ГОСУДАРСТВЕННЫЙ КОМИТЕТ СВЯЗИ, ИНФОРМАТИЗАЦИИ И ТЕЛЕКОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ
РЕСПУБЛИКИ УЗБЕКИСТАН

САМАРКАНДСКИЙ ФИЛИАЛ ТАШКЕНТСКОГО УНИВЕРСИТЕТА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ
Факультет «Компьютер инжиниринг»
Кафедра «Информационных технологий»

ВЫПУСКНАЯ
КВАЛИФИКАЦИОННАЯ РАБОТА
для получения академической степени бакалавриата по направлению
5521900- “Информатика и информационные технологии”

ТЕМА: Разработка программных обеспечение стохастических моделей дефицита информации о выборе весовых коэффициентов в сводном показателе

Рассмотрена на заседании кафедры	Исполнитель:
(протокол №___ от “___” ______ 2014 г.)	студент группы 405
и допущена к защите.	_________ Виктореико К.
Заведующий кафедрой	Научный руководитель:
_____ доц. Махмудов З.	_____ асс. Сафарова Г.
“_____” ___________2014	

С А М А Р К А Н Д – 2014

Содержание

Введение ... 3
Глава 1. Основные понятия математической статистики ... 6
1.1. Генеральная совокупность и выборка. Вариационный ряд, статистический ряд ... 6
1.2. Полигон частот. Выборочная функция распределения и гистограмма .. 8
1.3. Числовые характеристики статистического распределения. Статистическое описание и вычисление оценок параметров двумерного случайного вектора ... 10
Глава 2. Стохастические модели дефицита информации о выборе весовых коэффициентов в сводном показателе ... 32
2.1. Синтез аддитивного показателя при рандомизированных весовых коэффициентах ... 32
2.2. Методы разработка программных обеспечение ... 41
2.3. Требования к микроклимату и освещенности на рабочих местах, оборудованных ПЭВМ ... 43
Заключение ... 46
Литературы ... 47
Приложение ... 49

Введение

Актуальность работы. Республика Узбекистан начала активную работу в сфере ИКТ 30 мая 2002 года, когда Президент подписал Указ № УП-3080 «О дальнейшем развитии компьютеризации и внедрении информационно-коммуникационных технологий». Принятие данного указа также послужило поводом для институционализации ИКТ, образовав Координационный совет по развитию компьютеризации и информационно-коммуникационных технологий, преобразовав Узбекское агентство почты и телекоммуникаций в Узбекское агентство связи и информатизации, создав Центр развития и внедрения компьютерных и информационных технологий и преобразовав Ташкентский электротехнический институт связи в Ташкентский университет информационных технологий. В конце 2003 года последовало принятие двух основных законов: Закона «Об информатизации» и Закона «Об электронной цифровой подписи».

2007 год стал знаковым событием и особенно продуктивным годом с точки зрения сосредоточения на важных вопросах взаимодействия государственных органов между собой, а также с гражданами и предприятиями через предоставление услуг ИКТ (ПКМ №181 от 23 августа 2007 года)[1,2]. Это является результатом эволюционного развития ИКТ в Узбекистане, когда ранее принятые решения повлекли за собой создание разнообразных информационных ресурсов в государственном секторе, и повышение внимания к вопросу эффективного взаимодействия между государственными органами. Было принято специальное положение для создания стандартов для интерактивных государственных услуг с использованием ИКТ, предоставляемых государственными органами, а также был создан перечень основных интерактивных услуг.

Также далее был уточнен перечень государственных информационных ресурсов (ПКМ № 34 от 16 февраля 2007 года). Было логично, что в том же году Правительство Узбекистана приняло Постановление № 259 от 17 декабря 2007 года для усовершенствования представления информационных ресурсов и оказания интерактивных услуг на Правительственном портале.

Вместе с тем, можно сказать, что именно 2012 год стал революционным для электронного правительства в Узбекистане. В 2012 году были приняты 4 закона с целью значительного ускорения развития ИКТ в Республике Узбекистан. В первую очередь, Президент указал в своем Постановлении № ПП-1730 от 21 марта 2012 года на дополнительные меры по созданию Национальной Информационной Системы (НИС), охватывающей 32 государственные информационные системы, через углубление интеграции действующих информационных систем, контролируемых государством и других общественных органов и лиц, на основе общей технической политики. Президентом была утверждена Программа на 2012-2014 годы, исполнение которой возложено на Координационный Совет по развитию компьютеризации и информационно-коммуникационных технологий. При Совете создана постоянно действующая экспертная группа, которая будет встречаться на постоянной основе[1,2].

Указ Президента № УП-4475 от 16 октября 2012 года преобразовал Узбекское агентство связи и информатизации в Государственный комитет связи, ИКТ Республики Узбекистан.

В последние десятилетия получили широкое распространение прикладные программы, использующие алгоритмы генерации, селекции и оценивания различных комбинаторных объектов. Такие алгоритмы являются важными составными частями, например, методов целочисленного программирования, в которых они применяются для генерации опорных планов. Особенно следует отметить роль комбинаторных алгоритмов перебора структур разного вида (графы, деревья, разбиения и т.п.) в методах многокритериальной оптимизации, в процедурах принятия многокритериальных решений, в системах искусственного интеллекта, в современных интерактивных системах анализа данных, в экспертных системах и т.д[3,4,5,6,7,8,9].

Особенно важную роль рассматриваемые комбинаторные" алгоритмы играют в СППР, ориентированных на использование нечисловой, неточной и неполной информации о весовых коэффициентах и/или вероятностей. Эти СППР в последнее время широко применяются для синтеза сводных оценок сложных объектов разной природы, например, для построения оценок многопараметрических технических систем военного назначения и НИР, направленных на разработку таких систем, для сводной оценки сложных финансово-экономических объектов (коммерческих банков, страховых компаний и т.п.), для формирования рациональных вариантов инвестиционного портфеля, для оценивания экологического состояния и устойчивости геосистем, для построения непараметрических байесовских оценок распределений случайных величин и т.д.

Цель и задачи исследования. Цель работы заключается в исследовании и программной реализации комплекса комбинаторных методов генерации, оценки целочисленных композиций для разработки программного обеспечения прикладной системы поддержки принятия решений (СППР), реализующей метод построения сводных показателей в условиях неопределенности.

Структура и объем работы. Работа состоит из введения, двух глав, заключения и приложений.

Глава 1. Основные понятия математической статистики

1.1. Генеральная совокупность и выборка. Вариационный ряд, статистический ряд

Математическая статистика занимается установлением закономерностей, которым подчинены массовые случайные явления, на основе обработки статистических данных, полученных в результате наблюдений. Двумя основными задачами математической статистики являются:

- определение способов сбора и группировки этих статистических данных;
- разработка методов анализа полученных данных в зависимости от целей исследования, к которым относятся:

а) оценка неизвестной вероятности события; оценка неизвестной функции распределения; оценка параметров распределения, вид которого известен; оценка зависимости от других случайных величин и т.д.;

б) проверка статистических гипотез о виде неизвестного распределения или о значениях параметров известного распределения[9].

Для решения этих задач необходимо выбрать из большой совокупности однородных объектов ограниченное количество объектов, по результатам изучения которых можно сделать прогноз относительно исследуемого признака этих объектов.

Определим основные понятия математической статистики.

Генеральная совокупность – все множество имеющихся объектов.

Выборка – набор объектов, случайно отобранных из генеральной совокупности.

Объем генеральной совокупности N и объем выборки n – число объектов в рассматриваемой совокупности.

Виды выборки:

Повторная – каждый отобранный объект перед выбором следующего возвращается в генеральную совокупность;

Бесповторная – отобранный объект в генеральную совокупность не возвращается.

Замечание. Для того, чтобы по исследованию выборки можно было сделать выводы о поведении интересующего нас признака генеральной совокупности, нужно, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была репрезентативной (представительной). Учитывая закон больших чисел, можно утверждать, что это условие выполняется, если каждый объект выбран случайно, причем для любого объекта вероятность попасть в выборку одинакова.

Первичная обработка результатов.

Пусть интересующая нас случайная величина X принимает в выборке значение x_1 n_1 раз, x_2 – n_2 раз, …, $x_к$ – $n_к$ раз, причем $\sum_{i=1}^{k} n_k = n,$ где n – объем выборки. Тогда наблюдаемые значения случайной величины x_1, x_2,…, $x_к$ называют вариантами, а n_1, n_2,…, $n_к$ – частотами. Если разделить каждую частоту на объем выборки, то получим относительные частоты $w_i = \frac{n_i}{n}$.

Последовательность вариант, записанных в порядке возрастания, называют вариационным рядом, а перечень вариант и соответствующих им частот или относительных частот – статистическим рядом:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k
w_i	w_1	w_2	…	w_k

Пример.

При проведении 20 серий из 10 бросков игральной кости число выпадений шести очков оказалось равным 1,1,4,0,1,2,1,2,2,0,5,3,3,1,0,2,2,3,4,1.Составим вариационный ряд: 0,1,2,3,4,5. Статистический ряд для абсолютных и относительных частот имеет вид:

x_i	0	1	2	3	4	5

n_i	3	6	5	3	2	1
w_i	0,15	0,3	0,25	0,15	0,1	0,05

Если исследуется некоторый непрерывный признак, то вариационный ряд может состоять из очень большого количества чисел. В этом случае удобнее использовать группированную выборку. Для ее получения интервал, в котором заключены все наблюдаемые значения признака, разбивают на несколько равных частичных интервалов длиной h, а затем находят для каждого частичного интервала n_i – сумму частот вариант, попавших в i-й интервал. Составленная по этим результатам таблица называется группированным статистическим рядом[9]:

Номера интервалов	1	2	…	k
Границы интервалов	$(a, a + h)$	$(a + h, a + 2h)$	…	$(b – h, b)$
Сумма частот вариант, попавших в интервал	n_1	n_2	…	n_k

1.2. Полигон частот. Выборочная функция распределения и гистограмма

Для наглядного представления о поведении исследуемой случайной величины в выборке можно строить различные графики. Один из них – полигон частот: ломаная, отрезки которой соединяют точки с координатами (x_1, n_1), (x_2, n_2),…, (x_k, n_k), где x_i откладываются на оси абсцисс, а n_i – на оси ординат. Если на оси ординат откладывать не абсолютные (n_i), а относительные (w_i) частоты, то получим полигон относительных частот (рис.1).

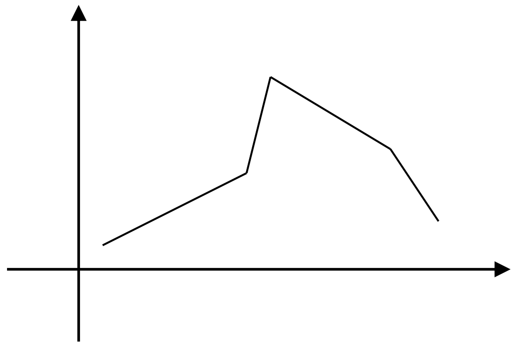

Рис. 1.

По аналогии с функцией распределения случайной величины можно задать некоторую функцию, относительную частоту события $X < x$.

Определение. Выборочной (эмпирической) функцией распределения называют функцию $F^*(x)$, определяющую для каждого значения x относительную частоту события $X < x$. Таким образом,

$$F^*(x) = \frac{n_x}{n}, \qquad (1.1)$$

где n_x – число вариант, меньших x, n – объем выборки.

Замечание. В отличие от эмпирической функции распределения, найденной опытным путем, функцию распределения $F(x)$ генеральной совокупности называют *теоретической функцией распределения*. $F(x)$ определяет вероятность события $X < x$, а $F^*(x)$ – его относительную частоту. При достаточно больших n, как следует из теоремы Бернулли, $F^*(x)$ стремится по вероятности к $F(x)$.

Из определения эмпирической функции распределения видно, что ее свойства совпадают со свойствами $F(x)$, а именно:

1. $0 \le F^*(x) \le 1$.
2. $F^*(x)$ – неубывающая функция.
3. Если x_1 – наименьшая варианта, то $F^*(x) = 0$ при $x \le x_1$; если $x_к$ – наибольшая варианта, то $F^*(x) = 1$ при $x > x_к$.

Для непрерывного признака графической иллюстрацией служит гистограмма, то есть ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиной h, а высотами –

отрезки длиной n_i $/h$ (гистограмма частот) или w_i $/h$ (гистограмма относительных частот). В первом случае площадь гистограммы равна объему выборки, во втором – единице (рис.2).

