**ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ-
ВЫСШАЯ ШКОЛА ЭКОНОМИКИ**

САНКТ-ПЕТЕРБУРГСКИЙ ФИЛИАЛ

А.И. Иванов, Р.С. Минвалеев

ИНФОРМАЦИОННЫЙ ПОДХОД К АНАЛИЗУ СИСТЕМ

ЧАСТЬ I.

ОСНОВЫ МОДЕЛИРОВАНИЯ И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ

Санкт-Петербург
2006

УДК 519.2
ББК 22.17

Иванов А.И., Минвалеев Р.С.
Информационный подход к анализу систем: Основы моделирования и первичная обработка данных: учебное пособие для студентов вузов. Ч.I. СПб филиал ГУ – ВШЭ, 2006 - 100 с.

Под ред. Ю.И. Рейнова, к.ф-м.н.

В пособии изложены отдельные ставшие классическими методы первичной обработки результатов наблюдений, применимые к решению задач естествознания и экономики.

Пособие может быть использовано при изучении курсов «Математическая статистика» и «Экономическая статистика».

Пособие рекомендовано студентам старших курсов вузов, аспирантам и научным работникам.

Рекомендовано к печати в качестве учебного пособия Учебно-методическим советом СПб филиала ГУ-ВШЭ для студентов и слушателей программ высшего профессионального образования

Содержание

Предисловие

Настоящее пособие содержит первую часть материала, входящего в раздел «Информационный подход к анализу систем» годового курса дисциплины «Теория систем и системный анализ». Пособие включает материал как теоретического, так и рецептурного характера. В силу ориентации содержания на освоение читателем методов решения прикладных задач, внимание акцентировано на рецептурной части. В расширенном виде курс «Информационный подход к анализу систем» читался А.И. Ивановым на факультете прикладной математики – процессов управления СПбГУ. В сокращенном виде курс читался Р.С. Минвалеевым на курсах постдипломного образования в СПбГУ. Материалы пособия служат основой для самостоятельного и под руководством научного руководителя выполнения научно-исследовательской работы, оформления ее результатов и публичной защиты. Пособие должно облегчить обращение к специальной литературе, в которой затронутый материал изложен подробнее.

Дисциплина «Информационный подход к анализу систем» достаточно автономна и может читаться в любом семестре на старших курсах. Однако предпочтительно – на этапе обучения, на котором читатель знаком с содержанием курса «Высшей математики» и «Теории вероятностей и математической статистики», прочитанных в объеме, утвержденном в Государственных требованиях для естественнонаучных и экономических факультетов вузов.

В главе 1 содержатся краткие сведения по теории вероятностей и математической статистике, необходимые для успешного решения задач первичной обработки данных. Изложены сведения о часто применяемых в процессе первичной обработки данных законах распределения и методах нахождения значений характеристик по данным выборок.

В главу 2 помещены краткие сведения по часто применяемым статистическим критериям и методам их использования при решении задач первичной обработки данных.

В главе 3 изложены сведения о традиционных и новых методах нахождения законов распределения использованием выборочного материала. Особое внимание уделено применению результатов теорем В.И. Зубова.

Глава 4 содержит начальные сведения, необходимые для успешного применения при решении прикладных задач метода статистических испытаний.

В продолжение и развитие подхода П.Л. Чебышева изложенный в пособии материал доведен до уровня возможностей вычислительного применения.

Для более глубокого ознакомления читателей с содержащимся в пособии материалом в процессе изложения даны многочисленные ссылки на научную литературу, список которой помещен в конце книги. В целях удобства читателей в необходимых случаях указаны страницы работ.

Нередко процесс решения рассматриваемых задач подразумевает значительные объемы вычислительной работы. В таких случаях мы рекомендуем пользоваться компьютером. В целях улучшения качества обучения мы отказались от рекомендаций по применению мощных систем символьной математики, позволяющих без должного изучения методов математического моделирования находить всевозможные решения, использование которых логично, но не убедительно. Мы ограничились рекомендациями по применению достаточно мощной, но требующей самостоятельного введения в компьютер подходящих аналитических выражений системой символьной математики «Derive». Возможно, по указанной причине системой не достаточно популярной. Вместе с тем найти *«Derive»* на рынке т. н. «софта» в СПб не представляет трудностей, т. е. общедоступность приобретения в личное пользование очевидна. Свидетельством в пользу правомерности нашего подхода – использования в процессе обучения математическому моделированию т. н. «нестатистических» систем служит содержание учебника [2]. Весьма подробные сведения по применению рекомендуемой нами системы символьной математики *«Derive»* можно найти в работе [16].

Авторы благодарны сотруднику Библиотеки Российской Академии Наук С.Ю. Угольниковой за помощь в поиске книг и работ, использованных в процессе написания пособия. Авторы признательны заведующему кафедрой высшей математики СПб филиала Высшей Школы экономики Ю.И. Рейнову за плодотворное обсуждение материалов в процессе написания пособия.

Вводные замечания к курсу

В современной математике отправным пунктом служит понятие множества. Понятие множества не поддается точному определению, поэтому ограничимся описанием, заимствованным из пособия [17].

Определение 1. ▎*Множеством называется совокупность объектов произвольной природы, объединенных по какому-нибудь признаку.* ▎

Например, можно говорить о множестве всех натуральных чисел, множестве всех точек прямой, множестве всех многочленов с вещественными коэффициентами и т. п. Говоря о множестве, мы считаем, что относительно всякой вещи верно одно и только одно из двух: вещь либо входит в наше множество в качестве его элемента, либо не входит. В качестве второго отправного пункта будем использовать понятие математической модели. Дадим полезное определение.

Определение 2. ▎*Математической моделью будем называть совокупность аналитических выражений, достаточно адекватно описывающих объект исследования.* ▎

В интересах читателя уточним понятие модель, заимствовав разъяснение из монографии [2, с.12]. В настоящее время различают три уровня моделирования: концептуальный, структурный, математический. Разъясним содержание уровней подробнее, воспользовавшись общеизвестными примерами.

Концептуальный и концептуально-структурный уровень моделирования заключается в том, что по мере накопления знаний человек выдвигает различные принципы (концепции), применение которых, по его мнению, позволяет более кратко, просто и в более доступной форме объяснить наблюдаемые явления. Классическим примером концептуально-структурной модели является геоцентрическая модель К. Птолемея (ок. 178-100 гг. до н. э.), согласно которой Земля является центром Вселенной; Солнце, звезды и планеты вращаются вокруг Земли. Другим, тоже классическим, примером концептуально-структурной модели является гелиоцентрическая модель Н. Коперника (1473-1543) – Земля и Солнце меняются местами.

Очертим границы применимости концептуально-структурного моделирования в естествознании и в экономике. Разработка модели начинается со сбора, накопления и обобщения сведений о процессах, происходящих в системе, описания признаков и характеристик элементов, выяснения объективных законов взаимодействия элементов друг с другом и внешней средой, выбора совокупностей компонент векторов внутренних и входных воздействий, вектора состояний системы, выходного вектора, разработки схемы системы и т. д. Выдвигаются концепции функционирования системы, принимаются допущения и пренебрежения. И только после выполнения достаточно большого объема работы по сбору сведений и их первичной обработке, первичному анализу концепций функционирования, приступают к поиску математических моделей. Известно, что в результате работ (см., например, книгу [3]) Н. Винера (1894-1964) по усовершенствованию структурных моделей найдена схема с обратной связью (схема Винера. – А. И.). Отличительной чертой схемы с обратной связью является объединение объекта управления, измерительных устройств и исполнительных органов в одну подсистему. Подтверждением продуктивности схемы Винера является найденное с ее помощью доказательство общности процессов управления, происходящих в машинах и живых организмах. Это позволило по-новому подойти к решению задачи синтеза автоматических систем управления, в том числе решить задачу создания компьютеров.

Наличие концептуальной и структурной модели позволяет перейти к этапу нахождения модели математической. Воспользовавшись содержанием монографии [2], (см. с.16), запишем одно из возможных определений математической модели.

Определение 3. ▌ *Математической моделью называется приближенное количественное описание системы в определенных условиях и принятых допущениях, выполненное с помощью аналитических выражений.* ▌

Классическими методами нахождения математических моделей в настоящее время принято считать методы, в которых используются объективные законы естествознания, выраженные в математической форме, т. е. в форме дифференциальных, разностных, интегральных, алгебраических и др. уравнений и неравенств, логических и функциональных соотношений. Адекватно описывающие объект исследования математические модели могут быть найдены только после предварительного достаточно тщательного качественного и количественного изучения свойств объекта. Изложим одну из возможных классификаций математических моделей по степени общности и детализации, заимствованную из книги [2].

Математические модели можно разделить на следующие классы:

1. Математические модели процессов и ситуаций.
2. Прикладные математические модели.
3. Математические задачи.

Поясним подробнее содержание классов моделей. Модели класса «Математическая задача» содержат формулировку задачи, в которой указаны известные и неизвестные величины и соотношения их связывающие, численные значения известных величин, точную формулировку того, что требуется найти, установить или определить. Например, дана система линейных алгебраических уравнений с постоянными коэффициентами. Требуется найти ее решение.

Модели класса «Прикладные математические модели» также содержат входные и выходные величины и связывающие их соотношения. Но при этом не имеется указаний, какие именно величины являются известными, а какие неизвестными. Дополнительно, в общем виде дано указание на предполагаемый перечень задач, которые можно сформулировать и решить использованием данной прикладной математической модели.

Модели процессов и ситуаций содержат достаточно полный и общий набор математических соотношений, описывающих физические, химические, экономические и др. законы, применение которых позволяет на их основе разработать прикладную математическую модель для постановки и решения требуемого комплекса задач. Классическими примерами моделей класса «Математические модели процессов и ситуаций» в экономике являются модели линейной, нелинейной и динамической оптимизации процессов управления. В отличие от популярных заявлений, авторы которых пытаются связать фактические явления более или менее правдоподобными концепциями, применение математического моделирования предполагает длительную научную работу по отбрасыванию противоречивых концепций и выдвижению новых идей, позволяющих доказательно описывать ход реальных процессов. Многочисленные примеры успешно найденных моделей процессов и ситуаций, применяемых при решении экономических задач в развитых странах, можно найти в книгах [4], [5].

В свою очередь каждый из классов математических моделей можно разделить на модели детерминированные и стохастические (вероятностные). Ознакомится с весьма общими методами нахождения математических моделей детерминированных систем можно, воспользовавшись, например, книгой [7] и мн. др. В частности, весьма подробное описание методов нахождения математических моделей систем управления изложено в учебном пособии [12]. Обзор методов математического моделирования стохастических систем выполнен в книге [7]. В сжатом виде принципы нахождения большинства стохастических моделей описаны в математической энциклопедии по теории вероятностей и математической статистике [8], более подробно в трехтомном справочном издании «Прикладная статистика» [9 – 11] под ред. С.А. Айвазяна. В доступном для широких кругов читателей изложении основы математического моделирования стохастических систем можно найти в учебных пособиях [12], [13]. Примеры применения результатов стохастического моделирования при решении задач обследования населения – в учебном пособии [14].

Сообщим краткие сведения о границах применимости математического моделирования. Перечень типичных возражений против применения математических моделей можно найти в книге [9, с.66]. Следуя содержанию книги [9], перечислим некоторые из них и ответы на возражения. От некоторых представителей, например, таких областей, как медицина, социальные и экономические науки, до сих пор иногда можно услышать, что изучаемые ими явления слишком сложны для адекватного отражения их математическими средствами. Не отрицая определенной специфичности и сложности явлений, изучаемых в этих областях, следует твердо сказать, что продуктивной альтернативы использованию языка математических моделей и связанному с ними определенному упрощению и схематизации действительности нет.

При этом основным лимитирующим фактором является не ограниченность возможностей современного математического аппарата, но ограниченные возможности человеческого разума в изучении сложных ситуаций. Действительно, не стоит строить слишком сложные математические модели, следствия из которых мы просто не в состоянии охватить и осмыслить.

Одним из методических приемов, позволяющих обойти сверхупрощение при изучении сложных явлений, является использование нескольких моделей одновременно. При этом каждая из моделей может рассматриваться как частный случай более общей модели, необходимой для адекватного описания действительности. Второе возражение против использования математических моделей – применение в моделях вероятностного описания. Среди недостаточно информированных оппонентов до сих пор бытует мнение, что рассмотрение человека как представителя некоторой массовой совокупности, как индивидуума, чьи реакции описываются вероятностными законами, эквивалентно лишению его индивидуальности и свободы выбора.

В действительности же именно индивидуальность человека и непредсказуемость выбора диктует то, что поведение его должно описываться в вероятностных терминах. Единственной альтернативой вероятностному подходу является подход детерминистский, но именно он лишает человека индивидуальности. Вместе с тем при описании массовых явлений в больших коллективах детерминистские модели, упрощенно описывающие поведение одного субъекта, оказываются чрезвычайно продуктивными в применении к группе. В качестве примера можно привести результаты модели распространения эпидемии гриппа [15] в СССР. Третье возражение, по нашему мнению самое коварное, сформировалось вследствие массового распространения компьютеров. Отрицаются не математические модели сами по себе, а целесообразность исследования математическими средствами.

В качестве исчерпывающей характеристики сложившейся ситуации процитируем фрагмент из введения к книге [16]: «Основная причина пониженного интереса студентов к знаниям – невостребованность знаний обществом. Считается, что в условиях рыночной экономики и бизнеса глубокие знания не нужны, достаточно четырех арифметических действий и умения работать на персональном компьютере в среде так называемых офисных программ. По этой причине... знания многих студентов можно охарактеризовать, как посредственные знания о многом».

Значительное место в содержании возражений против изучения методов математического моделирования занимают подмены интерпретаций трактовками и авторскими суждениями. С целью избежать недоразумений, воспользовавшись «Математическим энциклопедическим словарем» [1], запишем определение.

Определение 4. ▍ *Интерпретация – сопоставление всем исходным понятиям и отношениям данной аксиоматической теории T некоторых математических объектов и отношений между ними.* ▍

В результате интерпретации каждому утверждению F теории T естественным образом соответствует высказывание F^* об объектах системы S. Конечной целью интерпретации является задание смысла формул. При этом всякой формуле F данного языка сопоставляется по определенным правилам некоторая формула F^* другого языка и считается, что формула F означает то же, что и формула F^*. Освежить память по алгебре высказываний заинтересованный читатель может, воспользовавшись учебным пособием [17].

Хрестоматийным примером успешной интерпретации экономических явлений применением математического моделирования может служить модель Леонтьева многоотраслевой экономики. Описание модели можно найти в учебнике [19].

Перейти к этапу нахождения математической модели можно только при наличии концептуальной и структурной модели. Однако сам процесс поиска адекватных концептуальных и структурных моделей включает в себя сбор достаточно большого количества сведений об изучаемом явлении. Одним из ключевых методов, применение которого позволяет принять адекватное решение о достаточности количества сведений, является вероятностно-статистический метод. В качестве меры в вероятностно-статистическом методе используется мера, называемая вероятность. Подробные сведения по теории меры заинтересованный читатель может найти, воспользовавшись учебником [21]. Необходимые для изучения нашей дисциплины сведения о вероятностной мере будут изложены во второй главе. Пока же дадим полезное для понимания дальнейшего изложения определение.

Определение 5. ▌ *Назовем область изучения достигшей количественной зрелости, если количество собранных в ней результатов наблюдений и экспериментов использованием принятой меры можно оценить, как достаточное для решения задачи математического моделирования.* ▌

В настоящее время можно считать, что ряд областей экономики достиг необходимого для применения математического моделирования уровня количественной зрелости. Примером служит известная балансовая модель Леонтьева. В целях иллюстрации изложим содержание отправных пунктов модели [22, с.72].

Рассматривается экономическая система, состоящая из n отраслей. Обозначим через $x = (x_1,...,x_n)$ вектор валовой продукции системы, через $y = (y_1,...,y_n)$ – вектор конечной продукции. Тогда систему уравнений материального баланса при условии линейности функций производственных издержек можно записать, как

$$E_1 - \sum_{j=1}^{n} a_{ij} x_j = y_i, \quad i = 1,2,....,n \tag{1}$$

или в векторно-матричной форме

$$(E - A)\, x = y. \tag{2}$$

Матрицу $A = (a_{ij})$ называют матрицей затрат или технологической матрицей; E – единичная матрица. Коэффициенты a_{ij}, $i=1,...,n$, $j=1,...,n$ называют коэффициентами прямых затрат; физически они представляют собой затраты продукции i-й отрасли на изготовление единицы валовой продукции k-й отрасли. В наиболее простом случае принято считать, что справедливо равенство $a_{ik} = const$. Уравнение (2) называют моделью Леонтьева.

Сформулируем задачу. В рамках модели (2) Леонтьева при заданном векторе y конечного продукта найти вектор x валовой продукции.

В рамках изложенной выше классификации моделей, описанная модель относится к классу математические задачи.

Изложим в общих чертах два по сути тождественных и весьма важных понятия, часто применяемых при решении задач математического моделирования интуитивно, которым, по нашему мнению, уделено недостаточно внимания в доступной учебной литературе. Это понятия правильной интерпретации и доказательной модели.

Пусть имеется некоторая аксиоматическая теория T, в которой выполнены требования непротиворечивости, независимости и полноты (последнее нестрого обязательно). И пусть всем исходным понятиям и отношениям аксиоматической теории T сопоставлены объекты системы S, т. е. выполнена интерпретация, в результате которой каждому утверждению F теории T соответствует высказывание F^* об объектах системы S.

Определение 6. ▌ *Если высказывание F^* истинно, то говорят, что утверждение F истинно (true) в данной интерпретации, в противном случае утверждение F в этой интерпретации ложно (false).* ▌

Определение 7. ▌*Если все аксиомы теории T истинны в данной интерпретации, то эта интерпретация называется правильной (доказательной).*▌

Из определений *6, 7,* заимствованных из математической энциклопедии [1] непосредственно следует, что только математические модели могут быть доказательными. Примерами доказательных моделей в экономике могут служить изложенная выше модель Леонтьева, модели из книг [4], [5] и др. Большое количество доказательных моделей, результаты которых применимы не только в экономике, но и в естествознании в целом, можно найти в пособии [6].

Глава 1. Краткие сведения из теории вероятностей и математической статистики

1.1. Аксиомы теории вероятностей

Теория вероятностей есть математическая наука, выясняющая на теоретико-модельном уровне закономерности, которые возникают при взаимодействии большого числа случайных факторов. В доступном изложении ключевые идеи, лежащие в основании теории вероятностей, математической статистики и методов обработки результатов наблюдений можно найти в книге [47].

Первые результаты теории вероятностей принято связывать с работами Л. Пачоли («Сумма арифметики, геометрии, учения о пропорциях и отношениях», 1494 г.), Д. Кардано («Книга об игре в кости», 1526 г.) и Н. Тартальи («Общий трактат о числе и мере», 1556-1560 г.). Первым учебником по теории вероятностей – книга П. Лапласа (1749-1827) «Аналитическая теория вероятностей», изданная в 1812 г., первым русским учебником по теории вероятностей – книга В.Я. Буняковского (1804-1889) «Основания математической теории вероятностей» [31], изданная в 1846 г.

Принято считать, что задача создания системы аксиом теории вероятностей впервые сформулирована Д. Гильбертом и известна в настоящее время как шестая проблема Д. Гильберта – математическое изложение аксиом физики, аксиоматика теории вероятностей и механики (см. кн. [1, с.151]). История становления аксиоматики теории вероятностей изложена в учебнике [23]. В нашем изложении будем придерживаться системы аксиом А.Н. Колмогорова [25,26]. В основе математических моделей, используемых в теории вероятностей, лежат три понятия: пространство Ω элементарных событий, класс событий F (подмножеств Ω) и определенная на этом классе функция множеств P, называемая вероятностной мерой или распределением вероятностей. Элементы класса F, называемые событиями, будем обозначать большими латинскими буквами.

Замечание. ° Читателю, обнаружившему затруднения с пониманием материала, мы рекомендуем следовать рецепту Ж.Л. Даламбера (1717-1783): «Идите вперед, а понимание придет потом». ●

Пусть $\Omega = \{ \omega_1, \omega_2, \omega_3,..., \omega_n,...\}$– пространство элементарных событий, т.е. множество элементов ω_i, которые будем называть элементарными событиями, а F – множество подмножеств из Ω. Элементы множества F будем называть случайными событиями (или просто – событиями). Следуя содержанию книги [25] А.Н. Колмогорова, напомним полезное для дальнейшего изложения определение.

Определение 1.1. ▌ *Система F подмножеств Ω называется алгеброй, если*

a) $\Omega \in F$,

b) $A, B \in F \Rightarrow A+B \in F$,

c) $A \in F \Rightarrow \overline{A} \in F$. ▌

Если в условии b) рассматривать бесконечное число событий, то система F называется сигма-алгеброй.

Определение 1.2. ▌ *Два события A и B несовместны, если справедливо равенство* $AB = \varnothing$' (т.е. произведение событий является невозможным событием). ▌

Определение 1.3. ▌ *События $A_1, A_2,...,A_n$ образуют полную группу событий, если справедливо равенство $A_1+A_2+...+A_n = \Omega$, т. е. сумма событий есть достоверное событие.* ▌

В ряде прикладных задач рассматривают полную группу *попарно несовместных* событий $A_1, A_2,...,A_n$.

Напомним аксиомы теории вероятностей. Ограничимся их перечислением. Следуя А.Н. Колмогорову, имеют место аксиомы.

Аксиома 1. ▌ *F является алгеброй множеств.* ▌

Аксиома 2. ▌ *Каждому множеству A из F поставлено в соответствие неотрицательное действительное число P(A). Это число называется вероятностью события A.* ▌

Аксиома 3. ▌ $P(\Omega) = 1$. ▌

Аксиома 4. ▌ *Если A и B не пересекаются, то*

$$P(A + B) = P(A) + P(B). ▌$$

Определение 1.4. ▌ *Совокупность объектов* (Ω, F, P), *удовлетворяющую определению 1.1 и аксиомам 1-4 будем называть полем вероятностей (вероятностным пространством).* ▌

Доказано (подробнее см., например, в кн. [38, с.11], что система аксиом 1-4 непротиворечива. Для бесконечной последовательности событий $A_1, A_2,...,A_n,...$ кроме аксиом 1-4 постулируется аксиома 5 (непрерывности).

Аксиома 5 (непрерывности). ▌ *Если последовательность событий $A_1,A_2,...,A_n...$ такова, что каждое последующее влечет за собой предыдущее и произведение всех событий есть невозможное событие, то*

$$\lim_{n\to\infty} P(A_n) \to 0 ▌$$

Известно (см., например, в [25, с.27]), что если система множеств F конечна, аксиома 5 следует из аксиом 1-4.

Система аксиом Колмогорова *непротиворечива*, так как существуют реальные объекты, которые всем этим аксиомам удовлетворяют. Система аксиом Колмогорова *неполна* – даже для одного и того же множества Ω вероятности в множестве F мы можем выбирать различными способами.

Перечислим свойства вероятности P, следующие из содержания аксиом. Числовая функция P, определенная на *F*

1) нормирована, т. е. $P(\Omega) = 1$,

2) неотрицательна, т. е. $P(A) \geq 0$ для всех $A \in F$,

3) счетно-аддитивна, т. е. $P(\Sigma A_i) = \Sigma P(A_i)$, $i = 1,...,n$.

Весьма подробное изложение обоснования справедливости аксиом теории вероятностей и вытекающих из них свойств вероятности можно найти в учебнике [24].

Одним из ключевых понятий теории вероятностей является понятие независимости.

Определение 1.6. ▎*События A и B называются независимыми, если справедливо равенство P(AB) = P(A) P(B).* ▎

Из изложенных выше аксиом, наряду с понятием вероятности, следует и понятие условной вероятности.

Определение 1.7. ▎*Если P(A)# 0, то частное P(B|A), задаваемое равенством*

$$(B\,|\,A) = \frac{P(A\,B)}{P(A)} \tag{1.1}$$

называют условной вероятностью события B при условии A. ▎

Весьма полезна формула, применение которой позволяет по вероятности *P(B|A)* найти вероятность *P(A|B)*

$$(B\,|\,A) = \frac{P(A\,|\,B)P(A)}{P(B)} \tag{1.2}$$

Воспользовавшись определением 1.7, можно сформулировать несколько отличающееся по форме от определения 1.6 понятие независимости событий.

Определение 1.8. ▎*События A и B называются независимыми, если справедливо равенство P(B) = P(B|A).* ▎

Уместно задаться вопросом: каковы границы применимости изложенных аксиом и определений при решении прикладных задач ?

Следуя А.Н. Колмогорову, обозначим через ***S*** комплекс условий, допускающий неограниченное число испытаний и будем изучать события, которые могут наступить или не наступить в результате осуществления ***S***-комплекса. Если после реализации ***S***-комплекса произошло какое-либо событие, то будем говорить, что произошло событие *A*. Этому событию *A*

можно поставить в соответствие число *P(A)*, удовлетворяющее неравенству $0 \leq P(A) \leq 1$. Можно быть практически уверенным, что если ***S***-комплекс будет повторен *n* раз, *n* – натуральное число, которое можно считать достаточно большим, и если при этом событие *A* наступило *m* раз, то отношение *m/n* будет весьма мало отличаться от числа *P(A)*. Изложенное обстоятельство служит основой классического определения вероятности.

Определение 1.9. Классическое определение вероятности.

