Теория вероятности и математическая статистика

## Модель эксперимента с конечным числом исходов

Элементарные события — все возможные исходы эксперимента.

Пространство элементарных событий — совокупность элементарных событий.

Событие — подмножество множества .

Объединение событий — произошло хотя бы одно из событий :

Пересечение событий — событие, состоящее из исходов, входящих и в A, и в B:

Дополнение — множество исходов из , не входящих в .

Алгебра — система множеств, для которой выполняются свойства:



## Определение вероятности события

Вероятность исхода — «вес» элементарного исхода:

* (неотрицательность),
* (нормированность).

Вероятность — сумма вероятностей элементарных событий, составляющих событие :

Вероятностная модель — тройка (Ω, 𝔄, P).



### Классический способ задания вероятностей

Если пространство элементарных событий

Вероятность — отношение числа элементарных событий, составляющих событие , к числу всех элементарных событий.

### Статистическое определение вероятностей

Относительная частота события A — отношение числа испытаний, в которых появилось событие A, к общему числу испытаний:

### Геометрическое определение вероятностей

Вероятность события A равна отношению объема множества А к объему всего пространства Ω:

# Условная вероятность

Условная вероятность события A при условии, что произошло событие B:

События A и B — независимые, если

## Формула полной вероятности

События образуют полную группу, если

Если образуют полную группу, то вероятность события B:

## Формула Байеса

Если образуют полную группу

## Схема Бернулли

Пространство всех элементарных исходов имеет следующую структуру:

События, означающие, что в испытаниях произойдет ровно «успехов»:

Вероятность каждого элементарного события:

## Биноминальное распределение

Схема Бернулли — тройка (Ω, 𝔄, P), определяющая вероятностную модель, отвечающую независимым испытаниям с двумя исходами.

Вероятность события равна

Наивероятнейшее число появления «успехов» в испытаниях удовлетворяет неравенствам

Если не является целым числом, то единственно.

Если — целое число, то наивероятнейших значений два:

## Локальная предельная теорема

применяется, если надо вычислить вероятность и .

## Интегральная предельная теорема Муавра–Лапласа

Применяется для приближенного нахождения сумм вероятностей

— вероятность того, что событие A наступит число раз:

Теорема Муавра–Лапласа утверждает, что

## Теорема Пуассона

Теорема Пуассона применяется, если велико, а мало (то есть )

## Случайная величина

* — пространство элементарных событий
* — алгебра подмножеств
* *—* набор вероятностей

Образуют вероятностную модель или вероятностное пространство эксперимента с конечным числом исходов.

Случайная величина — любая функция на множестве .

## Функция распределения

принимает значения .

Распределение вероятностей случайной величины — набор чисел

Функция распределения случайной величины — функция

определенная для .

## Дискретная случайная величина

Случайная величина — дискретная, если ее множество значений не более чем счетное. ДСВ представима в виде суммы

Функция распределения дискретной случайной величины — кусочно-постоянная и равна

Свойства функций распределения

— ограниченная функция,

— неубывающая функция

непрерывна справа, кусочно-постоянная и возрастает в точках скачками величиной

## Независимые случайные величины

— независимые, если для

Математическое ожидание

## Свойства математического ожидания

* Еслинезависимы, то

## Дисперсия

Для дискретной случайной величины

## Свойства дисперсии



Среднее квадратическое отклонение случайной величины :

Если — случайные величины

## Ковариация

## Корреляция случайных величин

Коэффициент корреляции

— некоррелированные, если

Свойства:

Если , то линейно зависимы, то есть

где , если , и , если

## Вероятностное пространство с бесконечным числом исходов

— множество объектов – элементарных исходов эксперимента. Количество элементарных исходов бесконечно.

Сигма-алгебра 𝔍:



Событие — подмножества , входящие в .

Вероятность события — числовая функция, определенная на и ставящая в соответствие каждому событию число . Функция — вероятностная мера, обладающая следующими свойствами:

Вероятностное пространство :

Ω — пространство элементарных событий,

— некоторая сигма-алгебра подмножеств Ω

— вероятностная мера.

Сигма-алгебра борелевских множеств

— наименьшая сигма-алгебра, содержащая все интервалы виды .