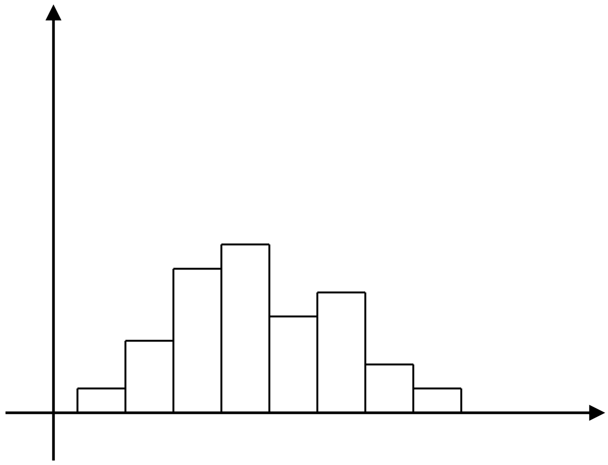

Рис.2.

1.3. Числовые характеристики статистического распределения. Статистическое описание и вычисление оценок параметров двумерного случайного вектора

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины.

Определение 16.1. Выборочным средним называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\bar{x}_B = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{n_1 x_1 + n_2 x_2 + \ldots + n_k x_k}{n} = \frac{\sum_{i=1}^{k} n_i x_i}{n}, \qquad (1.2)$$

где x_i – варианты, n_i - частоты.

Замечание. Выборочное среднее служит для оценки математического ожидания исследуемой случайной величины. В дальнейшем будет рассмотрен вопрос, насколько точной является такая оценка.

Определение. Выборочной дисперсией называется

$$D_B = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_B)^2}{n} = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n}, \qquad (1.3)$$

а выборочным средним квадратическим отклонением –

$$\sigma_B = \sqrt{D_B}. \qquad (1.4)$$

Так же, как в теории случайных величин, можно доказать, что справедлива следующая формула для вычисления выборочной дисперсии[9]:

$$D = \overline{x^2} - (\bar{x})^2. \qquad (1.5)$$

Пример 1. Найдем числовые характеристики выборки, заданной статистическим рядом

x_i	2	5	7	8
n_i	3	8	7	2

$$\bar{x}_B = \frac{2\cdot 3 + 5\cdot 8 + 7\cdot 7 + 8\cdot 2}{20} = 5{,}55;\quad D_B = \frac{4\cdot 3 + 25\cdot 8 + 49\cdot 7 + 64\cdot 2}{20} - 5{,}55^2 = 3{,}3475;\quad \sigma_B = \sqrt{3{,}3475} = 1{,}83.$$

Другими характеристиками вариационного ряда являются:

- мода M_0 – варианта, имеющая наибольшую частоту (в предыдущем примере $M_0 = 5$).

- медиана m_e - варианта, которая делит вариационный ряд на две части, равные по числу вариант. Если число вариант нечетно ($n = 2k + 1$), то $m_e = x_{k+1}$, а при четном $n = 2k$ $m_e = \frac{x_k + x_{k+1}}{2}$. В частности, в примере 1 $m_e = \frac{5+7}{2} = 6$.

Оценки начальных и центральных моментов (так называемые эмпирические моменты) определяются аналогично соответствующим теоретическим моментам:

- начальным эмпирическим моментом порядка k называется

$$M_k = \frac{\sum n_i x_i^k}{n}. \qquad (1.6)$$

В частности, $M_1 = \frac{\sum n_i x_i}{n} = \bar{x}_B$, то есть начальный эмпирический момент первого порядка равен выборочному среднему.

- центральным эмпирическим моментом порядка k называется

$$m_k = \frac{\sum n_i (x_i - \bar{x}_B)^k}{n}. \qquad (1.7)$$

В частности, $m_2 = \frac{\sum n_i (x_i - \bar{x}_B)^2}{n} = D_B$, то есть центральный эмпирический момент второго порядка равен выборочной дисперсии.

4.Статистическое описание и вычисление характеристик двумерного случайного вектора

При статистическом исследовании двумерных случайных величин основной задачей является обычно выявление связи между составляющими. Двумерная выборка представляет собой набор значений случайного вектора: (x_1, y_1), (x_2, y_2), …, (x_n, y_n). Для нее можно определить выборочные средние составляющих: $\bar{x}_B = \frac{\sum x_i}{n}$, $\bar{y}_B = \frac{\sum y_i}{n}$ и соответствующие выборочные дисперсии и средние квадратические отклонения. Кроме того, можно вычислить условные средние: $\bar{y}_x$ - среднее арифметическое наблюдавшихся значений Y, соответствующих $X = x$, и $\bar{x}_y$ - среднее значение наблюдавшихся значений X, соответствующих $Y = y$.

Если существует зависимость между составляющими двумерной случайной величины, она может иметь разный вид: функциональная зависимость, если каждому возможному значению X соответствует одно значение Y, и статистическая, при которой изменение одной величины приводит к изменению распределения другой. Если при этом в результате изменения одной величины меняется среднее значение другой, то статистическую зависимость между ними называют корреляционной.

5.Основные свойства статистических характеристик параметров распределения

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть Θ* - статистическая оценка неизвестного параметра Θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема *n* и вычислим для каждой из них оценку параметра Θ: $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$. Тогда оценку Θ* можно рассматривать как случайную величину, принимающую возможные значения $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$. Если математическое ожидание Θ* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\Theta^*) > \Theta$, и с недостатком, если $M(\Theta^*) < \Theta$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\Theta^*) = \Theta$.

Определение. Статистическая оценка Θ* называется несмещенной, если ее математическое ожидание равно оцениваемому параметру Θ при любом объеме выборки:

$$M(\Theta^*) = \Theta. \qquad (1.8)$$

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Однако несмещенность не является достаточным условием хорошего приближения к истинному значению оцениваемого параметра. Если при этом возможные значения Θ* могут значительно отклоняться от среднего значения, то есть дисперсия Θ* велика, то значение, найденное по данным одной выборки, может значительно отличаться от оцениваемого параметра. Следовательно, требуется наложить ограничения на дисперсию.

Определение. Статистическая оценка называется эффективной, если она при заданном объеме выборки *n* имеет наименьшую возможную дисперсию.

При рассмотрении выборок большого объема к статистическим оценкам предъявляется еще и требование состоятельности.

Определение . Состоятельной называется статистическая оценка, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру (если эта оценка несмещенная, то она будет состоятельной, если при $n \to \infty$ ее дисперсия стремится к 0).

Убедимся, что $\bar{x}_B$ представляет собой несмещенную оценку математического ожидания $M(X)$.

Будем рассматривать $\bar{x}_B$ как случайную величину, а $x_1, x_2, \ldots, x_n$, то есть значения исследуемой случайной величины, составляющие выборку, – как независимые, одинаково распределенные случайные величины $X_1, X_2, \ldots, X_n$, имеющие математическое ожидание a. Из свойств математического ожидания следует, что

$$M(\bar{X}_B) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = a.$$

Но, поскольку каждая из величин $X_1, X_2, \ldots, X_n$ имеет такое же распределение, что и генеральная совокупность, $a = M(X)$, то есть $M(\bar{X}_B) = M(X)$, что и требовалось доказать. Выборочное среднее является не только несмещенной, но и состоятельной оценкой математического ожидания. Если предположить, что $X_1, X_2, \ldots, X_n$ имеют ограниченные дисперсии, то из теоремы Чебышева следует, что их среднее арифметическое, то есть $\bar{X}_B$, при увеличении n стремится по вероятности к математическому ожиданию a каждой их величин, то есть к $M(X)$. Следовательно, выборочное среднее есть состоятельная оценка математического ожидания.

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Можно доказать, что

$$M(D_B) = \frac{n-1}{n} D_Г, \qquad (1.9)$$

где $D_Г$ – истинное значение дисперсии генеральной совокупности. Можно предложить другую оценку дисперсии – исправленную дисперсию s^2, вычисляемую по формуле

$$s^2 = \frac{n}{n-1} D_B = \frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n-1} . \qquad (1.10)$$

Такая оценка будет являться несмещенной. Ей соответствует исправленное среднее квадратическое отклонение

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \bar{x}_B)^2}{n-1}} . \qquad (1.11)$$

Определение. Оценка некоторого признака называется асимптотически несмещенной, если для выборки $x_1, x_2, \ldots, x_n$

$$\lim_{n\to\infty} \frac{x_1 + x_2 + \ldots + x_n}{n} = X , \qquad (1.12)$$

где X – истинное значение исследуемой величины.

6.Способы построения оценок

1. Метод наибольшего правдоподобия.

Пусть X – дискретная случайная величина, которая в результате n испытаний приняла значения x_1, x_2, …, x_n. Предположим, что нам известен закон распределения этой величины, определяемый параметром Θ, но неизвестно численное значение этого параметра. Найдем его точечную оценку.

Пусть $p(x_i, \Theta)$ – вероятность того, что в результате испытания величина X примет значение x_i. Назовем функцией правдоподобия дискретной случайной величины X функцию аргумента Θ, определяемую по формуле:

$$L\ (x_1, x_2, \ldots, x_n; \Theta) = p(x_1,\Theta)p(x_2,\Theta)\ldots p(x_n,\Theta).$$

Тогда в качестве точечной оценки параметра Θ принимают такое его значение $\Theta^* = \Theta(x_1,\ x_2,\ \ldots,\ x_n)$, при котором функция правдоподобия достигает максимума. Оценку Θ* называют оценкой наибольшего правдоподобия.

Поскольку функции L и $\ln L$ достигают максимума при одном и том же значении Θ, удобнее искать максимум $\ln L$ – логарифмической функции правдоподобия. Для этого нужно:

1) найти производную $\frac{d \ln L}{d\Theta}$;
2) приравнять ее нулю (получим так называемое *уравнение правдоподобия*) и найти критическую точку;
3) найти вторую производную $\frac{d^2 \ln L}{d\Theta^2}$; если она отрицательна в критической точке, то это – точка максимума.

Достоинства метода наибольшего правдоподобия: полученные оценки состоятельны (хотя могут быть смещенными), распределены асимптотически нормально при больших значениях n и имеют наименьшую дисперсию по сравнению с другими асимптотически нормальными оценками; если для оцениваемого параметра Θ существует эффективная оценка Θ^*, то уравнение правдоподобия имеет единственное решение Θ^*; метод наиболее полно использует данные выборки и поэтому особенно полезен в случае малых выборок.

Недостаток метода наибольшего правдоподобия: сложность вычислений.

Для непрерывной случайной величины с известным видом плотности распределения $f(x)$ и неизвестным параметром Θ функция правдоподобия имеет вид:

$$L\ (x_1, x_2, \ldots, x_n; \Theta) = f(x_1,\Theta)f(x_2,\Theta)\ldots f(x_n,\Theta).$$

Оценка наибольшего правдоподобия неизвестного параметра проводится так же, как для дискретной случайной величины.

2. Метод моментов.

Метод моментов основан на том, что начальные и центральные эмпирические моменты являются состоятельными оценками соответственно начальных и центральных теоретических моментов, поэтому можно приравнять теоретические моменты соответствующим эмпирическим моментам того же порядка.

Если задан вид плотности распределения $f(x, \Theta)$, определяемой одним неизвестным параметром Θ, то для оценки этого параметра достаточно иметь одно уравнение. Например, можно приравнять начальные моменты первого порядка:

$$\bar{x}_B = M(X) = \int_{-\infty}^{\infty} xf(x;\Theta)dx = \varphi(\Theta),$$

получив тем самым уравнение для определения Θ. Его решение Θ* будет точечной оценкой параметра, которая является функцией от выборочного среднего и, следовательно, и от вариант выборки:

$$\Theta = \psi\ (x_1, x_2, \ldots, x_n).$$

Если известный вид плотности распределения $f(x, \Theta_1, \Theta_2)$ определяется двумя неизвестными параметрами Θ_1 и Θ_2, то требуется составить два уравнения, например

$$\nu_1 = M_1,\ \ \mu_2 = m_2.$$

Отсюда $\begin{cases} M(X) = \bar{x}_B \\ D(X) = D_B \end{cases}$ - система двух уравнений с двумя неизвестными Θ_1 и Θ_2.

Ее решениями будут точечные оценки Θ_1* и Θ_2* - функции вариант выборки:

$$\Theta_1 = \psi_1\ (x_1, x_2, \ldots, x_n),$$

$$\Theta_2 = \psi_2(x_1, x_2, \ldots, x_n).$$

3. Метод наименьших квадратов.