▎*Вероятностью P(A) события A называется число, заданное через аналитическое выражение*

$$P(A) = \frac{m}{n}, \tag{1.3}$$

где n – общее количество испытаний, m – количество испытаний, в котором произошло событие A. В другой терминологии n – общее число исходов испытания, m – число т. н. благоприятных исходов. ▎

Заслуга введения в науку классического определения понятия вероятность принадлежит Я. Бернулли. Выражение (1.3) послужило основой для создания аксиоматики А.Н. Колмогорова. Если число *P(A)*, т. е. вычисленное применением выражения (1.3) можно считать чрезвычайно малым, то можно быть практически уверенным в том, что при однократной реализации ***S***-комплекса событие *A* не произойдет. Изложенный подход позволяет выполнить эмпирическую дедукцию аксиом.

Выполним дедукцию. Известно, что алгебра *F* событий, которым приписаны определенные вероятности, содержит в качестве элемента множество Ω (аксиома 1 и первая часть аксиомы 2). Очевидно, что $0 \leq m/n \leq 1$, поэтому и вторая часть аксиомы 2 оказывается вполне естественной. Для события Ω всегда $m = n$, благодаря чему $P(\Omega) = 1$ – это обеспечивает выполнение аксиомы 3. Если, наконец, положить, что происходит два события A_1 и A_2 такие, что $A_1A_2 = \varnothing$ (т. е. A_1 и A_2 несовместны), то $m = m_1 + m_2$ и $m/n = m_1/n + m_2/n$. Следовательно, уместно положить

$$P(A_1 + A_2) = P(A_1) + P(A_2),$$

т. е. аксиома 4 тоже выполняется. В силу конечности числа *n*, выполняется и аксиома 5.

Изложенная интерпретация – эмпирическая дедукция аксиом является правильной, что дает право применения аксиом и законов вероятностного пространства (Ω, F, P) для решения прикладных задач. Следовательно, все изложенные выше аксиомы и определения применимы не только к абстрактным (математическим) множествам, но и к физическим системам, подлежащим изучению. В том числе, к экономическим системам.

Отдельный интерес представляет случай, в котором число элементов множества Ω конечно, т. е. эксперимент, результаты которого при заданном комплексе условий **S** описываются конечным числом исходов. Подробное рассмотрение такого случая выполняется в рамках элементарной теории вероятностей. В целях полноты изложения запишем определение, заимствованное из книги [25] Колмогорова.

Определение 1.10. ▌ *Мы называем элементарной теорией вероятностей ту часть теории вероятностей, в которой приходиться иметь дело с вероятностями лишь конечного числа событий.* ▌

В рамках элементарной теории вероятностей нетрудно доказать теоремы, результаты которых часто применяются при решении вероятностных задач – теорему сложения, теорему умножения, формулу полной вероятности и теорему Байеса [18, 25].

1.2. Случайные величины

Будем рассматривать конечное или счетное пространство элементарных событий Ω, элементам которого ω отвечают вероятности $P(\omega)$.

Определение 1.11. ▌ *Случайной величиной* X *называется однозначная функция* X(ω), *определенная на пространстве элементарных событий* Ω. ▌

Очевидно, что функция X(ω) отображает множество Ω на множество чисел. Ограничимся рассмотрением случайных величин с вещественными значениями. Обозначим возможные значения случайной величины X через $x_1, x_2, \ldots, x_n, \ldots$, через p_i вероятности $p_i = P(X = x_i)$. Запишем полезное определение.

Определение 1.12. ▌ *Таблица вида*

$$\begin{pmatrix} x_1, E_2, \ldots, x_n, \ldots \\ p_1, p_2, \ldots\ldots, p_n, \ldots \end{pmatrix} \qquad (1.4)$$

называется распределением случайной величины X. ▌

В статистических исследованиях принято разделять случайные величины на дискретные и непрерывные.

Определение 1.13. ▌ *Величина X, заданная на* $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_n, \ldots\}$ *и принимающая не более чем счетное число значений, называется дискретной случайной величиной. Если величина X принимает на Ω континуум значений, то ее называют непрерывной.* ▌

Для изучения случайных величин, как и для изучения вероятностных законов для событий, необходимо определить понятие независимости.

Определение 1.14. ▌ Дискретные случайные величины *X и Y называются независимыми, если для всех пар (i, j) выполняются соотношения*

$$P(X{=}x_i, Y{=}y_j) = P(X{=}x_i)P(Y{=}y_j)$$ ▌

В приложениях теории вероятностей, как правило, имеют дело не с самими случайными величинами, а их распределениями. Возможных значений $x_1, x_2, \ldots, x_n, \ldots$ случайной величины и их вероятностей может оказаться слишком много, во многих случаях – континуум. Это создает необходимость охарактеризовать распределение несколькими параметрами, избежав записи в табличной форме (1.12), которая к тому же невозможна для непрерывных случайных величин, т. к. они имеют континуум значений. Выход из ситуации был найден основателем математической статистики А. Кетле (1796-1874), сведения о работах которого можно найти в работе [28, с.84].

Ключевой идеей Кетле была идея введения функций, полностью описывающих свойства случайных величин. Другое название этих функций – законы распределения. Аналитические выражения, являющиеся законами распределения были известны и до результатов Кетле. Заслугой Кетле является нахождение физической интерпретации этих выражений, применение которой позволило успешно находить решения статистических задач.

Закон распределения случайной величины может быть записан в трех видах: в виде функции распределения, плотности вероятности и характеристической функции. Остановимся на первых двух из них.

Замечание. ° В качестве прототипа изложения материала по функции распределения использовано содержание учебника [23]. ●

Определение 1.15. ▌ *Пусть X – случайная величина и x – произвольное действительное число. Вероятность того, что X примет значение, меньшее чем x, называется функцией распределения вероятностей случайной величины X*:

$$F(x) = P(X<x). \qquad (1.5)$$ ▌

Хрестоматийным примером функции распределения дискретной случайной величины является функция распределения Бернулли.

Для ее описания обозначим через μ – число появлений события *A* в последовательности *n* независимых испытаний, в каждом из которых вероятность его появления постоянна и равна *p*. В зависимости от случая число μ может принимать все целочисленные значения от нуля до *n* включительно, *n* – общее количество испытаний. Известно (результат впервые найден Бернулли), что вероятность $P_n(m)$ появления события *m* раз в *n* испытаниях может быть найдена применением выражения

$$P_n(m) = P(\mu = m) = C_n^m p^m (1-p)^{n-m} . \tag{1.6}$$

Воспользовавшись определением 1.12, распределение случайной величины μ можно записать в виде таблицы. Но, применив определение 1.15, можно записать

$$F(x) = \begin{cases} 0 & 4;\,O \quad x \le 0 \\ \sum\limits_{k<x} P_n(k) & 4;\,O \quad 0 < x \le n \\ 1 & 4;\,O \quad x > n \end{cases} \tag{1.7}$$

Графической интерпретацией функции (1.7) является ступенчатая линия с разрывами первого рода (т. н. скачками) в точках $x = 0, 1, 2, \ldots, n$; скачок в точке $x = k$ равен $P_n(k)$.

Другим распространенным хрестоматийным примером функции распределения случайной величины является функция распределения нормально распределенной случайной величины. Запишем ее аналитическое выражение.

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt \, , \tag{1.8}$$

где $a = const$ может быть любым действительным числом, число σ удовлетворяет неравенству $\sigma > 0$.

В дальнейшем, в целях краткости записи, функцию (1.8) будем записывать как $F_N(x, a, \sigma)$.

Перечислим отдельные свойства функции распределения, применение которых часто используется при первичной обработке опытных данных.

1. Функция распределения любой случайной величины есть неубывающая функция.

2. Функция распределения может иметь не более чем счетное множество скачков.

3. Функция распределения непрерывна слева.

4. Функция распределения принимает значения на множестве [0, 1].

Подводя итог свойствам 1-4 можно сказать, что любая функция распределения является неубывающей, непрерывной слева и удовлетворяющей условиям $F(-\infty) = 0$, $F(+\infty) = 1$ функцией. Верно и обратное: любая функция, удовлетворяющая перечисленным условиям, может рассматриваться как функция распределения некоторой случайной величины.

Наряду с применением в качестве описания закона распределения случайной величины функции распределения, используется описание закона распределения в виде плотности вероятности случайной величины. Определим понятие плотности вероятности.

Определение 1.16. ▎*Неотрицательная функция f(x), удовлетворяющая при любых возможных значениях аргумента x равенству*

$$F(x) = \int_{-\infty}^{x} f(x)dx \tag{1.9}$$

называется функцией плотности вероятности случайной величины. ▎

Понятие плотность вероятности имеет смысл только для непрерывных случайных величин. Нетрудно увидеть, что функция распределения $F(x)$ и плотность вероятности $f(x)$ связаны выражением

$$\frac{dF(x)}{dx} = F_x'(x) = f(x) \tag{1.10}$$

Кроме того, для плотности вероятности $f(x)$ справедливо выражение

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{1.11}$$

Например, воспользовавшись выражением (1.10), выполнив дифференцирование правой части выражения (1.8), находим плотность вероятности нормально распределенной случайной величины:

$$(F_N(x,a,\sigma))_x' = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-a)^2}{2\sigma^2}} . \tag{1.12}$$

В дальнейшем, в целях краткости, выражение для плотности вероятности нормально распределенной случайной величины будем записывать, как $f_N(x, a, \sigma)$.

В целях иллюстрации на рис.1,2 расположены графические интерпретации функции распределения (1.8) и соответствующей ей плотности вероятности (1.12) при значениях параметров $a = 0$, $\sigma = 1$.

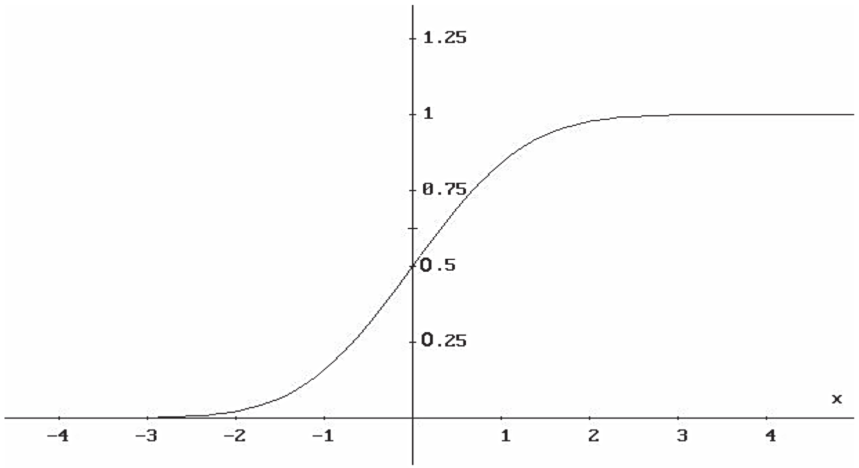

Рис.1.

На рис.1: по оси абсцисс – значения случайной величины, по оси ординат – значения функции распределения $F_N(x, 0, 1)$.

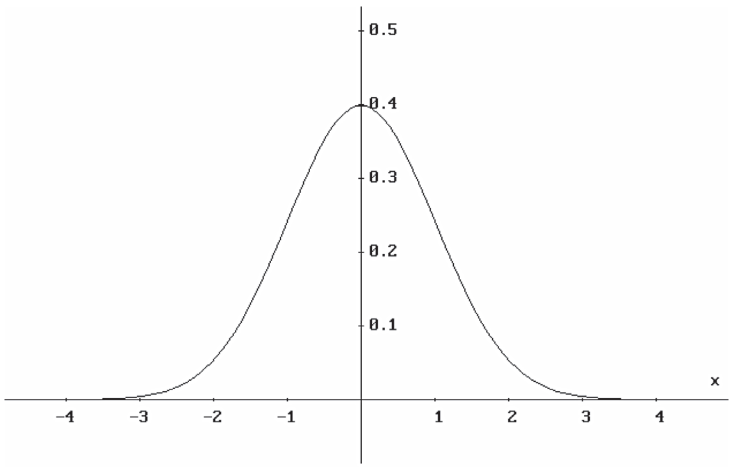

Рис.2.

На рис.2: по оси абсцисс – значения случайной величины, по оси ординат – значения плотности вероятности $f_N(x, 0, 1)$.

В процессе дальнейшего изложения наряду с терминами функция распределения и плотность вероятности мы будем пользоваться термином закон распределения, в зависимости от контекста подразумевая при этом либо функцию распределения, либо плотность вероятности.

Закон распределения – функция распределения или плотность вероятности полностью характеризует свойства случайной величины. Вместе с тем, при решении задач моделирования весьма часто возникает необходимость получить представление об отдельных свойствах случайной величины, воспользовавшись числовыми характеристиками. Весьма часто применяемыми числовыми характеристиками случайных величин являются числовые характеристики, называемые моментами случайных величин.

1.3. Числовые характеристики случайных величин

Моменты

Дадим определения числовых характеристик случайных величин, называемых моментами. В целях удобства сформулируем отдельно определения моментов для дискретных и непрерывных случайных величин.

Для дискретных случайных величин.

Пусть дана дискретная случайная величина, принимающая численные значения $x_1, x_2, \ldots, x_n$ с соответствующими вероятностями $p_1, p_2, \ldots, p_n$. Имеют место определения.

Определение 1.17. ❙ *Начальным моментом k-го порядка дискретной случайной величины называется характеристика m_k, определяемая выражением*

$$m_k = \sum_{i=1}^{n} x_i^k p_i\,, \qquad k = 1,2,.... \tag{1.13}$$

При k = 1, т. е. начальный момент первого порядка, называется математическим ожиданием. Обозначается M[X]. ❙

Определение 1.18. ❙ *Центральным моментом k-го порядка дискретной случайной величины называется характеристика m_k, определяемая выражением*

$$\mu_k = \sum_{i=1}^{n} (x_i - M[X])^k p_i\,, \qquad k = 1,2,..... \tag{1.14}$$

При k = 2, т. е. центральный момент второго порядка, называется дисперсией. Обозначается D[X]. ❙

Для непрерывных случайных величин.

Определение 1.19. ▌*Начальным моментом k-го порядка непрерывной случайной величины называется характеристика m_k, определяемая выражением*

$$m_k = \int_{-\infty}^{\infty} x^k f(x)dx\,, \qquad k=1,2,..... \tag{1.15}$$

где f(x) — плотность вероятности,

При k = 1, т. е. начальный момент первого порядка, называется математическим ожиданием. Обозначается M[X]. ▌

Определение 1.20. ▌*Центральным моментом k-го порядка непрерывной случайной величины называется характеристика* m_k, *определяемая выражением*

$$\mu_k = \int_{-\infty}^{\infty} (x - M[X])^k f(x)dx\,, \qquad k=1,2,..... \tag{1.16}$$

где f(x) — плотность вероятности,

При k = 2, т. е. центральный момент второго порядка, называется дисперсией. Обозначается D[X]. ▌

Определение 1.21. ▌*Значения моментов, вычисленные по выборочным данным численных значений случайных величин, называются эмпирическим.* ▌

Например, начальный эмпирический момент первого порядка (который часто называют оценкой среднего арифметического), начальный эмпирический момент второго порядка (который часто называют оценкой дисперсии) и т. д.

Наряду с моментами при изучении случайных величин применяется характеристика, называемая среднеквадратическим отклонением.

Определение 1.22. ▌*Среднеквадратическим отклонением случайной величины называется положительное значение квадратного корня из дисперсии.* ▌

Математическое ожидание и дисперсия существует не у всех случайных величин.

Доступное изложение свойств математического ожидания и дисперсии можно найти в пособии [12].

1.4. Построение формул для вычисления эмпирических моментов случайных величин

1.4.1. Метод моментов

Метод найден К. Пирсоном, опубликован впервые в 1894 г. в работе [48]. Изложим рекомендацию по применению, заимствованную из книги [8]: найденные использованием закона распределения моменты в количестве, равном числу оцениваемых параметров, приравниваются к соответствую-

щим выборочным моментам. Найденная система уравнений решается относительно параметров. Решения системы являются требуемыми оценками. Известно, что найденные методом моментов оценки являются состоятельными. Напомним определение состоятельности оценки Θ^*.

Определение 1.23. ▌ *Оценка Θ_n^* параметра Θ называется состоятельной, если для любого ε > 0 выполняется равенство*

$$\lim_{n\to\infty} P\Big(\,|\Theta_n^* - \Theta| < \varepsilon\Big) = 1\,. \qquad ▌$$

Пример. Найти методом моментов оценку математического ожидания нормально распределенной случайной величины.

Решение. □ Известно, что математическое ожидание *M[X]* случайной величины с законом распределения $f_n(x,a,\sigma)$ равен параметру *a*. Приравняв *M[X]* начальному эмпирическому моменту первого порядка (см. (1.13)), запишем

$$m_1 = \sum_{i=1}^{n} x_i^1 p_i = a^* = \frac{1}{n}\sum_{i=1}^{n} x_i\,. \;■$$

1.4.2. Метод максимального правдоподобия

Метод считается наиболее важным в теоретическом смысле. Для частных случаев найден еще К. Гауссом (1777-1855), но как общий метод для нахождения оценок впервые предложен в 1912 г. Р. Фишером (1890-1962) и опубликован в работе [44]. Доступное описание метода можно найти в пособии [12, с.229], учебнике [45] и мн. др. книгах, развитие – в монографии [49]. Свойства оценок, найденных применением метода, изложены в книге [80]. Изложим кратко ключевую идею метода и его вычислительную часть.

Будем обозначать через *X* – случайную величину, через $x_1, x_2, \ldots, x_n$ – численные значения, которые величина приняла в результате *n* испытаний. Будем считать, что закон распределения величины *X* известен, но неизвестно аналитическое выражение, применением которого используя данные выборки можно найти значение параметра Θ. Обозначим вероятность того, что в результате испытания величина *X* примет значение x_i, $i=1,\ldots,n$, через $p(x_i;\Theta)$. Дадим необходимые определения.

Определение 1.24. ▌ *Функцией правдоподобия дискретной случайной величины X называют функцию аргумента Θ :*

$$L(x_1, \ldots\ldots, x_n; \Theta) = \prod_{i=1}^{n} p(x_i, \Theta) \qquad ▌$$

Определение 1.25. ▌ *Функцией правдоподобия непрерывной случайной величины X называют функцию аргумента Θ :*

$$L(x_1,....,x_n;\Theta)=\prod_{i=1}^{n} f(x_i,\Theta),$$

где f(x; Θ) — плотность вероятности. ▌

А. Вальдом (1902-1950) в 1949 г. доказано, что значение аргумента Θ^*, в котором функция правдоподобия L достигает максимума, является состоятельной оценкой параметра Θ.

В вычислительных целях нередко удобнее использовать не функцию *L*, а ее логарифм *ln L*. Правомерность замены обоснована тем, что функции *L* и *ln L* достигают максимума при одном и том же значении аргумента Θ. Изложим процедуру нахождения оценки Θ^*.

1. Найти производную $\dfrac{d\ln L}{d\Theta}$.
2. Приравнять производную нулю и найти стационарную точку — корень уравнения правдоподобия
   $$\frac{d\ln L}{d\Theta}=0$$
3. Найти вторую производную $\dfrac{d^2\ln L}{d\Theta^2}$

Если вторая производная в точке $\Theta=\Theta^*$ отрицательна, то Θ^* — точка максимума. Найденную точку максимума Θ^* принимают в качестве оценки наибольшего правдоподобия параметра Θ.

В случае оценки нескольких параметров $\Theta_1,..., \Theta_k$, нужно найти производные по каждому из параметров. Затем, приравняв производные нулю, решить найденную систему уравнений. Если нахождение корней системы затруднительно, следует применить численные методы, подробное изложение которых можно найти в пособии [46].

Если для оцениваемого параметра Θ существует эффективная оценка Θ^*, то уравнение правдоподобия имеет единственное решение. Оценка наибольшего правдоподобия далеко не всегда совпадает с оценкой, найденной методом моментов.

1.5. Законы распределения случайных величин

Напомним основные законы распределения дискретных и непрерывных случайных величин.

Распределения дискретных случайных величин

1.5.1. Биномиальное распределение

Ряд распределения $P(x=m)=C_n^m p^m(1-p)^{n-m}$.

Математическое ожидание *M[X]=np.*

Дисперсия *D[X] = p (1 – p)n.*

Оценивание параметров: $p^*=m/n$. ММП

Типичная интерпретация: *m* — число успехов в *n* независимых испытаниях с вероятностью *p* успеха в единственном испытании.

1.5.2. Геометрическое распределение 1

Ряд распределения $P(X)=p(1-p)^x,\ x=0,1,2,...$

Математическое ожидание $M[X]=(1-p)/p$.

Дисперсия $D[X]=(1-p)/p^2$.

Типичная интерпретация: *X* — число неудачных испытаний, предшествующих первому успеху; *p* — вероятность успеха в одиночном испытании.

1.5.3. Геометрическое распределение 2 (распределение Фарри)

Ряд распределения $P(Y)=p(1-p)^{y-1},\ y=1,2,....$

Математическое ожидание $M[Y]=1/p$. Дисперсия $D[Y]=(1-p)/p^2$.

Типичная интерпретация: *Y* — число независимых испытаний до первого успеха, включая и первый успех; *p* — вероятность успеха в одиночном испытании.

1.5.3. Распределение Пуассона

Ряд распределения $P(X)=\dfrac{\mu^x}{x!}e^{-\mu},\ \mu>0$.

Функция распределения

$$F(x)=\begin{cases}0, & x\le 0\\ e^{-\mu}\sum\limits_{i=0}^{k}\dfrac{\mu^l}{i!}, & \underset{k=0,1,2,....}{k<x\le k+l}\end{cases}$$

Математическое ожидание и дисперсия $M[X]=\mu$, $D[X]=\mu$.

Типичная интерпретация: *X* — число событий стационарного ординарного потока (пуассоновского) без последействия в интервале фиксированной длины; μ — математическое ожидание числа событий в рассматриваемом интервале.

Оценивание параметров

$$\mu^* = \frac{1}{n}\sum_{i=1}^{n} x_i \,. \qquad \text{ММ, ММП.}$$

Подробные рекомендации по вычислению оценок и доверительных границ для параметра μ распределения изложены в документе [86].

1.5.4. Гипергеометрическое распределение

Ряд распределения

$$P(X) = \frac{C_M^x C_{N-M}^{n-x}}{C_N^n}, \qquad x = \alpha, \alpha+1, \ldots, \beta \qquad ,$$

где *N, M, n* – целые неотрицательные числа ($M \le N$, $n \le N$), $\alpha = max(0, M + n\ \ N)$, $\beta = min(M, n)$.

Функция распределения

$$F(x) = \begin{cases} 0, & x \le \alpha \\ \dfrac{1}{C_N^n}\sum_{i=\alpha}^{\alpha+k} C_M^i C_{N-M}^{n-i}, & \underset{k=0,1,\ldots,\beta-\alpha-1;}{\alpha+k < x \le \alpha+k+1}, \\ 1, & x > \beta\,. \end{cases}$$

Математическое ожидание $M[X] = M\,n/N$.

Дисперсия

$$D[X] = \frac{nM(N-M)(N-n)}{N^2(N-1)} \quad .$$

Типичная интерпретация: *X* – число меченых элементов в выборке без возвращения объема *n*, извлеченной из генеральной совокупности объема *N*, содержащей *M* меченых элементов.

Оценивание параметров

$$M = (x/n)\,N. \qquad \text{ММП.}$$

Свойства случайной величины, распределенной по гипергеометрическому закону, широко применяются при решении задач одновыборочного контроля качества продукции (см. например, работу [87]), а также задач выборочного оценивания в социологии и экономике (см., например, диссертацию [88]).

Распределения непрерывных случайных величин

Законы распределения с возможными значениями на всей числовой оси.

1.5.5. Нормальное распределение (Гаусса-Лапласа)

Плотность вероятности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \,,$$

a – параметр положения, σ – параметр масштаба.

Функция распределения

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt \,.$$

Математическое ожидание $M[X] = a$, Дисперсия $D[X] = \sigma^2$.

Нормально распределенная случайная величина с математическим ожиданием $a = 0$ и дисперсией $\sigma^2 = 1$ называется стандартной нормально распределенной. Ее функция распределения обозначается через $\Phi(x)$. Если в функции $\Phi(x)$ нижнюю границу интегрирования заменить с $-\infty$ на 0, то такая функция называется функцией Лапласа и обозначается $\Phi_0(x)$.

Оценивание параметров

$$a^* = x^* = \frac{1}{n}\sum_{i=1}^{n} x_i\,, \qquad \sigma^{2*} = S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - x^*\right)^2 \quad \text{ММ. ММП.}$$

Нормально распределенная случайная величина является единственной, для которой оценки параметров, найденные ММ и ММП совпадают.

1.5.6. Распределение Коши

Плотность вероятности $f(x) = \dfrac{\lambda}{\pi[\lambda^2 + (x-\mu)^2]}$,

μ – параметр положения, λ – параметр масштаба.

Функция распределения $F(x) = \dfrac{1}{2} + arctg\dfrac{x-\mu}{\lambda}$.

Математическое ожидание Не существует

Дисперсия Не существует

Оценивание параметров Методом ММ и ММП невозможно.

1.5.7. Распределение Гумбеля 1 (минимального значения)

Плотность вероятности

$$f(x) = \frac{1}{\lambda}\exp\left(\frac{x-\mu}{\lambda} - e^{(x-\mu)/\lambda}\right),$$

μ – параметр положения, λ – параметр масштаба ($\lambda > 0$).

