

Введение В настоящее время существует множество пакетов прикладных программ, предназначенных для решения задач математической статистики, таких, как Statistica, Mathematica, Math, SPCC, Stata, SYSTAT, SPSS, Maple и др. Однако довольно часто возникающая на практике задача может быть решена проще и нагляднее с использованием хорошо описанных процедур, без изучения соответствующего сложного и к тому же дорогостоящего лицензионного программного продукта. В данной работе авторы на простых примерах, без использования перечисленных пакетов прикладных программ, показали, как, пользуясь не очень сложным математическим аппаратом, решать некоторые задачи статистического характера. Описанная методика может быть использована как инструмент в научных исследованиях и в программных модулях систем искусственного интеллекта [1].

Постановка задачи Часто на практике возникает задача следующего характера: имеется объект исследования, который характеризуется набором входных переменных и выходной переменной y . Требуется установить зависимость между входными и выходной переменными, то есть найти функцию, наилучшим образом описывающую исходные данные. Рассмотрим простейший случай, с одной входной переменной x . Требуется найти зависимость: . Решение задачи моделирования в данном случае состоит из четырех этапов: 1) проведение эксперимента; 2) выбор вида зависимости; 3) нахождение параметров для выбранной зависимости; 4) исследование модели и анализ результатов. Пусть на первом этапе исходные данные получены в виде статистического ряда (x_i, p_i) , . Здесь p_i – частота появления значения x_i . Требуется выбрать кривую, наилучшим образом описывающую исходные данные. Второй этап достаточно сложен, так как не существует универсальной методики, позволяющей выбрать вид зависимости для любых исходных данных. Исследователь, как правило, руководствуется имеющимся опытом. Первым шагом может стать построение по исходным данным гистограммы и ее анализ. Для дальнейшего более точного установления зависимости между входными и выходными параметрами можно рассчитать значения коэффициентов асимметрии и эксцесса и по "близости" значений оценок параметров и диапазонов их теоретических значений выбрать вид распределений [2, 3]. В настоящей работе авторы ограничились рассмотрением следующих видов распределения: гамма-распределение, распределение хи-квадрат, показательное распределение, распределение Вейбулла, распределение Максвелла. На третьем этапе необходимо подобрать параметры для выбранной кривой распределения таким образом, чтобы она была оптимальной в данном классе кривых. Существует множество методик оценки параметров распределения: метод максимального правдоподобия (ОМП), метод моментов (ОММ), метод квантилей и другие. В последнее время часто применяют одношаговые оценки (ОШ-оценки) [2]. Оценки, полученные на основе применения различных методов, различаются. Универсального ответа на вопрос,

какой из рассмотренных методов лучше или можно ли положиться на данный метод при решении любой задачи, нет. Значение оценки в каждом конкретном случае (для разных выборок) отличается от истинного значения параметра на неизвестную величину. Точность оценок можно косвенно определить на основе проверки согласованности эмпирических и теоретических данных.

Преимущество применения типовых законов распределения состоит в их хорошей изученности и возможности получения состоятельных, несмещенных и относительно высоко эффективных оценок параметров. Для оценки параметров в данной работе применяли метод максимального правдоподобия и метод моментов [4,5]. Для получения оценок были выведены уравнения, для которых существуют численные методы решения. Поэтому все, что требуется – это обладание небольшими навыками программирования и умение применять численные методы решения уравнений. В некоторых видах распределений требуется вычислить значения гамма-функции и ее производной, что является довольно сложной задачей. Вместо непосредственного вычисления гамма-функции ее значения можно рассчитывать по известным аппроксимирующим функциям, например, по формуле Стирлинга [6]. В данной работе для получения оценок использовали свободно распространяемую версию интерактивной системы Scilab, обладающую всеми необходимыми возможностями в области математических вычислений и программирования. На четвертом этапе необходимо проверить гипотезу о согласованности эмпирических данных с теоретическими данными, рассчитанными на основе выбранного теоретического закона распределения. Данную проверку авторы проводили на основе применения критериев согласия [5]. Показательное (экспоненциальное) распределение Непрерывная случайная величина X распределена по показательному (экспоненциальному) закону, если ее плотность распределения вероятностей имеет вид: Показательное распределение часто встречается в теории массового обслуживания (например, X – время ожидания при техническом обслуживании) и в теории надежности (например, X – срок службы радиоэлектронной аппаратуры). Графики плотности распределения показательного распределения для различных λ приведены в [4]. Некоторые характеристики СВХ Математическое ожидание . Дисперсия . Функция распределения . Оценка параметра λ методом максимального правдоподобия Функция правдоподобия имеет вид: Здесь m – число различных значений x , . Логарифмическая функция правдоподобия Для нахождения λ имеем одно уравнение: Отсюда , где . Метод моментов Используется эмпирический начальный момент первого порядка (оценка математического ожидания). Имеем . Откуда . Таким образом, оценки, полученные методом максимального правдоподобия и методом моментов совпадают и с точки зрения трудоемкости вычислений расчет оценок на основе применения указанных методов не вызывает сложностей. Распределение Максвелла Распределение Максвелла—

