

Сведения о выполненных работах и
полученных научных результатах в 2022 году

по проекту «**Эффективные статистические методы анализа
эпидемиологических моделей**»,
поддержанному Российским научным фондом

Соглашение № 22-21-00302

Руководитель: Воробейчиков Сергей Эрикович, д-р физ.-мат. наук
Пергаменщиков Сергей Маркович, д-р физ.-мат. наук

Работы первого этапа проекта проводились по двум основным направлениям вероятностного и статистического анализа динамики распространения эпидемий. В первом направлении развивались статистические методы скорейшего выявления начала распространения и локализации очага распространения эпидемий, во втором разрабатывались робастные методы эффективного статистического анализа стохастических динамических систем, определяющих распространение эпидемий. Задачи по раннему выявлению и локализации эпидемий изучались на основе методов последовательного анализа, обеспечивающих оптимальные статистические решения при минимальном числе наблюдений. Задачи обнаружения рассматривались в двух постановках в асимптотической – при наличии неограниченного числа наблюдений, когда вероятности ошибок идентификации стремятся к нулю и в точной постановке – на ограниченном временном интервале при произвольных фиксированных верхних границах для вероятностей ложных тревог. Для асимптотических задач обнаружения и идентификации эпидемий невозможно применять классические процедуры типа КУСУМ или Робертса–Ширяева, так как соответствующие дивергенции Кульбака–Лейблера равны нулю. Поэтому для таких задач предлагается использовать условно-гауссовские эпидемиологические модели, построенные на основе нормальной аппроксимации сумм бернуллиевских случайных величин с помощью ЦПТ. Для таких моделей подсчитаны и изучены соответствующие дивергенции Кульбака–Лейблера, определяющие информационные нижние границы для среднего времени запаздывания процедур обнаружения. Получены неасимптотические равномерные оценки для вероятностей отклонений в эргодических теоремах для логарифма отношения правдоподобия в данных моделях, обеспечивающих робастное минимаксное обнаружение для случая неизвестных распределений после разладок. На основе усредненных по неизвестным параметрам модели базовых статистик Робертса разработаны последовательные процедуры Робертса–Ширяева обнаружения и идентификации скорейшего выявления начала эпидемий, для которых с помощью полученных нижних границ устанавливается асимптотическая оптимальность в минимаксном смысле для рисков, определяемых средним временем запаздывания в случае неизвестных вероятностных распределений, определяющих стохастическую составляющую в динамике распространения эпидемий. Установлено также, что свойства оптимальности сохраняются и для робастных рисков, построенных на основе супремума среднего времени запаздывания, взятого по всевозможным значениям неизвестных параметров, определяющих основные вероятностные

характеристики динамики распространения эпидемий. Алгоритмы, разработанные на базе полученных в проекте процедур, являются помехозащитными, устойчивыми по отношению к изменению вероятностных характеристик эпидемиологической динамики в достаточно широких пределах. Для уточнения асимптотических методов выявления эпидемий на основе байесовского подхода разработаны методы для скорейшего выявления начала эпидемий в неасимптотической постановке. В данном случае задача скорейшего обнаружения представляется как задача об оптимальной остановке для однородных марковских процессов в байесовской постановке на ограниченном временном горизонте. При этом в качестве априорного распределения выбиралось равномерное. Это принципиально новый прорывной результат в задачах обнаружения, потому что ранее такие задачи решались только для геометрического априорного распределения, которое, во-первых, нельзя использовать на ограниченных временных интервалах, а, во-вторых, это распределение содержит неизвестный параметр, который невозможно оценить, что делает невозможным использование байесовских процедур, построенных для геометрических априорных распределений в большинстве практически важных задачах, в частности, для анализа эпидемий. В проекте эта задача представляется как задача об оптимальной остановке для процесса, построенного на основе базовых статистиках Робертса. Установлено, что для эпидемиологических моделей статистики Робертса, построенные для равномерного априорного распределения в совокупности с наблюдениями, образуют двумерный однородный марковский процесс, для которого задача скорейшего обнаружения формулируется как задача об оптимальной остановке для полученного марковского процесса. Отметим, что в данном случае, задача обнаружения рассматривается на конечном временном интервале и оптимальные процедуры ищутся для задачи минимизации среднего времени запаздывания в классе последовательных процедур с вероятностями ложных тревог, не превосходящих фиксированных значений. Найдено точное решение соответствующей задачи об оптимальной остановке на основе техники огибающих Снэлля и, как следствие, развит оптимальный последовательный метод обнаружения начала эпидемий в смысле минимума среднего времени запаздывания в классе последовательных процедур с вероятностями ложных тревог, не превосходящих заданного фиксированного произвольного значения. С целью существенного сокращения числа операций в разрабатываемых статистических алгоритмах, в проекте развиваются последовательные методы раннего выявления для статических условно-гауссовских эпидемиологических моделей, построенных на основе нормальной аппроксимации биномиальных моделей. Для таких моделей, за счет замены громоздких сумм бернуллиевских случайных величин гауссовскими, существенно сокращается число операций необходимых для вычисления огибающей Снэлля, что значительно ускоряет работоспособность соответствующих алгоритмов. Далее, в проекте разрабатываются эффективные статистические методы анализа скорости распространения эпидемий. При этом за основу берется классическая модель «восприимчивые – инфицированные – выздоровевшие» (SIR). Согласно этой модели, популяция делится на три группы: (S) восприимчивые, (I) инфицированные и (R) выздоровевшие. Для описания эволюции этих трех переменных в классической

модели используется система обыкновенных дифференциальных уравнений. В проекте, с целью учета влияния непредсказуемых случайных факторов на динамику числа инфицированных (I), изучается стохастическая модель SIR, построенная на основе стохастических дифференциальных уравнений, определяемых с помощью скачкообразных процессов Леви с неизвестным распределением. Для таких моделей на основе улучшенных методов оценивания развиты адаптивные статистические методы выбора моделей для оценивания скорости распространения эпидемий по наблюдениям в дискретные моменты времени. Разработаны новые адаптивные процедуры выбора моделей, основанные на взвешенных улучшенных оценках наименьших квадратов со специально подобранными коэффициентами, обеспечивающими асимптотическую эффективность. Для задачи непараметрического оценивания скорости распространения эпидемий найдены достаточные условия на частоту поступления наблюдений, обеспечивающие получение точных неасимптотических оракульных неравенств для робастных рисков. Проведен теоретический анализ эффекта улучшения неасимптотической точности оценивания. Разработаны также статистические алгоритмы для практической реализации улучшенных процедур выбора модели и на основе метода Монте-Карло проведены численные эксперименты, подтверждающие их работоспособность на практике.

Основные результаты находятся на следующих интернет ресурсах:

<https://hal.archives-ouvertes.fr/hal-03871920>

<https://lmrs.univ-rouen.fr/fr/content/quickest-changepoint-detection-and-identification-general-stochastic-models>

https://ideas.repec.org/a/spr/sistpr/v25y2022i3d10.1007_s11203-021-09258-0.html

https://www.sciencedirect.com/science/article/pii/S0047259X22000203?casa_token=mf84rKMqguMAAAAA:PFiPnnPsq8aI3ShqFXp6N46cO3AaoUPuQoLe3Hxuhx0JrMePuOSpkZJhCLbOYi430ByVNs64m8k

<https://www.elibrary.ru/item.asp?id=49802686>

<https://www.elibrary.ru/item.asp?id=49807100>