

УДК 004.21

## ИССЛЕДОВАНИЕ МНОГОУРОВНЕВОГО РАСЩЕПЛЕНИЯ ДЛЯ ОЦЕНКИ МАЛЫХ ВЕРОЯТНОСТЕЙ В ДИСКРЕТНЫХ ПРОСТРАНСТВАХ

Голиков Н.Р. (ИТМО), Сухов В.Д. (ИТМО)

Научный руководитель – кандидат технических наук, доцент Сергушичев А.А. (ИТМО)

**Введение.** Непараметрические тесты часто используются для проверки статистических гипотез. Одним из таких тестов является тест Колмогорова-Смирнова, который проверяет что две выборки сделаны из одного вероятностного распределения. Ключевым моментом теста является вычисление р-значения для полученного результата. Точно вычислить р-значение обычно тяжело, поэтому его считают приближенно, например, используя алгоритм Монте-Карло. Одной из проблем использования алгоритма Монте-Карло является низкая относительная точность. Для решения этой проблемы можно использовать алгоритм многоуровневого расщепления Монте-Карло [1], заключающийся в разбиении вероятности на несколько уровней, и подсчете условной вероятности на каждом из них. Например, такой подход был использован ранее для решения задачи анализа представленности генов [2]. Такой алгоритм плохо работает, если вероятностное пространство сильно дискретно, как например в тесте Колмогорова-Смирнова.

**Основная часть.** Проведено исследование возможности применения многоуровневого расщепления Монте-Карло в дискретных вероятностных пространствах. В таких случаях у алгоритма возникают проблемы со сходимостью, если какое-то значение встречается очень много раз. В таком случае, может произойти ситуация, когда значение для уровня в алгоритме совпадает со значением на предыдущем уровне, и алгоритм не может продолжать работу. На примере теста Колмогорова-Смирнова данная проблема была показана в работе [3]. Для решения данной проблемы разработан алгоритм на основе хеширования, позволяющий расширить вероятностное пространство. Для каждого элемента вводится хеш, и значением уровня считается упорядоченная пара значения статистики и значения хеша, используя лексикографическое сравнение. При использовании данного подхода удастся избежать проблемы с совпадением значений у уровней.

Также было проведено исследование возможности обобщения данного алгоритма для использования в пространствах, где значение статистики является вещественным числом.

**Выводы.** Разработан алгоритм многоуровневого расщепления Монте-Карло в дискретных вероятностных пространствах на основе хеширования. Проведен анализ качества разработанного алгоритма, а также алгоритм был обобщен для использования в пространствах с вещественным значением статистики.

### Список использованных источников:

1. Multilevel Splitting for Estimating Rare Event Probabilities / P. Glasserman [et al.] // *Operations Research*. — 1999. — Vol. 47, no. 4. — P. 585–600. — DOI: 10.1287/opre.47.4.585.
2. Fast gene set enrichment analysis / G. Korotkevich [et al.] // *bioRxiv*. — 2021. — Feb. — P. 060012. — eprint: 060012. — URL: <https://doi.org/10.1101/060012>
3. Сухов В.Д., Короткевич Г.В., Сергушичев А.А. Многоуровневое расщепление в методе Монте-Карло для оценки вероятностей редких событий в пермутационных тестах // *Научно-технический вестник информационных технологий, механики и оптики*. – 2024.