Распознавание образов представляет собой задачу обработки данных, в процессе которой делается вывод о принадлежности распознаваемого образа к определённому классу. Этот класс и определяет вид искомой модели.

Наиболее типовыми задачами построения математических моделей на основе статистических методов являются задачи оценивания параметров (например, параметров законов распределения случайных объектов), оценивания неизвестных функциональных зависимостей (например, законов распределения), проверки гипотез, построения уравнений регрессии и распознавания образов.

В задачах анализа моделей производится оценка влияния множества факторов на конечный результат и выбор наиболее важных факторов, а также исследуется структура экспериментальных данных и построенных на их основе математических моделей.

Один из возможных вариантов перечня задач, решаемых при первичной и вторичной обработке показан на рис.1.2.

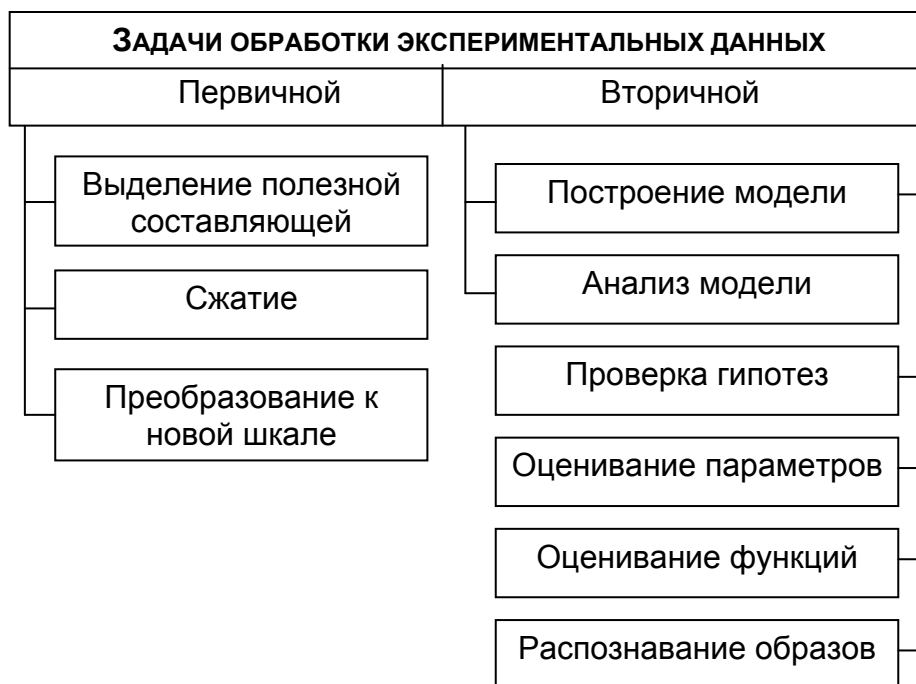


Рис.1.2. Классификация задач обработки экспериментальных данных