Если требуется оценить зависимость величин y и x, причем известен вид связывающей их функции, но неизвестны значения входящих в нее коэффициентов, их величины можно оценить по имеющейся выборке с помощью метода наименьших квадратов. Для этого функция $y = \varphi\ (x)$ выбирается так, чтобы сумма квадратов отклонений наблюдаемых значений $y_1, y_2, \ldots, y_n$ от $\varphi(x_i)$ была минимальной:

$$\sum_{i=1}^{n}(y_i - \varphi(x_i))^2 = \min.$$

При этом требуется найти стационарную точку функции φ(x; a, b, c…), то есть решить систему:

$$\begin{cases} \sum_{i=1}^{n}(y_i-\varphi(x_i;a,b,c...))\left(\frac{\partial\varphi}{\partial a}\right)_i=0 \\ \sum_{i=1}^{n}(y_i-\varphi(x_i;a,b,c...))\left(\frac{\partial\varphi}{\partial b}\right)_i=0 \\ \sum_{i=1}^{n}(y_i-\varphi(x_i;a,b,c...))\left(\frac{\partial\varphi}{\partial c}\right)_i=0 \\ \text{.......................................} \end{cases}$$

(решение, конечно, возможно только в случае, когда известен конкретный вид функции φ).

Рассмотрим в качестве примера подбор параметров линейной функции методом наименьших квадратов.

Для того, чтобы оценить параметры a и b в функции $y = ax + b$, найдем $\left(\frac{\partial\varphi}{\partial a}\right)_i=x_i;\left(\frac{\partial\varphi}{\partial b}\right)_i=1.$ Тогда $\begin{cases} \sum_{i=1}^{n}(y_i-(ax_i+b))x_i=0 \\ \sum_{i=1}^{n}(y_i-(ax_i+b))=0 \end{cases}$. Отсюда $\begin{cases} \sum_{i=1}^{n}x_iy_i-a\sum_{i=1}^{n}x_i^2-b\sum_{i=1}^{n}x_i=0 \\ \sum_{i=1}^{n}y_i-a\sum_{i=1}^{n}x_i-bn=0 \end{cases}$. Разделив оба полученных уравнения на n и вспомнив определения эмпирических моментов, можно получить выражения для a и b в виде:

$a=\frac{(K_{xy})_B}{(D_x)_B}$, $b=\bar{y}_B-\frac{(K_{xy})_B}{(D_x)_B}\bar{x}_B$. Следовательно, связь между x и y можно задать в виде:

$$y-\bar{y}_B=\frac{(K_{xy})_B}{(D_x)_B}(x-\bar{x}_B).$$

4. Байесовский подход к получению оценок.

Пусть (Y, X) – случайный вектор, для которого известна плотность $p(y|x)$ условного распределения Y при каждом значении $X = x$. Если в результате эксперимента получены лишь значения Y, а соответствующие значения X неизвестны, то для оценки некоторой заданной функции $\varphi(x)$ в качестве ее приближенного значения предлагается искать условное математическое ожидание $M(\varphi(x)|Y)$, вычисляемое по формуле:

$\psi(Y) = \dfrac{\int \varphi(x) p(Y \mid x) p(x) d\mu(x)}{q(Y)}$, где $q(y) = \int p(y \mid x) p(x) d\mu(x)$, $p(x)$ – плотность безусловного распределения X, $q(y)$ – плотность безусловного распределения Y. Задача может быть решена только тогда, когда известна $p(x)$. Иногда, однако, удается построить состоятельную оценку для $q(y)$, зависящую только от полученных в выборке значений Y.

6. Интервальное оценивание неизвестных параметров

При выборке малого объема точечная оценка может значительно отличаться от оцениваемого параметра, что приводит к грубым ошибкам. Поэтому в таком случае лучше пользоваться *интервальными оценками*, то есть указывать интервал, в который с заданной вероятностью попадает истинное значение оцениваемого параметра. Разумеется, чем меньше длина этого интервала, тем точнее оценка параметра. Поэтому, если для оценки Θ^* некоторого параметра Θ справедливо неравенство $|\Theta^* - \Theta| < \delta$, число $\delta > 0$ характеризует точность оценки (чем меньше δ, тем точнее оценка). Но статистические методы позволяют говорить только о том, что это неравенство выполняется с некоторой вероятностью.

Определение. Надежностью (доверительной вероятностью) оценки Θ^* параметра Θ называется вероятность γ того, что выполняется неравенство $|\Theta^* - \Theta| < \delta$. Если заменить это неравенство двойным неравенством $-\delta < \Theta^* - \Theta < \delta$, то получим:

$$p(\Theta^* - \delta < \Theta < \Theta^* + \delta) = \gamma.$$

Таким образом, γ есть вероятность того, что Θ попадает в интервал $(\Theta^* - \delta, \Theta^* + \delta)$.

Определение. Доверительным называется интервал, в который попадает неизвестный параметр с заданной надежностью γ.

Построение доверительных интервалов[9].

1. Доверительный интервал для оценки математического ожидания нормального распределения при известной дисперсии.

Пусть исследуемая случайная величина X распределена по нормальному закону с известным средним квадратическим σ, и требуется по значению выборочного среднего $\bar{x}_B$ оценить ее математическое ожидание a. Будем рассматривать выборочное среднее $\bar{x}_B$ как случайную величину $\bar{X}$, а значения вариант выборки x_1, x_2,…, x_n как одинаково распределенные независимые случайные величины X_1, X_2,…, X_n, каждая из которых имеет математическое ожидание a и среднее квадратическое отклонение σ. При этом $M(\bar{X}) = a$, $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$ (используем свойства математического ожидания и дисперсии суммы независимых случайных величин). Оценим вероятность выполнения неравенства $|\bar{X} - a| < \delta$. Применим формулу для вероятности попадания нормально распределенной случайной величины в заданный интервал:

$p\ (|X - a| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$. Тогда , с учетом того, что $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$, $p\ (|\bar{X} - a| < \delta)$ = $2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right)$=2Φ(t), где $t = \frac{\delta\sqrt{n}}{\sigma}$. Отсюда $\delta = \frac{t\sigma}{\sqrt{n}}$, и предыдущее равенство можно переписать так:

$$p\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}} < a < \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma. \qquad (1.13)$$

Итак, значение математического ожидания a с вероятностью (надежностью) γ попадает в интервал $\left(\bar{x}_B - \frac{t\sigma}{\sqrt{n}}; \bar{x}_B + \frac{t\sigma}{\sqrt{n}}\right)$, где значение t определяется из таблиц для функции Лапласа так, чтобы выполнялось равенство 2Φ(t) = γ.

Пример. Найдем доверительный интервал для математического ожидания нормально распреде-ленной случайной величины, если объем выборки n = 49, $\bar{x}_B = 2{,}8$, σ = 1,4, а доверительная вероятность γ = 0,9.

Определим t, при котором $\Phi(t) = 0{,}9:2 = 0{,}45$: $t = 1{,}645$. Тогда

$2{,}8 - \frac{1{,}645\cdot 1{,}4}{\sqrt{49}} < a < 2{,}8 + \frac{1{,}645\cdot 1{,}4}{\sqrt{14}}$, или $2{,}471 < a < 3{,}129$. Найден доверительный интервал, в который попадает a с надежностью 0,9.

2. Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии.

Если известно, что исследуемая случайная величина X распределена по нормальному закону с неизвестным средним квадратическим отклонением, то для поиска доверительного интервала для ее математического ожидания построим новую случайную величину

$$T = \frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}}, \qquad (1.14)$$

где $\bar{x}_B$ - выборочное среднее, s – исправленная дисперсия, n – объем выборки. Эта случайная величина, возможные значения которой будем обозначать t, имеет распределение Стьюдента (см. лекцию 12) с $k = n - 1$ степенями свободы.

Поскольку плотность распределения Стьюдента $s(t,n) = B_n\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$,

где $B_n = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)}\Gamma\left(\frac{n-1}{2}\right)}$, явным образом не зависит от a и σ, можно задать вероятность ее попадания в некоторый интервал (- t_γ , t_γ), учитывая четность плотности распределения, следующим образом: $p\left(\left|\frac{\bar{x}_B - a}{\frac{s}{\sqrt{n}}}\right| < t_\gamma\right) = 2\int_0^{t_\gamma} s(t,n)dt = \gamma$.

Отсюда получаем:

$$p\left(\bar{x}_B - \frac{t_\gamma s}{\sqrt{n}} < a < \bar{x}_B + \frac{t_\gamma s}{\sqrt{n}}\right) = \gamma. \qquad (1.15)$$

Таким образом, получен доверительный интервал для a, где t_γ можно найти по соответствующей таблице при заданных n и γ.

Пример. Пусть объем выборки n = 25, $\bar{x}_B$ = 3, s = 1,5. Найдем доверительный интервал для a при γ = 0,99. Из таблицы находим, что t_γ (n = 25, γ = 0,99) = 2,797. Тогда $3-\frac{2{,}797\cdot 1{,}5}{\sqrt{25}}<a<3+\frac{2{,}797\cdot 1{,}5}{\sqrt{25}}$, или 2,161< a < 3,839 – доверительный интервал, в который попадает a с вероятностью 0,99.

3. Доверительные интервалы для оценки среднего квадратического отклонения нормального распределения.

Будем искать для среднего квадратического отклонения нормально распределенной случайной величины доверительный интервал вида (s – δ, s +δ), где s – исправленное выборочное среднее квадратическое отклонение, а для δ выполняется условие: p (|σ – s| < δ) = γ.

Запишем это неравенство в виде: $s\left(1-\frac{\delta}{s}\right)<\sigma<s\left(1+\frac{\delta}{s}\right)$ или, обозначив $q=\frac{\delta}{s}$,

$$s(1-q)<\sigma<s(1+q). \qquad (1.16)$$

Рассмотрим случайную величину χ, определяемую по формуле

$$\chi=\frac{s}{\sigma}\sqrt{n-1},$$

которая распределена по закону «хи-квадрат» с n-1 степенями свободы.

Плотность ее распределения

$$R(\chi,n)=\frac{\chi^{n-2}e^{-\frac{\chi^2}{2}}}{2^{\frac{n-3}{2}}\Gamma\left(\frac{n-1}{2}\right)}$$

не зависит от оцениваемого параметра σ, а зависит только от объема выборки n. Преобразуем неравенство (6.4) так, чтобы оно приняло вид $\chi_1 < \chi < \chi_2$. Вероятность выполнения этого неравенства равна доверительной вероятности γ, следовательно, $\int_{\chi_1}^{\chi_2} R(\chi, n)d\chi = \gamma$. Предположим, что q < 1, тогда неравенство (6.4) можно записать так:

$$\frac{1}{s(1+q)} < \frac{1}{\sigma} < \frac{1}{s(1-q)},$$

или, после умножения на $s\sqrt{n-1}$, $\frac{\sqrt{n-1}}{1+q} < \frac{s\sqrt{n-1}}{\sigma} < \frac{\sqrt{n-1}}{1-q}$. Следовательно, $\frac{\sqrt{n-1}}{1+q} < \chi < \frac{\sqrt{n-1}}{1-q}$. Тогда $\int_{\frac{\sqrt{n-1}}{1+q}}^{\frac{\sqrt{n-1}}{1-q}} R(\chi, n)d\chi = \gamma$. Существуют таблицы для распределения «хи-квадрат», из которых можно найти q по заданным n и γ, не решая этого уравнения. Таким образом, вычислив по выборке значение s и определив по таблице значение q, можно найти доверительный интервал (6.4), в который значение σ попадает с заданной вероятностью γ.

Замечание. Если q > 1, то с учетом условия σ > 0 доверительный интервал для σ будет иметь границы

$$0 < \sigma < s(1+q) \tag{1.17}$$

Пример.

Пусть n = 20, s = 1,3. Найдем доверительный интервал для σ при заданной надежности γ = 0,95. Из соответствующей таблицы находим q (n = 20, γ = 0,95) = 0,37. Следовательно, границы доверительного интервала: 1,3(1-0,37) = 0,819 и 1,3(1+0,37) = 1,781. Итак, 0,819 < σ < 1,781 с вероятностью 0,95.