## Случайная величина

Случайна величина — действительная функция , определённая на , если для любого борелевского подмножества его прообраз относительно функции 𝜉 является событием :

## Функция распределения случайных величин

Распределение вероятностей случайной величины 𝜉 — вероятностная мера на , определённая для любого борелевского множества равенством

Функция распределения случайной величины 𝜉 — функция,

определенная для любого .

## Свойства функций распределения

*—* неубывающая функция

непрерывна справа и имеет пределы слева в каждой точке .

Случайная величина 𝜉 — дискретная, если ее множество значений не более чем счетное.

Случайная величина 𝜉 — непрерывная, если ее функция распределения непрерывна по .

Случайная величина 𝜉 — абсолютно непрерывная, если существует такая неотрицательная функция – плотность, что

Плотность случайной величины

удовлетворяет равенству

Вероятность попадания случайной величины случайной величины 𝜉 на участок

Если случайная величина 𝜉 имеет плотность, то

## Числовые характеристики непрерывных случайных величин

Если случайная величина , где — непрерывная функция, то

Дисперсия:

## Многомерные функции распределения

Если на одном и том же вероятностном пространстве определены случайные величины , то говорят, что задан случайный вектор

## Функция n переменных

— многомерная функция распределения случайных величин

Свойства двумерной функции распределения:

* — неубывающая функция переменных

Если величины имеют функцию распределения , то можно найти вероятность попадания случайной точки с координатами на плоскости в прямоугольник, ограниченный прямыми

## Плотность совместного распределения

двух случайных величин — вторая смешанная частная производная от функции распределения:

Функция — неотрицательная и обладает свойством:

Зная плотность совместного распределения, можно найти функцию

Плотности распределения отдельных величин выражаются через двумерную плотность

## Независимые случайные величины

— независимые, если для любых борелевских множеств имеет место равенство:

Случайные величины независимы тогда и только тогда, когда их функция распределения представима в виде

Пусть распределение случайного вектора абсолютно непрерывно. Тогда необходимым и достаточным условием независимости случайных величин служит равенство

## Функции распределения суммы, произведения и частного случайных величин

Пусть — независимые случайные величины, имеющие плотности .

Плотность их суммы вычисляется с помощью формулы свертки:

Плотность их произведения находится по формуле:

Плотность их частного находится по формуле:

## Дискретные цепи Маркова

Пусть — некоторое конечное или счетное множество,

Заданы неотрицательные функции

такие, что

Вероятность одного исхода равна

*—* система всех подмножеств *.*

Модель испытаний, связанных в цепь Маркова — вероятностное пространство

## Определение цепи Маркова

— последовательность случайных величин, определенная на вероятностном пространстве со значениями в множестве *X*.

## Свойство условных вероятностей (Марковское свойство)

Марковская цепь — последовательность случайных величин , определенная в конечном вероятностном пространстве со значением во множестве X, для которой соблюдается Марковское свойство.

Начальное распределение — набор вероятностей

Матрица переходных вероятностей из состояния в состояние в момент времени — матрица где

Цепь Маркова — однородная, если для вероятность  не зависит от .

Матрица вероятностей перехода за один шаг — матрица с элементами .

Эта матрица является стохастической, то есть

## Уравнение Колмогорова-Чепмена

Матрица вероятностей перехода за шагов — матрица с элементами

Для любых справедливо уравнение Колмогорова-Чепмена

или в матричном виде

## Классификация состояний цепи Маркова по свойствам переходных вероятностей

Несущественное состояние

достижимо из :

Сообщающиеся состояния — два состояния взаимно достижимы.

Период — наибольший общий делитель чисел .

## Первое возвращение

Вероятность ПЕРВОГО возвращения в состояние :

для — вероятность первого попадания в состояние в момент времени , когда :

Тогда — вероятность того, что частица, находящая в начальный момент в состоянии , рано или поздно попадет в состояние .

Состояние — возвратное, если

## Эргодическая теорема

Эргодическая однородная цепь Маркова

Пусть — матрица переходных вероятностей марковской цепи с конечным числом состояний .

Всякое неотрицательное решение системы , удовлетворяющее условию — стационарное распределение вероятностей для цепи Маркова с матрицей переходных вероятностей .

