

Для сокращения объема вычислений можно попытаться ограничить область перебора следующим образом:

$$\{h_i \mid i = \overline{1, N}, h_i \in N, 0 \leq h_i \leq m, m < n\}. \quad (3)$$

Предложенное ограничение приводит нас к рассмотрению поведения статистики максимальной частоты M_n ,

$$M_n = \max_{i=1}^N h_i.$$

Опираясь на формулу (в частности, см. [6]) $\mathbf{P}\{A\} = \mathbf{P}\{AB\} + \mathbf{P}\{A\bar{B}\}$, выпишем равенство для вероятностей значений симметрично аддитивно разделимой статистики S_n , с учетом ограничений (3):

$$\mathbf{P}\{S_n \geq c\} = \mathbf{P}\{S_n \geq c, M_n > m\} + \mathbf{P}\{S_n \geq c, M_n \leq m\}.$$

Теперь, если предположить, что m подобрано так, чтобы $\mathbf{P}\{S_n \geq c, M_n > m\} \leq \Delta$, то

$$|\mathbf{P}\{S_n \geq c\} - \mathbf{P}\{S_n \geq c, M_n \leq m\}| = \mathbf{P}\{S_n \geq c, M_n > m\}$$

и по нашему предположению

$$|\mathbf{P}\{S_n \geq c\} - \mathbf{P}\{S_n \geq c, M_n \leq m\}| \leq \Delta.$$

Следовательно, вместо вычисления вероятности $\mathbf{P}\{S_n \geq c\}$, которое требует перебора по области (2), можно вычислять вероятность $\mathbf{P}\{S_n \geq c, M_n \leq m\}$, используя ограниченную область (3). Вычисленные значения будут отличаться не более, чем на Δ . В отличие от точного значения вероятности $\mathbf{P}\{S_n \geq c\}$, вероятность $\mathbf{P}\{S_n \geq c, M_n \leq m\}$, будем называть Δ -точной, а соответствующее ей распределение вероятностей по аналогии с [1] Δ -точным распределением.

Предложенное направление модернизации частотного метода представляется перспективным и необходимо исследовать его точность, вычислительную и временную сложность.

СПИСОК ЛИТЕРАТУРЫ

1. Мельников А. К., Ронжин А. Ф. Обобщенный статистический метод анализа текстов, основанный на расчете распределений вероятности значений статистик. — Информатика и ее примен., 2016, т. 10, в. 4, с. 89–95.
2. Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling. — Philos. Mag. Ser. 5., 1900, v. 50. № 302, p. 157–170.
3. Зелокин Н. Б., Мельников А. К. Сложность расчета точных распределений вероятности значений статистик и область применения предельных распределений. В сб.: Электронные средства и системы управления: Материалы докладов XIII Международной научно-практической конференции. (29 ноября – 1 декабря 2017 г.) Томск: В-Спектр, 2017, Ч. 2, с. 84–90.
4. Мельников А. К. Сложность расчета точных распределений вероятности симметричных аддитивно разделяемых статистик и область применения предельных распределений. — Доклады ТУСУРа. 2017, т. 20, № 4, с. 126–130.
5. Холл М. Комбинаторика. М.: Мир, 1970, 424 с.
6. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т. 1: Пер. с англ. М.: Мир, 1984, 528 с.