7. Статистическая проверка статистических гипотез

Определение. Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Определение. Нулевой (основной) называют выдвинутую гипотезу H_0. Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Пример. Пусть H_0 заключается в том, что математическое ожидание генеральной совокупности $a = 3$. Тогда возможные варианты H_1: а) $a \neq 3$; б) $a > 3$; в) $a < 3$.

Определение. Простой называют гипотезу, содержащую только одно предположение, сложной – гипотезу, состоящую из конечного или бесконечного числа простых гипотез.

Пример. Для показательного распределения гипотеза H_0: $\lambda = 2$ – простая, H_0: $\lambda > 2$ – сложная, состоящая из бесконечного числа простых (вида $\lambda = c$, где c – любое число, большее 2).

В результате проверки правильности выдвинутой нулевой гипотезы (такая проверка называется статистической, так как производится с применением методов математической статистики) возможны ошибки двух видов: ошибка первого рода, состоящая в том, что будет отвергнута правильная нулевая гипотеза, и ошибка второго рода, заключающаяся в том, что будет принята неверная гипотеза.

Замечание. Какая из ошибок является на практике более опасной, зависит от конкретной задачи. Например, если проверяется правильность выбора метода лечения больного, то ошибка первого рода означает отказ от правильной методики, что может замедлить лечение, а ошибка второго рода (применение неправильной методики) чревата ухудшением состояния больного и является более опасной.

Определение. Вероятность ошибки первого рода называется уровнем значимости α.

Основной прием проверки статистических гипотез заключается в том, что по имеющейся выборке вычисляется значение некоторой случайной величины, имеющей известный закон распределения.

Определение. Статистическим критерием называется случайная величина K с известным законом распределения, служащая для проверки нулевой гипотезы.

Определение 19.6. Критической областью называют область значений критерия, при которых нулевую гипотезу отвергают, областью принятия гипотезы – область значений критерия, при которых гипотезу принимают.

Итак, процесс проверки гипотезы состоит из следующих этапов:

1) выбирается статистический критерий K;
2) вычисляется его наблюдаемое значение $K_{набл}$ по имеющейся выборке;
3) поскольку закон распределения K известен, определяется (по известному уровню значимости α) критическое значение $k_{кр}$, разделяющее критическую область и область принятия гипотезы (например, если $p(K > k_{кр}) = \alpha$, то справа от $k_{кр}$ располагается критическая область, а слева – область принятия гипотезы);
4) если вычисленное значение $K_{набл}$ попадает в область принятия гипотезы, то нулевая гипотеза принимается, если в критическую область – нулевая гипотеза отвергается.

Различают разные виды критических областей:

- правостороннюю критическую область, определяемую неравенством $K > k_{кр}$ ($k_{кр} > 0$);
- левостороннюю критическую область, определяемую неравенством $K < k_{кр}$ ($k_{кр} < 0$);
- двустороннюю критическую область, определяемую неравенствами $K < k_1$, $K > k_2$ $\quad (k_2 > k_1)$.

Определение. Мощностью критерия называют вероятность попадания критерия в критическую область при условии, что верна конкурирующая гипотеза.

Если обозначить вероятность ошибки второго рода (принятия неправильной нулевой гипотезы) β, то мощность критерия равна 1 – β. Следовательно, чем больше мощность критерия, тем меньше вероятность совершить ошибку второго рода. Поэтому после выбора уровня значимости следует строить критическую область так, чтобы мощность критерия была максимальной.

7. Критерий для проверки гипотезы о вероятности события

Пусть проведено n независимых испытаний (n – достаточно большое число), в каждом из которых некоторое событие A появляется с одной и той же, но неизвестной вероятностью p, и найдена относительная частота $\frac{m}{n}$ появлений A в этой серии испытаний. Проверим при заданном уровне значимости α нулевую гипотезу H_0, состоящую в том, что вероятность p равна некоторому значению p_0.

Примем в качестве статистического критерия случайную величину

$$U = \frac{\left(\frac{M}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}}, \qquad (1.18)$$

имеющую нормальное распределение с параметрами $M(U) = 0$, $\sigma(U) = 1$ (то есть нормиро-ванную). Здесь $q_0 = 1 - p_0$. Вывод о нормальном распределении критерия следует из теоремы Лапласа (при достаточно большом n относительную частоту можно приближенно считать нормально распределенной с математическим ожиданием p и средним квадратическим отклонением $\sqrt{\frac{pq}{n}}$).

Критическая область строится в зависимости от вида конкурирующей гипотезы.

1) Если H_0: $p = p_0$, а H_1: $p \neq p_0$, то критическую область нужно построить так, чтобы вероятность попадания критерия в эту область равнялась заданному уровню значимости α. При этом наибольшая мощность критерия достигается тогда, когда критическая область состоит из двух интервалов, вероятность попадания в каждый из которых равна $\frac{\alpha}{2}$. Поскольку U симметрична относительно оси Оy, вероятность ее попадания в интервалы (-∞; 0) и (0; +∞) равна 0,5, следовательно, критическая область тоже должна

быть симметрична относительно Оy. Поэтому $u_{кр}$ определяется по таблице значений функции Лапласа из условия $\Phi(u_{кр}) = \frac{1-\alpha}{2}$, а критическая область имеет вид $(-\infty; -u_{кр}) \cup (u_{кр}; +\infty)$.

Замечание. Предполагается, что используется таблица значений функции Лапласа, заданной в виде $\Phi(x) = \int_0^x e^{-\frac{t^2}{2}} dt$, где нижний предел интегрирования равен 0, а не -∞. Функция Лапласа, заданная таким образом, является нечетной, а ее значения на 0,5 меньше, чем значения стандартной функции $\Phi(x)$.

Далее нужно вычислить наблюдаемое значение критерия:

$$U_{набл} = \frac{\left(\frac{m}{n} - p_0\right)\sqrt{n}}{\sqrt{p_0 q_0}} \quad . \qquad (1.19)$$

Если $|U_{набл}| < u_{кр}$, то нулевая гипотеза принимается.

Если $|U_{набл}| > u_{кр}$, то нулевая гипотеза отвергается.

2) Если конкурирующая гипотеза H_1: $p > p_0$, то критическая область определяется неравенством $U > u_{кр}$, то есть является правосторонней, причем $p(U > u_{кр}) = \alpha$. Тогда $p(0 < U < u_{кр}) = \frac{1}{2} - \alpha = \frac{1-2\alpha}{2}$. Следовательно, $u_{кр}$ можно найти по таблице значений функции Лапласа из условия, что $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$.

Вычислим наблюдаемое значение критерия по формуле (1.19).

Если $U_{набл} < u_{кр}$, то нулевая гипотеза принимается.

Если $U_{набл} > u_{кр}$, то нулевая гипотеза отвергается.

3) Для конкурирующей гипотезы H_1: $p < p_0$ критическая область является левосторонней и задается неравенством $U <- u_{кр}$, где $u_{кр}$ вычисляется так же, как в предыдущем случае.

Если $U_{набл} > - u_{кр}$, то нулевая гипотеза принимается.

Если $U_{набл} < - u_{кр}$, то нулевая гипотеза отвергается.

Пример. Пусть проведено 50 независимых испытаний, и относительная частота появления события A оказалась равной 0,12. Проверим при уровне значимости α = 0,01 нулевую гипотезу H_0: p = 0,1 при конкурирующей гипотезе H_1: $p > 0{,}1$. Найдем $U_{набл} = \dfrac{(0{,}12 - 0{,}1)\sqrt{50}}{\sqrt{0{,}1 \cdot 0{,}9}} = 0{,}471.$ Критическая область является правосторонней, а $u_{кр}$ находим из равенства $\Phi(u_{кр}) = \dfrac{1 - 2 \cdot 0{,}01}{2} = 0{,}49.$ Из таблицы значений функции Лапласа определяем $u_{кр}$ = 2,33. Итак, $U_{набл} < u_{кр}$, и гипотеза о том, что p = 0,1, принимается.

8. Критерий для проверки гипотезы о математическом ожидании

Пусть генеральная совокупность X имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу a_0. Рассмотрим две возможности.

1) Известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема n найдем выборочное среднее $\bar{x}_в$ и проверим нулевую гипотезу H_0: $M(X) = a_0$.

Учитывая, что выборочное среднее $\bar{X}$ является несмещенной оценкой $M(X)$, то есть $M(\bar{X}) = M(X)$, можно записать нулевую гипотезу так: $M(\bar{X}) = a_0$. Для ее проверки выберем критерий

$$U = \frac{\bar{X} - a_0}{\sigma(\bar{X})} = \frac{(\bar{X} - a_0)\sqrt{n}}{\sigma}. \qquad (1.20)$$

Это случайная величина, имеющая нормальное распределение, причем, если нулевая гипотеза справедлива, то $M(U) = 0$, $\sigma(U) = 1$.

Выберем критическую область в зависимости от вида конкурирующей гипотезы:

- если H_1: $M(\bar{X}) \neq a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \dfrac{1-\alpha}{2}$, критическая область двусторонняя, $U_{набл} = \dfrac{(\bar{x} - a_0)\sqrt{n}}{\sigma}$, и, если $|U_{набл}| < u_{кр}$, то нулевая гипотеза принимается; если $|U_{набл}| > u_{кр}$, то нулевая гипотеза отвергается.

- если H_1: $M(\overline{X}) > a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$, критическая область правосторонняя, и, если $U_{набл} < u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} > u_{кр}$, то нулевая гипотеза отвергается.

- если H_1: $M(\overline{X}) < a_0$, то $u_{кр}$: $\Phi(u_{кр}) = \frac{1-2\alpha}{2}$, критическая область левосторонняя, и, если $U_{набл} > - u_{кр}$, то нулевая гипотеза принимается; если $U_{набл} < - u_{кр}$, то нулевая гипотеза отвергается.

2) Дисперсия генеральной совокупности неизвестна.

В этом случае выберем в качестве критерия случайную величину

$$T = \frac{(\overline{X} - a_0)\sqrt{n}}{S}, \tag{1.21}$$

где S – исправленное среднее квадратическое отклонение. Такая случайная величина имеет распределение Стьюдента с $k = n - 1$ степенями свободы. Рассмотрим те же, что и в предыдущем случае, конкурирующие гипотезы и соответствующие им критические области. Предварительно вычислим наблюдаемое значение критерия:

$$T_{набл} = \frac{(\bar{x}_B - a_0)\sqrt{n}}{S}. \tag{1.22}$$

- если H_1: $M(\overline{X}) \neq a_0$, то критическая точка $t_{двуст.кр.}$ находится по таблице критических точек распределения Стьюдента по известным α и $k = n - 1$.

Если $| T_{набл} | < t_{двуст.кр.}$, то нулевая гипотеза принимается.

Если $| T_{набл} | > t_{двуст.кр.}$, то нулевая гипотеза отвергается.

- если H_1: $M(\overline{X}) > a_0$, то по соответствующей таблице находят $t_{правост.кр.}(\alpha, k)$ – критичес-кую точку правосторонней критической области. Нулевая гипотеза принимается, если

$T_{набл} < t_{правост.кр.}$.

- при конкурирующей гипотезе H_1: $M(\overline{X}) < a_0$ критическая область является левосторон-ней, и нулевая гипотеза принимается при условии $T_{набл} > - t_{правост.кр.}$. Если $T_{набл} < - t_{правост.кр.}$., нулевую гипотезу отвергают.

9. Критерий для проверки гипотезы о сравнении двух дисперсий

Пусть имеются две нормально распределенные генеральные совокупности *X* и *Y*. Из них извлечены независимые выборки объемов соответственно n_1 и n_2, по которым вычислены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется при заданном уровне значимости α проверить нулевую гипотезу H_0: $D(X) = D(Y)$ о равенстве дисперсий рассматриваемых генеральных совокупностей. Учитывая несмещенность исправленных выборочных дисперсий, можно записать нулевую гипотезу так:

$$H_0: M(s_X^2) = M(s_Y^2). \qquad (1.23)$$

Замечание. Конечно, исправленные дисперсии, вычисленные по выборкам, обычно оказываются различными. При проверке гипотезы выясняется, является ли это различие незначимым и обусловленным случайными причинами (в случае принятия нулевой гипотезы) или оно является следствием того, что сами генеральные дисперсии различны.