## Математическая статистика

Математическая статистика — раздел прикладной математики, занимающийся методами и правилами обработки статистических данных.

Основные задачи:

Указать *способы сбора и группировки* статистических сведений, полученных в результате наблюдений или поставленных экспериментов.

Разработка *методов анализа статистических данных* в зависимости от целей исследования:

* Оценка неизвестной вероятности события,
* Оценка параметров распределения, вид которого неизвестен,
* оценка зависимости случайной величины от других случайных величин,
* проверка статистических гипотез о виде или величине параметров неизвестного распределения.

Основная задача математической статистики — создание методов сбора и обработки статистических данных для получения научных и практических выводов.

## Совокупности

Генеральная совокупность — совокупность всех объектов, подчиненных данному признаку.

Выборочная совокупность — совокупность случайно отобранных объектов.

Объем совокупности — число объектов этой совокупности.

Репрезентативность — выборка должна быть моделью генеральной совокупности, правильно представлять ее пропорции.

Типы выборок

Простая случайная выборка — отбор элементов случайным образом.

* Бесповторный отбор
* Повторный отбор

Механическая выборка — генеральную совокупность делят на столько групп, сколько объектов должно войти в выборку и из каждой группы отбирают один объект (через равные интервалы).

Серийный отбор — объекты отбираются сериями, а внутри каждой серии производится сплошное обследование.

Типический отбор — объекты выбираются из генеральной совокупности, предварительно разделенной на возможно более однородные группы, пропорционально размеру каждой группы.

Комбинированный отбор — сочетаются указанные выше способы.

## Статистическое распределение выборки

Из генеральной совокупности извлечена выборка

Варианты — наблюдаемые значения.

Вариационный ряд — последовательность вариант, записанных в порядке возрастания.

Частота — число наблюдения данной варианты.

Относительная частота  — отношение частоты варианты к объему выборки.

Медиана — варианта, которая делит поровну вариационный ряд.

Мода — варианта, которая имеет максимальную частоту в вариационном ряду.

Размах выборки  — показатель вариации, разность между крайними значениями вариационного ряда.

Статистическое распределение выборки — перечень вариант и соответствующих им частот или относительных частот, расположенных в порядке возрастания значений вариант.

## Графическое представление статистической совокупности

Полигон частот — ломаная линия, вершинами которой являются точки , определяемые элементами статистического ряда.

Гистограмма частот — ступенчатая фигура, составленная из прямоугольников, построенных на интервалах длины так, что площадь каждого прямоугольника численно равна частотеварианты , расположенной в середине -го интервала. Высоты прямоугольников равны отношению (плотность частоты). Площадь гистограммы равна объему выборки n.

Эмпирическая функция распределения — функция , которая каждому значению ставит в соответствие сумму относительных частот вариант выборки, меньших :

Эмпирическая функция распределения позволяет составить представление об интегральной функции распределения всей генеральной совокупности.

## Статистические оценки параметров распределения

Требуется оценить неизвестный количественный признак генеральной совокупности. Имеются лишь данные выборки, рассматриваемые как значения независимых случайных величин.

Статистическая оценка неизвестного параметра генеральной совокупности — функция от выборочных значений. Точечная оценка определяется одним числом.

Несмещённая — точечная оценка, математическое ожидание которой равно оцениваемому параметру при любом объеме выборки, т.е. .

Выборочная средняя — несмещенная оценка генеральной средней

Выборочная дисперсия — смещенная оценка генеральной дисперсии. Характеризует рассеяние наблюдаемых значений количественного признака выборки вокруг своего среднего значения.

Исправленная выборочная дисперсия — несмещенная оценка генеральной дисперсии

Среднее квадратическое отклонения — характеристика рассеяния значений признака выборочной совокупности вокруг своего среднего значения:

Коэффициент вариации — оценка сопоставимости различных исследований:

## Интервальные оценки

Интервальная — оценка, которая определяется двумя числами — концами интервала, покрывающего оцениваемый интервал.

Оценка математического ожидания (генеральной средней) количественного признака X генеральной совокупности — выборочная средняя .