Функция распределения
$F(x) = 1 - \exp\left(-e^{(x-\mu)/\lambda}\right)$.

Математическое ожидание
$M[X] = \mu - \lambda\gamma$, γ – постоянная Эйлера, $\gamma \approx 0.57722$.

Дисперсия $D[X] = \frac{\pi^2}{6}\lambda^2 \approx 1.64\cdot\lambda^2$.

Оценивание параметров
$\mu^* = x^* + 0.4501\cdot S$, $\lambda^* = 0.7797\cdot S$. ММ.

1.5.8. Распределение Гумбеля 2 (максимальные значения)

Плотность вероятности $f(x) = \frac{1}{\lambda}\exp\left(-\frac{x-\mu}{\lambda} - e^{(x-\mu)/\lambda}\right)$,

μ – параметр положения, λ – параметр масштаба ($\lambda > 0$).

Функция распределения $F(x) = \exp\left(-e^{-(x-\mu)/\lambda}\right)$.

Математическое ожидание
$M[X] = \mu + \lambda\gamma$, γ – постоянная Эйлера, $\gamma \approx 0.57722$.

Дисперсия $D[X] = \frac{\pi^2}{6}\lambda^2 \approx 1.64\cdot\lambda^2$.

Оценивание параметров: $\mu^* = x^* - 0.4501\cdot S$, $\lambda^* = 0.7797\cdot S$. ММ.

1.5.9. Логистическое распределение

Плотность вероятности

$$f(x) = \frac{\exp\left(\frac{x-\mu}{\lambda}\right)}{\lambda\left[1 + \exp\left(\frac{x-\mu}{\lambda}\right)\right]^2}, \quad (\lambda > 0)$$

где μ – параметр положения, λ – параметр масштаба.

Функция распределения $F(x) = \frac{1}{1 + \exp\left(-\frac{x-\mu}{\lambda}\right)}$.

Математическое ожидание $M[X] = \mu$. Дисперсия $D[X] = (\lambda\pi)^2/3$.

Оценивание параметров $\mu^* = x^*$, $\lambda^* = S\sqrt{3}/\pi$. ММ.

Законы распределения с положительными значениями аргумента

1.5.10. Показательное (экспоненциальное) распределение

Плотность вероятности $f(x) = \lambda e^{-\lambda x}$, $x > 0$,
где λ – параметр масштаба ($\lambda > 0$).

Функция распределения $F(x) = 1 - e^{-\lambda x}$.

Математическое ожидание $M[X] = 1/\lambda$. Дисперсия $D[X] = 1/\lambda^2$.

Оценивание параметров $\lambda^* = 1/x^*$. ММП. ММ.

1.5.11. Логнормальное распределение

Плотность вероятности

$$f(x) = \frac{1}{xa\sqrt{2\pi}} e^{-\frac{(\ln(x/m))^2}{2a^2}},$$

где m – параметр масштаба; a – параметр формы.

Математическое ожидание $M[X] = m\, e^{a^2/2}$.

Дисперсия $D[X] = m^2 e^{a^2}(e^{a^2} - 1)$.

Оценивание параметров

$$m^* = \frac{x^*}{\left(1 + \left(\frac{S}{x^*}\right)^2\right)^{0.5}}, \quad a^* = \left[\ln\left[1 + \left(\frac{S}{x^*}\right)^2\right]\right]^{0.5}. \text{ ММ.}$$

$$m^* = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln x_i\right), \quad a^* = \left(\frac{1}{n}\sum_{i=1}^{n}\left(\ln x_i - \ln m^*\right)^2\right)^{0.5}. \text{ ММП.}$$

Распределения с возможными значениями на ограниченном интервале

1.5.12. Равномерное распределение

Плотность вероятности

$$f(x)=\frac{1}{b-a},\quad a\le x\le b$$

где a,b – границы области возможных значений случайной величин.

Функция распределения $F(x)=\dfrac{x-a}{b-a}$.

Математическое ожидание $M[X]=(a+b)/2$.

Дисперсия $D[X]=(b-a)^2/12$.

Оценивание параметров $a^*=x^*-S\sqrt{3}$, $b^*=x^*+S\sqrt{3}$. ММ.

$$a^*=x_{(1)}-\frac{x_{(n)}-x_{(1)}}{n-1},\ b^*=x_{(1)}+\frac{x_{(n)}-x_{(1)}}{n-1},\qquad \text{ММП.}$$

где $x_{(1)}$ – наименьший из элементов выборки, $x_{(n)}$ – наибольший из элементов выборки.

1.5.13. Усеченное нормальное распределение. Двухстороннее усечение

Плотность вероятности

$$f(x)=\frac{1}{\gamma a\sqrt{2\pi}}\exp\left(-\frac{(x-m)^2}{2a^2}\right),\qquad \alpha\le x\le\beta \qquad ,$$

где α, β – границы области возможных значений случайной величины, т. е. левая и правая границы усечения; m – параметр положения, a – параметр масштаба ($a>0$) и

$$\gamma=\Phi_0\left(\frac{\beta-m}{a}\right)-\Phi_0\left(\frac{\alpha-m}{a}\right).$$

Число $1-\gamma$, равное вероятности того, что исходная (неусеченная) нормально распределенная случайная величина окажется вне интервала (α,β), называется степенью усечения.

Функция распределения

$$F(x)=\frac{1}{\gamma}\left[\Phi_0\left(\frac{x-m}{a}\right)-\Phi_0\left(\frac{\alpha-m}{a}\right)\right].$$

Математическое ожидание

$$M[X]=m+\frac{a}{\gamma}\left[\varphi(u_1)-\varphi(u_2)\right]\qquad ,$$

где $\varphi(u)$ – плотность вероятности стандартной нормальной случайной величины; $u_1=(\alpha-m)/a$, $u_2=(\beta-m)/a$ – стандартизированные точки усечения.

Дисперсия

$$D[X]=a^2\left\{1+\frac{1}{\gamma}\left[u_1\varphi(u_1)-u_2\varphi(u_2)\right]-\frac{1}{\gamma^2}\left[u_1\varphi(u_1)-u_2\varphi(u_2)\right]^2\right\}.$$

Оценивание параметров. В том случае, когда точки усечения α и β неизвестны, а объем выборки достаточно велик, в качестве α можно взять число несколько меньшее наименьшего элемента $x_{(1)}$ выборки, в качестве β – число, несколько превышающее наибольший $x_{(n)}$ элемент выборки.

Распределения, наиболее часто используемые в математической статистике

При проведении ряда статистических исследований (например, оценивание параметров или проверка гипотез), используют так называемые «специальные распределения», основные из которых мы и рассмотрим.

1.5.14. χ^2-распределение Пирсона

Плотность вероятности

$$f(x)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}x^{n/2-1}e^{-x/2},\qquad x>0,$$

где $n>0$ – параметр формы, число степеней свободы, Γ – символ гамма-функции (интеграла Эйлера второго рода)

$$\Gamma(x)=\int_0^\infty t^{x-1}e^t\,dt.$$

1.5.15. F-распределение Фишера-Снедекора

Плотность вероятности

$$f(x)=\frac{\Gamma\left(\frac{k_1+k_2}{2}\right)k_1^{k_1/2}k_2^{k_2/2}}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)}x^{k_1/2-1}\left(k_1x+k_2\right)^{-\frac{k_1+k_2}{2}}\qquad ,$$

k_1,k_2 – числа степеней свободы.

F-распределение Фишера-Снедекора называют также распределением дисперсионного отношения.

1.5.16. t-распределение Стьюдента

Плотность вероятности

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-\frac{v+1}{2}}, \qquad -\infty < t < \infty,$$

где v – параметр, число степеней свободы, положительное целое число.

Перейдем к рассмотрению некоторых общих методов нахождения оценок для параметров распределений по выборочным значениям.

1.6. Оценка функции распределения по результатам выборки

Изучим подробнее функцию распределения. С этой целью изложим краткие сведения о том, как по результатам наблюдений и экспериментов составить представление о функции распределения случайной величины. Обозначим через X – множество результатов измерений. Расположим значения результатов измерений в порядке неубывания и перенумеруем их

$$x_1, x_2,...,x_n. \tag{1.17}$$

Найденный ряд (1.17) называют ранжированным по неубыванию. Введем обозначения: n_x – число наблюдений, при котором наблюдалось значение признака, меньшее x, n – общее число наблюдений (объем выборки). При таком обозначении относительная частота события $X<x$ будет равна n_x/n. Запишем определение.

Определение 1.25. ▎*Эмпирической функцией распределения называют функцию $F_n(x)$, определяющую для каждого значения x относительную частоту события $X<x$*

$$F_n(x) = \frac{n_x}{n} \;▎ \tag{1.18}$$

Введенная эмпирическая функция распределения $F_n(x)$ является статистическим аналогом функции распределения $F(x)$. Различие между функциями $F_n(x)$ и $F(x)$ состоит в том, что функция $F(x)$ определяет вероятность события $P(X < x)$, а функция $F_n(x)$ определяет относительную частоту этого же события. Такое заключение подтверждается тем, что функция $F_n(x)$ обладает всеми свойствами функции $F(x)$:

1) $F_n(x) \in [0,1]$;

2) $F_n(x)$ – неубывающая функция;

3) если x_1 – наименьшее значение из ряда (1.17), то $F_n(x) = 0$ при $x \le x_1$; если x_n – наибольшее из ряда (1.17), то $F_n(x) = 1$ при $x > x_n$.

Для иллюстрации приведем пример нахождения эмпирической функции распределения.

Пример. Дана ранжированная по неубыванию выборка численных значений случайной величины

-2, –1, –1, 0, 0, 1, 1, 1, 1, 2.

Найти эмпирическую функцию распределения этой величины.

Решение. ☐ Ранжированный ряд содержит пять различных значений: 2, 1, 0, 1, 2. Их частоты соответственно равны 0.1, 0.2, 0.2, 0.4, 0.1. Воспользовавшись определением 1.25, запишем значения эмпирической функции в виде таблицы 1.

ТАБЛИЦА 1

X	< 2	- 2	- 1	0	1	2	2 + c (c > 0)
$F_n(x)$	0	0.1	0.3	0.5	0.9	1	1

На рис.3 помещена графическая интерпретация найденной функции $F_n(x)$.

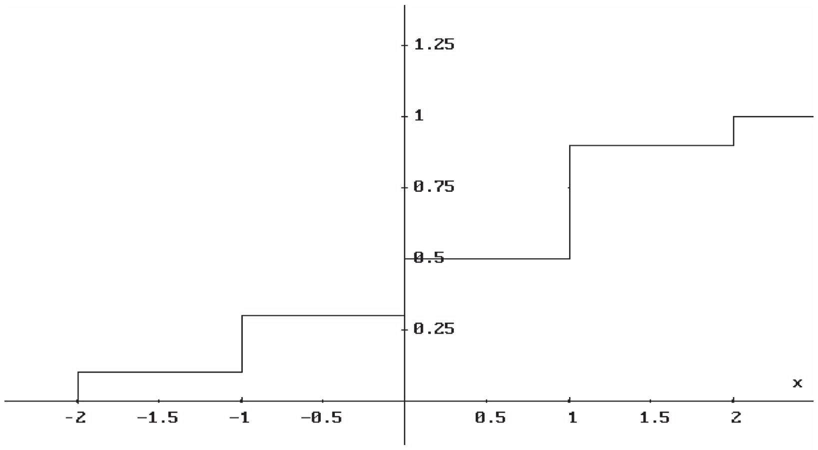

Рис.3.

На рис.3: по оси абсцисс – значения случайной величины X, по оси ординат – значения эмпирической функции $F_n(x)$. ■

Замечание. ° Несмотря на простоту, строгое определение эмпирической функции дано только в XX в. В книгах, изданных в начале XX века, например, в книге [32] Э. Юла и Дж. Кендала, изданной в 1911 г. и считающейся одним из лучших руководств того времени по математической статистике, эмпирическая функция распределения не описана. Принято считать, что теория эмпирических функций построена А.Н. Колмогоровым, впервые опубликована в работе [33] (1933) ●.

Продолжим изложение сведений о свойствах эмпирической функции распределения. После того, как функция $F_n(x)$ найдена, возникает очередная задача: сколь адекватно функция описывает свойства изучаемой случайной величины. Исчерпывающие сведения по решению задачи можно найти в работах [34], [35]. Весьма приблизительную оценку надежности того, что функция распределения содержится в интервале $F_n(x) \pm \varepsilon$ с вероятностью $1 - Q$ можно найти, воспользовавшись формулой

$$Q \approx e^{-2n\varepsilon^2}. \tag{1.19}$$

В доступном изложении вывод формулы (1.19) можно найти в книге [28, с.86]. В приведенном ниже примере изложен результат применения формулы (1.19).

Пример. Объем выборки равен 10. Эмпирическая функция $F_n(x)$ найдена. С какой вероятностью неизвестная функция $F(x)$ распределения исследуемой случайной величины X окажется в интервале $F_n(x)\pm 0.1$? Можно ли по результатам данной выборки адекватно судить о свойствах случайной величины?

Решение. □ Подставив в правую часть формулы (1.19) значения $n = 10$ и $\varepsilon = 0.1$, выполнив вычисления, запишем

$$Q \approx e^{-2\cdot 10\cdot 0.1^2} = 0.8,\quad 1-0.8=0.2$$

Ответ. С вероятностью 0.2 неизвестная функция $F(x)$ будет находиться в интервале $F_n(x) \pm 0.1$. Найденное значение вероятности мы считаем незначительным. Для нахождения закона распределения объем выборки недостаточен. ■

Применив формулу (1.19) при том же значении $\varepsilon = 0.1$, но объеме выборки $n = 100$ нетрудно вычислить, что в таком случае вероятность нахождения неизвестной функции $F(x)$ в интервале $F_n(x)\pm 0.1$ равна 0.86. Найденный результат является свидетельством в пользу известной рекомендации о том, что по возможности нужно стремиться к увеличению объемов выборок. Например, в вычислительной части работы [36] использована выборка, превышающая объем 50 тысяч численных данных. Вместе с тем, при решении отдельных прикладных задач возможно использование выборок малых объемов.

Правомерность применения выборочных данных при нахождении неизвестной функции распределения доказана Гливенко и Кантелли. Впервые найденное Гливенко доказательство опубликовано в работе [37] в 1933 г. В 1939 г. теорема помещена в учебник [38]. Напечатанный ниже текст теоремы заимствован из учебника [23, с.282], по которому можно ознакомиться и с содержанием доказательства.

Теорема (Гливенко-Кантелли). Пусть $F(x)$ — функция распределения случайной величины ξ и $F_n(x)$ — эмпирическая функция распределения результатов n независимых наблюдений над величиной ξ. Тогда

$$P\left(\lim_{n\to\infty}\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0\right)=1 \tag{1.20}$$

Результаты теоремы могут быть интерпретированы, как мера достижения количественной зрелости для применения математического моделирования (см. определение 1.4).

1.7. Неравенство Чебышева

В качестве средства оценки вероятностей уклонений значений случайной величины от математического ожидания используется неравенство Чебышева, которое дадим в двух видах

$$\begin{aligned} &P\left(\left|X-M[X]\ge\varepsilon\right|\right)\le D[X]/\varepsilon^2\\ &P\left(\left|X-M[X]<\varepsilon\right|\right)\ge 1-D[X]/\varepsilon^2 \end{aligned} \tag{1.21}$$

При решении прикладных задач использование неравенства Чебышева может вызывать значительные трудности. Например, пусть мы захотели найти численную оценку значения математического ожидания $M[X]$ цены одного и того же товара, продаваемого в разных магазинах по разным ценам. Тогда $X - M[X]$ означает отклонение значения цены от ее математического ожидания. Предположим, что колебание значений цен в разных магазинах столь незначительное, что можно считать значение дисперсии равным единице, т. е. $D[X] = 1$. Чему равно n — количество магазинов, в которых нужно узнать цену товара, чтобы с вероятностью, большей 0.99, заявить о том, что средняя цена товара в разных магазинах отличается не больше, чем на 0.1 у.е.?

Решим задачу. □ Известно, что среднее значение цены товара $x_{ср.}$ можно найти, воспользовавшись формулой

$$x_{ср.}=(x_1+x_2+\ldots+x_n)/n.$$

Из результатов центральной предельной теоремы А.М.Ляпунова (см. пособия [12, с.135], [13, с.130] и др.) известно, что дисперсию математического ожидания $D[M[X]]$ можно найти применением формулы $D[M[X]]=D[X]/n$. Применив неравенства (1.21) Чебышева, запишем

$$P(|x_{ср.}-M[X]<0.1|)>0.99 \text{ или } P(|x_{ср.}-M[X]>0.1|)<0.01.$$

Снова применив неравенство Чебышева, запишем

$$P(|x_{ср.}-M[X]>0.1|)\leq D[X]/(n\cdot 0.1^2).$$

Значит, если число n выбрать таким, чтобы выполнилось неравенство $D[X]/(n\cdot 0.1^2)<0.01$, то $n > 10^4$. Т. е. требуется обследовать больше 10000 магазинов. ■

По нашему мнению найденный результат является свидетельством в пользу того, что популярные оценки средней заработной платы, среднего прожиточного минимума, средней цены потребительской корзины и т. п., без привлечений огромных массивов исходных данных могут оказаться не вполне адекватно описывающими фактическое состояние экономики. Более точных результатов можно достигнуть, воспользовавшись неравенством Колмогорова. Сведения о неравенстве Колмогорова можно найти в книге [29, с.240]. Кроме того, повысить точность и сократить объем выборки можно, воспользовавшись результатами следствия из неравенства Чебышева (см. пособие [30, с.86]).

Однако в ряде случаев возможность применения неравенства Колмогорова и следствия из неравенства Чебышева ограничена дополнительными трудностями. Проще всего найти вероятность отклонения значений случайной величины от ее математического ожидания в случае, в котором аналитическое выражение закона распределения известно. Например, если случайная величина распределена по закону $f(x)$, то вероятности отклонений нетрудно вычислить интегрированием. Однако, закон распределения исследуемой случайной величины известен далеко не всегда. В этом случае его нахождение может быть связано с преодолением значительных трудностей. Возникает задача оценки закона распределения по выборочным данным.

Такая задача не имеет решения в рамках теории вероятностей. Для ее решения требуется применение аппарата математической статистики. Одной из ключевых теорем математической статистики является теорема В.И. Гливенко (1897-1940) и Ф. Кантелли (р. в 1875, дата смерти неизвестна).

Глава 2. Статистические критерии

2.1 Вводные замечания

Известно, что в процессе решения экономических и др. задач появляются разные предположения. Многие из них остаются в уме исполнителя, т. к. не заслуживают письменной фиксации. Другие фиксируются при оформлении решения. В областях деятельности, требующих доказательности принятых решений, любые предположения могут быть изложены только с помощью аналитических выражений. Например, предполагая скорость v процесса постоянной, мы должны записать $v = const$. Наоборот, предполагая скорость зависящей от времени t, мы должны обозначить ее через $v(t)$ и записать аналитически ее фактическую или предполагаемую зависимость. Только при таком подходе справедливость предположения может быть проверена применением результатов известных теорем или законов и может претендовать называться гипотезой. Справедливость утверждений, не содержащих аналитических выражений, не может быть ни доказана, ни опровергнута в рамках математического моделирования. Поэтому такие утверждения можно рассматривать только как бездоказательные. Например, утверждение «Сонечка Мармеладова чистая душа» может быть рассмотрено, как предположение, но не как гипотеза. Такое предположение не может быть ни доказано, ни опровергнуто в рамках требований, предъявляемых к результатам научных исследований. По этой причине мы не будем рассматривать задачи о предположениях, выраженных в концептуальной или иной, отличной от аналитической формах.

2.2. Статистические гипотезы

Определение 2.1 ▌ *Статистической называют гипотезу об аналитическом выражении закона распределения или о значении параметров закона.* ▌

В процессе проверки справедливости гипотезы наряду с выдвинутой гипотезой рассматривается гипотеза, противоречащая выдвинутой. Если выдвинутая гипотеза в результате проверки отвергается, то можно говорить о справедливости противоречащей гипотезы. Будем обозначать через H_0 основную (нулевую) гипотезу, через H_1 – альтернативную (конкурирующую). Гипотезы могут быть простыми и сложными.

Определение 2.2. ▌ *Простой называется гипотеза, содержащая только одно предположение.* ▌

Определение 2.3. ▎*Сложной называется гипотеза, состоящая из конечного или бесконечного числа простых гипотез.* ▎

Пример. □

1. Простая гипотеза. Математическое ожидание нормально распределенной случайной величины равно 5. Записывается: H_0: $a = 5$.
2. Наличие альтернативной гипотезы. Нулевая гипотеза H_0: $a = 5$, альтернативная гипотеза: H_1: $a > 5$. Записывается: H_0: $a = 5$, H_1: $a > 5$.
3. Гипотеза H_0: $a > 5$ сложная, т. к. содержит бесконечное множество простых гипотез. Например, гипотез H_0: $a = 6$, H_0: $a = 7$ и т. д.
4. Гипотеза H_0: $a = 5$ – математическое ожидание равно 5 при неизвестном значении дисперсии – сложная (более одного предположения).

Используя наблюдения выборки X, нужно либо принять, либо отвергнуть нулевую гипотезу. Правило принятия одного из этих решений называется статистическим критерием или просто критерием.

Мерой адекватности заключений, являющихся результатами проверки гипотезы по критерию, служит вероятностная мера. Применение вероятностной меры позволяет найти значение вероятности ошибки окончательного заключения. Принято учитывать два вида ошибок – ошибки первого и второго рода.

Определение 2.4. ▎*Вероятность* α *отвергнуть гипотезу* H_0 *при условии, что она верна, называется вероятностью ошибки первого рода.* ▎

Вероятность ошибки 1 рода также называют уровнем значимости критерия.

Определение 2.5. ▎*Вероятность* β *принять гипотезу* H_0 *при условии, что верна гипотеза* H_1, *называется вероятностью ошибки второго рода Значение* $\gamma = 1 - \beta$ *называют мощностью критерия.* ▎

Поскольку при фиксированном объеме выборки вероятности этих ошибок одновременно сделать достаточно малыми нельзя, то обычно α задают заранее, а β при уже выбранном α стараются сделать как можно меньше (хотя применимы и другие подходы – смотри ниже п. 2.2).

Адекватное (правильное) решение может быть принято в двух случаях:

1) гипотеза принимается, причем и в действительности она верна,

2) гипотеза отвергается, причем и в действительности она не верна.

При решении задач проверки гипотез используются критерии, применение которых позволяет найти вероятности α и β.

2.3. Критерии как средства проверки гипотез

В доступном изложении полезные сведения можно найти в учебном пособии [51, с.206], в более строгом – учебном пособии [52] и монографии [53]. Известно большое количество книг, в которых весьма подробно изложен аппарат обоснования методов проверки гипотез. Например, в монографии [54].

Согласно содержанию статьи из энциклопедии [8]: "*Статистический критерий – правило принятия решения о справедливости одной из статистических гипотез».*

Запишем определение критерия, ориентированное на решение прикладных задач.

Определение 2.6. ▎*Статистическим (в теории вероятностей и математической статистике) критерием называют случайную величину K, которая служит для проверки гипотезы.* ▎

Для умения успешно применять статистические критерии необходимо знание следующих ниже определений, заимствованных нами из пособия [51].

Определение 2.7. ▎ *Наблюдаемым (эмпирическим) значением* $K_н$ *называют значение критерия, вычисленное по выборкам.* ▎

Определение 2.8. ▎*Критической областью называют совокупность значений критерия, при которых нулевую гипотезу* H_0 *отвергают.* ▎

Определение 2.9. ▎ *Областью принятия гипотезы (областью допустимых значений) называют совокупность значений критерия, при которых нулевую гипотезу принимают.* ▎

Определение 2.10. ▎ *Критическими точками (границами)* $k_{кр.}$ *называют точки, отделяющие критическую область от области принятия гипотезы.* ▎

Определение 2.11. ▎*Правосторонней называют критическую область, определяемую неравенством* $K > k_{кр.}$ ▎

Определение 2.12. ▎*Левосторонней называют критическую область, определяемую неравенством* $K < k_{кр}$. ▎

Определение 2.13. ▎*Двухсторонней называют критическую область, определяемую неравенством* $K < k_1$, $K > k_2$ *где* $k_2 > k_1$. ▎

Определение 2.14. ▎*Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Другими словами, мощность критерия есть вероятность того, что нулевая гипотеза будет отвергнута, если верна конкурирующая гипотеза.* ▎

Сформулируем основной принцип применения критериев при проверке гипотез, рекомендуемый в подавляющем большинстве учебников, пособий и руководств. Например, в книгах [7 – 12] и др.

Принцип 2.1. ▎*Если наблюдаемое значение критерия принадлежит критической области — гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы — гипотезу принимают (точнее — не отвергают).* ▎

Не умаляя достоинств применения принципа 2.1, в целях полноты дополним, что принцип 2.1 – не единственный принцип применения критериев при проверке гипотез. Существует и другая точка зрения, изложенная, например, в пособии [52, с.301]. Она состоит в том, что нет надобности задавать фиксированный уровень значимости, т. к. для выбора его предварительного численного значения неизвестны недвусмысленные правила. Альтернативный принципу 2.1 принцип состоит в том, что уровень значимости α не задается заранее. Вместо этого вычисляется фактически достигаемый уровень значимости и область определения критерия разделяется на область допустимых значений и критическую точкой $K=k_{кр.}$. Примеры применения такого подхода можно найти в пособии [52, с.301]. В нашем изложении при проверке гипотез мы будем применять как принцип 2.1, так и альтернативный.

