Міністерство освіти і науки України

Київський національний університет імені Тараса Шевченка

Механіко-математичний факультет

кафедра теорії ймовірностей,

статистики та актуарної математики

Кваліфікаційна робота для здобуття ступеня магістра на тему:

Реалізація сайту для аналізу моделей сумішей зі змінними концентраціями

Студента 2 року магістратури

Денної форми навчання

спеціальності актуарна

та фінансова астматика

Лук’янова Іллі Богдановича

Науковий керівник:

професор, доктор фізико-математичних наук

Майборода Ростислав Євгенович

«Допустити до захисту у ДЕК»

Завідувач кафедри теорії ймовірностей,

Статистики та акутарної математики

д.ф.-м.н., проф. Мішура Ю.С.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис)

Київ – 2015

Зміст

# ВСТУП

Моделі сумішей зустрічаються в сучасному житті надзвичайно часто. Вони виникають природно в соціологічних, хімічних, біологічних та інших дослідженнях протягом вже декількох століть. Відповідно розвивається і математична теорія статистичної оцінки відповідних даних.

На сьогоднішній день в Інтернеті немає жодного сервісу, яким би могли скористуватися науковці, аби проаналізувати результати власних вибірок. Тому тема роботи є надзвичайно **актуальною**, адже дослідники можуть мати недостатньо досвіду для роботи з такими великими статистичними пакетами як Statistica, який до того ж не є безкоштовним, або з такими складними, як R, і не мати необхідних знань і навичок, аби отримати адекватні результати статистичного аналізу свої вибірок.

**Об’єктом** дослідження є методи статистичного аналізу сумішей у класичній моделі та методи їх реалізації мовою програмування.

**Метою** дипломної роботи є створення функціонуючого, сучасно оформленого он-лайн сервісу для статистичної обробки даних сумішей, який буде мати інтуїтивно зрозумілий та дружній інтерфейс, не буде вимагати регістрації, буде працювати достатньо швидко, та виводити всі необхідні результати (квантилі, оцінки моментів, щільності та функцій розподілу компонент).

# РОЗДІЛ 1. Теоретична частина

## Загальна постановка задачі аналізу сумішей.

Вважаємо, що спостерігаються деякі об’єкти , кожен з яких може належати одній з популяцій (компонент). Номер популяції, якій належить об’єкт позначимо . Справжнє значення його вважаємо невідомим, але відомі ймовірності . Ці ймовірності називають концентраціями або ймовірностями перемішування. Концентрації компонент повинні задовольняти наступні умови: , .

У всіх об’єктів спостерігається один і той самій набір характеристик . Для цей набір позначимо . Множину всіх можливих значень характеристик позначимо . У роботі, як правило, Х це дійсна прямо або дійснозначний вектор. Взагалі кажучі, Х може бути будь-яким вимірним простором, тобто простором, на якому задана -алгебра вимірниз підмножин. Спостережувані характеристики вважаємо випадковими елементами Х, незалежними для різних Розподіл цих характеристик залежить від того, якій компоненті належить об’єкт. Розподіл характеристик k-ої компоненти будем позначати , тобто

для всіх вимірних множин з Х. Надалі, – скінченно-вимірний векторний прості, то ж будемо позначати відповідну функцію розподілу:

для всіх Нерівності для векторів слід розуміти покоординатно.

Таким чином, розподіл спостережуваних характеристик має вигляд

Надалі ми будемо використовувати схему серій для опису асимптотичної поведінки наших оцінок при необмеженому зростанні обсягу вибірки, тобто коли . Тому спостережувані дані розглядаються як один рядок трикутного масиву .

Зрозуміло, що реально статистик має справу лише з вибіркою фіксованого обсягу, з одним рядочком. Відповідно і концентрації кожнох компоненти можна трактувати як трикутні масиви.

Крім концентрацій будуть використовуватися і інші масиви аналогічної структури. Часто до таких масивів буде застосовуватись оператор усереднення по рядочках Ми будемо позначати його

Аналогічно, якщо , то

Визначений таким чином функціонал можна розглядати як скалярний добуток N-тих рядочків наших масивів. Якщо границя

існує, то ми будемо позначати її <a.> = <a>.