Надежность, доверительная вероятность — вероятность выполнения неравенства

Доверительный интервал — — интервал, который с заданной надежность. покрывает заданный параметр.

## Доверительный интервал для оценки математического ожидания генеральной совокупности

Заранее известна величина среднего квадратического отклонения генеральной совокупности:

— выборочная средняя

— объем выборки

— известное среднее квадратическое отклонение генеральной совокупности

— величина, определяемая по таблице Лапласа из условия

*— предельная ошибка.*

для доли

Величина среднего квадратического отклонения генеральной совокупности неизвестна, используется выборочное среднее квадратическое отклонение, определяемое по данным выборки.

— выборочная средняя

— объем выборки

— известное среднее квадратическое отклонение генеральной совокупности

— величина, определяемая по таблице Стьюдента при заданном объеме и доверительной вероятности

## Виды зависимостей

Функциональная зависимость

— связь между переменными величинами, при которой зависимая величина-функция полностью определяется значениями влияющих независимых величин-аргументов.

Статистическая зависимость

— каждому значению одной переменной соответствует определенное (условное) распределение другой переменной. Каждому значению одной переменной соответствует не какое-то определенное, а множество возможных значений другой переменной.

Корреляционная зависимость

— каждому значению переменной X соответствует условное (т.е. при условии, что переменная X приняла значение x) математическое ожидание другой переменной Y.

## Регрессия

Модельное уравнение регрессии — односторонняя зависимость случайной переменной от одной неслучайной независимой переменной

Первая задача теория корреляции — установить форму корреляционной зависимости, то есть вид функции регрессии: линейная или нелинейная.

Вторая задача теории корреляции — оценить тесноту (силу) корреляционной связи. Оценивается по величине рассеяния значений вокруг условного среднего .

## Линейная модель парной регрессии

Поле корреляции или диаграмма рассеяния — отображение пар значений точками на плоскости XOY.

Уравнение регрессии ищется в виде линейного уравнения, где

— уравнение выборочной регрессии, а

— уравнение в генеральной совокупности

## Коэффициент корреляции Пирсона

Характеристики коррелируют друг с другом, если две характеристики имеют тенденцию изменяться совместно так, что создается возможность предсказать величину одной из них по значению другой.

Положительная корреляция — высоким значениям одной переменной соответствуют высокие значения другой, а низким — низкие.

Набор значений двух переменных

*—* средние значения

— выборочные квадратические отклонения

Коэффициент корреляции Пирсона

Уравнение линейной регрессии:

Свойства:

коэффициент корреляции принимает значение на отрезке [-1, 1]. Чем ближе к единице, тем теснее связь.

При корреляционная связь представляет линейную функциональную зависимость. При этом все наблюдаемые значения располагаются на прямой линии.

При линейная корреляционная связь отсутствует.

## Ранговая корреляция

применяется для определения тесноты связи между признаками, измеренными в порядковых шкалах.

Коэффициенты ранговой корреляции Спирмена и Кендалла используются для определения тесноты связи между двумя величинами

Коэффициент конкордации устанавливает статистическую связь между несколькими признаками.

Методы ранговой корреляции могут быть использованы для определения тесноты связи не только между количественными переменными, но и между качественными признаками при условии, что их значения можно упорядочить и проранжировать.

## Коэффициент ранговой корреляции Спирмена

Каждая из двух совокупностей располагается в виде вариационного ряда с присвоением каждому члену ряжа соответствующего порядкового номера (ранга), выраженного натуральным числом. Одинаковым значениям ряда присваивают среднее ранговое число.

Сравниваемые признаки можно проранжировать в любом направлении: как в сторону ухудшения качества, так и наоборот. Главное, чтобы обе переменные были проранжированы одинаковым способом.

Формула ранговой корреляции Спирмена:

— разность рангов для каждой -ой пары из наблюдений.

## Коэффициент ранговой корреляции Кендалла

Имеет те же свойства, что и коэффициент Спирмена

* изменяется от -1 до 1
* для независимых случайных величин равен нулю

считается более информативным.

Для расчета коэффициента Кендалла ряды переменных ранжируются (одинаковым значениям ряда присваивают среднее ранговое число).

Первая переменная должна быть упорядочена по возрастанию рангов.