2.4. Критерии, часто применяемые при решении экономических задач

2.4.1. Критерий χ^2-Пирсона

Правомерность применения критерия χ^2-Пирсона следует из того, что многие случайные величины, свойства которых используются при решении экономических задач, распределены по закону χ^2.

Наблюдаемое (эмпирическое) значение χ^2-критерия (см. *определение 2.7*) вычисляется применением формулы

$$\chi^2=\sum_{i=1}^{m}\frac{\left(m_i-np_i\right)^2}{np_i}, \qquad (2.1)$$

в которой n — объем выборки, m — количество интервалов разбиения ранжированной по неубыванию выборки, m_i — количество элементов выборки, попавших в i-й интервал разбиения, $i=1,...,m$, p_i — теоретическая вероятность попадания элемента выборки в i-й интервал разбиения.

При вычислении фактически достигаемого уровня значимости α число степеней свободы k величины χ^2 вычисляется применением формулы

$$k = m - (\text{число параметров закона} + 1).$$

В целях удобства читателей опишем способы применения критерия, используя примеры решений типичных задач.

Пример. В целях решения задачи планирования размеров выпускаемой одежды, в 18891890 г., на одной из фабрик г. Москвы, выполнено измерение роста 1000 взрослых рабочих мужчин.

Результаты измерений расположены в таблице 2.1.

ТАБЛИЦА 2.1

Рост в см.		Число мужчин	Доля	$x_{ср}^*$(см) в группе
143	153	11	0.011	147.5
153	164	392	0.392	158.5
164	175	491	0.491	170
175	183	102	0.102	179
183	188	4	0.004	185.5

Проверить гипотезу о том, что случайная величина — рост взрослого мужчины распределена по нормальному закону.

Решение. ☐ Известно, что закон нормального распределения записывается с помощью аналитического выражения $f_N(x, a, \sigma)$. В нашем случае значения параметров a и σ неизвестны. Вычислим оценки a^* и σ^* неизвестных значений параметров a и σ. Воспользуемся формулами для оценки параметров по выборочным данным, найденными методом максимального правдоподобия (см. гл. 1). Перепишем формулы.

$$a^*=x^*=\frac{1}{n}\sum_{i=1}^{n}x_i\,,\qquad \sigma^{2*}=S^2=\frac{1}{n}\sum_{i=1}^{n}\left(x_i-x^*\right)^2$$

Подставив необходимые значения из таблицы 2.1, запишем

$a^*= 0.011\cdot 147.5 + 0.392\cdot 158.5 + 0.491\cdot 170 +0.102\cdot 179 + 0.004\cdot 185.5\approx 166.22.$

При вычислении точечной оценки дисперсии σ^{2*} необходимо проявить некоторую осторожность. А именно, непосредственное применение записанной выше формулы невозможно вследствие того, что из таблицы 2.1 нам не известны все значения x_i, $i=1,...,n$, т. к. данные сгруппированы. Поэтому для выполнения вычисления значения σ^{2*} читателю рекомендуется освежить в памяти темы «Свойства математического ожидания», «Свойства дисперсии», «Групповая, внутригрупповая и межгрупповая дисперсии» из курса теории вероятностей (см., например, в пособии [12] с. 207-209). Напомним, что между центральным моментом второго порядка и начальными моментами первого и второго порядков справедливо соотношение

$$\mu_k = D[X] = m_2 - m_1^2.$$

Известно, что $m_1 = M[X]$. Вычислим значение начального момента второго порядка m_2. Воспользовавшись сведениями из таблицы 2.1, запишем

$m_2^* = 0.011 \cdot 147.5^2 + 0.392 \cdot 158.5^2 + 0.491 \cdot 170^2 + 0.102 \cdot 179^2 + 0.004 \cdot 185.5^2 \sim 27600.$

$S^2 = 27600 - 166.22^2 = 53.87,\ S = (53.87)^{0.5} = 7.34.$

Запишем предполагаемый закон распределения роста рабочих: $f_N(x,\ 166.22,\ 7.34)$. Сформулируем гипотезу и запишем ее аналитически

$H_0: f(x) = f_N(x,\ 166.22,\ 7.34),\ H_1: f(x) \neq f_N(x,\ 166.22,\ 7.34).$

Запись читается: нулевая гипотеза H_0: случайная величина – рост рабочих распределена по нормальному закону $f_N(x,\ 166.22,\ 7.34)$. Альтернативная гипотеза H_1: случайная величина не распределена по нормальному закону.

Выполним проверку гипотезы H_0 применением критерия χ^2-Пирсона. Вычислим наблюдаемое значение критерия (см. определение 2.7). Для этого воспользуемся формулой 2.1. Для ее применения вычислим значения вероятностей $p_i,\ i = 1,...,5$.

$$p_1 = \int_{-\infty}^{153} f_N(x, 166.22, 7.34)\,dx = 0.035\,,\ p_2 = \int_{153}^{164} (.)dx = 0.345\,,$$

$$p_3 = \int_{164}^{175} (.)dx = 0.503\,,\ p_4 = \int_{175}^{183} (.)dx = 0.104\,,\ p_5 = \int_{183}^{\infty} (.)dx = 0.011.$$

Перепишем формулу (2.1), подставив в ее правую часть найденные численные значения вероятностей $p_i,\ i = 1,...,5$ и требующиеся значения из таблицы 2.1.

$$\chi^2 = \frac{(11 - 1000\cdot 0.035)^2}{1000\cdot 0.035} + \frac{(392 - 1000\cdot 0.34)^2}{1000\cdot 0.34} + \frac{(491 - 1000\cdot 0.503)^2}{1000\cdot 0.503} +$$

$$+ \frac{(102 - 1000\cdot 0.104)^2}{1000\cdot 0.104} + \frac{(4 - 1000\cdot 0.01)^2}{1000\cdot 0.01} = 27.6.$$

Найдем число степеней свободы k. Число параметров нормального закона равно двум. Число интервалов разбиения – пяти. Следовательно, $k = 5 - (2 + 1) = 2.$

Вычислим значение фактически достигаемого уровня значимости, обозначив его через α. Воспользовавшись законом распределения величины χ^2 с двумя степенями свободы (см. гл. 1), запишем

$$\alpha = \int_{27.6}^{\infty} \frac{1}{2^{\frac{2}{2}}\Gamma\left(\frac{2}{2}\right)} x^{\frac{2}{2}-1} e^{-x/2} dx = 1.02\cdot 10^{-6} \quad .$$

Выполним вероятностную интерпретацию результата. Отвергая гипотезу H_0: $f(x) = f_N(x,\ 166.22,\ 7.34)$ в пользу альтернативной гипотезы H_1: $f(x) \neq f_N(x,\ 166.22,\ 7.34)$ мы рискуем совершить ошибку первого рода, т. е. ошибку отвержения правильной гипотезы с вероятностью $\alpha = 1.02\cdot 10^{-6}$. Вероятность ошибки мала, поэтому нулевую гипотезу отвергаем.

Сформулируем окончательный результат. На основании опыта результатов измерений доказано, что случайная величина – рост рабочих мужчин не распределена по нормальному закону. Вероятность ошибочности заключения $\alpha = 1.2710^{-6}$.

Сформулируем экономический результат. Если при изготовлении мужской одежды доли продукции разных размеров вычислить использованием нормального закона, то выпущенная продукция не удовлетворит требованиям многих потребителей, что в итоге принесет экономический ущерб. ■

Замечание. ° Во многих изложениях заявление о том, что случайная величина – рост мужчины распределена по нормальному закону используется в качестве хрестоматийного примера. В рамках изложенного выше решения мы доказали, что такое заявление не вполне согласуется с результатами фактических измерений. В качестве социально-экономической интерпретации найденного решения мы, например, не исключаем, что, несмотря на изобилие товаров, читатель не всегда удовлетворен ассортиментом имеющейся в продаже одежды. ●

Продолжим изучение возможностей применения критерия χ^2Пирсона при решении задач экономики. Одной из таких задач является задача контроля качества в случае, в котором в качестве меры качества готовой продукции применяется распределение по сортам. Класс таких задач относится к классу задач об однородности выборок.

Определение 2.15. ▌ *Выборки называются однородными, если они извлечены из одной и той же генеральной совокупности.* ▌

Пример. Продукт, закупаемый фирмой, распределен по сортам 1, 2, 3 и 4. Из результатов предыдущего опыта известно, что в среднем требованиям 1-го сорта удовлетворяет 53.4% продукции, 2-го – 26.6%, 3-го – 13.3%, 4-го – 6.7%. Сотрудникам фирмы стало известно, что поставщик изменил технологию изготовления. Поэтому для проверки качества было взято 600 единиц новой продукции. В результате проверки оказалось, что первому сорту соответствует 340 единиц, второму – 130, третьему – 100 и четвертому 30 единиц.

Требуется выяснить, изменилось ли распределение качества продукции.

Решение. □ Если считать, zчто качество продукции не изменилось, то распределение продукции по сортам должно как и раньше соответствовать таблице

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0.534 & 0.266 & 0.133 & 0.067 \end{pmatrix}.$$

Сформулируем гипотезу H_0: распределение новой партии продукции соответствует напечатанной выше таблице, которую назовем первой. Альтернативная гипотеза – H_1: распределение новой партии не соответствует первой таблице.

Вычислим ожидаемое число единиц продукции. Ожидается, что число единиц продукции 1-го сорта будет $600\cdot 0.534 = 320.4$, 2-го сорта – $600\cdot 0.266 = 159.6$, 3-го сорта – $600\cdot 0.133 = 79.8$, 4-го – $600\cdot 0.067 = 40.2$. Вычисленные значения ожидаемого числа отличны от значений, найденных проверкой. Это вызывает сомнение в справедливости утверждения поставщиков о том, что качество продукции не изменилось.

Воспользовавшись формулой (2.1), вычислим эмпирическое значение величины χ^2.

$$\chi^2 = \frac{(340-600\cdot 0.534)^2}{600\cdot 0.534} + \frac{(130-600\cdot 0.266)^2}{600\cdot 0.534} + $$
$$+ \frac{(100-600\cdot 0.133)^2}{600\cdot 0.133} + \frac{(30-600\cdot 0.067)^2}{600\cdot 0.067} = 14.4$$

Найдем число степеней свободы величины χ^2. Минуя весьма длительные обоснования, сообщим в виде рецепта, как найти число степеней свободы в задачах такого типа: для этого нужно из числа строк таблицы вычесть единицу и из числа столбцов таблицы вычесть единицу. Найденные числа перемножить. В нашем случае число степеней свободы $k = (2\ \ 1)\cdot(4\ \ 1) = 3$. Вычислим значение фактически достигаемого уровня значимости α.

$$\alpha = \int_{14.4}^{\infty} f\left(\chi^2, k=3\right) dx = 0.0024$$

Сформулируем результат. Отвергая гипотезу H_0: «качество продукции не изменилось», мы рискуем совершить ошибку первого рода с вероятностью 0.0024. Такое значение вероятности можно считать малым. Поэтому гипотезу H_0 отвергаем в пользу альтернативной гипотезы H_1: «качество изменилось».

Сформулируем экономический результат. В результате изменения технологии качество закупаемой продукции изменилось. Вероятность ошибки заключения равна 0.0024. Для решения задачи о том, в какую сторону изменилось качество, лучшую или худшую, необходимо дополнительное исследование.■

Сообщим краткие сведения о границах применимости критерия χ^2Пирсона. Обязательным условием применения критерия является группировка значений исследуемой случайной величины X на классы, которые по числу, размерам и началу более или менее произвольны, чем вызывают изменения в вычисленном использованием формулы (2.1) значении величины χ^2. Согласно содержанию заключения, напечатанного в монографии [65, с.663], это может привести к ненадежным и даже противоречивым выводам при использовании критерия. Краткое изложение результатов разбиения ряда значений исследуемой случайной величины на разное количество классов при применении критерия χ^2Пирсона можно найти в книге [66]. Вместе с тем закон, применением которого должно быть выполнено разбиение на классы, найден еще в 1933 г. В.И. Гливенко и опубликован в работе [37]. Применение найденного В.И. Гливенко закона требует использования весьма громоздких вычислительных процедур, которые не реализованы вплоть до настоящего времени в виде вычислительных программ. В качестве весьма удобной при решении прикладных задач рекомендации можно использовать результат, напечатанный в книге [66, с.587]: количество интервалов разбиения должно быть возможно меньшим, если нас интересует использование критерия при наибольшей возможной мощности.

Другим критерием, применение которого позволяет решать задачи об однородности выборок, является критерий Н.В. Смирнова.

2.4.2. Критерий Смирнова однородности выборок

Критерий найден в 1939 г. и впервые опубликован в работе [55]. Изложим теорему из работы [55], доказанную Н.В. Смирновым.

Теорема 2.1. ▌ *Если объемы выборок n_1 и n_2 неограниченно возрастают так, что отношение $n_1/n_2 = t$ остается постоянным, то*

$$P_{n_1,n2}(\lambda) = \left\{D(n_1,n_2) \le \lambda/\sqrt{n}\right\} \to \Phi(\lambda) \tag{2.2}$$

где $\lambda > 0$, $n = n_1 n_2/(n_1 + n_2) = n_2/(1+\tau)$,

$$\Phi(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2}, \tag{2.3}$$

каков бы ни был непрерывный закон распределения $F(x)$ величины X. ▌

В качестве меры $D(n_1,n_2)$ расхождения эмпирических кривых $Fn_1(x)$ и $Fn_2(x)$ в выражении (2.2) принята мера, определенная через аналитическое выражение

$$D(n_1,n_2) = \sup_{-\infty<x<\infty} \left|F_{n_1}(x) - F_{n_2}(x)\right|. \tag{2.4}$$

Функция *Ф(λ)*, как это было установлено в работе [56] А.Н. Колмогорова, представляет предельный закон распределения для величины

$$\sqrt{n}\,\sup\left|F_n(x)-F(x)\right|. \qquad (2.5)$$

где $F_n(x)$ – эмпирическая функция распределения.

Значения функции *Ф(λ)*, записанной выражением (2.3), табулированы. Таблицы можно найти во многих учебниках. Наиболее полные таблицы напечатаны в книге [57]. В целях удобства читателей в таблице 2.2 напечатан удобный для ориентировочного применения фрагмент таблицы значений функции *Ф(λ)*.

ТАБЛИЦА 2.2

Значение α	0.40	0.30	0.20	0.10	0.05	0.01	0.005	0.001
Значение λ_0	0.89	0.97	1.07	1.22	1.36	1.63	1.73	1.95

С точностью достаточной для решения большинства прикладных задач, значения λ_0 и соответствующие им значения α нетрудно найти, воспользовавшись линейной интерполяцией. Для значений λ_0 меньших, чем в таблице 2.2, А.Н. Колмогоровым найдена приблизительная асимптотическая формула (взята нами из монографии [65, с.661])

$$\alpha \approx 1-\frac{\sqrt{2\pi}}{\lambda}e^{-\pi^2/8\lambda^2}. \qquad (2.6)$$

Изложим кратко вычислительную схему применения критерия.

1. Использованием данных двух выборок ищутся эмпирические функции распределения Fn_1 и Fn_2.

2. Вычисляется значение меры расхождения $D(n_1,n_2)$ между функциями Fn_1 и Fn_2.

3. Вычисляется значение величины λ

$$\lambda = D(n_1,n_2)\sqrt{n}. \qquad (2.7)$$

4. Проверяется гипотеза H_0: $Fn_1 = Fn_2$, H_1: $Fn_1 \neq Fn_2$. Решение о принятии или отклонении гипотезы H_0 принимается применением правила: если вычисленное значение величины λ окажется больше значения λ_0, гипотеза H_0 отклоняется.

В качестве примера применения критерия рассмотрим пример

Пример. В течение месяца выборочно выполнялась проверка торговых точек города по продаже овощей. Результаты двух проверок по недовесам помещены в таблицу 2.3.

ТАБЛИЦА 2.3

Номер Интервала	Недовесы В г	Количества случаев недовеса	
		Для выборки 1	Для выборки 2
1	0-10	3	5
2	10-20	10	12
3	20-30	15	8
4	30-40	20	25
5	40-50	12	10
6	50-60	5	8
7	60-70	25	20
8	70-80	15	7
9	80-90	5	5
		$n_1 = 110$	$n_2 = 100$

Можно ли считать, что недовесы овощей описываются одним и тем же законом распределения.

Решение. ☐ Обозначим через $n_1^{нак}$ и $n_2^{нак}$ накопленные количества случаев недовесов, обнаруженных при первой и второй проверке. Найдем эмпирические функции распределения $Fn_1(x)$ и $Fn_2(x)$ и разность $|Fn_1(x)-Fn_2(x)|$. Результаты поместим в таблицу 2.4.

ТАБЛИЦА 2.4

x_i	$n_1^{нак}$	$n_2^{нак}$	$Fn_1(x)$	$Fn_2(x)$	$|Fn_1(x)-Fn_2(x)|$
10	3	5	0.027	0.050	0.023
20	13	17	0.118	0.170	0.052
30	28	25	0.254	0.250	0.004
40	48	50	0.436	0.500	0.064
50	60	60	0.545	0.600	0.055
60	65	68	0.591	0.680	0.089
70	90	88	0.818	0.880	0.072
80	105	95	0.955	0.950	0.005
90	110	100	1.000	1.000	0.000

Находим в последнем столбце таблицы наибольшее значение. Оно равно 0.089. Это означает, что имеет место равенство

$$\sup|Fn_1(x) - Fn_2(x)| = 0.089.$$

Воспользовавшись формулой (2.7), вычислим значение λ.

$$\lambda = \sqrt{(110\cdot 100)/(110+100)}\cdot 0.089 = 0.644.$$

Обратимся к содержанию таблицы 2.2. Вычисленное нами значение $\lambda = 0.644$ меньше, чем любое значение λ_0 из таблицы. Воспользовавшись асимптотической формулой (2.6), выполнив вычисления, находим, что $\alpha \approx 0.8$. Это означает, что, отвергая гипотезу H_0: $Fn_1 = Fn_2$ в пользу гипотезы H_1: $Fn_1 \approx Fn_2$, мы рискуем совершить ошибку первого рода с вероятностью $\alpha \approx 0.8$. Вероятность ошибки можно считать весьма большой. В частности, значительно превышающей принятую во многих областях вероятность, равную 0.05. Следовательно, гипотезу H_0: $Fn_1 = Fn_2$ не отвергаем.

Сформулируем окончательное заключение. Значения случайной величины – недовесов покупателям, обнаруженным в двух разных проверках, описываются одной и той же функцией распределения. Найденный результат можно интерпретировать, как одну из социальнопсихологических особенностей продавцов: отношение к покупателям у разных продавцов одинаковое. ■

Ограничением к применению результатов теоремы Смирнова в качестве критерия однородности выборок является требование непрерывности величины *X*.

2.5. Критерии зависимости и независимости результатов наблюдений

Изложим весьма полезные при выполнении процедуры первичной обработки данных способы, применение которых позволяет решить задачу зависимости или независимости результатов наблюдений. В доступном для приложений изложении методы решения можно найти в книге [59, с.106]. В более строгом – пособии [60, с.128,134] и монографии [61]. Особое значение задача приобретает при анализе нестационарных данных.

Известно, что результаты наблюдений могут быть результатами реализации случайных величин с неизвестным законом распределения. В таких случаях отдельные исследования удобно выполнять на основе свободных от распределений непараметрических методов. Запишем определение, заимствованное из энциклопедии [8].

Определение 2.16. ▍ *Статистический критерий называется непараметрическим, если статистические выводы, получаемые с помощью этого критерия, не зависят от распределений вероятностей случайных величин, по результатам наблюдений которых проверяют гипотезу* H_0 *против гипотезы* H_1. ▍

Изложим краткие сведения об отдельных непараметрических критериях и способах их применения.

2.5.1. Критерий серий

Предложен в 1940 г. Вальдом и Вольфовитцем.

Рассмотрим последовательность *N* наблюдений значений случайной величины *X*, причем каждое наблюдение отнесено к одному из двух взаимно исключающих классов, которые обозначим через (+) и (-). В результате чего последовательность наблюдений приобретает вид, типичный случай которого можно записать, как

++	–	++	–	+++	–	+	– –	+	– –	+	– – –
1	2	3	4	5	6	7	8	9	10	11	12

Определение 2.17. ▍ *Серией будем называть последовательность однотипных наблюдений, перед и после которой следуют наблюдения противоположного типа или же нет никаких наблюдений.* ▍

В нашем примере в последовательности из $N = 20$ наблюдений имеется $r = 12$ серий.

Число серий, появившихся в последовательности наблюдений, позволяет выяснить, являются ли отдельные результаты независимыми наблюдениями одной и той же случайной величины или в последовательности присутствует тренд. Проверяется гипотеза H_0: «тренда нет» при альтернативной гипотезе H_1: «тренд есть». Если вероятность отдельных исходов (+) или (-) не меняется от наблюдения к наблюдению, то выборочное распределение числа серий в последовательности является случайной величиной *r* с математическим ожиданием

$$\mu_r = \frac{2N_1N_2}{N} + 1$$

и дисперсией

$$D[r] = \frac{2N_1N_2(2N_1N_2 - N)}{N^2(N-1)},$$

где N_1 – число исходов (+), N_2 – число исходов (-).

В монографии [61] имеется доказательство того, что при достаточно больших значениях чисел N_1 и N_2 ($N_1+N_2 \geq 20$) значения критерия распределены по закону $f_N(r, m_r, D[r])$. Для меньших значений чисел N_1 и N_2 значения критерия можно найти в таблицах [57].

2.5.2. Критерий инверсий

Как и критерий серий применяется при решении задачи проверки независимости результатов наблюдений.

Пусть дана последовательность из N наблюдений случайной величины X, обозначенных x_i, $i = 1,...,N$. Сосчитаем, сколько раз в последовательности имеют место неравенства $x_i > x_j$ *при* $i < j$. Выполнение каждого из таких неравенств называется инверсией. Обозначим через A общее число инверсий. Для уяснения понятия инверсии рассмотрим пример.

Пример. Дана последовательность наблюдений

$$x_1 = 5, x_2 = 3, x_3 = 8, x_4 = 9, x_5 = 4, x_6 = 1, x_7 = 7, x_8 = 5.$$

Найти число инверсий в последовательности.

Решение. □ В этой последовательности $x_1 > x_2$, $x_1 > x_5$, $x_1 > x_6$, т. е. число инверсий A_1 для элемента x_1 равно трем ($A_1 = 3$). Теперь возьмем элемент x_2 и сравним его с последующими наблюдениями. Обнаруживаем, что имеется только один случай инверсии: $x_2 > x_6$. Следовательно, число инверсий A_2 для элемента x_2 равно единице ($A_2 = 1$). Продолжая процедуру сравнения для элементов x_3 находим, что $A_3 = 4$, $A_4 = 4$, $A_5 = 1$, $A_6 = 0$, $A_7 = 1$. Общее число A инверсий равно

$$A = A_1 + A_2 + ... A_7 = 3 + 1 + 4 + 4 + 1 + 0 + 1 = 14. \blacksquare$$

Продолжим описание критерия. Доказано (см. например, в пособии [93, с.134]), что если последовательность из N наблюдений состоит из независимых исходов одной и той же случайной величины, то число инверсий является случайной величиной A с математическим ожиданием

$$\mu = \frac{N(N-1)}{4}$$

и дисперсией

$$D[A] = \frac{2N^3 + 3N^2 - 5N}{72}.$$

Доказано (см. пособие [60, с.134]), что при больших значениях N случайная величина A распределена приблизительно по нормальному закону $f_N(A, m_A, D[A])$. Точные значения критерия можно найти в таблицах [57].

Сообщим дополнительные сведения о применимости результатов, найденных использованием критериев серий и инверсий. Доказано (см. монографию [61]), что критерий инверсий является более мощным по сравнению с критерием серий при решении задач обнаружения монотонного тренда в последовательности наблюдений. Критерий серий является более мощным по сравнению с критерием инверсий при обнаружении немонотонного тренда, часто называемого в приложениях трендом флуктуаций.

2.6. Доверительный интервал как мера адекватности

2.6.1. Доверительный интервал оценивания параметра

Метод доверительных интервалов разработан Ю. Нейманом (18941981), исходя из идей Р. Фишера. Краткое изложение теоретических основ теории доверительного оценивания можно найти в монографии [62, с.420], весьма подробное — в монографии [64].

Определение 2.18. ▌*Доверительным интервалом оценки параметра* Θ^* *называют интервал* $(\Theta^*-\Delta,\Theta^*+\Delta)$, который накрывает значение Θ^* *параметра с заданной надежностью* γ. ▌

Наряду с термином доверительный интервал используется равносодержательный термин *доверительные пределы*. В таком случае принято говорить о нижнем и верхнем доверительном пределе. Доверительный интервал связан с двухсторонней критической областью (см. *определение 2.13*). Действительно, отыскивая двухстороннюю критическую область при уровне значимости α, тем самым находят и доверительный интервал с надежностью $\gamma = 1 - \alpha$. Вместе с тем, статистическая интерпретация результатов нахождения доверительного интервала и двухсторонней критической области различна: двухсторонняя критическая область интерпретируется как границы, между которыми заключена доля (1 α) численных значений критерия, найденная при повторении опытов; доверительный интервал — как границы, между которыми в $(\gamma = 1 - \alpha)\%$ опытов заключено истинное значение оцениваемого параметра.