распределение вероятностей, встречающееся в физике и химии. Оно лежит в основании кинетической теории газов, которая объясняет многие фундаментальные свойства газов, включая давление и диффузию.

Распределение Максвелла также применимо для электронных процессов переноса и других явлений. Распределение Максвелла применимо и к множеству свойств индивидуальных молекул в газе. Оно применяется при статистическом анализе качества технологических процессов. Случайная величина имеет распределение Максвелла, если ее функция плотности имеет вид [4]: . Графики функции плотности этого распределения показаны на рис. 1

Рис. 1 – Графики функций плотности распределения Максвелла

Функция распределения , где .

Другие характеристики случайной величины[4]: Математическое ожидание: . Мода . Дисперсия . Коэффициент асимметрии . Коэффициент эксцесса . Оценка параметра распределения Максвелла

Оценка методом максимального правдоподобия

Для простоты предположим, что все частоты равны единице. Запишем функцию правдоподобия для закона Максвелла

Переходя к статистическому ряду (\sum), получим уравнение для нахождения a . Оценка методом моментов

Поскольку по выборке оценивается лишь один параметр, то для нахождения a используем оценку математического ожидания. , где , Отсюда .

Распределение хи-квадрат

Гамма-распределение с параметрами $\nu=n/2$ (n – натуральное число), $\lambda=1/2$ называется распределением . Распределение играет большую роль в математической статистике при нахождении доверительных интервалов для дисперсии, при проверке статистических гипотез. Если X подчиняется закону , то ее функция плотности распределения вероятностей записывается в виде: , n – число степеней свободы. С увеличением n распределение хи-квадрат медленно приближается к нормальному распределению. Числовые характеристики распределения можно рассчитать по формулам [7]: Математическое ожидание $M[X]=n$. Мода $M_0=n-2$, $n \geq 2$. Дисперсия $D[X]=2n$. Асимметрия . Эксцесс .

Графики плотности распределения хи-квадрат с 4 и 10 степенями свободы показаны на рис. 2.

Рис. 2 – Графики функций плотности распределения хи-квадрат

Оценка методом максимального правдоподобия

Функция правдоподобия для распределения хи-квадрат имеет вид: где t - неизвестный параметр, k – объем выборки, m – число различных значений, принимаемых случайной величиной. Логарифмическая функция правдоподобия имеет вид: Для нахождения параметра t вычислим частную производную от этого выражения и приравняем ее к нулю: Решая полученное уравнение, находим неизвестный параметр t .

Метод моментов

Поскольку по выборке оценивается лишь один параметр n , то для его оценки можно использовать как оценку математического ожидания (первый начальный момент), так и оценку дисперсии (второй центральный момент).

Гамма-распределение

Гамма-распределение находит применение при решении задач об отказах сложных систем с резервированием, когда величина,

характеризующая отказы отдельных элементов системы, подчиняется экспоненциальному закону с интенсивностью λ , а отказ всего объекта наступает после возникновения ν отказов отдельных элементов. Гамма-распределение применяют при решении статистических задачах стандартизации и управления качеством. Оно описывает широкий класс эмпирических распределений проницаемости и широко применяется в практике проектирования и анализа разработки нефтяных месторождений. Говорят, что случайная величина имеет гамма-распределение, если функция плотности распределения вероятностей имеет вид: .