— число рангов во втором вариационном ряду, больших, чем данное ранговое число и расположенных ниже него.

## Проверка статистических гипотез

Статистическая гипотеза — гипотеза о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Нулевая гипотеза — выдвинутая гипотеза .

— конкурирующая (альтернативная) гипотеза, которая противоречит нулевой.

Ошибка первого рода — отвергнута правильная гипотеза.

Уровень значимости — вероятность ошибки первого рода

Ошибка второго рода — принята неправильная гипотеза. Вероятность ошибки — .

Статистический критерий — случайная величина , распределение которой известно, и которая служит для проверки нулевой гипотезы.

Наблюдаемое значение критерия — значение критерия, вычисленное по данным выборок.

Критическая область — совокупность значений критерия, при которых гипотезу отвергают.

Область принятия гипотезы — совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки гипотез — если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если принадлежит области принятия гипотезы – гипотезу принимают.

## Критические области

Правосторонняя критическая область — Находим из таблицы критическую точку , чтобы при условии справедливости выполнялось

При гипотеза может оказаться верной, и, отвергая , совершаем ошибку первого рода с вероятностью .

Левосторонняя критическая область — Находим из таблицы критическую точку , чтобы при условии справедливости выполнялось

Двусторонняя критическая область — находим критические точки , чтобы при условии справедливости выполнялось

Если распределение симметрично относительно нуля, то и находится из условия

## Гипотезы о равенстве средних при известных и при неизвестных дисперсиях

Есть 2 выборки и объемом из различных генеральных совокупностей.

Генеральные совокупности имеют нормальное распределение с параметрами соответственно

## Гипотеза о равенстве средних при известных дисперсиях

известны

Выбираем уровень значимости

находим — выборочные средние

Для проверки гипотезы используется критерий

В случае справедливости критерий имеет нормальное распределение с параметрами . Критическое значение находим из равенства

— удвоенная функция Лапласа.

Тогда область принятия гипотезы:

## Гипотеза о равенстве средних при неизвестных дисперсиях

Дисперсии генеральных совокупностей неизвестны, но равны между собой:

Для оценки можно воспользоваться любой из исправленных выборочных дисперсий:

Критерий:

Критерий при верной гипотезе имеет распределение Сттьюдента с числом степеней свободы , критическая точка находится из условия:

Область принятия гипотезы:

## Гипотеза о равенстве дисперсий при неизвестных средних

Есть 2 выборки и объемом из различных нормально распределенных генеральных совокупностей, средние значения которых неизвестны. Задан уровень значимости

Проверяем гипотезу

Находим исправленные выборочные дисперсии:

Составляем критерий:

Если верна гипотеза , то критерий *F* имеет распределение Фишера с степенями свободы.

Гипотезу H0 отвергаем, когда *F* существенно отличается от 1, то есть или.

Здесь — критические точки распределения Фишера,

Если нашли критерий , то нужно проверить только выполнение неравенств

Гипотезу принимают, если .

Если проверяется гипотеза

то

## Гипотеза о равенстве дисперсий при известных средних

Есть 2 выборки и объемом из различных нормально распределенных генеральных совокупностей, средние значения которых равны . Задан уровень значимости

Проверяем гипотезу

Находим исправленные выборочные дисперсии:

Составляем критерий:

Если верна гипотеза , то критерий *F* имеет распределение Фишера с степенями свободы.

Гипотезу отвергаем, когда *F* существенно отличается от 1, то есть или.

Здесь — критические точки распределения Фишера,

Если нашли критерий , то нужно проверить только выполнение неравенств

Гипотезу принимают, если .

Если проверяется гипотеза

то

## Гипотеза о равенстве относительной частоты с гипотетической вероятностью появления события

Пусть по достаточно большому числу независимых испытаний, в каждом из которых вероятность появления события постоянна, но неизвестна, найдена относительная частота события . Требуется при заданном уровне значимости проверить нулевую гипотезу, состоящую в том, что неизвестная вероятность равна гипотетической вероятности :

Вычисляем наблюдаемое значение критерия

и по таблице функции Лапласа находим критическую точку из равенства

Если , то нулевую гипотезу принимают.