Перечислим ключевые отличия интервального оценивания от точечного.

1) Доверительный интервал возможных значений известного значения параметра Θ может быть интерпретирован как оценка в геометрическом смысле менее точная, т. к. его характеристики описывают не единственное значение параметра, а множество возможных значений.

2) С другой стороны, утверждение $P\{\Theta \in (\Theta^*-\Delta,\Theta^*+\Delta)\} \geq 1 - \alpha$ является истинным, в то время как событие $\Theta = \Theta^*$, как правило, имеет вероятность равную нулю, т. е. может быть интерпретировано как событие невозможное.

Для нахождения с надежностью γ нижнего и верхнего пределов доверительного интервала $(\Theta^*-\Delta,\Theta^*+\Delta)$ параметра Θ весьма желательно знание аналитического выражения закона распределения параметра Θ. Для отдельных случаев такие законы известны. Например, известно, что если случайная величина X распределена нормально, то выборочная средняя $x_{ср.}$, найденная по независимым наблюдениям, также распределена нормально. Однако во многих случаях аналитическое выражение закона распределения значений оцениваемого параметра неизвестно. Тогда численные характеристики искомого доверительного интервала находят косвенно, вводя в рассмотрение другую случайную величину с известным законом распределения. Например, при нахождении доверительного интервала для оценки математического ожидания нормального закона при неизвестном значении среднеквадратического отклонения σ в рассмотрение вводят случайную величину T, определяемую через выражение

$$\mathrm{T} = \frac{E_{ср.} - M[X]}{S\sqrt{n-1}},$$

которая распределена по закону Стьюдента с $k = n - 1$ степенями свободы; здесь $x_{ср.}$ – выборочная средняя, S – точечная оценка исправленного среднеквадратического отклонения.

Тогда вероятность

$$P\left(x_{ср.} - t_\gamma S\sqrt{n-1} < M[X] < x_{ср.} - t_\gamma S\sqrt{n-1}\right) = \gamma$$

нетрудно найти, воспользовавшись выражением

$$\gamma = 2\int_0^{t_\gamma} S(t,n)dt$$

в котором $S(t, n)$ – плотность распределения Стьюдента (см. гл.1).

2.6.2. Доверительный интервал для эмпирической функции распределения

Понятие доверительного интервала для эмпирической функции распределения фактически уже использовано нами при описании вероятностного смысла формулы (1.19). Вместе с тем границы доверительного интервала для эмпирической функции распределения $F_n(x)$ можно найти, воспользовавшись критерием А.Н. Колмогорова, сведения в доступном изложении о способах применения которого можно найти в учебнике [67, с.172]. В качестве прототипа описания применения критерия для нахождения доверительных границ эмпирической функции распределения мы использовали содержание книги [66, с.].

Известно, что какова бы ни была истинная функция распределения $F(x)$, мы имеем, обозначая через $d(\alpha)$ критическое значение D_n при размере критерия α,

$$P\left\{D_n = \sup_x\left|F_n(x) - F(x)\right| > d(\alpha)\right\} = \alpha$$

Записанное аналитическое выражение нетрудно переписать в виде доверительного утверждения

$$P\left\{F_n(x) - d(\alpha) \le F_n(x) + d(\alpha)\,\forall x\right\} = 1 - \alpha$$

Таким образом, мы располагаем полосу шириной $\pm\, d(\alpha)$ вокруг эмпирической функции $F_n(x)$, и с вероятностью $1 - \alpha$ функция $F(x)$ лежит целиком внутри этой полосы.

Никакой другой критерий согласия не позволяет такого обращения критерия в доверительный интервал, т. к. ни один критерий не использует такой прямой и просто интерпретируемой меры расхождения, как D_n. Вместе с тем существенным ограничением возможностей применения критерия является необходимость априорного знания аналитического выражения для функции $F(x)$, в том числе знание значений входящих в нее параметров. При решении прикладных задач такими сведениями мы располагаем весьма редко.

Распространенной ошибкой применения критерия является вычисление значений оценок параметров функции $F(x)$ по данным выборки. Примеры такой ошибки можно найти в литературе, в том числе, учебной.

Глава 3. Интерполяция и аппроксимация вероятностных распределений

3.1. Вводные замечания

Цель главы – ознакомить читателя с содержанием отдельных современных результатов в теории вероятностей и математической статистике.

Из содержания общего курса теории вероятностей и математической статистики известно, что статистическим аналогом функции распределения *F(x)* случайной величины *X* является эмпирическая функция распределения $F_n(x)$, которую нетрудно найти, воспользовавшись выборкой численных значений исследуемой случайной величины. Из результатов теоремы Гливенко-Кантелли (см. главу 1) известно, что при неограниченном возрастании объема выборки *n* функция $F_n(x)$ приближается к функции *F(x)*. Изложенный закон принято записывать, как

$$P\left(\lim_{n\to\infty}\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0\right)=1 \tag{3.1}$$

По мнению многих известных специалистов, например, автора ставшего классическим образцом учебника «Курс теории вероятностей» [23], теорема Гливенко-Кантелли и некоторые примыкающие к ней теоремы являются основной теоремой математической статистики. Несмотря на интуитивную понятность содержания теоремы Гливенко-Кантелли, изучение ее строгого доказательства возможно только при наличии у читателя основательных знаний современной математики. Возможно, что по этой причине теорема не включена в содержание многих учебников и пособий. Например, известного и многократно переизданного пособия [12].

Вернемся к содержанию теоремы Гливенко-Кантелли. Другими словами соотношение (3.1) означает, что отклонение D_n, определенное через аналитическое выражение

$$D_n=\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|, \tag{3.2}$$

эмпирической функции $F_n(x)$ от теоретической *F(x)*, на всей оси, при неограниченном возрастании объема выборки *n*, с вероятностью единица будет сколь угодно мало. Применение результатов известной теоремы А.Н. Колмогорова (1933) позволяет для конечных значений объемов выборок *n* оценивать вероятность отклонений случайной величины D_n от нуля.

По нашему мнению к основной теореме математической статистики примыкает теорема В.И. Зубова, опубликованная в 1991 г. в работе [71].

3.2. Теорема В.И. Зубова

В излагаемом ниже разделе сохранены обозначения, примененные В.И. Зубовым.

Рассмотрим множество F_1 всех непрерывных распределений *f(x)*, $x\in(-\infty,+\infty)$, *f(x)* – монотонная функция, $f(x)\to +1$ при $x\to+\infty$, $f(x)\to 0$ при $x\to-\infty$. Выберем какую-либо функцию распределения $f_0(x)\in F_1$ и построим семейство функций

$$f_j(x)=f_0\left(\frac{x-\xi_j}{\sigma_j}\right),\qquad j=1,\dots,n.$$

Ясно, что каждая из этих функций при любом выборе вещественного числа ξ_j и любого положительного числа σ_j будет функцией распределения.

Теорема 3.1 (В.И. Зубов). ▎*Множество всевозможных линейных комбинаций* $\sum_{j=1}^{n}\alpha_j f_j(x)$ *будут всюду плотными в* F_1, *где* α_j – *неотрицательные константы такие, что* $\sum_{j=1}^{n}\alpha_j=1$. *Иначе говоря, для любой функции распределения* $f(x)\in F_1$ *для каждого* $\varepsilon>0$ *можно указать число n, вещественные числа* ξ_j, *положительные числа* σ_j, *неотрицательные числа* α_j, $\sum_{j=1}^{n}\alpha_j=1$, *такие, что*

$$\left|f(x)-\sum_{j=1}^{n}\alpha_j f_j(x)\right|<\varepsilon \tag{3.3}$$

при $x\in(-\infty,+\infty)$. ▎

Следствие 1 из теоремы 3.1. ▎*Предположим*

$$f_0(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\xi^2/2}d\xi.$$

При таком выборе функции $f_0(x)$ *функции* f_j *будут нормальными распределениями с математическими ожиданиями* ξ_j *и дисперсиями* σ^2_j

$$f_j=\frac{1}{\sigma_j\sqrt{2\pi}}\int_{-\infty}^{x}e^{-(\xi-\xi)^2/2\sigma^2}d\xi.$$

Из содержания теоремы 3.1 и ее следствия следует, что в этом случае всевозможные смеси нормальных распределений образуют всюду плотное подмножество пространства F_1. Иначе говоря, любое заданное распределение $f(x)$ может быть сколь угодно точно аппроксимировано смесью нормальных распределений в равномерной метрике в целом. А именно, найдутся математические ожидания ξ_j, дисперсии σ^2_j и неотрицательные величины α_j, $\Sigma a_j = 1$, такие, что

$$\left| f\left(x\right) - \sum_{j=1}^{n} \alpha_j \frac{1}{\sigma_j \sqrt{2\pi}} \int_{-\infty}^{x} e^{-(\xi-\xi)^2/2\sigma^2} d\xi \right| < \varepsilon \qquad (3.4)$$

при $x \in (-\infty, +\infty)$.

Доказательство теоремы В.И. Зубова можно найти в дополнении к книге [72]. Развитие результатов теоремы до возможностей численного применения – в работе [73], опубликованной в приложении к книге [72], методическом пособии [74] и др. работах. В работах [72], [73], [75] и др. опубликован метод, применение результатов которого позволяет использованием выборочных данных найти аналитическое выражение закона распределения в виде смеси нормальных законов.

Отдельное место среди смесей занимают смеси нормальных распределений. Использованием свойств смесей нормальных распределений В.И. Зубовым найдены результаты, применение которых позволило сформулировать новый подход к проблеме обращения центральной предельной теоремы А.М. Ляпунова (см. работу [76]).

3.3. Статистическая схема возникновения смесей

В справочнике [40, с.27] изложена статистическая схема возникновения случайной величины X с функцией распределения смесь вероятностных распределений.

Пусть даны случайные величины Y и Z с известными плотностями вероятности $f_1(y)$ и $f_2(z)$. Пусть случайная величина Y связана с некоторым комплексом условий «1», а случайная величина Z – с комплексом условий «2», т. е. всякий раз, когда в процессе испытания реализуется комплекс условий «1», реализуется одно из возможных значений величины Y, а при реализации комплекса условий «2» реализуется одно из возможных значений величины Z. Предположим, что в бесконечной последовательности независимых испытаний реализации комплекса условий «1» чередуются случайным образом с реализациями комплекса условий «2». Причем вероятность того, что при очередном испытании будет реализован комплекс условий «1», равна α_1, а вероятность того, что будет реализован комплекс условий «2», равна α_2

$(\alpha_1 + \alpha_2 = 1)$. Выполнив бесконечное число таких испытаний, получим смесь реализаций случайных величин Y и Z, в которой доля реализаций случайной величины Y равна α_1, а доля реализаций случайной величины Z равна α_2. Будем рассматривать полученную таким образом смесь как генеральную совокупность значений новой случайной величины X. Тогда плотность вероятности величины X (т. е. плотность вероятности смеси случайных величин Y и Z) определяется через аналитическое выражение

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x).$$

Функция распределения смеси двух распределений запишется, как

$$F(x) = \alpha_1 F_1(x) + \alpha_2 F_2(x).$$

В общем случае, в котором генеральная совокупность значений случайной величины X представляет собой смесь значений n независимых случайных величин $X_1, X_2, \ldots, X_n$, функция распределения запишется, как

$$F\left(x\right) = \sum_{j=1}^{n} \alpha_j F_j\left(x\right), \qquad j = 1, \ldots, n.$$

Примеры успешного применения смесей нормальных законов при решении задачи нахождения функции распределения случайной величины использованием выборок можно найти в работах [73], [75].

Остановимся подробнее на отдельных свойствах функции распределения смесь нормальных распределений. Будем обозначать эту функцию через $F_{см}(x)$, соответствующую плотность вероятности – через $f_{см}(x)$. Запишем аналитические выражения для функций $F_{см}(x)$ и $f_{см}(x)$:

$$F_{см}\left(x\right) = \sum_{j=1}^{n} \alpha_j F_N\left(x, a_j, \sigma_j\right), \quad f_{см}\left(x\right) = \sum_{j=1}^{n} \alpha_j F_N\left(x, a_j, \sigma_j\right) \qquad (3.5)$$

Можно доказать, что существуют такие значения характеристик α_j, a_j и σ_j, $j = 1,\ldots,n$ функции $f_{см}(x)$, при которых функция $f_{см}(x)$ n-модальна. В качестве примера на рис.4 помещена графическая интерпретация плотности вероятности $f_{см}(x)$, записанной аналитическим выражением

$$f_{см}(x) = 0.3 \cdot f_N(x,1,0.238) + 0.39 \cdot f_N(x,2,0.238) +$$
$$+ 0.21 \cdot f_N(x,3,0.238) + 0.1 \cdot f_N(x,4,0.238). \qquad (3.6)$$

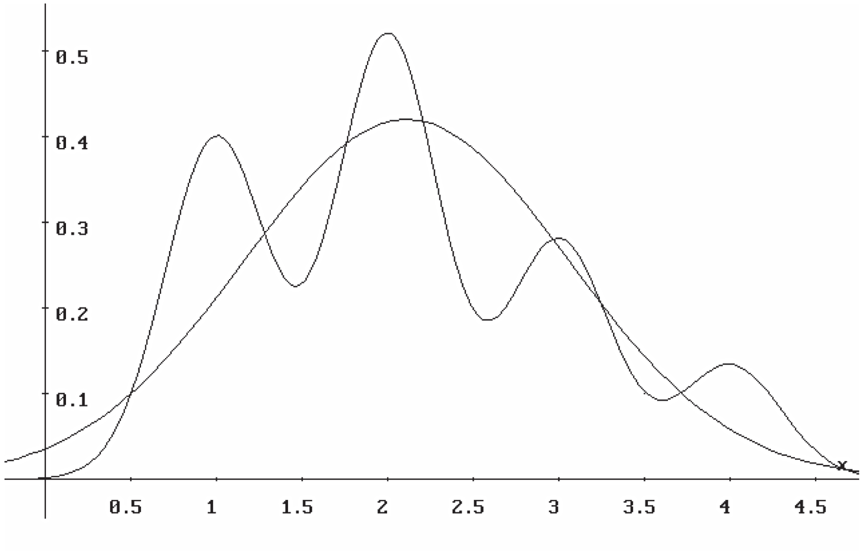

Рис.4.

На рис.4: по оси абсцисс – значения аргумента функции (3.6), по оси ординат – значения функции плотности вероятности смеси нормальных распределений (3.6) и плотности вероятности $f_N(x, 2.11, 0.95)$.

Правомерность применения аппроксимации искомой функции распределения смесью нормальных законов использованием выборок численных значений непрерывной случайной величины произвольной природы следует из результатов теоремы 5.2, опубликованной в работе [76].

Теорема 3.2 (В.И. Зубов). ▌ ***Каждое непрерывное, строго монотонное распределение вероятностей имеет семейство случайных величин, являющихся непрерывными функциями, заданными на всей вещественной оси, изменяющихся строго монотонно от*** $-\infty$ ***до*** $+\infty$***. При этом каждая случайная величина этого семейства имеет одну и ту же функцию распределения, которая представляет упомянутые выше распределения вероятностей.*** ▌

В целях иллюстрации преимуществ применения смесей нормальных законов при решении задачи аппроксимации закона распределения использованием данных выборки, приведем пример 3.1. Для сравнения результатов аппроксимации в примере также выполнена аппроксимация нормальным законом с использованием результатов из той же выборки. Перед изложением содержания примера напомним еще одно определение критерия, заимствованное нами из пособия [12, с.287, 314], которое в интересах читателя дано в пособии в двух формулировках.

Определение 3.1. ▌***1. Мощностью критерия называется вероятность того, что нулевая гипотеза будет отвергнута, если справедлива конкурирующая гипотеза.***

2. Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. ▌

Пример 3.1. Известно, что отпускная цена одного и того же товара в разных торговых предприятиях может значительно отличаться. В таблице 5.1 помещены сведения о ценах (в у.е.) одного и того же товара и количества предприятий торговли из 100 обследованных предприятий, продающих товар.

ТАБЛИЦА 3.1

Цена в у.е.	0.5-1	1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4	4-4.5
Число предприятий	15	15	20	19	11	10	5	5

Найти. Закон распределения цены товара а) в виде нормального закона; б) в виде смеси нормальных распределений. Проверить результаты применением критерия χ^2-Пирсона.

Решение. □ Выполнив очевидные вычисления находим, что значение точечной оценки среднего $x_{ср} = 2.11$, значение точечной оценки дисперсии $S^2 = 0.898$, оценки среднеквадратического отклонения $S = 0.95$. Сформулируем гипотезу

$$H_0\text{: } F(x) = F_N(x, 2.11, 0.95),\ H_1\text{: } F(x) \neq F_N(x, 2.11, 0.95).$$

Воспользовавшись данными из таблицы 3.1, выполнив разбиение на 4 интервала (с целью достижения критерием наибольшей мощности), вычислим значение величины χ^2. С этой целью вычислим значения интегралов

$$p_1 = \int_{-\infty}^{1.5} f_N\left(x, 2.11, 0.95\right) dx = 0.26\,,$$

$$p_2 = \int_{1.5}^{2.5} (\cdot) = 0.4\,,$$

$$p_3 = \int\limits_{2.5}^{3.5} (.) = 0.27\,,$$

$$p_4 = \int\limits_{3.5}^{\infty} (.) = 0.07\,.$$

Вычислим значение χ^2

$$\chi^2 = \frac{(30-100\cdot 0.26)^2}{100\cdot 0.26} + \frac{(39-100\cdot 0.4)^2}{100\cdot 0.4} + \frac{(21-100\cdot 0.27)^2}{100\cdot 0.27} +$$

$$+\frac{(10-100\cdot 0.07)^2}{100\cdot 0.07} = 3.26.$$

Вычислим значение уровня значимости α при 1 степени свободы

$$\alpha = \int\limits_{3.26}^{\infty} f\left(\chi^2, k=1\right) dx = 0.07\,.$$

Сформулируем окончательный результат. Отклоняя гипотезу H_0: $F(x) = F_N(x, 2.11, 0.95)$ мы рискуем совершить ошибку первого рода с вероятностью $\alpha = 0.07$. Считая значение вероятности 0.07 значительным, гипотезу H_0 не отклоняем.

Сформулируем экономическую интерпретацию результата. Случайная величина – цена на товар распределена по нормальному закону.

Дадим математические комментарии. В силу того, что вероятность 0.07 можно интерпретировать, как относительно небольшую, неотрицание гипотезы H_0 вовсе не означает, что закон распределения в действительности нормальный. Правильнее сказать, что закон нормального распределения не исключен. Вместе с тем такое (более строгое) заключение вряд ли удовлетворит заказчика.

Выполним аппроксимацию закона распределения по данным выборки применением результатов развития теорем 3.1 и 3.2 (В.И. Зубова) методом, найденным А.И. Ивановым. Применив результаты из работ [73], [74], [75], за неимением места не излагая весьма громоздкие вычисления, запишем окончательное аналитическое выражение

$$f_{см}(x) = 0.3\cdot f_N(x, 1, 0.238) + 0.39\cdot f_N(x, 2, 0.238) +$$

$$+ 0.21\cdot f_N(x, 3, 0.238) + 0.1\cdot f_N(x, 4, 0.238).$$

Записанное выражение уже использовалось нами в процессе объяснения понятия смесей в качестве примера (см. выражение (3.6)). Выполним проверку гипотезы H_0: $F(x) = F_{см}(x)$ при альтернативе H_1: $F(x) \neq F_{см}(x)$. Исходя из количества параметров найденной функции $F_{см}(x)$ приходим к заключению о том, что интервал значений цен для применения критерия χ^2-Пирсона нужно разбить на 8 частей. Приступим к вычислению значения χ^2. Вычислим значения теоретических вероятностей.

$$p_1 = \int\limits_{-\infty}^{1} f_{см}(x)\,dx = 0.15\,,\quad p_2 = \int\limits_{1}^{1.5} (.) = 0.152\,,\quad p_3 = \int\limits_{1.5}^{2} (.) = 0.194\,,$$

$$p_4 = \int\limits_{2}^{2.5} (.) = 0.19\,,\quad p_5 = \int\limits_{2.5}^{3} (.) = 0.11\,,\quad p_6 = \int\limits_{3}^{3.5} (.) = 0.1\,,$$

$$p_7 = \int\limits_{3.5}^{4} (.) = 0.052\,,\quad p_8 = \int\limits_{4}^{\infty} (.) = 0.05.$$

Запишем выражение для значения χ^2 и выполним вычисление.

$$\chi^2 = \frac{(15-100\cdot 0.15)^2}{100\cdot 0.15} + \frac{(15-100\cdot 0.152)^2}{100\cdot 0.152} + \frac{(20-100\cdot 0.194)^2}{100\cdot 0.194} +$$

$$+\frac{(19-100\cdot 0.19)^2}{100\cdot 0.19} + \frac{(11-100\cdot 0.11)^2}{100\cdot 0.11} + \frac{(10-100\cdot 0.1)^2}{100\cdot 0.1} + \frac{(5-100\cdot 0.052)^2}{100\cdot 0.052} +$$

$$+\frac{(5-100\cdot 0.05)^2}{100\cdot 0.05} = 0.045$$

Вычислим значение уровня значимости α при 1 степени свободы

$$\alpha = \int\limits_{0.045}^{\infty} f\left(\chi^2, k=1\right) dx = 0.83\,,$$

т. е. отвергая гипотезу H_0: $F(x) = F_{см}(x)$ мы рискуем совершить ошибку первого рода с вероятностью 0.83, поэтому гипотеза H_0: $F(x) = F_{см}(x)$ принимается.

Выполним численное сравнение найденных результатов. С этой целью вычислим значения вероятностей ошибок второго рода β, характеризующих первый и второй результат. Воспользуемся асимптотической формулой для вычисления приблизительного значения величины β при относительно больших значениях объема выборки n (n – десятки и больше), рекомендованной в книге [9, с.314]. Подробные сведения о правомерности применения формулы нетрудно найти в книге [66, с.310].

$$\beta = 1 - \int\limits_{\left(\frac{k+\lambda}{k+2\lambda}\right)\chi^2(\alpha,k)}^{\infty} \chi^2\left(k + \frac{\lambda^2}{k+2\lambda}\right) dx\,, \tag{3.7}$$

где k – число степеней свободы, λ – размер критерия, в нашем случае $\lambda = \alpha$, $\chi^2(\alpha, k)$ – вычисленное в результате проверки гипотезы значение величины χ^2. В той же книге можно найти формулы для вычисления вероятности β, возникающей при применении некоторых других критериев.

Вычислим значение вероятности ошибки второго рода β для случая, в котором проверялась гипотеза

$$H_0\text{: } F(x) = F_N(x,\ 2.11,\ 0.95),\ H_1\text{: } F(x) \neq F_N(x,\ 2.11,\ 0.95).$$

Найдено (см. выше), что в таком случае $\lambda = \alpha = 0.07, k = 1, \chi^2(0.07,1) = 3.26$. Подставив перечисленные значения λ, k и χ^2 в правую часть формулы (3.7), выполнив вычисления, находим, что $\beta = 0.92$.

Вычислим значение вероятности ошибки второго рода β для случая, в котором проверялась гипотеза

$$H_0\text{: } F(x) = F_{см},\ H_1\text{: } F(x) \neq F_{см}.$$

Найдено (см. выше), что в таком случае $\lambda = \alpha = 0.83, k = 1, \chi^2(0.07,1) = 0.045$. Подставив перечисленные значения в правую часть формулы (3.7), выполнив вычисления, находим $\beta = 0.08$.

Сравнив найденные значения вероятностей ошибок второго рода, приходим к заключению о том, что, приняв гипотезу о нормальности распределения исследуемой случайной величины – цены, мы рискуем совершить ошибку принятия неверной гипотезы с вероятностью 0.98. Приняв гипотезу о том, что случайная величина – цена распределена по закону смесь нормальных распределений, мы рискуем совершить ошибку принятия неверной гипотезы с вероятностью 0.08.

Замечание. ° У читателя может сложиться впечатление, что справедливо равенство $\alpha + \beta \approx 1$. Однако, в общем случае, это не так. ●

Сформулируем окончательный результат. Закон распределения цены продукта найден. Он записывается аналитическим выражением (3.6). Найденный закон – смесь нормальных распределений значительно адекватнее описывает вероятностные свойства исследуемой случайной величины – цены, чем закон нормального распределения. Вероятность того, что принятая гипотеза о законе распределения смесь нормальных распределений не верна равна 0.08.