Будемо використовувати зважені емпіричні міри вигляду:

як оцінки для за . Тут а є деяким невипадковим трикутним масивом вагових коефіцієнтів. Під не випадковістю мається на увазі незалежність від , але не від . Ці вігові коефіцієнти часто будуть залежати від деякого параметра . У таких випадках будемо писати

)

## Незміщені мінімаксні оцінки.

Якщо і то

є оцінкою для функції розподілу і зветься емпіричною функцією розподілу.

Якщо вимагати незміщенності як оцінки , то ми отримуємо

для всіх можливих наборів . При N > M ця умова виконується тоді і тільки тоді, коли

*для всіх*

Цю умову будемо називати умовою незміщеності.

Помітимо, що коли вона виконується, то

отже, якщо є незміщенною оцінкою для то

Однак, оскільки для всіх m, то з умови незміщеності випливає, що повинні приймати негативні значення для деяких j. Легко бачити, що у цьому випадку не може бути ймовірнісною мірою на Х, якщо всі різні. З іншого боку, для всіх А

Отже, якщо виконано умову незміщенності, то є зарядом зі скінченною варіацією.

З класу всіх можливих незміщених оцінок вигляду

доцільно обрати одну, в певному розумінні найкращу. Яку міру якості обрати визначає науковець, але в роботі ми обираємо гарантований ризик при квадратичній функцї витрат.

## Обчислення оптимальної оцінки відносно гарантованого ризику при квадратичній функцї витрат.

Нехай є деяка оцінка для за спостереженнями . Ми будемо вважати, що витрати від використання неточної оцінки замість справжнього значення задаються квадратичною функціює ризику . Відповідно, середні витрати становлять

Тоді

являє собою гарантований ризик оцінки , тобто максимальні витрати, які в середньому можна понести при використанні оцінки при найгірших значеннях характеристик моделі. Ми будемо брати sup по всіх можливих ймовірнісних розподілах оскільки розглядається непараметрична задача оцінювання. Розглянемо зважену емпіричну міру як оцінку для . Тоді гарантований ризик буде функцією вектора коефіцієнтів а,

Знайдемо якщо виконані умови незміщенності з попереднього пункту. Використовуючи ці умови, отримуємо:

оскільки sup не більше ¼ причому це значення досягається коли для всіх m.

Отже ми повинні знайти вектор який мінімізує

при виконанні умов незміщенності

*для всіх*

Використаємо метод множників Лагранжа для розв’язання цієї задачі мінімізаціхю Як відомо, необхідною умовою того, що а є точкою умовного екстремуму J при відповідних обмеженнях є

,де j = 1,…N

– невизначені множники Лагранжа

Ця умова рівносильна

де – довільні константи.

Тобто вектор оптимальниз ваговиз коефіцієнтів є лінійною комбінацією векторів навантажень. Підставлючси цей розклад в умову незміщенності

*для всіх*

отримуємо систему лінійних рівнянь для :

Припустимо, що матриця є не виродженою.

Тоді це рівняння має єдиний розв’язок

де – lk мінори .

Відповідний оптимальний (мінімаксний) вектор вагових коефіцієнтів визначається як

Умова, що матриця є не виродженою еквівалентна лінійній незалежності системи векторів оскільки вона є матрицею Грама цієї системи у скалярному добутку .

Щоб підкреслити, що це є скалярним добутком векторів з N-вимірного дійснозначного простору будемо записувати

Підставивши

в

отримуємо найменше можливе значення гарантованого ризику:

Дійсно,

Помітимо, що всі

де =

Отже маємо

Отже, ми показали, що вагові коефіцієнти забезпечують найкращий з точки зору гарантованого квадратичного ризику результат при використанні як оцінки для . Оцінити її якою-небудь незміщеною оцінкою так, щоб вона мала б менший гарантований ризик неможливо.