Министерство образования и науки Российской Федерации

Санкт-Петербургский государственный политехнический университет

—

**Институт Кибербезопасности и Защиты Информации**

Отчёт по курсовой работе

«**Обобщённое нормальное распределение**»

По дисциплине «Теория вероятностей и математическая статистика»

Выполнил

студент гр. 4851003/10001 Шелин И.Я.

<подпись>

Проверил

ассистент преподавателя Штыркина A.A.

<подпись>

Санкт-Петербург

2023

**Входные данные**

Дана выборка X, подчиняющаяся обобщённому нормальному распределению с параметрами формы и смещения

**Ход работы**

1. История появления, назначение и применение обобщённого нормального распределения в научных исследованиях.

Обобщенное нормальное распределение или Обобщенное нормальное распределение (GGD) является одним из два семейства параметрических непрерывных распределений вероятностей на действительной линии. Оба семейства добавляют параметр формы к нормальному распределению.

В данной работе рассматривается скошенное нормальное распределение. Это семейство непрерывных распределений вероятностей, в которых параметр формы может использоваться для введения перекоса. Когда параметр формы равен нулю, получается нормальное распределение. Положительные значения параметра формы приводят к распределению с уклоном влево, ограниченному вправо, а отрицательные значения параметра формы приводят к распределению с уклоном вправо, ограниченному влево. Только когда параметр формы равен нулю, функция плотности для этого распределения является положительной по всей действительной оси.

Нормальное распределение является одним из наиболее важных и широко используемых распределений в математической статистике[2]. Также его называют «распределением Гаусса» в честь математика Карла Фридриха Гаусса, который разработал формулу для нормального распределения в 1809 году (независимо от Адрейна в 1908) и показал, что ошибки, допущенные в ходе измерений в силу несовершенности приборов измерения, подчиняются нормальному закону распределения.

Важность открытия нормального распределения обусловлена тем, что распределения многих природных явлений, с некоторой долей приближённости, нормально распределены. Поскольку рассматриваемое распределение является общим случаем нормального распределения, сфера его применения тесно пересекается с последним.

Нормальные распределения применяются при анализе погрешностей измерений, контроле технологических процессов и режимов, а также при анализе и прогнозировании и моделировании различных явлений в биологии, медицине и других областях знаний[1].

1. Были написаны функции для вычисления статистик выборки X из варианта. В результате были получены следующие значения:

Сумма: 1663.3444190000005

Среднее: 5.544481396666669

Медиана: 5.55874

Мода: 4.053289

Размах: 1.4832519999999993

Смещённая дисперсия: 0.3536406874108258

Несмещённая дисперсия: 0.3548234321847751

Центральный момент 4-го порядка: 0.31778239695687927

Начальный момент 4-го порядка: 1010.4427830454232

1. Понятие эмпирической функции распределения.

Функция распределения случайной величины () – это вероятность того, что случайная величина примет значение меньше, чем , которая определяется для каждого значения .

В свою очередь, Эмпирическая функция распределения () – это функция, которая определяет для каждого x частоту события X <x и предназначена для оценки теоретической функции распределения генеральной совокупности. В данной записи X – это элементы выборки из генеральной совокупности.

Где – это количество элементов меньших .

Для построения кусочной эмпирической функции распределения сначала необходимо рассчитать минимальное количество интервалов по правилу Стерджиса:

Где N – количество элементов выборки.

Для подвыборок были рассчитаны следующие значения:

|  |  |  |  |
| --- | --- | --- | --- |
| N | 10 | 100 | 200 |
| n | 4 | 7 | 8 |

Далее необходимо найти длину интервала по формуле:

После этого выборку необходимо отсортировать в порядке возрастания.

Далее считаем граничные точки:

Для каждой точки считаем количество элементов выборки, что xi <= yj

Каждое такое значение делим на количество элементов выборки, и получаем значения эмпирической функции распределения.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Yi | yi | y2 | … | yn |
| w*i* | w*i* | w*i* | … | wn |

Для построения графика эмпирической функции распределения была реализована соответствующая функция на языке python. Ниже представлены графики таких функций для случайных выборок из 10, 100 и 200 элементов.

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

*Рис.1 график эмпирической функции подвыборки из 10 элементов*

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

*Рис.2 график эмпирической функции подвыборки из 100 элементов*

Изображение выглядит как текст, снимок экрана, диаграмма, дисплей

Автоматически созданное описание

*Рис.3 график эмпирической функции подвыборки из 200 элементов*

Для рассуждений о виде эмпирической функции распределения из варианта, была также построена кумулятивная функция распределения, которая имеет следующий вид:

– *полная гамма-функция*

– *неполная нижняя гамма-функция*

Функция плотности вероятности

*- гамма-функция*

----------------------------------------------------------------------------------------------------

Перебором параметров было установлено, что эмпирическая функция распределения выборки из 300 элементов практически совпадает с кумулятивной. Для получения наложения были подобраны параметры:

μ(смещение) = Me(Xn) = 5,56

β(форма) = 5,0

σ(масштабирование) = 1,0

Результаты представлены на рисунке ниже.

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

*Рис.4 график эмпирической и кумулятивной функций распределения*

1. Понятие гистограммы

Гистограммой частот называют ступенчатую фигуру, состоящую из прямоугольников, основаниями которых служат частичные интервалы длиною h, а высоты пропорциональны числу элементов выборки, попадающих в конкретный интервал.

Для построения гистограммы сначала, аналогично предыдущему пункту, определяется минимальное количество интервалов, а также длина интервала. Далее элементы выборки сортируются и подсчитывается количество элементов, входящих в каждый интервал. Поделим каждое такое число на количество элементов выборки – получим частоты, которые и будут высотой столбцов.