Выполним экономическую интерпретацию результата. В рамках примера доказано, что при решении экономических задач нужно пользоваться функцией распределения смесь нормальных распределений. Применение аппроксимации нормальным законом неизбежно приведет к ошибочным результатам. ■

Продолжим рассмотрение содержания задачи из примера 3.1. Во многих изложениях нетрудно найти заявление о том, что многие величины, изучаемые в экономических науках, распределены по логнормальному закону. В рамках данных из примера 3.1 проверим адекватность заявления. С этой целью найдем значения точечных оценок m^* и a^* параметров логнормального закона распределения. Воспользовавшись содержанием гл.1, перепишем аналитическое выражение для логнормального закона, обозначив его через $f_{лог}(x)$

$$f_{лог}(x) = \frac{1}{xa\sqrt{2\pi}} e^{-\frac{(\ln(x/m))^2}{2a^2}}.$$

Перепишем из гл.1 формулы для вычисления значений точечных оценок m^* и a^*, для простоты вычислений ограничившись формулами, найденными методом моментов

$$m^* = \frac{x^*}{\left(1+\left(\frac{S}{x^*}\right)^2\right)^{0.5}},\quad a^* = \left[\ln\left[1+\left(\frac{S}{x^*}\right)^2\right]\right]^{0.5}. \tag{ММ}$$

Воспользовавшись тем, что значения точечных оценок x^* и S нами уже найдены при проверке адекватности аппроксимации данных из таблицы 3.1 законом нормального распределения, запишем $x^* = 2.11,\ S = 0.95$. Подставив значения $x^* = 2.11$ и $S = 0.95$ в формулы (ММ), выполнив вычисления, находим: $m^* = 1.92,\ a^* = 0.43$. Подставив найденные значения в аналитическое выражение логнормального закона, запишем

$$f_{лог}(x) = \frac{1}{x\cdot 0.43\cdot\sqrt{2\pi}} e^{-\frac{(\ln(x/1.92))^2}{2\cdot 0.43^2}}.$$

Для краткости будем записывать найденный закон как $f_{лог}(x,\ 1.92,\ 0.43)$. Применив критерий χ^2-Пирсона выполним проверку гипотезы о том, что цена товара (см. таблицу 3.1) есть случайная величина, распределенная по закону $f_{лог}(x,\ 1.92,\ 0.43)$. Запишем гипотезу аналитически

$$H_0\text{: } f(x) = f_{лог}(x,\ 1.92,\ 0.43),\ H_1\text{: } f(x) \neq f_{лог}(x,\ 1.92,\ 0.43).$$

Вычислим значение величины χ^2. Для этого в начале вычислим значения вероятностей p_i, $i = 1,\ldots,4$, входящих в выражение для вычисления значения величины χ^2.

$$p_1 = \int_0^{1.5} f_{лог}(x, 1.92, 0.43)\,dx = 0.283,\qquad p_2 = \int_{1.5}^{2.5} (.) = 0.447,$$

$$p_3 = \int_{2.5}^{3.5} (.) = 0.188,\qquad p_4 = \int_{3.5}^{\infty} (.) = 0.081.$$

Воспользовавшись найденными значениями вероятностей, вычислим значение величины χ^2

$$\chi^2 = \frac{(30-100\cdot 0.283)^2}{100\cdot 0.283} + \frac{(39-100\cdot 0.477)^2}{100\cdot 0.477} + \frac{(21-100\cdot 0.188)^2}{100\cdot 0.188} + $$
$$+ \frac{(10-100\cdot 0.081)^2}{100\cdot 0.081} = 2.39.$$

В силу того, что количество параметров в логнормальном законе равно двум, а число интервалов разбиения для вычисления значения величины χ^2 равно четырем, воспользовавшись аналитическим выражением закона распределения величины χ^2 с одной степенью свободы, вычислим достигаемое критерием значение вероятности ошибки первого рода α

$$\alpha = \int_{2.39}^{\infty} f\left(\chi^2, k=1\right) dx = 0.12 .$$

Следовательно,
отвергая гипотезу H_0:*f(x) =* $f_{лог}$*(x, 1.92, 0.43)*
в пользу гипотезы H_1: *f(x)* $\neq f_{лог}$*(x, 1.92, 0.43)*
мы рискуем совершить ошибку отвержения правильной гипотезы с вероятностью $\alpha = 0.12$. Считая вероятность 0.12 достаточно большой, гипотезу H_0 не отклоняем. Вычислим вероятность того, что принятая гипотеза H_0 ошибочная. Воспользовавшись формулой (3.7), выполнив вычисления, находим: $\beta = 0.86$. Выполним сравнение результатов аппроксимации.

1. Нормальным законом: $\alpha = 0.07$, $\beta = 0.92$.
2. Логнормальным законом: $\alpha = 0.12$, $\beta = 0.86$.
3. Законом смесь нормальных законов: $\alpha = 0.83$, $\beta = 0.08$.

Найденный результат – очередное свидетельство в пользу того, что исследуемая случайная величина – цена товара распределена по закону смесь нормальных распределений.

Продолжим изложение материала по методам нахождения значения ошибки второго рода β в процессе проверки гипотез. По нашему мнению, в большинстве учебников и пособий алгоритм нахождения значений вероятности β если и изложен, то весьма кратко. По нашему мнению, одним из немногих пособий, в котором описана схема проверки гипотез, доведенная до возможностей численного применения, является пособие [77]. Кроме того, в доступном изложении метод нахождения значения вероятности β можно найти в справочнике [89, с.27]. Изложим вычислительную схему.

1. Исходя из содержания прикладной экономической (физической) задачи формулируют основную гипотезу H_0.

2. Ориентировочно задаются величиной уровня значимости критерия α, т. е. вероятностью отвергнуть гипотезу H_0 в случае, в котором она верна. В наших примерах уровень значимости α мы вычисляли, что значительно облегчает принятие адекватного в смысле численного обоснования решения. При наличии столь же правдоподобной альтернативы H_1 уровень значимости следует выбрать, считая его равным мощности критерия. Из чего нетрудно найти вероятность ошибки второго рода β (отвергнуть конкурирующую гипотезу H_1, когда она верна.

3. Выясняют, в какую из выбранных областей попадает значение критерия. Если значение находится в области, в которой правдоподобна гипотеза H_0, то считают, что эксперимент не противоречит гипотезе H_0, т. е. гипотеза H_0 принимается. Этому выводу всегда соответствуют вероятности ошибок α и β. В частности, в случае $\alpha = \beta$ гипотезы H_0 и H_1 признаются одинаково правдоподобными.

В изложенных выше примерах при принятии гипотез мы стремились к тому, чтобы вычисленная вероятность α оказывалась по возможности большей. В таких случаях, при принятии гипотезы H_0, мы добивались меньших значений вероятности β.

Глава 4. Метод статистических испытаний

4.1. Замечания о особенностях применения математического моделирования

Тема главы, по нашему мнению, весьма доступна для чтения, но чрезвычайно сложна для осознания. Причина тому – отсутствие однозначных рекомендаций, следуя которым читатель обязательно достигнет цели – найдет адекватную математическую интерпретацию (модель) исследуемого явления. При решении фактических прикладных задач исследователь, как правило, сталкивается со значительно большим количеством трудностей, чем при решении задач учебных, условия и решения которых можно найти в учебно-методической литературе. Нередко, возможно исходя из искусственных педагогических соображений, попадаются и учебные задачи, решение которых тривиально, но при этом лишено прикладного смысла. Например: «Охотник сидит в засаде и ждет медведя. Медведь может выскочить из-за первого куста с вероятностью 0.1, из-за второго куста – с вероятностью 0.2 и из-за третьего куста – с вероятностью 0.3. В первом случае охотник убивает медведя с вероятностью 0.5, во втором случае – с вероятностью 0.4, в третьем случае – с вероятностью 0.3. Найти вероятность того, что охотник убьет медведя».

Не подвергая сомнению педагогическую ценность задачи, можно гарантировать то, что такой эксперимент, давший значения содержащимся в условии вероятностям, никогда на самом деле не проводился. Многие задачи с военным или экономическим контекстом ничем не содержательнее «задачи про медведя». При выполнении фактических исследований в распоряжении, в лучшем случае, оказывается достаточно большого объема выборка численных значений результатов измерений исследуемой величины. Поясним понятие достаточно большого объема выборки. При выполнении математического моделирования может оказаться, что количество элементов n выборки кажется значительным – многие десятки или сотни.

Однако вычисленное значение точечной оценки дисперсии тоже значительное, что свидетельствует в пользу справедливости подозрения о недостаточности объема выборки для решения отдельных задач математического моделирования. По нашему мнению значительно чаще исследователь оказывается и в более драматичной ситуации – объем выборки чрезвычайно мал – единицы элементов, но заказчик желает выяснить неоправданно большое количество сведений. Ответ на вопрос о том, как именно поступать в таких случаях, содержится в результатах так называемого «разведочного» анализа – молодой математической дисциплины, созданной Дж. Тьюки [68]. Разведочный анализ является своеобразной предпервичной обработкой результатов наблюдений, осуществляемой посредством простейших средств – карандаша, бумаги и калькулятора для научных расчетов. Задача разведочного анализа – представить наблюдения в наглядной форме с помощью схем, таблиц и графиков, облегчающих выявление закономерностей и подбор способов более глубокой статистической обработки. В рамках разведочного анализа решается задача о том, как привести имеющиеся данные в такой вид, в котором они будут легче и эффективнее восприниматься исследователем. В связи с этим, следуя рекомендациям Дж. Тьюки, нужно иметь в виду:

1) все, что упрощает описание, облегчает его восприятие нами;

2) все, что позволяет заглянуть глубже какого-то ранее достигнутого уровня понимания, делает описание более эффективным;

3) содержание графиков, предназначенных для разведки данных, должны вынуждать нас заметить то, что они могли бы нам сказать. Графики, подчеркивающие лишь то, что нам уже известно, не стоят места, которое они занимают. График имеет ценность только тогда, когда его содержание вынуждает нас заметить то, что мы совсем не ожидали увидеть.

Считая очень важным упрощение описания в результате любого разумного изменения подхода, мы тем самым утверждаем свою веру в количественный характер знания. Мы верим, что большинство ключевых вопросов в нашем мире рано или поздно обязательно потребуют ответа на вопрос: «Насколько?», а не только лишь: «В каком направлении?». Вместе с тем, знакомство с методами разведочного анализа у недостаточно подготовленного знанием прикладной математики читателя может вызвать сомнение в целесообразности кропотливого и тщательного ее изучения. Мы предостерегаем читателя от поспешных заключений, т. к. успешное применение методов разведочного анализа неразрывно связано с необходимостью глубокого знания известных статистических способов проверки гипотез и оценивания параметров. Любые результаты, найденные с помощью приемов разведочного анализа, требуют подтверждения применением результатов известных в теории вероятностей и математической статистике теорем [69], [70].

4.2. Идея метода статистических испытаний (метод Монте-Карло)

В задачах, рассмотренных в гл. 1-3, требовалось найти аналитическое решение. Однако в некоторых задачах естествознания найти аналитическое решение затруднительно. В таких случаях используется метод статистических испытаний. Принято считать, что метод найден в 1949 г. Н. Метрополисом и

С. Улэмом и впервые опубликован в работе [79]. Другое распространенное название метода статистических испытаний – метод Монте-Карло. Весьма обширный список работ по методу Монте-Карло и его применению в разных областях естествознания можно найти в книгах [78], [85]. Следуя содержанию монографии [85], будем называть методами Монте-Карло *численные методы решения математических задач при помощи моделирования случайных величин и статистической оценки их характеристик*. При таком определении приходится к методам Монте-Карло причислить и некоторые другие методы, как, например, стохастические приближения или случайный поиск, которые по традиции рассматриваются отдельно. Однако специалисты, занимающиеся этими вопросами, нередко сами называют свои приемы методами Монте-Карло.

В то же время в определении подчеркивается что:

а) речь идет о численных методах (и конкурировать они могут с классическими численными методами, а не с аналитическими методами решения задач);

б) решать методами Монте-Карло можно любые математические задачи (а не только задачи вероятностного происхождения, связанные со случайными величинами).

Изложим идею метода. Вместо поиска аналитического решения либо выполняется большое количество экспериментов, предложенных в задаче, либо испытания заменяют другими, имеющими с исходными одинаковую вероятностную структуру, или рассматриваемые в задаче явления имитируют другими случайными явлениями. После чего найденные по результатам испытаний характеристики принимают в качестве приблизительных решений задачи. Обоснованием правомерности изложенного подхода служит содержание закона больших чисел.

Замечание. ° В некоторых изложениях метод статистических испытаний назван методом статистического моделирования, что, на наш взгляд, не совсем корректно. Известно, что содержание термина «моделирование» регламентировано (см. гл. 1) в рамках понятия «математического моделирования» для описания математических моделей явлений. В методе статистических испытаний, напротив, речь идет о статистической имитации фактических испытаний (опытов), а не о создании статистических моделей опытов. ●

Для иллюстрации идеи метода статистических испытаний рассмотрим типичные примеры.

Пример 4.1. Из опыта работы торгового предприятия известно, что вероятность того, что продавщица в течение времени T не заболеет и выйдет на работу, равна 5/6. Продавщицы заболевают независимо друг от друга. Отдел обслуживается тремя продавщицами. При невыходе на работу хотя бы одной продавщицы отдел признается неработоспособным.

Не используя аналитические методы найти вероятность P того, что отдел за время T окажется неработоспособным.

Решение. ☐ Заменим описанные в задаче испытания другими, имеющими одинаковую с условиями задачи вероятностную структуру. Будем подбрасывать три игральные кости. При этом будем считать, что выпадение на кости одного очка соответствует событию «продавщица заболела». Повторяя подбрасывания костей много раз подряд, нетрудно найти число m – количество случаев, в которых хотя бы на одной из трех костей выпала единица и число n – общее количество выполненных испытаний. Точечной оценкой P^* вероятности P будет $P^* = m/n$. При достаточно большом числе испытаний n оценка P^* станет близка к искомой вероятности P. ■

Пример 4.2. По большому торговому павильону ходит покупатель.

Не применяя аналитических методов найти схему статистических испытаний, имитирующих траекторию перемещения покупателя по павильону.

Решение. ☐ В целях упрощения ограничим возможности движения покупателя четырьмя направлениями: север, юг, запад, восток. Будем считать, что за один этап передвижения до временной остановки покупатель проходит одно и то же расстояние. Направление перемещения будем задавать, используя результат двукратного бросания монеты. Пусть P – событие «Выпала решка», O – «выпал орел». Результаты двукратного бросания монеты будем интерпретировать, как: OO – передвижение покупателя на один этап на север, PP – на юг, OP – на запад, PO – на восток. Каждую точку конца движения в одном направлении – как остановку движения. Точку начала движения выберем произвольно. Выполнив многократные бросания монеты сериями по два бросания, найдем одну из возможных траекторий движения покупателя по торговому павильону. ■

Хрестоматийным примером успешного применения метода статистических испытаний является процесс вычисления значения определенного интеграла, описанный в примере 4.3.

Пример 4.3. Требуется вычислить значение определенного интеграла

$$\int_0^1 f(x)\,dx .$$

Дополнительно известно, что на области значений аргумента (0,1) значение интеграла не превосходит единицы.

Решение. ☐ Очевидно, что значение интеграла равно площади области G, ограниченной кривой $y = f(x)$, осью Ox и прямыми $x = 0$, $x = 1$. Будем многократно бросать на единичный квадрат случайно точку. После этого сосчитаем число m – бросаний, при которых точка попала в область G и n – общее число бросаний. Отношение m/n будет оценкой значения интеграла. С увеличением количества бросаний n точность результата будет расти, т. е. при возрастании n справедливо выражение

$$\int_0^1 f(x)\,dx \approx \frac{m}{n}. \qquad \blacksquare$$

В настоящее время метод статистических испытаний обладает весьма развитым теоретическим аппаратом. Применение результатов теоретического аппарата метода позволяет находить ответы на вопросы о том, как именно целесообразно выбрать случайную величину, наиболее подходящую при реализации статистических испытаний и как найти возможные значения этой величины.

Определение 4.1. ▌*Отыскание возможных значений случайной величины, выбранной для применения метода статистических испытаний, называется разыгрыванием случайной величины.* ▌

Отправным пунктом решения большинства задач нахождения нужных значений случайной величины (разыгрывание случайной величины) в методе статистических испытаний является нахождение нужных значений (разыгрывание) равномерно распределенной случайной величины. Вместе с тем существует немало задач, для успешного решения которых требуемся знание методов разыгрывания дискретной случайной величины (см. примеры 4.1, 4.2).

4.3. Разыгрывание случайных величин

4.3.1. Общие положения

Сформулируем полезное для дальнейшего изложения определение, воспользовавшись содержанием энциклопедии [8].

Определение 4.2. ▌*Случайными числами называются числа* x_n*, последовательность которых обладает теми или иными статистическими закономерностями.* ▌

Различают случайные числа, генерируемые какими-либо стохастическими устройствами и псевдослучайные числа, конструируемые с помощью арифметических алгоритмов. При этом с большим или меньшим основанием принимают, что найденная последовательность чисел обладает комплексом частотных свойств, типичным для последовательности независимых реализаций какой-либо случайной величины X с функцией распределения $F(x)$. В таком случае говорят о независимых случайных или псевдослучайных числах, распределенных по закону $F(x)$. Для проверки качества последовательности случайных или псевдослучайных чисел используют критерии согласия. Например, критерий χ^2-Пирсона.

4.3.2. Разыгрывание равномерно распределенной случайной величины

Для нахождения последовательности численных значений равномерно распределенной случайной величины воспользуемся доступным устройством – монетой. Договоримся, что при выпадении «герба» при бросании монеты будем считать, что некоторая случайная величина X приобретает значение $x = 1$, при выпадении «решки» – приобретает значение $x = 0$. Будем использовать последовательность значений $x_1, x_2, \ldots, x_k, \ldots$ случайной величины X. Через Y обозначим некоторую дробь, заданную выражением

$$Y = x_1 \cdot 2^{-1} + x_2 \cdot 2^{-2} + \ldots + x_k \cdot 2^{-k} + \ldots. \qquad (4.1)$$

Доказано и известно, что численные значения дроби (4.1) при любой последовательности $x_1, x_2, \ldots, x_k, \ldots$, состоящей из нулей и единиц, принадлежат отрезку [0,1]. Доказано, что величина Y непрерывная и, рассмотренная как величина случайная, распределена равномерно на отрезке [0,1] при условии бесконечного количества слагаемых в правой части выражения (4.1) и бесконечного количества раз повторений процесса нахождения значений дроби Y. Изложенное свойство дроби (4.1) широко применяется при решении задачи разыгрывания равномерно распределенной случайной величины.

Нетрудно догадаться, что практическая реализация изложенного метода при бесконечных значениях k невозможна. Однако доказано, что всегда можно найти значение числа k для дроби

$$Y = x_1 \cdot 2^{-1} + x_2 \cdot 2^{-2} + \ldots + x_k \cdot 2^{-k}, \qquad (4.2)$$

при котором распределение величины Y будет приблизительно равномерным. При этом точность приближения регулируется изменением значения числа k. Очевидно, что если все значения x_i, $i = 1,\ldots,k$ равны нулям, то значение выражения (4.2) равно нулю. Нетрудно вычислить, что если все значения x_i, $i = 1,\ldots,k$ равны единицам, то уже при $k = 24$ наибольшее значение выражения (4.2) достигает приблизительно $1 - 10^{-7} \sim 1$. Поэтому, в рамках заданной условиями прикладной задачи точности, в дальнейшем можно не различать случайную величину Y, заданную выражением (4.1) и случайную величину Y, заданную выражением (4.2). Нетрудно догадаться, что при использовании выражения (4.2) для вычисления значений случайной величины, распределенной по закону $R[0,1]$, при конечных значениях числа k найденным значениям равномерно распределенной случайной величины будет фактически соответствовать не функция распределения $F_r[0,1]$, а статистически эквивалентная эмпирическая функция $F_{эr}[0,1]$, связанная с функцией $F_r[0,1]$ через выражение

$$|F_r[0,1] - F_{эr}[0,1]| < 1/2^k.$$

Записанное выражение нетрудно применить в качестве решения задачи оценки точности псевдослучайных чисел. Например, при $k = 24$ правая часть выражения приблизительно равна $6 \cdot 10^{-8}$, что обеспечивает достаточную точность при решении большинства прикладных задач.

Воспользуемся изложенным методом разыгрывания равномерно распределенной случайной величины для решения задачи из примера 4.1. В задаче любая из трех обслуживающих отдел продавщиц за время Т могла заболеть с вероятностью 1/6 и не заболеть с вероятностью 5/6. Для решения задачи в примере 4.1 использовалась игральная кость. Теперь используем для решения задачи значения равномерно распределенной случайной величины, найденные применением изложенного выше метода. В целях краткости обозначим случайную величину, распределенную равномерно на отрезке $[0,1]$ через R. Очевидно, что для величины R справедливы равенства

$$P(0 \le R \le 1/6) = 1/6,\ P(1/6 \le R \le 1) = 5/6.$$

Содержание равенств дает нам право события «продавщица заболела» и «продавщица не заболела» определять так:

1) находят значение r_i случайной величины R;
2) если $r_i \le 1/6$, то фиксируют правомерность события «продавщица заболела», в противном случае – «продавщица не заболела».

Чтобы принять решение о работоспособности отдела, нужно получить три значения величины R, и если хотя бы одно из них меньше 1/6, то принимается решение о неработоспособности отдела, в противном случае – о работоспособности.

Продолжим изложение сведений по разыгрыванию равномерно распределенной случайной величины. Нетрудно догадаться, что даже при относительно небольших значениях числа k в правой части формулы (4.2) нахождение значений величины Y из левой части формулы с помощью бросания монеты является весьма трудоемким процессом. Очевидно, что вместо бросания монеты возможно использование процесса получения последовательностей из нужного количества нулей и единиц с помощью компьютера. Также очевидно, что последующие арифметические вычисления с использованием результатов полученных последовательностей можно выполнять с помощью компьютера. В настоящее время большинство доступных систем символьной математики, например, система «*Derive*», снабжены программными модулями, способными выполнять процесс разыгрывания случайных величин с разными законами распределения. В частности, с равномерным законом. Это позволяет выполнять успешные имитации случайных испытаний с помощью компьютеров. Дадим полезное определение.

Определение 4.3. ▌ *Программные модули, обеспечивающие генерацию последовательностей псевдослучайных чисел с заданным законом распределения, называются* RND-*датчиками или датчиками псевдослучайных чисел.* ▌

Как правило, в системах символьной математики название RND сохраняется. В системе «*Derive*» содержится функция генерации псевдослучайных чисел *RANDOM(n)*, которая в зависимости от значения аргумента n выполняет разные действия по генерации значений равномерно распределенной случайной величины. Весьма удобным в применении является образование системой «*Derive*» [16, с.75] вектора случайных чисел по программе

VECTOR(RANDOM(n),k,1,m),

в которой при $n = 1$ возвращается случайное число из интервала $R[0,1]$, k – размерность вектора. Например, выполнение программы *VECTOR(RANDOM(1),k,1,10)* создает вектор из 10 псевдослучайных чисел закона распределения $R[0,1]$.

Кроме изложенного метода разыгрывания равномерно распределенной случайной величины известно большое количество других методов. Наиболее часто употребляемые из них заинтересованный читатель может найти в книге [78, с.225].

Известно, что, воспользовавшись каким-либо из методов разыгрывания равномерно распределенной случайной величины, нетрудно перейти к освоению методов разыгрывания случайных величин как непрерывных, так и дискретных с практически любыми законами распределения.

В целях практического ознакомления с возможностями RND-датчиков, рекомендуем читателю самостоятельно найти последовательность объемом 100-200 шт. псевдослучайных чисел, являющихся реализациями равномерно распределенной на отрезке [0,1] случайной величины. После этого, считая найденную последовательность выборкой, применив критерий χ^2-Пирсона или K-S-критерий, проверить гипотезу о принадлежности выборки к генеральной совокупности численных значений равномерно распределенной случайной величины. В помощь читателю в качестве образца решения дан пример 4.4.

Пример 4.4. Использованием RND-датчика с помощью системы символьной математики «*Derive*» выполнена генерация 52 псевдослучайных чисел, которые должны являться 52 значениями непрерывной случайной величины, распределенной равномерно на отрезке *[0.1]*. Значения псевдослучайных чисел помещены в таблицу 4.1. Инструкцию по применению RND-датчика можно найти в книге [41, с.88].

ТАБЛИЦА 4.1

ЗНАЧЕНИЯ ПСЕВДОСЛУЧАЙНЫХ ЧИСЕЛ ИЗ R(0,1)

0.35 0.13 0.77 0.32 0.58	0.13 0.27 0.56 0.73 0.54	0.77 0.07 0.30 0.66 0.09	0.07 0.80 0.20 0.55 0.84	0.77 0.20 0.96 0.42 0.29	0.85 0.43 0.19 0.87 0.74	0.51 0.96 0.83 0.64 0.15	0.74 0.01 0.85 0.11 0.45	0.33 0.82 0.65 0.93	0.83 0.34 0.73 0.94	0.37 0.48 0.53 0.86

Воспользовавшись критерием χ^2-Пирсона проверить качество генерации RND-датчиком псевдослучайных чисел с помощью системы *«Derive»*.

Решение. □ Разобьем отрезок *[0,1]* на четыре интервала: *[0, 0.25], [0.25, 0.5], [0.5, 0.75], [0.75, 1]*. В случае попадания числа в границу интервала, будем относить его к интервалу, стоящему слева.

В интервал *[0, 0.25]* вошли 11 чисел из таблицы 4.1:

0.13, 0.13, 0.07, 0.09, 0.07, 0.20, 0.20, 0.19, 0.01, 0.11, 0.15.

В интервал *[0.25, 0.5]* вошли 12 чисел:

0.35, 0.27, 0.30, 0.29, 0.42, 0.43, 0.33, 0.34, 0.37, 0.48, 0.32, 0.45.

В интервал *[0.5, 0.75]* – 13 чисел:

0.58, 0.56, 0.73, 0754, 0.66, 0.55, 0.74, 0.51, 0.64, 0.74, 0.65, 0.73, 0.53.