В качестве критерия примем случайную величину

$$F = \frac{S_\sigma^2}{S_M^2} \text{ -} \qquad (1.24)$$

- отношение большей выборочной дисперсии к меньшей. Она имеет распределение Фишера-Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объем выборки, по которой вычислена большая исправленная дисперсия, а n_2 – объем второй выборки. Рассмотрим два вида конкурирующих гипотез:

- пусть H_1: $D(X) > D(Y)$. Наблюдаемым значением критерия будет отношение большей из исправленных дисперсий к меньшей: $F_{набл} = \frac{s_\sigma^2}{s_M^2}$. По таблице критических точек распределения Фишера-Снедекора можно найти критическую точку $F_{набл}(\alpha;\ k_1;\ k_2)$. При $F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

- если H_1: $D(X) \neq D(Y)$, то критическая область является двусторонней и определяется неравенствами $F < F_1$, $F > F_2$, где $p(F < F_1) = p(F > F_2) = \alpha/2$.

При этом достаточно найти правую критическую точку $F_2 = F_{кр}\left(\frac{\alpha}{2}, k_1, k_2\right)$.

Тогда при $F_{набл} < F_{кр}$ нулевая гипотеза принимается, при $F_{набл} > F_{кр}$ отвергается.

Глава 2. Стохастические модели дефицита информации о выборе весовых коэффициентов в сводном показателе

2.1. Синтез аддитивного показателя при рандомизированных весовых коэффициентах

Пусть качество сложного объекта описывается вектором $q=(q_1,\ldots,q_m)$ нормированных отдельщик показателей, каждый из которых принимает значения из промежутка [0,1]. При этом значение $q_i=0$ соответствует наихудшей оценке качества сложного объекта , а значение $q_i=1$ - наилучшей оценке качества объекта по I - мю критерию. Для характеризации качества объектов в целом используется сводный показатель $Q=Q(q)$, обладающий свойство[3,4,5,6,7,8]

$$(\forall\, i,\ q_i \geq q_i') \rightarrow (Q(q) \geq Q(q')), \qquad (2.1)$$

где $q=(q_1,\ldots,q_m)$, $q'=(q'_1,\ldots,\ q'_m)$. Дополнительно положим , что наихудшее значение сводного показателей $Q=0$, а наилучшее - $Q=1$ (см., например , работы [8,10]). При этом должно выполняться условие

$$Q(0,\ldots,0)=0\ ,\ \ Q(1,..,1)=1. \qquad (2.2)$$

Наибольшее распространение получил простейший сводный показатель вида

$$Q(q)=Q(q;p)=\sum_{i=1}^{m} q_I\ p_i, \qquad (2.3)$$

где $P=(p_1,\ldots,p_m)$есть вектор весовых коэффициентов $p_1,\ldots,p_m$ $p_i \geq 0, i=1\ ,\ \ldots,\ m,$ $p_1+\ldots.+p_m=1$ каждый из которых указывает значимость соответствующего отдельного показателя, степень его влияния на сводный показатель Q[12,13].

При практическом использовании сводных показателей для оценки качества сложных объектов зачастую имеет место дефицит информации , выражающийся в том , что исследователь знает только класс допустимых векторов $p=(p_{1,\ldots,}\ p_m)$ весовых коэффициентов. Естественной моделью такой неопределенности выбора весовых коэффициентов служит случайная величина р=($р_{1,\ldots,}$ $р_м$), принимающая значения из класса допустимых векторов

$p=(p_{1,\ldots,}\ p_м)$. Исходя из принципа максимальной энтропии, можно задавать m - мерную случайную величину p равномерным распределением на области допустимых наборов весовых коэффициентов[9]. При этом переход к дискретным значениям весов $p_i \in \{0, n^{-1}, \ldots, n^{-1}\ (n{-}{-}1), 1\}$ позволяет осуществить моделирование случайной величины p на ЭВМ путем перебора ее возможных значений. Однако такой подход к моделированию случайных весов $p=(p_{1,\ldots,}\ p_m)$ влечет использование переборных алгоритмов, требующих значительного расхода машинного времени. Поэтому мы попытаемся для простейшего случая сводного показателя найти явные формулы для статистических характеристик рандомизированного сводного показателя $\tilde{Q} = Q(q, \tilde{p})$.

Положим, что компоненты $\tilde{P}_1, \ldots, \tilde{P}_m$ рандомизированного вектора весов $\tilde{P} = (\tilde{P}_1, \ldots, \tilde{P}_m)$ равномерно распределены в m - мерном симплексе $S'' = \{(P_1, \ldots, P_m) ; P_i \geq 0,\ i=1, \ldots, m,\ P_1 + \ldots + P_m = 1\}$. Переменные $P_1, \ldots, P_m$, линейно зависимы, т.к. $P_1 + \ldots + P_m = 1$. Используя это равенство, перейдем к линейно независимым переменным $P_1, \ldots, P_{m-1}$, $(P_1 + \ldots + P_{m-1} \leq 1)$. Очевидно, что случайные величины $\tilde{P}_1, \ldots, \tilde{P}_{m-1}$ равномерно распределены в (m -1) - мерном симплексе $S = \{(P_1, \ldots, P_{m-1}) ; P_i \geq 0,\ i=1, \ldots, m-1,\ P_1 + \ldots + P_{m-1} \leq 1\}$. Единичный симплекс S расположен в *(m -1)* - мерном пространстве и имеет объем $V(S)=[(m-1)!]$.

Иными словами, вектор $\tilde{P} = (\tilde{P}_1, \ldots, \tilde{P}_m)$, равномерно распределенный на симплексе S', можно представить при помощи *(m -1)* - мерном случайной величины $(\tilde{P}_1, \ldots, \tilde{P}_{m-1})$, равномерно распределенной на единичном симплексе S, причем последняя компонента P_m определена равенством $\tilde{P} = 1 - \tilde{P}_1 - \ldots - \tilde{P}_{m-1}$

Итак, компоненты $\tilde{P}_1, \ldots, \tilde{P}_{m-1}$ рандомизированного вектора весов $\tilde{P} = (\tilde{P}_{1,\ldots,}\ \tilde{P}_m)$ равномерно распределены в симплексе $S = \{(P_1, \ldots, P_{m-1}) ; P_i \geq 0,\ i=1, \ldots, m-1,\ P_1 + \ldots + P_{m-1} \leq 1\}$ и имеют совместную плотность распределения

$$F_{\tilde{P}}(P_1,\ldots,P_{m-1}) = \begin{cases} (m-1)!, & (P_1,\ldots,P_{m-1}) \in S \\ 0, & (P_1,\ldots,P_{m-1}) \in S \end{cases} \qquad (2.4)$$

Заметим, что равномерная плотность $F_{\tilde{P}}(P_1,\ldots,P_{m-1})$ есть частный случай (при $V_1=V_2=\ldots=V_m=1$) плотность распределения Дирихле $\tilde{D}(V_1,V_2,\ldots,V_m)$

$$F_{\tilde{P}}(P_1,\ldots,P_{m-1}) = \begin{cases} \left(\left(\Gamma\left(\sum_{i=1}^{m} V_i\right)\right)/\left(\prod_{i=1}^{m}\Gamma(V_i)\right)\right)\left(1-\sum_{i=1}^{m} P_i\right)^{V_m-1} & (P_1,\ldots,P_{m-1}) \in S \\ 0, & (P_1,\ldots,P_{m-1}) \in S \end{cases} \qquad (2.5)$$

Маргинальное распределение $(P_1,\ldots,P_k)$, $k \le m-1$, случайной величины с распределением Дирихле, имеет распределения Дирихле вида $D(V_1,\ldots,V_k,V_{k+1}+\ldots+V_m)$. Поэтому отдельный рандомизированный все P_i имеет плотность распределения:

$$f_{p_{\tilde{i}}}(P) = (m-1)(1-p)^{m-2}, 0 \le x \le 1 \qquad (2.6)$$

представляющую собой частный вид плотность случайной величины $\tilde{\beta}$ (V_1,V_2), имеющей бета-распределение с плотностью

$$f_{\tilde{\beta}}(x,v_1,v_2) = ((\Gamma(V_1+V_2))/(\Gamma(V_1)\Gamma(V_2)))\, x^{v_1-1}(1-x)^{v_2-1},\ 0 \le x \le 1 \quad (2.7)$$

т.е. $\tilde{P}_i = \tilde{\beta}(1, m-1)$.

Следовательно, статистические характеристики случайных величин $\tilde{P}_1,\ldots,\tilde{P}_{m-1}$ могут быть легко найдены как частный случай соответствующих характеристик компонент $(m-1)$ - мерном случайной величины $\tilde{\delta} = (\tilde{\delta}_1,\ldots,\tilde{\delta}_{m-1})$ распределенной по закону Дирихле. Из определения среднего значения и дисперсии имеем

$$\mathrm{M}\tilde{P}_i = \int_0^1 p f_{\tilde{p}_i}(p)\,dp = (m-1)\int_0^1 p(1-p)^{m-2}\,dp = 1/m \qquad (2.8)$$

$$\mathrm{M}\,\tilde{P}_i^2=\int_0^1 p^2 f_{\tilde{p}_i}(p)dp=(m-1)\int_0^1 p^2(1-p)^{m-2}dp=2/m(m+1) \qquad (2.9)$$

$$\mathrm{D}\,\tilde{P}_i=\int_0^1 (p-1/m)^2 f_{\tilde{p}_i}(p)dp=M\tilde{P}^2{}_i-(M\tilde{P}_i)^2=(m-1)/m(m+1) \qquad (2.10)$$

Двумерная случайная величина $(\tilde{p}_i,\tilde{p}_j)$, i≠j, I,j=1,…,m-1, имеет плотность распределения

$$f_{ij}(p_i,p_j)=m(m-1)(1-p_i-p_j)^{m-3},\ p_i\neq p_j\leq 1, \qquad (2.11)$$

знание которой позволяет найти начальный и центральный смешанные моменты

$$\tilde{P}_i\tilde{P}_j=\iint\limits_{p_i+p\leq 1_j} P_iP_j f_{ij}(P_iP_j)dp_idp_j=m(m-1)\int_0^{1-p_j}\left[\int_0^1 p_ip_j(1-p_i-p_j)^{m-3}dp_i\ \right]dp_j=1/m(m+1), i\neq j, \qquad (2.12)$$

$$\mathrm{Cov}(\tilde{p}_i,\tilde{p}_j)=(\mathrm{cov}(p_i,p_j)/(\sqrt{D\tilde{P}_i}\sqrt{D\tilde{P}_j}))=-1/(m-1) \qquad (2.13)$$

Определив последнюю координату P_m случайного вектора $\tilde{P}=(\tilde{P}_1,\ldots,\tilde{P}_m)$ формулой $\tilde{p}_m=1-\tilde{p}_1-\ldots-p_{m-1}$ из (2.8)-(2.13) получим формулы

$$M\tilde{P}_m=1-\sum_{i=1}^{m-1}M\tilde{P}_i=1/m, \qquad (2.14)$$

$$M\tilde{P}_m{}^2=1+M(\sum_{i=1}^{m-1}\tilde{p}_i)^2-2\sum_{i=1}^{m-1}M\tilde{P}_i=1/m(m+1), \qquad (2.15)$$

$$D\tilde{P}_m=D(\sum_{i=1}^{m-1}\tilde{p}_i)=(m-1)/m^2(m+1) \qquad (2.16)$$

$$cov(\tilde{p}_i\tilde{p}_m)=1/m^2(m+1),\ \ i\neq m, \qquad (2.16)$$

$$M\tilde{P}_i,\tilde{p}_m=1/m(m+1), i\neq m, \qquad (2.17)$$

$$corr(\tilde{p}_i,\tilde{p}_m)=-1/(m-1), i\neq j, \qquad (2.18)$$

из которых следует, что формулы (1.1.8)-(1.1.14) имеют место для всех *I=1,…,m* .

