

А. П. Ковалевский, Н. С. Закревская (Новосибирск, НГТУ, НГУ). **Элементарная вероятностная модель литературных текстов и пуассонизация.**

Бахадур (1960) изучал бесконечную урновую схему, в которой вероятности убывают в соответствии со степенным законом. Он нашел асимптотику роста математического ожидания числа непустых урн и доказал закон больших чисел. Отметим, что убывание вероятностей слов в соответствии со степенным законом в естественных языках известно как закон Ципфа (1936), а рост числа различных слов в соответствии со степенным законом в лингвистике известно как закон Хипса (1978). В нашей модели выбор очередного слова из словаря моделируется выбором урны для очередного шара. Карлин (1967) предложил пуассонизацию бахадуровской модели, в соответствии с которой длина текста является случайной величиной, имеющей распределение Пуассона и не зависящей от процесса выбора слов. Используя этот подход, Карлин доказал центральную предельную теорему для числа разных слов. Барбур (2009) показал, что можно предложить более аккуратную пуассоновскую аппроксимацию. Чебунин и Ковалевский (2016) доказали функциональную центральную предельную теорему, используя пуассонизацию, предложенную Карлиным.

Мы используем эту функциональную теорему для анализа однородности текстов. Для этого мы строим такие оценки параметра Ципфа, для которых мы можем вычислить асимптотику их совместного распределения с процессом числа разных слов. Используя эти оценки, мы конструируем процесс, асимптотика которого известна в случае однородного текста. Однородность литературных текстов анализируется критерием типа омега-квадрат. Примеры применения включают анализ однородности текстов на французском, английском, русском, китайском языках.

Исследование поддержано грантом РФФИ 17-01-00683.

СПИСОК ЛИТЕРАТУРЫ

1. *Bahadur R. R.* On the number of distinct values in a large sample from an infinite discrete distribution. — *Proceedings of the National Institute of Sciences of India*, 1960, v. 26A, is. 2, c. 67–75.
2. *Barbour A. D.* Univariate approximations in the infinite occupancy scheme. — *Alea*, 2009, v. 6, c. 415–433.
3. *Chebunin M., Kovalevskii A.* Functional central limit theorems for certain statistics in an infinite urn scheme. — *Statistics and Probability Letters*, 2016, v. 119, c. 344–348.
4. *Heaps H. S.* *Information Retrieval: Computational and Theoretical Aspects*, Academic Press, 1978, 534 с.
5. *Karlin S.* Central Limit Theorems for Certain Infinite Urn Schemes. — *J. Mathematics and Mechanics*, 1967, v. 17, № 4, c. 373–401.
6. *Zipf G. K.* *The Psycho-Biology of Language*. Routledge, London, 1936, 236 с.