В интервал *[0.75, 1]* – 16 чисел: *0.77, 0.77, 0.80, 0.84, 0.77, 0.96. 0.85, 0.87, 0.96, 0.83, 0.85, 0.82, 0.93, 0.83, 0.94, 0.86.*

Выполним проверку гипотезы H_0: $F(x) = R[0,1]$ при альтернативной гипотезе H_1: $F(x) \neq R[0,1]$. Для этого вычислим значение величины χ^2.

$$\chi^2 = \frac{(11-52\cdot 0.25)^2}{52\cdot 0.25} + \frac{(12-52\cdot 0.25)^2}{52\cdot 0.25} + \frac{(13-52\cdot 0.25)^2}{52\cdot 0.25} +$$
$$+\frac{(10-100\cdot 0.081)^2}{52\cdot 0.25} = 1.08.$$

Воспользовавшись известной формулой

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-x/2}, \quad x > 0,$$

выполнив интегрирование при числе степеней свободы $n = 1$ и $\chi^2 = 1.08$ находим, что $\alpha = 0.3$. Сформулируем результат проверки. Отвергая гипотезу H_0: $F(x) = R[0,1]$ мы рискуем совершить ошибку отвержения правильной гипотезы с вероятностью $\alpha = 0.3$. Считая вероятность 0.3 значительной, гипотезу не отвергаем.

Сформулируем окончательный результат. В рамках выполненной проверки доказано, что RND-датчик системы *«Derive»* является вполне приемлемым устройством для решения прикладных задач методом статистических испытаний. ■

Замечание. ° Вычислив вероятность ошибки второго рода β – принятия неправильной гипотезы нетрудно убедиться в том, что заявление о непременной принадлежности выборки из таблицы 4.1 к генеральной совокупности численных значений случайной величины, распределенной по закону R[0,1], на самом деле является весьма торжественным. Это свидетельствует в пользу того, что задача повышения точности RND-датчиков продолжает оставаться актуальной. По нашему мнению при решении экономических задач высокой ответственности следует принимать во внимание не только значение вероятности ошибки первого рода α, но и значение вероятности ошибки второго рода β. ●

Будем считать, что задача разыгрывания равномерно распределенной случайной величины решена с требуемой для приложений точностью. В таком случае наличие решения позволяет в свою очередь успешно решать задачи имитационного моделирования событий.

4.4. Статистическое моделирование последовательностей случайных испытаний

4.4.1. Статистическое моделирование последовательности независимых испытаний

Будем проводить k независимых испытаний таких, что в результате каждого испытания может произойти только одно из противоположных событий A и B. Пусть известна вероятность $P(A) = p$. В таком случае, т. к. события A и B противоположные, вероятность события B тоже известна: $P(B) = 1 - p$. Опишем процесс статистического моделирования.

Применив метод разыгрывания равномерно распределенной случайной величины, изложенный в п.4.2.2, найдем последовательность значений $r_1, r_2, \ldots, r_k$ случайной величины, распределенной по закону $R[0,1]$. После этого будем считать, что если значение $r_i < p$, $i = 1,2,\ldots,k$, то наступило событие A;

если значение $r_i > p$, $i = 1,2,...,k$, то наступило событие *B*. В теоретических основах метода Монте-Карло доказано, что такое допущение правомерно.

Теперь предположим, что результатом каждого из *k* независимых испытаний может быть появление одного из *n* несовместных событий $A_1,A_2,...,A_n$, образующих полную группу. Предположим также, что вероятность появления $P(A_i) = p_i$, $i = 1,2,...,n$ каждого из событий, образующих полную группу, известна и не изменяется при переходе от одного испытания к другому. Очевидно, что в описываемом случай справедливо равенство $p_1 + p_2 +... + p_m = 1$. Опишем процесс моделирования такой последовательности независимых испытаний. Разделим отрезок $[0,1]$ на *n* участков $\Delta_1,\Delta_2,...,\Delta_n$, длины которых соответственно равны $p_1, p_2,..., p_n$. Таким образом, если $r_i \in \Delta_m$, то считаем, что в *i*-м испытании наступило событие A_m. Правомерность такого допущения следует из того, что

$$P(R \in \Delta_m) = \text{длине отрезка } \Delta_m = p_m = P(A_m).$$

Поясним изложенное численным примером 4.5.

Пример 4.5. (Заимствован из учебника [13, с.318]).

Пусть производится последовательность независимых испытаний. В результате каждого испытания может произойти одно из несовместных событий A_1, A_2, A_3, образующих полную группу. Известно, что вероятности $P(A_1) = 0.35$, $P(A_2) = 0.25$, $P(A_3) = 0.4$. Выполнить статистическое моделирование последовательности испытаний.

Решение. □ Разделим отрезок $[0,1]$ на три отрезка. Первый из них есть интервал $[0, 0.35]$, второй – интервал $[0.35, 0.6]$, третий – интервал $[0.6, 1]$. Предположим, что в результате первого срабатывания RND-датчика с законом распределения $R[0,1]$ выпало псевдослучайное число $r_1 = 0.15$. Это означает, что число принадлежит первому интервалу, т. е. в первом испытании произойдет событие A_1. Пусть второе число, найденное в результате срабатывания RND-датчика оказалось $r_2 = 0.34$. Число 0.34 принадлежит первому отрезку $[0, 0.35]$. Из чего следует, что в результате второго испытания вновь произошло событие A_1. Предположим, что в результате третьего срабатывания RND-датчика появилось число $r_3 = 0.71$. Это число принадлежит третьему интервалу. Что означает, что в третьем испытании произошло событие A_3. Испытания можно продолжить, тем самым найдя одну из возможных последовательностей из событий A_1, A_2, A_3. ■

Наряду с возможностью использования метода статистических испытаний для нахождения последовательностей независимых событий, представляет интерес использование метода статистических испытаний для нахождения последовательностей событий зависимых.

4.4.2. Статистическое моделирование последовательности зависимых испытаний

Пусть в результате испытания может произойти одно из противоположных событий *A* и *B*. Выполним моделирование возможной последовательности этих событий. Воспользуемся RND-датчиком с законом распределения $R[0,1]$.

Пусть в результате первого срабатывания датчика получено число r_1. Будем рассуждать так: если оказалось, что $r_1 < P_1(A)$, где $P_1(A)$ – вероятность наступления события *A* в первом испытании, то будем считать, что в первом испытании произошло событие *A*. Если $r_1 \geq P_1(A)$, то будем считать, что произошло событие *B*. Предположим, что в результате первого испытания появилось событие *A*. После этого с помощью RND-датчика получаем следующее число r_2. Теперь, если $r_2 < P_2(A|A)$, где $P_2(A|A)$ – условная вероятность появления события *A* во втором испытании при условии, что в первом испытании произошло событие *A*, то считаем, что во втором испытании произошло событие *A*; если $r_2 \geq P_2(A|A)$, считаем, что произошло событие *B*. Предположим, что во втором испытании произошло событие *B*. Снова с помощью RND-датчика получим псевдослучайное число. Обозначим его через r_3. Далее, если $r_3 < P_3(A|AB)$, где $P_3(A|AB)$ – вероятность появления события A в третьем испытании при условии наступления в 1-м и 2-м испытаниях событий *A* и *B*, то считаем, что в 3-м испытании произошло событие *A*; в противном случае – *B* и т.д.

Замечание. ° В качестве прототипа изложения использовано содержание книги [13]. ●

4.5. Нахождение последовательностей псевдослучайных чисел с заданным законом распределения

4.5.1. Использование статистической функции Кетле

Пусть требуется найти последовательность псевдослучайных чисел, являющихся значениями непрерывной случайной величины *X* с известным аналитическим выражением плотности $f(x)$. При этом будем считать, что в нашем распоряжении имеется последовательность значений случайной величины, распределенной по закону $R[0,1]$. В нашем случае эта последовательность будет состоять из псевдослучайных чисел. Математически это означает, что с помощью некоторой функции нужно выполнить преобразование $X = \varphi(R)$, связывающее значения равномерно распределенной случайной величины со значениями случайной величины, распределенной по закону $f(x)$. Известно (например, см. [80], с.63), что изложенной решение задачи нетрудно найти, воспользовавшись результатами теоремы 4.1.

Теорема 4.1. ▎ *Если случайная величина X распределена по закону f(x), f(x) – плотность вероятности, то случайная величина* $R=\int_{-\infty}^{x} f(u)du$ *распределена равномерно на* [0,1]. ▎

Доказательство теоремы в доступном изложении нетрудно найти в книге [80], пособии [12] и др.

Теорема 4.1 может быть сформулирована, как теорема 4.2.

Теорема 4.2. ▎*Если F(x) – функция распределения непрерывной случайной величины, а R – равномерно распределенная на* [0,1] *случайная величина, то случайная величина*

$$X = F^{-1}(R)$$

распределена по закону F(x), где F^{-1} *– функция, обратная к функции* F. ▎

Весьма часто метод нахождения численных значений случайной величины применением результатов теорем 4.1 и 4.2 называют методом обратных функций или стандартным методом моделирования. Применив результаты теорем 4.1 и 4.2, в некоторых случаях возможно найти аналитическое выражение, использование которого позволяет решить задачу нахождения последовательности псевдослучайных чисел. Изложим пример такого случая.

Пример 4.6. Пусть случайная величина *X* распределена по экспоненциальному закону

$$F(x)=1-e^{-\lambda x}.$$

Найти аналитическое выражение, применение которого позволяет находить последовательности псевдослучайных чисел, являющихся значениями распределенной по экспоненциальному закону случайной величины.

Решение. □ Использовав результаты теорем 4.1, 4.2, запишем $R = F(x) = 1 - e^{-\lambda x}$. Найдем функцию F^{-}, выразив *X*.

$$X=-\frac{1}{\lambda}\ln(1-R).$$

Следовательно, аналитическое выражение для нахождения искомой последовательности псевдослучайных чисел можно записать, как

$$x_i=-\frac{1}{\lambda}\ln(1-r_i) \quad ,$$

где r_i – числа из последовательности псевдослучайных чисел, являющихся значениями величины *R[0,1]*. ■

4.6. Статистическая имитация процессов измерений. Метод элиминации

Известно, что в процессе обработки результатов измерений естествоиспытатель нередко оказывается в двусмысленной ситуации: с одной стороны он должен выполнить требование к точности решения, с другой – для достижения требуемой точности он не располагает достаточным количеством сведений. Как правило, для таких случаев в рамках содержания многочисленных учебников, пособий и руководств дается рекомендация о необходимости увеличения объема выборки. По нашему мнению во многих случаях содержание такой рекомендации не является конструктивным, т. к. фактически естествоиспытатель не располагает выборкой необходимого объема по причине практической невозможности получения необходимых дополнительных сведений. Например, весьма затруднительно найти сведения о последствиях падения на реактор атомной электростанции самолета – такие случаи не зарегистрированы.

Затруднительно найти сведения, применение которых позволяет выполнить оценку вероятности распада государств, несмотря на то, что такие случаи известны. Затруднительно оценить вероятность повышения цены колбасы, т. к. для получения дополнительных сведений об изменении цен требуется время, а результат оценки, пусть и приблизительный, требуется сейчас. Одним из методов разрешения изложенной противоречивой ситуации является метод статистической имитации (Монте-Карло). Известно, что первоначально идея метода имитации принадлежит Дж. фон Нейману (1951). Модификации метода найдены Йонком (1964, Австрия), Д.И. Голенко (1965, СССР), В.И. Романовским (1966, СССР). В чрезвычайно доступном изложении ознакомиться с методом можно, воспользовавшись книгой [84]. Развитие идеи можно найти в работах [81], [82] и др. В некоторых изложениях метод назван «методом отказов», «методом элиминации», «методом jackknife (складного ножа)» и др. В последнем случае (jackknife) метод развит Дж. Тьюки.

Изложим кратко вычислительную часть метода. Условием применимости метода является требование к свойствам случайной величины состоящее в том, что закон ее распределения может быть задан с помощью функции плотности. В целях простоты объяснения будем считать, что закон *f(x)* известен и что значения случайной величины *X* ограничены интервалом (*a*,*b*). Вне интервала *f(x) = 0*. Кроме того, значения *f(x)* ограничены сверху, т. е. справедливо неравенство $f(x) \leq c = const$. Опишем алгоритм нахождения вероятных значений величины X в виде последовательности псевдослучайных чисел.

Алгоритм 4.1. □ 1. Находим два независимых псевдослучайных значения r_1 и r_2 величины $R[0,1]$.

2. Находим некоторую точку с координатами (z_1, z_2), где

$$z_1 = a + r_1(b - a),$$

$$z_2 = r_2 c.$$

3. Если $z_2 < f(z_1)$ полагаем, что величина X приняла значение z_1 и значение z_1 помещается в последовательность найденных псевдослучайных значений величины X. Если $z_2 \geq f(z_1)$, то точка $z = (z_1, z_2)$ не рассматривается, т. е. ее принято называть отброшенной или исключенной. В этом случае в последовательность псевдослучайных значений не записывается никакого числа.

4. Процесс нахождения значений псевдослучайных чисел повторяется, начиная с п.1.

Найденные выполнением п.1-4 значения являются псевдослучайными значениями случайной величины X. В качестве функции плотности $f(x)$ можно использовать ее статистические аналоги. Например, полигон распределения, что позволяет применением метода статистических испытаний увеличить объем выборки за счет найденных псевдослучайных значений.

Выполнив очевидные незначительные изменения алгоритм 4.1 нетрудно использовать для нахождения последовательности псевдослучайных чисел с использованием эмпирической функции распределения $F_э(x)$. При этом вероятности попадания псевдослучайных значений случайной величины в интервалы, в которых имеют место равенства $F_э(x) = 0$ и $F_э(x) = 1$ можно оценить, воспользовавшись распределением Гумбеля 1 и 2. ■

Многие эмпирические приемы, используемые при решении прикладных задач методом статистических испытаний, нуждаются в более строгих теоретических обоснованиях. Вместе с тем их применение обусловлено потребностью нахождения хотя бы весьма приблизительных решений отдельных задач. Изложим пример приблизительного решения типичной задачи.

Пример 4.7. В таблице 4.2 расположены данные о количествах читателей в разные календарные годы библиотеки им. Н.В. Гоголя в СПб.

ТАБЛИЦА 4.2

КОЛИЧЕСТВО ЧИТАТЕЛЕЙ В ЦБС им. ГОГОЛЯ

Год	1998	1999	2000	2001	2002	2003	2004
Количество читателей.	5500	5847	5336	6674	6851	6228	5914

Выполнить оценку закона распределения количества читателей.

Решение. □ Нетрудно заявить о том, что объем выборки количеств читателей недостаточен и что требуется существенное увеличение объема выборки. Вместе с тем довольно часто исследователь сталкивается с ситуацией, в которой получение дополнительных сведений невозможно, а заказчик непременно желает получить хотя бы приблизительные результаты.

Нетрудно вычислить, что среднее арифметическое значение количества читателей $x_{ср} = 6050$, несмещенная точечная оценка дисперсии $S^2 = 3.22 \cdot 10^5$, среднеквадратического отклонения $S = 568$. Приблизительную оценку вероятности значений количества читателей, как случайной величины, можно найти воспользовавшись неравенством Чебышева, в котором в качестве неизвестных значений математического ожидания и дисперсии использованы значения их точечных оценок.

$$P\left(\left|X - 6050\right| < k\varepsilon\right) \geq 1 - \frac{S^2}{(k\varepsilon)^2}.$$

Нетрудно вычислить, что при $k = 1.4$ имеет место неравенство $P(|X\ 6050\ | < 1.4 \cdot 568) \geq 0.49$. Из чего следует, что около половины возможных значений случайной величины удовлетворяют неравенствам $X < 5255$, $X > 6845$. Для уточнения возможных фактических вероятностных границ значений количества читателей воспользуемся свойствами распределений Гумбеля 1 и 2. Запишем их аналитические выражения.

$$f(x) = \frac{1}{\lambda}\exp\left(\frac{x-\mu}{\lambda} - e^{(x-\mu)/\lambda}\right), \tag{4.3}$$

где $\mu^* = x^* + 0.4501S$, $\lambda^* = 0.7797S$ – выборочные оценки параметров, найденные методом моментов.

$$f(x) = \frac{1}{\lambda}\exp\left(-\frac{x-\mu}{\lambda} - e^{(x-\mu)/\lambda}\right), \tag{4.4}$$

где $\mu^* = x^* - 0.4501S$, $\lambda^* = 0.7797S$ – выборочные оценки параметров, найденные методом моментов.

Известно, что выражение (4.3) используется для нахождения вероятностной оценки минимальных элементов, (4.4) – максимальных элементов выборки.

Найдем весьма вероятное наименьшее возможное значение величины X. Воспользовавшись данными из таблицы 1, вычислив значения μ^* и λ^* для выражения (4.3) находим, что $\mu^* = 6306$, $\lambda^* = 443$. Применением выражения (4.3) как распределением критерия, выполнив проверку гипотезы H_0: $X_{min} < 5336$ находим, что $\alpha = 0.1$. Гипотезу H_0: $X_{min} < 5336$ принимаем. Задав наименьшее значение, например, как *4000*, выполнив проверку гипотезы H_0: $X_{min} < 4000$ находим, что $\alpha = 0.005$. Гипотезу H_0: $X_{min} < 4000$ отклоняем. Сформулируем результат: доказано, что за наименьшее значение величины X можно принять $x = 4000$.

Найдем весьма вероятное наибольшее возможное значение величины X. Воспользовавшись данными из таблицы 4.1, вычислив значения μ^* и λ^* для выражения (4.4) находим, что $\mu^* = 5794$, $\lambda^* = 443$. Воспользовавшись законом распределения (4.4), выполнив проверку гипотезы H_0: $X_{max} > 6851$ находим, что $\alpha = 0.1$. Гипотезу H_0: $X_{max} > 6851$ принимаем. Задав в качестве наибольшего значения $x = 8000$, выполнив проверку гипотезы H_0: $X_{max} > 8000$ находим, что $\alpha = 0.007$. Гипотезу H_0: $X_{max} > 8000$ отклоняем.

Сформулируем окончательный результат. В рамках задачи доказано, что вероятное наименьшее и наибольшее значение случайной величины – возможного количества читателей находятся в границах $x_{min} = 4000$, $x_{max} = 8000$. Найденный результат можно использовать при решении задачи статистической имитации.

Приступим к решению задачи статистической имитации. Напомним, с какой целью мы будем выполнять статистическую имитацию. В имеющейся в нашем распоряжении выборке содержится всего семь элементов. Нас интересуют вероятностные свойства случайной величины, весьма ограниченным объемом выборочных значений которой мы располагаем. Но выборки значений объемом семь элементов недостаточно для выполнения заключений. Поэтому мы вынуждены прибегнуть к методу, применение которого позволяет использованием весьма ограниченного количества данных найти последовательность псевдослучайных чисел, обладающих вероятностными свойствами случайной величины, значения которой находятся в выборке. Для нахождения последовательности псевдослучайных чисел воспользуемся методами Монте-Карло, одним из которых является метод элиминации. С этой целью найдем эмпирическую функцию распределения $F_э(x)$ случайной величины X. В таблице 4.3 расположены значения величины X и соответствующие им значения функции $F_э(x)$.

ТАБЛИЦА 4.3

ЗНАЧЕНИЯ ФУНКЦИИ Fэ(x)

X<x	x<5336	x<5500	x<5847	x<5914	x<6228	x<6674	x<6851
$F_э(x)$	0	1/7	2/7	3/7	4/7	5/7	6/7
X<x	$x < 6851 + a,\ a > 0$						
$F_э(x)$	1						

Нетрудно увидеть, что в таблице 4.3 не отражены вероятностные свойства величины X состоящие в том, что минимальное значение величины $x_{min} = 4000$, максимальное $x_{max} = 8000$. Статистическая имитация, выполненная без учета этих свойств, не будет в возможно полной мере отражать свойства величины X. Для того, чтобы результаты статистической имитации содержали указанное вероятностное свойство величины X, А.И. Ивановым предложено в качестве прообраза процедуры элиминации использовать не эмпирическую функцию $F_э(x)$, а известный статистический аналог функции распределения кумулянту, которую обозначим через $F_к(x)$.

В отличии от функции $F_э(x)$ функция $F_к(x)$ не содержит разрывов первого рода, т. к. является не ступенчатой функцией, как функция $F_э(x)$, а состоящей из отрезков прямых ломаной, что при решении отдельных задач позволяет применением функции $F_к(x)$ более адекватно описывать вероятностные свойства выборки, чем функцией $F_э(x)$. В некоторых учебниках, например, в учебнике [118, с.131], функцию $F_к(x)$ называют статистической функцией распределения. Записав использованием выборочных данных таблицу значений функции $F_э(x)$, нетрудно воспользовавшись известным выражением для описания прямой

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1}$$

в котором (x_1, y_1), (x_2, y_2) – координаты двух разных лежащих на прямой точек, найти аналитические выражения описания отрезков прямых, составляющих функцию $F_к(x)$. Например, с учетом того, что найденное нами наименьшее возможное значение случайной величины X – количества читателей в библиотеке им. Н.В. Гоголя равно *4000*, а наибольшее равно *8000*, воспользовавшись значениями функции $F_э(x)$ из таблицы 4.3, нетрудно найти соответствующую функции $F_э(x)$ функцию $F_к(x)$. Воспользовавшись выражением (4.5), выполнив очевидные вычисления, запишем в таблице (4.4) значения функции $F_к(x)$, являющейся статистическим аналогом функции $F_э(x)$.

ТАБЛИЦА 4.4
ЗНАЧЕНИЯ ФУНКЦИИ $F_к(x)$

Интервал значений x	Значение $F_к(x)$	Интервал знач. $F_к(x)$
x < 4000	0	0
4000 < x < 5336	$1.07\cdot10^{-4}(x-4000)$	0, 1/7
5336 < x < 5500	$8.7\cdot10^{-4}(x-5172)$	1/7, 2/7
5500 < x < 5847	$4.1\cdot10^{-4}(x-4806)$	2/7, 3/7
5847 < x < 5914	$0.0021(x-5646)$	3/7, 4/7
5914 < x < 6228	$4.55\cdot10^{-4}(x-4658)$	4/7, 5/7
6228 < x < 6674	$3.2\cdot10^{-4}(x-3998)$	5/7, 6/7
6674 < x < 8000	$1.08\cdot10^{-4}(x+1282)$	6/7, 7/7
x > 8000 + c, c > 0	1	1

Обращаем внимание читателя на специфическое отличие функции $F_к(x)$ от функции $F_э(x)$. Функция $F_э(x)$ в точке $x = 6851$ претерпевает разрыв первого рода: $F_э(6851) = 6/7$ и $F_э(6851) = 1$. Функция $F_к(x)$ в точке $x = 6851$ не терпит разрыва, а достигает значения $F_к(x) = 1$ только в точке $x = 8000$. В интервале значений $x \in (6674, 8000)$ значения функции $F_к(x)$ возрастают линейно: $F_к(x) = 1.08\cdot10^{-4}(x + 1282)$.

В целях иллюстрации на рис.5 помешена графическая интерпретация функции $F_к(x)$ из таблицы 4.4.

На рис.5: по оси абсцисс — значения случайной величины X — количества читателей библиотеки им. Н.В. Гоголя. По оси ординат — значения статистической функции (кумулянты) $F_к(x)$. На рис.5 с целью демонстрации изменения значений функции $F_к(x)$ разными отрезками ломаной изображены разрывы. На самом деле функция $F_к(x)$ непрерывна. В интересах читателя в приложении 1 помещена программа нахождения функции $F_к(x)$ рассматриваемого примера, написанная на языке системы "Derive".

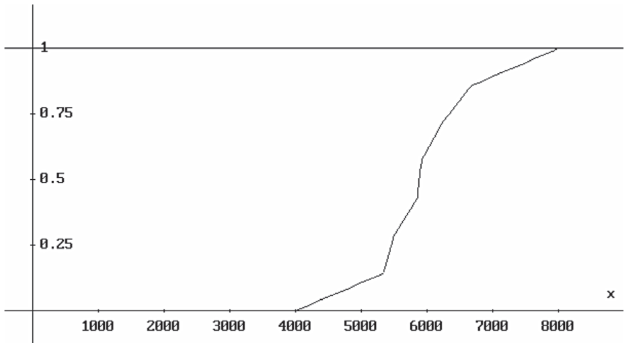

Рис.5.

Теперь, воспользовавшись содержанием таблицы 4.4, применив результаты теорем 4.1, 4.2, приступим к процессу статистической имитации. В интересах читателей опишем нахождение последовательности, состоящей из одного псевдослучайного числа с помощью системы *«Derive»*, инструкции по применению которой содержатся в работе [16, 41, 42, 43]. Опишем процесс нахождения одного псевдослучайного значения количества читателей.

1. Введем в *«Derive»* строку

#1: *RANDOM(1)*

Предположим, что в результате выполнения команды approx системой «Derive» сгенерировано псевдослучайное число $r_1 = 0.349687$.

#2: *0.349687*

Число $r_1=0.349687$ есть значение функции $F_к(x)$ из таблицы 4.4. Это значение соответствует интервалу значений функции $(2/7, 3/7) = (0.2857, 4286)$, на котором справедливо выражение

$F_к(x) = 4.1\ 710^{-4}(x - 4806),\ 5500 < x < 5847,$

т. е. $0.349687 = 4.1\cdot 10^{-4}(x - 4806)$. Из чего нетрудно вычислить, что $x = 5659$. Значение $x = 5659$ — новое значение количества читателей в библиотеке им. Гоголя. Процесс можно продолжить вплоть до нахождения последовательности желаемой длины.