Подставляя рандомизирование весовые коэффициенты $(\tilde{P}_1,\ldots,\tilde{P}_{m-1})$ в формулу (1.1.3), мы сводим задачу сравнения двух объектов с векторами отдельных показателей $q=(q_1,\dots,q_m)$, $q'=(q'_1,\dots,q'_m)$ к задаче сравнения случайных величин

$$\tilde{Q}=\tilde{Q}(q)=\sum_{i=1}^{m} q_i\tilde{p}_i,\quad \tilde{Q}'=\tilde{Q}(q')=\sum_{i=1}^{m} q'_i\tilde{p}_i, \tag{2.19}$$

Знание статистических характеристик случайных величин $\tilde{P}_1,\dots,\tilde{P}_m$ позволяет найти аналогичные характеристики рандомизированного сводного показателя Q :

$$M\tilde{Q}=M(\sum_{i=1}^{m} q_i\tilde{p}_i,)=\sum_{i=1}^{m} q_i M\tilde{p}_i=\sum_{i=1}^{m}\tilde{q}_i, \tag{2.20}$$

$$M\tilde{Q}^2=\sum_{i=1}^{m} q_i{}^2 M\tilde{p}_i{}^2+\sum_{i=1}^{m} q_i q_j M\tilde{P}_i\tilde{P}_j=(1/m(m+1))(2\sum_{i=1}^{m} q_i{}^2+\sum_{i=1}^{m} q_i q_j), \tag{2.21}$$

$$D\tilde{Q}=\sum_{i=1}^{m} q_i{}^2 D\tilde{p}_i+\sum_{i=1}^{m} q_i q_j \operatorname{cov}(\tilde{p}_i,\tilde{p}_j)=(1/m^2(m+1))((m-1)\sum_{i=1}^{m} q^2{}_i-\sum_{i=1}^{m} q_i q_j) \tag{2.22}$$

$$M\tilde{Q}\tilde{Q}'=\sum_{i=1}^{m} q_i q'_i M\tilde{p}_i{}^2+\sum_{i=1}^{m} q_i q' M\tilde{P}_i M\tilde{P}_j=(1/m(m+1))(2\sum_{i=1}^{m} q_i q_i{}'+\sum_{i,j=1}^{m} q_i q'_j), \tag{2.23}$$

$$\operatorname{Cov}(\tilde{Q},\tilde{Q}')=\sum_{i=1}^{m} q_i q_i{}' D\tilde{P}_i' +\sum_{i=1}^{m} q_i q'_j \operatorname{cov}(\tilde{p}_i,P_j)=(1/m^2(m+1))((m-1)\sum_{i=1}^{m} q_i q'_i-\sum_{i=1}^{m} q_i q'_j), \tag{2.24}$$

$$i\neq j$$

$$\operatorname{Corr}(\tilde{Q},\tilde{Q}')=$$

$$\operatorname{Corr}(\tilde{Q},\tilde{Q}')=\operatorname{cov}(\tilde{Q},\tilde{Q}')/\sqrt{D\tilde{Q}}\sqrt{D\tilde{Q}'}=((m-1)\sum q_i q_i{}'-\sum q_i q_j{}')/((m-1)\sum q^2{}_i+$$
$$+\sum q_i q_j)((m-1)\sum q^2{}_i+\sum q'_i q'_j))^{1/2} \tag{2.25}$$

Теперь, характеризуя общую оценку качества объектов $q=(q_1,\dots,q_m)$, $q'=(q'_1,\dots,q'_m)$ усредненными сводными показателями $\overline{Q}=M\tilde{Q}(q)$, $\overline{Q}=M\tilde{Q}(q')$, мы можем оценить достоверность доминирования объекта

$q' = (q'_1, ..., q'_m)$ над объектом q=(q1,…,qm), вероятностью $P\{\tilde{Q} \le \tilde{Q}'\}$ доминирования случайной величины Q` над случайной величиной Q`.

Вычислим вероятность $P\{\tilde{Q} \le \tilde{Q}'\}$. Преобразуем неравенство $\tilde{Q} \le \tilde{Q}'$ в эквивалентное неравенство

$$(q_m - q'_m) + \sum \tilde{P}_i (q_i - q'_i - q_m + q'_m) \le 0, \quad (2.28)$$

содержащее только случайные величины $\tilde{P}_1, \ldots, \tilde{P}_{m-1}$ равномерно распределенные в симплексе $S \le E^{m-1}$. Тогда вероятность события $\{\tilde{Q} \le \tilde{Q}'\}$ можно вычислить по формуле

$$P\{\tilde{Q} \le \tilde{Q}'\} = \int \ldots \int f_{p_i}(P_1, ..., P_{m-1}) dp_1 ... dp_{m-1}, \quad (2.29)$$

где $H=\{(p_1, \ldots, p_{m-1}): \alpha_0+\alpha_1 p_1+\ldots+\alpha_{m-1} p_{m-1} \le 0\}$, $\alpha_0=q_m-q_m^{`}$, $\alpha_i=q_i-q_i^{'}-q_m+q_m^{'}$. Для того чтобы из бежать интегрирования по сложной области $S \cap H \le E^{m-1}$, представляющей собой перечение симплекса S с полупространством H, воспользуемся известным представлением

$$\tilde{P}_i = \tilde{x}_i \left(\sum_{i=1}^{m} \tilde{x}_j\right)^{-1}, i = 1, ..., m-1 \quad (2.30)$$

компонент вектора ($\tilde{P}_1, \ldots, \tilde{P}_{m-1}$), равномерно распределенного на симплексе S, через независимые случайные величины $\tilde{x}_1, \ldots, \tilde{x}_{m-1}$, каждая из которых имеет экспоненциальную плотность распределения вида

$$f_i(x_i) = e^{-x_i}, \quad x_i \in [0,\infty), \quad i=1,\ldots,m \quad (2.31)$$

Подставим (1.1.30) в (1.1.28) и получим

$$\tilde{Q}(q) = \sum_{i=1}^{m} q_i \tilde{p}_i = q_m + \sum_{i=1}^{m} (q_i - q_m) \tilde{p}_i = \sum_{i=1}^{m} q_i \tilde{x}_i / \sum_{i=1}^{m} \tilde{x}_i \quad (2.32)$$

Т.е. неравенство $\{\tilde{Q} \le \tilde{Q}'\}$ принимает вид

$$\sum_{i=1}^{m} q_i \tilde{x}_i \le \sum_{i=1}^{m} q'_i \tilde{x}_i \quad (2.33)$$

или

$$\sum_{i=1}^{m}(q_i - q'_i)\tilde{x}_i \le 0 \qquad (2.34)$$

Перенумеруем q_i и $q_i^{`}$ так, чтобы были выполнены неравенства q_i- $q_i{}'$<…< q_k- $q_k{}'$<0< q_{k+1}- $q_{k+1}{}'$<…< q_m- $q_m{}'$.

Тогда неравенства (1.1.34) имеет вид

$$\sum_{j=k+1}^{m}(q_j - q'_j)\tilde{x}_j \le \sum(q'_i - q_i)\tilde{x}_i \qquad (2.35)$$

Отсюда, обозначив $\alpha_i=q'_i - q_i>0$, $i=1, ...,k$, $\alpha_j= q_j - q_j{}'>0$ $j=k+1,...,m$ получим неравенство

$$\sum_{j=k+1}^{m}\alpha_j\tilde{x}_j \le \sum_{i=1}^{k}\alpha_i\tilde{x}_i \qquad (2.36)$$

Так как

$$\alpha_i\tilde{x}_i = \tilde{t}_i, i = 1,...,k, \alpha_j\tilde{x}_j = \tilde{t}_j, j = k+1,...,m$$

и случайные величины $\tilde{x}_i$ распределены по экспоненциальному закону (1.1.31), то случайные величины распределены с плотностью

$$f_i(t_i) = (1/\alpha_i)\ell^{-(t_i/\alpha_i)}, \alpha_i > 0, t_i > 0 \qquad (2.37)$$

Пусть независимые случайные величины $z_1,\dots,z_n$ распределены по экспоненциальному закону с плотностями

$$f_i(z_i) = \alpha_i\ell^{-\lambda_i z_i}, z_i > 0, \lambda_i > 0 \qquad (2.38)$$

Тогда сумма случайных величин имеет обобщенное распределение Эрланга n - го порядка

$$f_i(z_i) = (-1)^{n-1}\prod_{i=1}^{n}\lambda_i\sum_{i=1}^{n}(\ell^{-\lambda_i z} / \prod_{t=1}^{n}(\lambda_i - \lambda_t)), z > 0, i \ne t, \qquad (2.39)$$

Лемма 1.1.1. Пусть

$$\tilde{\xi}_1,...,\tilde{\xi}_r \qquad \tilde{\eta}_1,...,\tilde{\eta}_s$$

независимые случайные величины, распределенные по экспоненциальному закону с параметрами

$$\lambda_1,...,\lambda_r \qquad \mu_1,...,\mu_s$$

соответственно. Тогда вероятность стохастического неравенства $\tilde{\xi}\langle\tilde{\eta}$, $\tilde{\xi}=\tilde{\xi}_1+...+\tilde{\xi}_r$, $\tilde{\eta}=\tilde{\eta}_1+...+\tilde{\eta}_s$ выражается формулой

$$\mathrm{P}\{\tilde{\xi}<\tilde{\eta}\}=(-1)^{r+s-2}\prod_{j=1}^{r}\lambda_j\prod_{i=1}^{s}\mu_i\sum_{j=k+1,i=1}^{r,s}(1/\mu_i(\lambda_j+\mu_i))\prod_{t\neq j,t=1}^{r}(1/\lambda_j-\lambda_t)\prod_{l\neq i,l=1}^{s}(1/\mu_i-\mu_l) \quad (2.40)$$

ДОКАЗАТЕЛЬСТВО. Пусть $\tilde{\xi}=\tilde{\xi}_1+...+\tilde{\xi}_r$. По формуле (2.39) находим плотность распределения случайной величины ξ:

$$f_\xi(\xi)=(-1)^{r-1}\prod_{i=1}^{r}\lambda i\sum_{i=1}^{r}(\ell^{-\lambda_i\xi}/\prod_{t\neq i,t=1}^{s}(\lambda_j-\lambda_t)),\xi>0 \quad (2.41)$$

Аналогичное представление и для случайной величины $\tilde{\eta}=\tilde{\eta}_1+...+\tilde{\eta}_s$:

$$f_\eta(\eta)=(-1)^{s-1}\prod_{j=1}^{s}\lambda j\sum_{j=1}^{s}(\ell^{-\mu_i r}/\prod_{t\neq j,t=1}^{s}(\lambda_j-\lambda_s)),\eta>0 \quad (2.42)$$

$\tilde{\xi},\tilde{\eta}$ - независимые случайные величины, то совместный закон распределения разности $\tilde{y}=\tilde{\xi}-\tilde{\eta}$ этих случайных величин находим по формуле

$$g_{\tilde{y}}(y)=\int_{-\infty}^{\infty}f_{\tilde{\xi}}(\xi)f_{\tilde{\eta}}(\xi-y)d\xi=\int_{-\infty}^{\infty}f_{\tilde{\xi}}(\eta-y)f_{\tilde{\eta}}(\eta)d\eta \quad (2.43)$$

По формуле (1.1.43) находим совместную плотность распределения разности двух независимых случайных величин $\tilde{\xi},\tilde{\eta}$, распределенных по показательным законам с параметрами $\lambda_1,...,\lambda_r$, $\mu_1,\ldots,\mu_s$ соответственно:

$$g_{\tilde{y}}(y)=\int_{0}^{\infty}(-1)^{r-1}\prod_{i=1}^{r}\lambda i\sum_{i=1}^{r}(\ell^{-\lambda_i\xi}/\prod_{t\neq i,t=1}^{s}(\lambda_j-\lambda_t)) *$$

$$*\prod_{j=1}^{s}\mu_j\sum_{j=1}^{s}(\ell^{-\mu_i(\xi-y)}/\prod_{t\neq j,t=1}^{s}(\mu_j-\mu_t))d\xi \quad (2.44)$$

Интегрируя, получаем

$$g_{\tilde{y}}(y) = (-1)^{r+s-2}\prod_{i=1}^{r}\lambda_i\prod_{j=1}^{s}\mu_j\sum_{i,j=1}^{r,s}\prod_{t\neq i,t=1}^{r}(\lambda_i-\lambda_t)\prod_{t=j,t=1}^{s}(\mu_j-\mu_t)^{-1}(\ell^{\mu_i y}/(\lambda_i+\mu_j)) \quad (2.45)$$

Вероятность неравенства $\tilde{\xi}\langle\tilde{\eta}$ определяется формулой

$$\mathrm{P}\{\tilde{\xi}<\tilde{\eta}\} = F_{\tilde{y}}(y) = \int_{-\infty}^{0} g_{\tilde{y}}(y)dy \qquad (2.46)$$

Подставляя (2.45) в (2.46), находим формулы (2.40).

Лемма доказана.

Теперь, используя вышепоказанную лемму, найдём вероятность события { $\tilde{Q}\leq\tilde{Q}'$ }.