Графики плотности вероятности этого распределения при различных значениях ν и λ показаны на рис. 3. Рис. 3 – Графики функций плотности гамма-распределения

Числовые характеристики гамма-распределения рассчитывают по формулам [8]:

Математическое ожидание M
Мода
Дисперсия
Коэффициент асимметрии
Эксцесс

Оценка параметров гамма-распределения

Оценка методом максимального правдоподобия

Функция правдоподобия имеет вид: где k – число различных значений случайной величины x . Логарифмическая функция правдоподобия имеет вид: Для нахождения неизвестных параметров находят частные производные от по λ и ν соответственно и приравнивают их к нулю. Получается система из двух уравнений с двумя неизвестными, решая которую находят λ и ν . Выразив из второго уравнения λ и подставив результат в первое уравнение, получим: . Тогда первое уравнение будет зависеть только от одной переменной. Полученное уравнение можно решить численным методом. В данной работе уравнение решалось в интерактивной среде Scilab с помощью функции $f_{solve}(x_0, f)$, где x_0 – начальное приближение, f – функция, описывающая левую часть уравнения $f(x)=0$. Для вычисления $\Gamma(x)$ использовалась стандартная функция Scilab. Производная от $\Gamma(x)$ вычислялась с помощью стандартной функции $numdiff$. При вычислении гамма-функции можно воспользоваться известными аппроксимациями этой функции [9]. Если ν не слишком мало, то можно воспользоваться приближением [4]: .

Полученное уравнение решалось методом деления отрезка пополам. Метод моментов

Для оценки неизвестных параметров рассмотрим оценки начального момента первого порядка (оценка математического ожидания): и центрального момента второго порядка (оценка дисперсии):

Поделив первое уравнение на второе, получим λ : . Тогда ν можно рассчитать по формуле: .

Распределение Вейбулла

Распределением Вейбулла можно описать случайные величины, характеризующие наработку на отказ многих невосстанавливаемых электронных приборов (электронные лампы, полупроводниковые приборы, некоторые приборы СВЧ). Оно характеризуется разнообразием форм кривых распределения. Случайная величина X распределена по закону Вейбулла, если ее функция плотности распределения вероятностей имеет вид: Графики плотности распределения Вейбулла ($\nu=5, \beta=3$) и ($\nu=2, \beta=6$) показаны на рис.4. Рис. 4 – Графики функций плотности

распределения Вейбулла Функция распределения вычисляется по формуле:
 Числовые характеристики случайной величины X [7] рассчитывают по формулам:
 Математическое ожидание Мода Дисперсия Оценка параметров распределения
 Вейбулла Оценка методом максимального правдоподобия Предположим, что все
 частоты равны единице, тогда функция правдоподобия распределения Вейбулла
 имеет вид: Найдем частные производные по ν и β и приравняем их нулю: Из
 последнего уравнения выразим β : В случае, когда задан статистический ряд,
 формулы для нахождения параметров β и ν приобретают вид: где m –
 количество интервалов, n – объем выборки. Метод моментов Для оценки
 неизвестных параметров используем оценку начального момента первого
 порядка (оценку математического ожидания): и центрального момента второго
 порядка (оценку дисперсии): . Возведем первое уравнение в квадрат и поделим
 второе уравнение на первое. Получим уравнение относительно ν : . Зная ν ,
 значение параметра β можно рассчитать по формуле: . Заключение Для каждого
 из рассмотренных распределений было смоделировано по три выборки. На
 основе применения метода максимального правдоподобия и метода моментов
 были рассчитаны оценки параметров распределений и относительные
 погрешности каждого из методов (табл. 1). Для получения оценок использован
 аппарат методов оптимизации, численных методов (численные методы решения
 уравнений, численное дифференцирование, аппроксимация) и
 программирования (вычисление сумм и произведений). Согласно данным,
 приведенным в таблице 1, оба метода дают близкие результаты, однако
 расчеты, сделанные методом моментов, как правило, менее трудоемки. Таблица
 1 – Оценка относительных погрешностей ММП и ММ (в случае оценки двух
 параметров указано максимальное значение погрешности)

Распределение	Относительная погрешность оценки ММП	Относительная погрешность оценки ММ
Распределение Максвелла	0,2% - 0,3%	0,2% - 0,3%
Распределение хи-квадрат	1% - 2%	0,3% - 3,5%
Гамма-распределение	1,1% - 6%	0,2% - 5%
Распределение Вейбулла	0,1% - 2,8%	0,1% - 3,1%