При помощи библиотеки *matplotlib* были построены гистограммы случайных подвыборок с n = 10, 100, 200.

Изображение выглядит как диаграмма, текст, линия, График

Автоматически созданное описание

*Рис.5 Гистограмма для подвыборки N=10*

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

*Рис.6 Гистограмма для подвыборки N=100*

Изображение выглядит как диаграмма, текст, График, линия

Автоматически созданное описание

*Рис.7 Гистограмма для подвыборки N=200*

При больших N видно, что гистограмма напоминает плотность вероятности обобщённого нормального распределения с параметрами μ = 5,56; β = 5.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

*Рис.8 Плотность вер-ти обобщённого нормального распределения.*

1. Описание параметров распределения

Исходя и варианта задания имеем обобщённое нормальное распределение с двумя параметрами: β - параметр формы и μ - параметр смещения. Также существует параметр σ, отвечающий за масштаб, но в случае σ = 1 его можно опустить.

Параметр β (shape) может быть любым вещественным положительным числом и контролирует форму распределения. Важно отметить, что при β = 1 распределение становится распределением Лапласа, при β = 2 – Нормальным распределением Гаусса, а при β → inf и вовсе напоминает равномерное распределение.

Параметр μ (loc) – также вещественное число, отвечающее за сдвиг распределения вдоль оси x. Параметр определяет точку, в которой находится пик графика функции плотности вероятности распределения.

В свою очередь, теоретическая функция распределения будет иметь следующий вид:

– *полная гамма-функция*

– *неполная нижняя гамма-функция*

Далее была написана функция построения теоретической функции распределения из варианта.

Зафиксировав параметр β, видим, что при малых значениях параметра, функция распределения на концах области определения стримится к значению 0,5, при этом точка перегиба в середине становится менее выраженной. Если и дальше уменьшать параметр, то функция примет вид прямой, параллельной оси X. Если увеличивать параметр β, то точка перегиба становится сначала более явной, на концах области определения функция стремится к 0 слева и 1 справа. Однако, если и дальше увеличивать значение параметра, функция примет вид прямой с углом наклона 45֯.

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

*Рис.9 Распределение при изменении параметра β*

Зафиксировав параметр μ, видим, что график сдвигается вдоль оси X, при этом в точке x = μ соответствует значение 0,5.

Изображение выглядит как текст, линия, График, диаграмма

Автоматически созданное описание

*Рис.10 Распределение при изменении параметра μ*

1. Понятие точечных оценок

Для оценки параметров распределения были использованы метод моментов и метод максимального правдоподобия. Все расчёты произведены для выборки n=300 (генеральная совокупность).

6.1) Метод моментов

В методе моментов применяется свойство выборочного момента k-го порядка, которое заключается в том, что он является состоятельной оценкой теоретического момента соответствующего порядка. Для оценки m неизвестных параметров, нам понадобится столько же теоретических моментов.

Для непосредственной оценки параметров распределения необходимо приравнять выборочные моменты к теоретическим.

*M(x)* – математическое ожидание, *D(x)* – дисперсия, – выборочное среднее, - выборочная дисперсия.

Далее имеем дифференциальное уравнение D(x) - = 0. При ранее вычисленной дисперсии выборки ~ 0,3536.. можем решить уравнение численным методом.

В результате получаем оценки параметров:

Оценка параметра μ – несмещённая, т.к.

Обе оценки являются состоятельными, т.к. так как они сходятся по вероятности к самим себе и имеют конечную дисперсию (свойство метода моментов).

6.2) Метод максимального правдоподобия:

Метод максимального правдоподобия состоит в составлении функции максимального правдоподобия, которая, в свою очередь, является плотностью совместного распределения случайных величин , при условии, что неизвестный параметр действительно равен . Для нахождения оценки неизвестного параметра, от полученной функции требуется найти максимум.

Функция максимального правдоподобия для выборки X={x1​,x2​,…,xn​} из обобщённого нормального распределения с параметрами μ и β:

Прологарифмируем полученную функцию.

Максимизируем полученную функцию, продифференцировав её по μ.

Приравняем полученную производную к 0:

= 0

Очевидно, что производная равна нулю, когда все члены () = 0, а следовательно:

= 0

Следовательно, = 5,4448

Оценка параметра μ является несмещённой и состоятельной (аналогично методу моментов)

Оценка параметра μ также является эффективной, т.к. она обладает минимальной дисперсией среди всех несмещённых оценок.

Для оценки параметра β необходимо минимизировать логарифмическую функцию максимального правдоподобия при условии, что μ = 5,4448. Сделать это аналитически не представляется возможным из-за сложности вычислений, поэтому минимизация была выполнена программно.

В итоге, β = 3,22506

Получившиеся значения параметра μ совпали для двух методов оценки, значения параметра β также отличаются незначительно. При построении графиков функций распределения видно, что теоретические функции распределения также практически совпадают с эмпирической.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

*Рис.11 График Теоретических и эмпирической функция распределения.*

1. Понятие интервальных оценок

**Список литературы**

1. Редакция Кодкампа: 6 реальных примеров нормального распределения // CODECAMP.RU: онлайн-институт цифровых навыков, 17.08.2022 (дата обращения: 20.05.2023).

URL: <https://www.codecamp.ru/blog/example-of-normal-distribution/>

1. Вячеслав Архипов: Правда, чистая правда и статистика. Сетевые сообщества // Хабр: Сообщество IT-специалистов., 30.09.2016.

URL: <https://habr.com/ru/articles/311092/>

1. Вентцель Е.С.: Теория вероятностей, 1969, 4-е издание [Г. 6] С. 116-127.