Введем обозначения

$$r=m\text{-}k\,,\ \lambda_j = 1/(q_j - q'_j), s = k, \mu_i = 1/(q'_i - q_i), \quad (2.47)$$

где $j=k+1,\dots,m$, $i=1,\dots,k$.

Подставляя (2.47) в (2.40), находим вероятность доминирования объекта $\tilde{Q}(q')$ над объектом $\tilde{Q}(q)$

$$P\{\tilde{Q}\leq\tilde{Q}'\} = (-1)^{m-2}\prod_{j=k+1}^{m}(1/(q-_j\, q_j{}'))\prod_{i=1}^{k}1/(q'_i-q_i)\sum_{j=k+1,i=1}^{m,k}[((q'_i-q_i{}^2)(q_j-q'_j)/(q'_i-q_i-q_j+q_j))$$

$$\times$$

$$\prod_{t\neq j,t=k+1}^{m}((q_j-q'_j)(q_t-q'_t)/(q_t-q'_t-q_j+q'_j))\prod_{l\neq j,l=1}^{k}((q'_i-q_i)(q'_l-q_l)/(q'_l-q_l-q'_i+q_i))]$$

Если в неравенстве (1.1.38) выполняются следующие условия

$$q_1-q'_1=\dots=q_{l_1}-q'_{l_1}, q_{l_1+1}-q'_{l_1+1}=\dots=q_{l_2}-q'_{l_l},\dots, q_{l_{r-1}+1}-q'_{l_{r-1}+1}=\dots=q_{l_l}-q'_{l_r},$$

и $q_{l_1}-q'_{l_1}<\dots<q_{l_k}-q'_{l_k}<0<q_{l_{k+1}}-q'_{l_{k+1}}<\dots<q_{l_r}-q'_{l_r}$ где 1<ĩ<m ,

1≤k≤ĩ, то получаем:

$$\alpha_{l_i} = q'_{l_i} - q_{l_i} > 0, \mu_i = 1/(q'_{l_i} - q_{l_i}), \quad r=r\text{-}k,\ \ s=k. \qquad (2.49)$$

Согласно (1.1.35) из (1.1.40) находим вероятность события

$$P\{\tilde{Q}\leq\tilde{Q}'\} = (-1)^{r-2}\prod_{j=k+1}^{r}(1/(q_{l_j} -\ q_{l_j}{}'))\prod_{i=1}^{k}1/(q'_{l_i}-q_{l_i})\sum_{j=k+1,i=1}^{r,k}[((q'_{l_i}-q_{l_i}{}^2)(q_{l_j}-q'_{l_{jt}})/(q'_{l_i}-q_{l_i}-q_{l_j}+q_{l_j}))$$

$$\times$$

$$\times \prod_{t \neq j, t=k+1}^{m} ((q_{l_j} - q'_{l_j})(q_{l_t} - q'_{t})/(q_{l_t} - q'_{l_t} - q_{l_j} + q'_{l_j})) \prod_{l \neq v, v=1}^{k} ((q'_{l_v} - q_{l_v})(q'_{l_i} - q_{l_i})/(q'_{l_v} - q_{l_v} - q'_{l_i} + q_{l_i}))] \tag{2.50}$$

Таким образом, в условиях дефицита информации о весовых коэффициентах $\tilde{P}$ 1,… ,$\tilde{P}$ m их моделирование при помощи случайного вектора $\tilde{P}=(\tilde{P}_1,\ldots,\tilde{P}_m)$ позволяет оценить средние значения $\overline{Q} = M\tilde{Q}$, $\overline{Q} = M\tilde{Q}'$ рандомизированных сводных показателей по формуле (2.22), оценить их разброс при помощи дисперсий $D\tilde{Q}$, $D\tilde{Q}'$, вычисляемых по формуле (2.24), и определить по формуле (2.48) и (2.50) вероятность $P\{\tilde{Q} \le \tilde{Q}'\}$ доминирования объекта с вектором отдельных показателей $q' = (q'_1,...,q'_m)$ над объектом с вектором отдельных показателей $q=(q_1,\ldots,q_m)$.

2.2. Методы разработка программных обеспечение

Для написания программы был выбран язык программирования C++. Он является полностью объектно-ориентированным, мощным языком программирования, поддерживающим практически все виды абстракций объектно-ориентированного программирования. Была использована среда программирования Borland C++ 6.0. Эта среда является визуальной средой программирования, позволяющей перетаскивать стандартные компоненты непосредственно в окно будущей программы[11,13].

Структурно программа представляет собой набор классов, часть из которых была создана разработчиком, а часть – самой средой программирования. Классы созданные средой программирования унаследованы от классов библиотеки VCL. Экземпляры одних классов создаются другими классами, и там происходит работа с ними. Среда программирования сама генерирует код главной функции и программисту он не доступен. Программист может работать только либо со своими классами, либо с классами, описывающими интерфейс программы и ее свойства. Также в исходниках программы присутствуют ресурсы, которые представляют из себя внешний объекты, подключаемые к программе во время связывания. Этими ресурсами являются диалоговые окна, картинки, иконки и специальные таблицы, определяющие системные и интерфейсные свойства программы.

Составим программу стохастического модели дефицита информации о выборе весовых коэффициентов в сводном показателе.

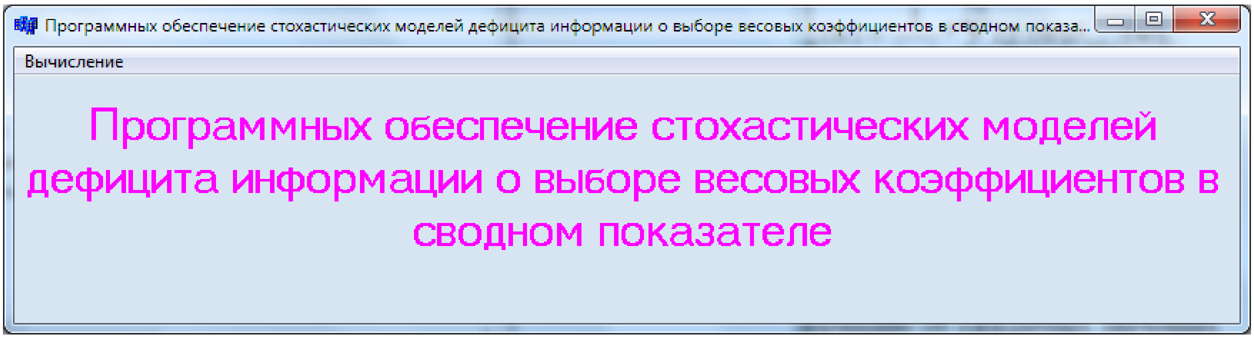

Рисунок 1. Главная форма.

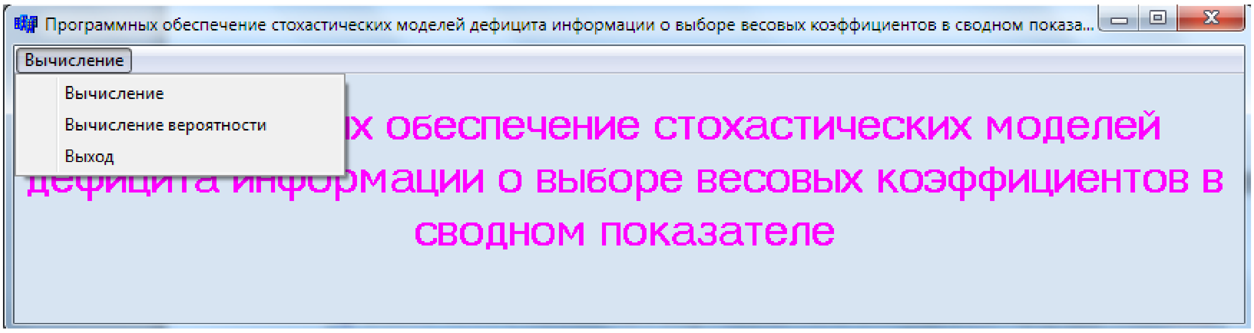

Рисунок 2. Главной меню программу.

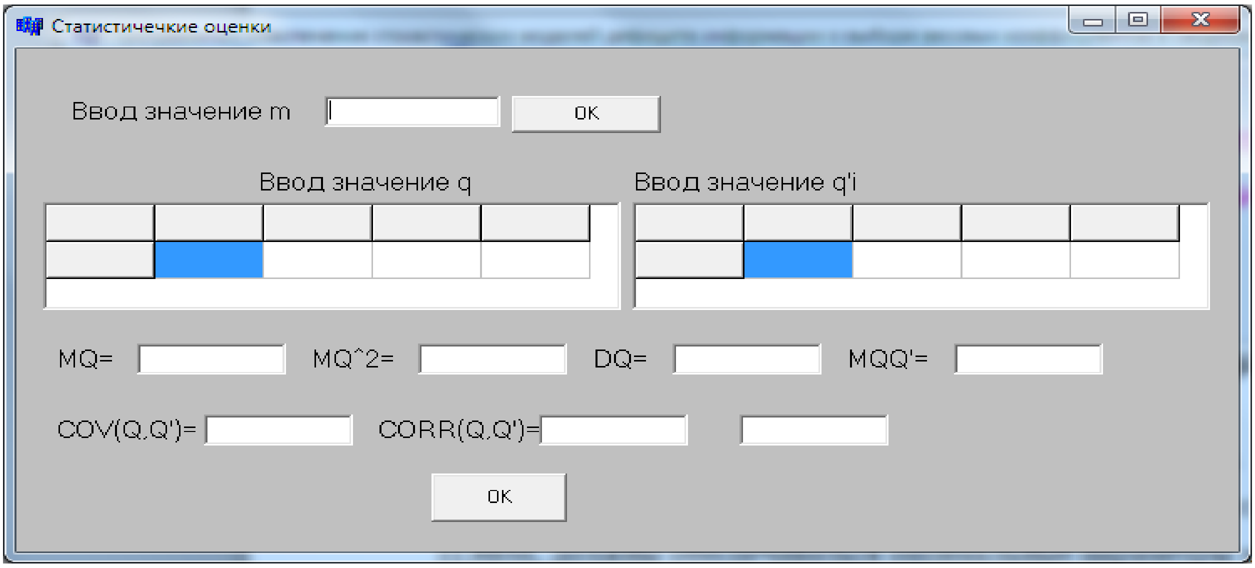

Рисунок 3. Форма для вычисления статические оценки сложных объектов.

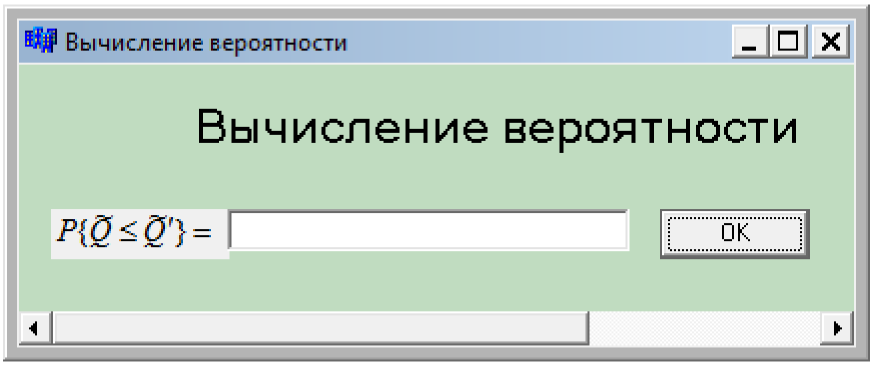

Рисунок 4. Форма для расчета вероятности доминируюшых объектов.

2.3. Требования к микроклимату и освещенности на рабочих местах, оборудованных ПЭВМ

В помещениях всех типов образовательных и культурно-развлекательных учреждений для детей и подростков, где расположены ПЭВМ, должны обеспечиваться оптимальные параметры микроклимата[17,18].

Оптимальные параметры микроклимата во всех типах учебных и дошкольных помещений с использованием ПЭВМ.

Таблица 1.

Температура, С°	Относительная влажность, %	Абсолютная влажность, г/м3	Скорость движения воздуха, м/с
19	52	10	〈0,1
20	58	10	〈0,1
21	55	10	〈0,1

В помещениях, оборудованных ПЭВМ, проводится ежедневная влажная уборка и систематическое проветривание после каждого часа работы на ПЭВМ. Уровни положительных и отрицательных аэроионов в воздухе помещений, где расположены ПЭВМ, должны соответствовать действующим санитарно-эпидемиологическим нормативам. Содержание вредных химических веществ в воздухе помещений, предназначенных для использования ПЭВМ во всех типах образовательных учреждений, не должно превышать предельно допустимых среднесуточных концентраций для атмосферного воздуха в соответствии с действующими санитарно-эпидемиологическими нормативами.

