

Секция «Теория вероятностей и математическая статистика»

**Свойства статистики Хмелёва**

**Научный руководитель – Ковалевский Артём Павлович**

***Попов Алексей Сергеевич***

*Студент (магистр)*

Новосибирский государственный университет, Механико-математический факультет,  
Новосибирск, Россия

*E-mail: a.popov11@g.nsu.ru*

Работа посвящена изучению свойств статистики Хмелёва. В рассматриваемой модели тексты на естественном языке трактуются как реализации некоторых цепей Маркова, где состояниями цепей выступают символы алфавита. Статистика Хмелёва была введена Д. В. Хмелёвым в статье 2000 года и выступает показателем того, насколько два текста похожи между собой. Этот анализ основывается на сравнении матриц переходных вероятностей цепей Маркова, соответствующих этим текстам.

В работе было показано, что если тексты порождены цепями Маркова с совпадающими матрицами переходных вероятностей, то нормализованная статистика Хмелёва для этих текстов стремится к нулю с вероятностью 1; если же матрицы различны, то нормализованная статистика будет стремиться к положительной константе с вероятностью 1. Аналогичное утверждение было доказано для случая, когда статистика вычисляется между некоторым текстом и конкатенацией двух других текстов, порождённых цепями Маркова с различными матрицами переходных вероятностей. Результаты проведённых статистических экспериментов показали, что статистика Хмелёва пригодна для определения, является ли некоторый текст подлинно написанным на определённом естественном языке или фальсификацией.