Изложим еще один предложенный нами прием нахождения последовательности псевдослучайных значений случайной величины использованием ограниченного объема выборки и функции $F_к(x)$. Будем считать, что использованием данных выборки статистическая функция $F_к(x)$ найдена. Например, в таблице 4.4 помещены значения статистической функции для количеств читателей библиотеки. Известно, что плотность вероятности $f(x)$ случайной величины X может быть найдена использованием выражения

$$f(x) = \frac{dF(x)}{dx}.$$

Это позволяет, воспользовавшись функцией $F_к(x)$, найти статистический аналог плотности $f(x)$. Например, воспользовавшись данными из таблицы 4.4, выполнив дифференцирование на каждом из участков значений аргумента, нетрудно вычислить необходимые численные значения для построения гистограммы.

Результаты вычислений для построения гистограммы с помощью таблицы 4.4 помещены в таблицу 4.5.

ТАБЛИЦА 4.5

Границы X	$F_к(x)$	$f(x)=\frac{dF(x)}{dx}$
x < 4000	0	0
4000 < x < 5336	$1.07{\cdot}10^{-4}(x-4000)$	$1.07{\cdot}10^{-4}$
5336 < x < 5500	$8.7{\cdot}10^{-4}(x-5172)$	8.7710^{-4}
5500 < x < 5847	$4.1{\cdot}10^{-4}(x-4806)$	$4.1{\cdot}10^{-4}$
5847 < x < 5914	$0.0021(x-5646)$	0.0021
5914 < x < 6228	$4.55{\cdot}10^{-4}(x-4658)$	$4.55{\cdot}10^{-4}$
6228 < x < 6674	$3.2{\cdot}10^{-4}(x-3998)$	$3.2\ {\cdot}10^{-4}$
6674 < x < 8000	$1.08{\cdot}10^{-4}(x+1282)$	$1.08{\cdot}10^{-4}$
x > 8000 + c, c > 0	1	0

В свою очередь, воспользовавшись найденной гистограммой и применив алгоритм 4.1, нетрудно найти последовательность псевдослучайных значений исследуемой случайной величины X, выборочными значениями которой мы располагаем. Применив изложенные методы нахождения последовательности псевдослучайных значений величины X, возможно найти последовательность любой требующейся длины. В связи с этим возникает задача о том, какую длину найденной последовательности псевдослучайных значений величины X считать достаточной.

Использование изложенного способа нахождения последовательности псевдослучайных значений величины X позволяет, минуя вычислительные трудности, решить задачу нахождения необходимой длины последовательности псевдослучайных значений. С этой целью обозначим через S_G – площадь под гистограммой. Известно, что она равна единице. Обозначим через S – площадь содержащего гистограмму прямоугольника с высотой, равной высоте наибольшего из столбцов гистограммы и длиной, равной разности между наибольшим и наименьшим значениями величины X. Отношение S_G/S нетрудно вычислить. Положим $S_G/S = p$. Обозначим через m – количество точек, найденных применением алгоритма 4.1, попавших на площадь гистограммы, через n – общее количество точек, попавших на площадь S. Обозначим через p^* отношение $m/n = p^*$. Нетрудно догадаться, что число p является значением геометрической вероятности, а число p^* – точечной оценкой известного значения геометрической вероятности p. При известном значении вероятности p нетрудно оценить точность значения оценки p^*, что позволяет обосновать решение о прекращении процесса статистической имитации. При решении отдельных задач оценка p^* может быть интерпретирована, как эффективность метода.

Определение 4.4. ▌*Эффективностью метода элиминации (отбора) называют вероятность того, что точка $z=(z_1, z_2)$ будет использована для нахождения псевдослучайного значения случайной величины X.* ▌

Очевидно, что при большом числе отброшенных точек метод элиминации малоэффективен. В научной литературе можно найти разнообразные обобщения метода элиминации [78, 85], позволяющие конструировать алгоритмы моделирования случайных величин с разными законами распределения.

Изложим отдельные сведения о другом способе оценки достаточности длины найденной последовательности псевдослучайных значений случайной величины. Одним из наиболее доступных критериев достаточности длины найденной последовательности псевдослучайных значений является критерий точности среднего арифметического. Изложим кратко способ применения критерия.

Обозначим через $X=(x_1, x_2,\ldots,x_n)$ – найденную последовательность длиной n псевдослучайных значений случайной величины X, через $x_{ср}$ – точечную оценку среднего арифметического. Известно, что

$$x_{cp}=\sum_{i=1}^{n}x_i\ .$$

Вычислим точечную оценку S_x^2 значения дисперсии величины X

$$S_x^2=\sum_{i=1}^{n}\left(x_i-x_{cp}\right)^2\ .$$

Из содержания центральной предельной теоремы А.М. Ляпунова известно, что при достаточно больших значениях n оценку дисперсии $S^2_{x.ср}$ среднего арифметического можно найти, воспользовавшись выражением

$$S^2_{x.cp}=\frac{S_x^2}{n}\ .$$

Известно, что при решении прикладных задач достаточно большими можно считать значения $n \geq 30$. Правомерность последнего утверждения следует из свойств случайной величины, распределенной по закону Стьюдента. Из результатов центральной предельной теоремы А.М. Ляпунова известно также, что величина $x_{ср}$, рассматриваемая как случайная, при весьма широких допущениях может считаться величиной, распределенной по закону $f_N(x, x_{ср}, S_{x.ср})$. В свою очередь применение знания закона $f_N(x, x_{ср}, S_{x.ср})$ позволяет найти значение границ доверительного интервала, с надежностью γ накрывающего значение $x_{ср}$. Допустимые границы доверительного интервала можно назначить, исходя из точности, предъявляемой к результату решения конкретной прикладной задачи.

Литература

1. Математический энциклопедический словарь./ Гл. ред. Ю.В. Прохоров; Ред. кол.: С.И. Адян, Н.С. Бахвалов, В.И. Битюцков, А.П. Ершов, Л.Д. Кудрявцев, А.Л. Онищик, А.П. Юшкевич. – М.: Сов. энциклопедия, 1988.- 847 с., ил.
2. Чернецкий В.И. Математическое моделирование динамических систем. Петрозаводск. Петрозаводское книжное издательство. 1996., 429 ст.
3. Винер Н. Кибернетика. М.: Советское радио. 1961.
4. Малафеев О.А., Муравьев А.И. Математические модели конфликтных ситуаций и их разрешение. Т.1. Общая теория и вспомогательные сведения. – СПб.: Издательство СПбГУЭФ. 2000.- 283 с.
5. Малафеев О.А., Муравьев А.И. Математические модели конфликтных ситуаций и их разрешение. Т.2. Математические основы моделирования процессов конкуренции и конфликтов в социально-экономических системах. – СПб.: Издательство СПбГУЭФ. 2001. – 294 с.
6. Зубов В.И. Лекции по теории управления: учебное пособие для студентов вузов. – М.: Издательство «Наука». Главная редакция физ.-мат. литературы. 1975. – 495 с.
7. Чернецкий В.И. Математическое моделирование стохастических систем. – Петрозаводск: Петрозаводское книжное издательство. 1994. – 488 с.
8. Вероятность и математическая статистика. Энциклопедия/ Гл. ред. Ю.В. Прохоров. М.: Большая Российская энциклопедия. 1999. – 910 с.
9. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд./ С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин. – М.: Финансы и статистика, 1983. – 471 с.
10. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей: Справ. изд. / С.А. Айвазян, И.С. Енюков, Л.Д. Мешалкин; Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1985.- 487 с., ил.
11. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л. Д. Прикладная статистика: Классификация и снижение размерности: Справ. изд. / С.А. Айвазян, В.М. Бухштабер, И.С. Енюков, Л.Д. Мешалкин; Под ред. С.А. Айвазяна. – М.: Финансы и статистика, 1989.- 607 с.: ил.

12. Гмурман В.Е. Теория вероятностей и математическая статистика: Учеб. пособие для втузов. Изд. 5-е, перераб. и доп. – М.: «Высшая школа». 1977.- 474 с.
13. Калинина В.Н., Панкин В.Ф. Математическая статистика: учебник для техникумов. Изд. 3-е, испр. – М.: Высш. шк. 2001.- 336 с.
14. Иванов А.И. Методы оценки наркоситуации: учебное пособие для служащих государственного аппарата управления. – СПб.: НПО им. Кузнецова. 2004. – 181 с.
15. Бароян О.В., Рвачев Л.А., Иванников Ю.Г. Моделирование и прогнозирование эпидемии гриппа на территории СССР. – М.: 1977.
16. Половко А.М. DERIVE для студента. – СПб.: БХВ-Петербург. 2005. – 352 с.
17. Рейнов Ю.И. Курс высшей математики. Вып. 1. Математический анализ: учебное пособие. – СПб.: Издательство «ОМ-Пресс». 2003. –64 с.
18. Рейнов Ю.И. Введение в курс теории вероятностей и математической статистики: учебное пособие. – СПб.: Издательство «ОМ-Пресс». 2006. –164 с.
19. Солодовников А.С., Бабайцев В.А., Браилов А.В. Математика в экономике: учебник для студентов экономических специальностей вузов. В двух частях. Часть I. – М.: «Финансы и статистика». 2000. – 224 с.
20. Советский энциклопедический словарь. / Ред. коллегия. Гл. ред. А.М. Прохоров. Изд. 3-е. – М.: «Советская энциклопедия». 1985. –1600 с.
21. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа: учебник для вузов. Изд. 3-е, перераб. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1972. – 496 с.
22. Справочник по математике для экономистов / В.Е. Барбаумов, В.И. Ермаков, Н.Н. Кривенцова и др.; Под ред. В.И. Ермакова. – М.: Высшая школа. 1987. – 336 с.
23. Гнеденко Б.В. Курс теории вероятностей: Учебник. – М.- Л.: Главное издательство технико-теоретической литературы. 1959. – 387 с.
24. Ширяев А.Н. Вероятность. В 2-х кн.: Учебник для вузов по физ.-мат. направлениям и спец. Вероятность-1: Элементарная теория вероятностей. Математические основания. Предельные теоремы. Изд. 3-е, перераб. и доп. – М.: Издательство МЦМНО. 2004. – 519 с.
25. Колмогоров А.Н. Основные понятия теории вероятностей / Теория вероятностей и математическая статистика. 2-е изд. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1974. – 119 с.

26. Kolmogoroff A. Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin, 1933.
27. Тутубалин В.Н. Теория вероятностей: краткий курс и научно-методические замечания. – М.: Издательство Московского университета. 1972. – 231 с.
28. Ван дер Варден Б.Л. Математическая статистика. / Пер. с нем. Л.Н. Большева. Под ред. Н.В. Смирнова. – М.: Издательство иностранной литературы. 1960. – 434 с.
29. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х т. Т.1. / Пер. с англ. Р.Л. Добрушина, А.А. Юшкевича и С.А. Молчанова. Под ред. Е.Б. Дынкина. С пред. А.Н. Колмогорова. Изд. 2-е, стереотипн. – М.: Издательство «Мир». Редакция литературы по математическим наукам. 1967. – 498 с.
30. Боровков А.А. Курс теории вероятностей. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1972. – 287 с.
31. Буняковский В.Я. Основания математической теории вероятностей. – СПб.: Типография Императорской Академии Наук. 1846.
32. Юл Э., Кендал Д. Теория статистики. Изд. 14-е, пересмотр. и расш. / Пер. с англ. Ф.Д. Лифшица. – М.: ГОССТАТИЗДАТ ЦСУ СССР. 1960. – 779 с.
33. Kolmogoroff A. Determinazione empirica di una legge di distribuzione, Giornale Istit. Ital. Attuari, 4 (1933), 83.
34. Смирнов Н.В. Об уклонениях эмпирической кривой распределения // Математический сборник, 1939, 6(48), №1, 3-24.
35. Смирнов Н.В. Приближение законов распределения случайных величин по эмпирическим данным // Успехи математических наук, (10), (1944), 179-206.
36. Иванов А.И. О глобальном похолодании климата Земли. // Проблемы окружающей среды и природных ресурсов. Обзорная информация. Вып. №1. – М.: 2003. – С.26-29.
37. Glivenko V., Sulla determinasione empirica delle leggidi probabilita. // Giorn. Attuari. IV (1933), 92-99.
38. Гливенко В.И. Курс теории вероятностей: учебник для ф.-м. ф-тов университетов. – М.-Л.: Государственное научно-техническое издательство. Редакция технико-теоретической литературы. 1939. – 220 с.
39. Губарев В.В. Вероятностные модели: Справочник. В 2-х ч./ Ч.1. – Новосибирск: Новосибирский электротехнический институт. 1992. – 196 с.
40. Вадзинский Р.Н. Справочник по вероятностным распределениям. – СПб.: Наука, 2001. 295 с., ил.

41. Дьяконов В.П. Справочник по системе символьной математики DERIVE. – М.: «СК Пресс», 1998. – 256 с., ил.
42. Лобанова О.В. Практикум по решению задач в математической системе Derive: Учеб. пособие. – М.: Финансы и статистика, 1999. – 544 с.: ил.
43. Половко А.М. DERIVE для студента. СПб.: БХВ-Петербург. 2005. – 352 с.
44. Fisher R. «Mess of Math», 1912, v. 41, p. 155-160.
45. Кремер Н.Ш. Теория вероятностей и математическая статистика: Учебник для вузов. 2-е изд., перераб. и доп. – М.: ЮНИТИДАНА. 2004. – 573 с.
46. Демидович Б.П., Марон И.А. Основы вычислительной математики: учебное пособие для втузов. Изд. 3-е, испр. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1966. – 664 с.
47. Колмогоров А.Н., Журбенко И.Г., Прохоров А.В. Введение в теорию вероятностей. М.: Наука. Главная редакция физ.-мат. литературы. 1982. – 160 с.
48. Pearson K. Contributions to the mathematical theory of evolution. PTRS, 185(1894)
49. Крамер Г. Математические методы статистики. Пер. с англ. А.С. Монина и А.А. Петрова. Под ред. А.Н. Колмогорова. Изд. 2-е, стереотипн. – М.: Издательство «Мир». 1976. – 648 с.
50. Дюге Д. Теоретическая и прикладная статистика. / Пер. с фр. В.М. Калинина. Под ред. Ю.В. Линника. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1972. – 383 с.
51. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учебное пособ. для студ. втузов. Изд. 3-е. – М.: «Высшая школа». 1979. – 400 с.
52. Боровков А.А. Математическая статистика. Оценка параметров. Проверка гипотез: учебное пособие для студ. математич. и физ. специальностей вузов. – М.: «Наука». Главная редакция физико-математической литературы. 1984. – 472 с.
53. Шметтерер Л. Введение в математическую статистику. Пер. с нем. под ред. Ю.В. Линника. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1976. – 520 с.
54. Леман Э. Проверка статистических гипотез. Изд. 2-е, испр. / Пер. с англ. Ю.В. Прохорова. – М.: «Наука». Главная редакция физ.-мат. литературы. 1979. – 408.

55. Смирнов Н.В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках // Бюллетень МГУ. 1939, 2, вып. 2, 314.
56. Kolmogoroff A. Sulla determinazione tmpirica di una leggi di distribuzione – Giorn. d. Att., 1933, 4, 83-91.
57. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. Изд. 3-е. М.: Наука. Главная редакция физико-математической литературы. 1983. – 416 с.
58. Смирнов Н.В. Теория вероятностей и математическая статистика. Избранные труды. М.: Издательство «Наука». 1970. – 290 с.
59. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. Пер. с англ. В.Е. Привальского и А.И. Кочубинского. Под ред. акад. И.Н. Коваленко. – М.: Издательство «Мир». 1989. – 540 с.
60. Ивченко Г.И., Медведев Ю.И. Математическая статистика: учебное пособие для втузов. – М.: «Высшая школа». 1981. – 248 с.
61. Уилкс С. Математическая статистика. / Пер. с англ. А.М. Кагана, Л.А. Халфина, О.В. Шалаевского. Под ред. Ю.В. Линника. М.: Издательство «Наука». Главная редакция физико-математической литературы. 1967. – 632 с.
62. Рао С.Р. Линейные статистические методы и их применения. Пер. с англ. А.М. Кагана, В.М. Калинина, К.П. Латышева. Под ред. Ю.В. Линника. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1968. – 548 с.
63. Абезгауз Г.Г., Тронь А.П., Копенкин Ю.Н., Коровина И.А. Справочник по вероятностным расчетам. – М.: Воениздат, 1970. – 536 с.
64. Кокс Д., Хинкли Д. Теоретическая статистика. – Пер. с англ. Е.В. Чепурина. Под ред. Ю.К. Беляева. – М.: Издательство «Мир». 1978. – 560 с.
65. Романовский В.И. Математическая статистика. Книга вторая: Оперативные методы математической статистики. – Ташкент: Издательство Академии Наук Узбекской ССР. 1963. – 794 с.
66. Кендалл М., Стьюдент А. Статистические выводы и связи. Пер. с англ. Л.И. Гальчука, А.Т. Терехина. Под ред. А.Н. Колмогорова. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1973. – 899 с.
67. Мацкевич И.П., Свирид Г.П. Высшая математика: Теория вероятностей и математическая статистика: Учебник для экономических специальностей вузов. Минск: «Вышэйшая школа». 1993. – 269 с.
68. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ. / Пер. с англ. А.Ф. Кушнира, А.Л. Петросяна, Е.Л. Резникова. Под ред.

В.Ф. Писаренко. – М.: Издательство «Мир». Редакция литературы по космическим исследованиям, астрономии и геофизике. 1981. – 693.
69. Вентцель Е.С. Введение в исследование операций. – М.: Издательство «Советское радио». 1964. – 388 с.
70. Чуев Ю.В., Мельников П.М., Петухов С.И., Степанов Г.Ф., Шор Я. Б. Основы исследования операций в военной технике. Под общ. ред. Ю.В. Чуева. – М.: Издательство «Советское радио». 1965. – 591 с.
71. Зубов В.И. Интерполяция и аппроксимация вероятностных распределений // Доклады Академии наук СССР. 1991. Т.316, №6. С.12981301.
72. Зубов В.И. Проблема устойчивости процессов управления. Изд. 2-е, испр. и доп. – СПб.: СПбГУ. Факультет ПМ-ПУ. НИИ Химии СПбГУ. 2001. – 353 с.
73. Иванов А.И. Приближение смесью нормальных распределений / Дополнение к кн. Зубов В.И. Проблема устойчивости процессов управления. Изд. 2-е, испр. и доп. – СПб.: СПбГУ. Факультет ПМ-ПУ. НИИ Химии СПбГУ. 2001.- С.336-351.
74. Зубова А.Ф., Иванов А.И., Григорьева И.А. Некоторые задачи теории и практики обработки результатов наблюдений. Методические указания. – СПб.: РОПИ Издательства СПбГУ. 2001. – 37 с.
75. Иванов А.И. Аппроксимация кумулянты // Процессы управления и устойчивость. Труды XXXIII научной конференции студентов и аспирантов. 15-18 апреля 2002 г. – Спб.: НИИ Химии СПбГУ. 2002.- С. 62-66.
76. Зубов В.И. Проблема обращения центральной предельной теоремы А.М. Ляпунова // Доклады Академии наук. 1995. Т.342, №1. С. 1516.
77. Колемаев В.А., Староверов О.В., Турундаевский В.Б. Теория вероятностей и математическая статистика: учебное пособие для студентов экономических специальностей вузов. / Под ред. В.А. Колемаева. – М.: «Высшая школа». 1991.- 400 с.
78. Бусленко Н.П., Голенко Д.И., Соболь И.М., Срагович В.Г., Шрейдер Ю. А. Метод статистических испытаний. (Метод Монте-Карло). / Под ред. Ю.А. Шрейдера/ Справочная математическая библиотека. Под общ. ред. Л.А. Люстерника и А.Р. Янпольского. – М.: Главное издательство физико-математической литературы. 1962. – 331 с.
79. Metropolis N., Ulam S., The Monte Carlo method // J. Amer. statistical assoc., 1949, 44, № 247, 335-341.
80. Иванова В.М. Случайные числа и их применение. Серия «Математическая статистика для экономистов». – М.: «Финансы и статистика». 1984.-112 с.

81. Голенко Д.И. Моделирование и статистический анализ псевдослучайных чисел на электронных вычислительных машинах. М.: Наука. 1965.
82. Zielinski R. Erzeugung von Zufallszahlen. – Leipzig: Fachbuchverlag. 1978.
83. Вентцель Е.С. Теория вероятностей: учебник для втузов. Изд. 2ое, перераб. и доп. – М.: Издательство «Наука». Государственное издательство физико-математической литературы. 1962. – 564 с.
84. Соболь И.М. Метод Монте-Карло. Изд. 3-е, доп. – М.: Издательство «Наука». Главная редакция физико-математической литературы. 1978.- 64 с. (Популярные лекции по математике. Вып. 46).
85. Соболь И.М. Численные методы Монте-Карло. – М.: Издательство «Наука». Главная редакция физ-мат. литературы. 1973. – 311 с.
86. ГОСТ 11.005-74 Прикладная статистика: Правила определения оценок и доверительных границ для параметров экспоненциального распределения и распределения Пуассона. М., Издательство стандартов., 1974.
87. Колмогоров А.Н. Статистический приемочный контроль при допустимом числе дефективных изделий, равном нулю. – Л.: Изд. Всесоюзного общества по распространению политических и научных знаний. 1951.
88. Иванов М.А. Применение моделирования статистических игр в задаче оценивания параметра гипергеометрического распределения: автореферат диссертации на соискание ученой степени к. ф.-м. н. – Великий Новгород: Новгородский государственный университет им. Ярослава Мудрого. 2000. – 23 с.
89. Рунион Р. Справочник по непараметрической статистике. Современный подход. / Пер. с англ. Е.З. Демиденко. Пред. Ю.Н. Тюрина. – М.: «Финансы и статистика». 1982. – 198 с.

ПРИЛОЖЕНИЕ 1

ПРОГРАММА НАХОЖДЕНИЯ ФУНКЦИИ $F_к(x)$

```
y=(x*(y1-y2)+x1*y2-x2*y1)/(x1-x2)
;#1'
(x*(y1-y2)+x1*y2-x2*y1)/(x1-x2)
;Sub(#2)
(x*(0-1/7)+4000*(1/7)-5336*0)/(4000-5336)
;Simp(#3)
1.06928*10^(-4)*(x-4000)
IF(4000<x<5336,1.06928*10^(-4)*(x-4000),0)
;Sub(#2)
(x*(1/7-2/7)+5336*(2/7)-5500*(1/7))/(5336-5500)
;Simp(#6)
8.7108*10^(-4)*(x-5172)
IF(5336<x<5500,8.7108*10^(-4)*(x-5172),0)
;Sub(#2)
(x*(2/7-3/7)+5500*(3/7)-5847*(2/7))/(5500-5847)
;Simp(#9)
4.11692*10^(-4)*(x-4806)
IF(5500<x<5847,4.11692*10^(-4)*(x-4806),0)
;Sub(#2)
(x*(3/7-4/7)+5847*(4/7)-5914*(3/7))/(5847-5914)
;Simp(#12)
0.00213219*(x-5646)
IF(5847<x<5914,0.00213219*(x-5646),0)
;Sub(#2) (x*(4/7-5/7)+5914*(5/7)-6228*(4/7))/(5914-6228)
;Simp(#15)
4.54959*10^(-4)*(x-4658)
IF(5914<x<6228,4.54959*10^(-4)*(x-4658),0)
;Sub(#2) (x*(5/7-6/7)+6228*(6/7)-6674*(5/7))/(6228-6674)
;Simp(#18)
3.20307*10^(-4)*(x-3998)
IF(6228<x<6674,3.20307*10^(-4)*(x-3998),0)
;Sub(#2) (x*(6/7-7/7)+6674*(7/7)-8000*(6/7))/(6674-8000)
;Simp(#21)
1.07735*10^(-4)*(x+1282)
IF(6674<x<8000,1.07735*10^(-4)*(x+1282),0)
1
```

ИвановА.И., к.ф-м.н.
Минвалеев Р.С.,

ИНФОРМАЦИОННЫЙ ПОДХОД К АНАЛИЗУ СИСТЕМ

Часть 1.

ОСНОВЫ МОДЕЛИРОВАНИЯ И ПЕРВИЧНАЯ ОБРАБОТКА ДАННЫХ

Учебно-методическое пособие

Рецензенты:	д.т.н., профессор Зубова А.Ф.; к.т.н., доцент Рыбакин А.С.
Тех. редактор	Ю. С. Ленкова
Верстка	Е. Е. Свежинцев

Издательский отдел Санкт-Петербургского филиала ГУ-ВШЭ
190008, Санкт-Петербург, ул. Союза Печатников, д.16,
тел. (812) 570-42-45;
e-mail: lenkova@hse.spb.ru
www.hse.spb.ru

Подписано в печать 05.08.2006. Формат 60х88/16
Гарнитура «Ньютон». Печать офсетная.
Объем: 6,25 печ. л., 7 учетно-издат. л.
Тираж 200. Заказ № 185

Отпечатано с готовых диапозитивов в типографии ООО «ЮТАС»
196105, Санкт-Петербург, ул. Рощинская, 36
тел./факс: (812) 388-03-21, e-mail: jutasprint@gmail.com