Освещенность. Рабочие столы следует размещать таким образом, чтобы видео дисплейные терминалы были ориентированы боковой стороной к световым проемам, чтобы естественный свет падал преимущественно слева.

Освещенность на поверхности стола в зоне размещения рабочего документа должна быть 300 - 500 лк. Освещение не должно создавать бликов на поверхности экрана. Освещенность поверхности экрана не должна быть более 300 лк. Следует ограничивать прямую блескость от источников освещения, при этом яркость светящихся поверхностей (окна, светильники и др.), находящихся в поле зрения, должна быть не более 200 кд/м2[17].

Следует ограничивать отраженную блескость на рабочих поверхностях (экран, стол, клавиатура и др.) за счет правильного выбора типов светильников и расположения рабочих мест по отношению к источникам естественного и искусственного освещения, при этом яркость бликов на экране ПЭВМ не должна превышать 40 кд/м2 и яркость потолка не должна превышать 200 кд/м2.

Показатель ослепленности для источников общего искусственного освещения в учебных помещениях не более 15.

Яркость светильников общего освещения в зоне углов излучения от 50 до 90 градусов с вертикалью в продольной и поперечной плоскостях должна составлять не более 200 кд/м2, защитный угол светильников должен быть не менее 40 градусов. Светильники местного освещения должны иметь не просвечивающий отражатель с защитным углом не менее 40 градусов.

Следует ограничивать неравномерность распределения яркости в поле зрения пользователя ПЭВМ, при этом соотношение яркости между рабочими поверхностями не должно превышать 3:1 - 5:1, а между рабочими поверхностями и поверхностями стен и оборудования 10:1.

В качестве источников света при искусственном освещении следует применять преимущественно люминесцентные лампы типа ЛБ и компактные люминесцентные лампы (КЛЛ). В светильниках местного освещения допускается применение ламп накаливания, в том числе галогенные.

Применение светильников без рассеивателей и экранирующих решеток не допускается.

Общее освещение при использовании люминесцентных светильников следует выполнять в виде сплошных или прерывистых линий светильников, расположенных сбоку от рабочих мест, параллельно линии зрения пользователя при рядном расположении видео дисплейных терминалов. При периметральном расположении компьютеров линии светильников должны располагаться локализовано над рабочим столом ближе к его переднему краю, обращенному к оператору.

Коэффициент запаса (Кз) для осветительных установок общего освещения должен приниматься равным 1,4.

Коэффициент пульсации не должен превышать 5%.

Для обеспечения нормируемых значений освещенности в помещениях для использования ПЭВМ следует проводить чистку стекол оконных рам и светильников не реже двух раз в год и проводить своевременную замену перегоревших ламп.

Учебные заведения или их базовые предприятия обеспечивают строительную готовность кабинета информатики под установку и монтаж КУВТ в соответствии с гигиеническими правилами и нормами, включая работы по пожарно-охранной сигнализации, проводке силовых кабелей, окраске стен и потолков, подготовке полов с укладкой металлических защитных труб или металлорукавов, по организации заземления.

Заключение

Перечислим в тезисной форме основные результаты работы:

1. Изучены методика и алгоритмов стохастические модели дефицита информации о выборе весовых коэффициентов в сводном показателе

1. Разработана общая модель дефицита информации о точных значениях весовых коэффициентов, в рамках которой рандомизированный вектор «весов» задается распределением Дирихле на симплексах, получающихся при сечении единичного симплекса полуплоскостями, проходящими через начало координат.

2. Построен алгоритм разбиения области задания весовых коэффициентов на симплексы, позволяющий находить статистические характеристики сводных рандомизированных показателей, чьи весовые коэффициенты удовлетворяют системам линейных неравенств.

3. Разработана программных средств система комбинаторных и статистических алгоритмов генерации рандомизированных весовых коэффициентов, математического обеспечения методики оценивания в условиях дефицита информации сложных объектов.

Метод рандомизированных сводных показателей позволяет определить одну сводную оценку, синтезирующую отдельные показатели качества объекта, как линейную "свертку" многих оценок исследуемого объекта, проводимых по различным критериям. С помощью метода рандомизированных сводных показателей можно оценивать качество разного рода различных многопараметрических объектов.

Литературы

1. Постановления Президента Республики Узбекистан «О мерах по дальнейшему внедрению и развитию современных информационно-коммуникационных технологий» от 21 марта 2012 года.
2. Постановления Президента Республики Узбекистан «О мерах по дальнейшему развитию Национальной информационно-коммуникационной системы Республики Узбекистан» от 27 июня 2013 года № ПП-1989.
3. Хованов Н.В. Анализ и синтез показателей при информационном дефиците. СПб., СПбГУ, 1996.
4. Хованов Н.В. Построение сводного показателя инвестиционной привлекательности акций российских предприятий // Материалы Всероссийской научной конференции "Инвестиционная политика России в современных условиях". СПб., СПбГУ, 1997. С. 2-3.
5. Хованов Н.В. Математические модели риска и неопределенности. СПб., СПбГУ, 1998.
6. Хованов Н.В. Три типа математических моделей неопределенности // Измерительная техника. 2005. N 9. С. 39-44.
7. Москвин Б.В. Теория принятия решений: Учебник / Б.В. Москвин. – СПб.: ВКА имени А.Ф. Можайского, 2005. – 383 с.
8. Ивченко Г.И., Медведев Ю. И. Введеиие в математическую статистику: Учебник. – М.: Издательство ЛКИ, 2010. 600 с.
9. Чернова Н. И. Математическая статистика: Учеб. пособие / Новосиб. гос. ун-т. Новосибирск, 2007. 148 с.
10. Махмудов З. М. Учет ограничений при моделирование неопределенности выбора весовых коээфициентов. //Вопросы выч. и пркладной матем. -Ташкент.: Институт кибернетики,1990. Вып. 90. стр. 150-159.
11. Стараструп Б. Програмирование по С++. –М.: Радио и связ, 1991.
12. Гулямов С.С Экономика и информатика.-Ташкент.: Мехнат, 1991.

13. Информатика.Учебник.-3-перераб.изд./Под ред.проф. Н.В.Макарова.- М.: Финансы и статистика, 1999.-768с.
14. Информационные системы в экономике. Под ред. В.В.Дика - М.: Финансы и статистика, 1996.-272с.
15. Введение в информационный бизнес. Под ред. В.П.Тихомирова, А.В.Хорошилова. - М.: Финансы и статистика, 1996.
16. Назаров С.В., В.И.Першиков, В.А.Тафинцев и др. Компьютерные технологии обработки информации. Учебное пособие/; Под ред. С.В.Назарова.- М.: Финансы и статистика, 1995.-248 с.
17. Сайдаминов С. С. Основы охраны окружающей среды. Ташкент, "Уқитувчи" 1989 г.
18. Белов С. В. и др. " Безопасность жизнедеятельности", "Высшая школа", Москва 1999 год.

Web страницы.

19. http://www.uza.uz (Последние новости по развитию ИКТ в Узбекистане).
20. http://tekhnosfera.com (Диссертации в Техносфере)
21. http://www.ci.ru/inform.htm (Информационные технологии Третьего Тысячелетия).
22. http://www.scool.ua/Htm/Technologie.htm (Информационные технологии в учебном процессе).
23. http://www.iite.ru/iite/russian/ (Институт ЮНЕСКО по информационным технологиям в образовании).
24. http://www.dl.uz/ (Спецкурс «Новые информаци-онные технологии в учебном про-цессе» в подготовке студентов).

Приложение

```
#include <vcl.h>
#pragma hdrstop

#include "DiplomMSH.h"
#include "Unit2.h"
#include "Unit3.h"
//---------------------------------------------------------------------------
#pragma package(smart_init)
#pragma resource "*.dfm"
TForm1 *Form1;
//---------------------------------------------------------------------------
__fastcall TForm1::TForm1(TComponent* Owner)
        : TForm(Owner)
{
}
//---------------------------------------------------------------------------

void __fastcall TForm1::MQ1Click(TObject *Sender)
{
  Form2->Show();
}
//---------------------------------------------------------------------------

void __fastcall TForm1::Chiqish1Click(TObject *Sender)
{
Close();
}
//---------------------------------------------------------------------------
```

```
void __fastcall TForm1::Button1Click(TObject *Sender)
{
Close();
}
//---------------------------------------------------------------------------

void __fastcall TForm1::Ehtimolliknihisoblash1Click(TObject *Sender)
{
Form3->Show();
}
#include <vcl.h>
#pragma hdrstop
//---------------------------------------------------------------------------
USEFORM("DiplomMSH.cpp", Form1);
USEFORM("Unit2.cpp", Form2);
USEFORM("Unit3.cpp", Form3);
//---------------------------------------------------------------------------
WINAPI WinMain(HINSTANCE, HINSTANCE, LPSTR, int)
{
        try
        {
                Application->Initialize();
                Application->CreateForm(__classid(TForm1), &Form1);
                Application->CreateForm(__classid(TForm2), &Form2);
                Application->CreateForm(__classid(TForm3), &Form3);
                Application->Run();
        }
        catch (Exception &exception)
```

```
        {
                Application->ShowException(&exception);
        }
        catch (...)
        {
                try
                {
                        throw Exception("");
                }
                catch (Exception &exception)
                {
                        Application->ShowException(&exception);
                }
        }
        return 0;
}
#include <vcl.h>
#pragma hdrstop
#include "Unit3.h"
#include "Unit2.h"
#include <math.h>
#pragma package(smart_init)
#pragma resource "*.dfm"
TForm3 *Form3;
//---------------------------------------------------------------------------
__fastcall TForm3::TForm3(TComponent* Owner)
        : TForm(Owner)
{
}
```

```
//---------------------------------------------------------------------------
void __fastcall TForm3::Button1Click(TObject *Sender)
{  float m;
  m=StrToFloat(Form2->Edit1->Text);
  float qq1,qqk,qqm,k,q,qi[500],q1i[500];
  for (int j=1; j<=m;j++)
  {
     qi[j]=StrToFloat(Form2->Q->Cells[j][1]);
     q1i[j]=StrToFloat(Form2->Q1->Cells[j][1]);
  }
  for (int j=1; j<=m;j++)
  {
  qq1=qi[j]-q1i[j];
  if (qq1>0)
   {
   k+=1;
    }
  else
   {
    k+=0;
    }
   }
  for (int i=1; i<=k;i++)
  {
   if (qi[i]-q1i[i]!=0)
   {
   qqk*=1/(qi[i]-q1i[i]);
    }
   else
```

```
 {
  qqk=1;
  }
 }
for (int j=k+1; j<=m;j++)
{
if (q1i[j]-qi[j]!=0)
    {
     qqm*=1/(q1i[j]-qi[j]);
     }
    else
    {
    qqm=1;
    }
 }
 //---------------------------------
double qqkl[500],qqkt[500];
for (int i=1;i<=k;i++)
{
  for (int l=1; l<=k;l++)
  {
      if (i!=l)
      {
   qqkl[i]*=(q1i[i]-qi[i])*(q1i[l]-qi[l])/(q1i[l]-qi[l]-q1i[i]+qi[i]);
       }
   }
 }

for (int j=k+1;j<=m;j++)
```

```
 {
   for (int t=k+1; t<=m;t++)
   {
       if ((j!=t)&&(qi[t]-q1i[t]-qi[j]+q1i[j])!=0)
       {
 qqkt[j]*=(qi[j]-q1i[j])*(qi[t]-q1i[t])/(qi[t]-q1i[t]-qi[j]+q1i[j]);
        }
    }
  }
 double qqmk,pqq;
 for (int i=1;i<=k;i++)
  {
   for (int j=k+1;j<=StrToFloat(Form2->Edit1->Text);j++)
   {
qqmk+=(q1i[i]-qi[i])*(qi[j]-q1i[j])/(q1i[i]-q1i[i]-q1i[j]+qi[j])*
       qqkl[j]*qqkt[i];
    }
  }
 float v=1;
 for (int i=1;i<=m-2;i++)
 {
  v=-1;
 }
Edit2->Text=FloatToStr(qqk);
pqq=v*qqk*qqm*qqmk;
Edit1->Text=FloatToStr(pqq);

}
```