Секция «Математика и механика»

Центральные предельные теоремы для массивов перестановочных случайных величин.

Ракитько Александр Сергеевич
Студент
Московский государственный университет имени М.В. Ломоносова, Механико-математический факультет, Москва, Россия
E-mail: rakitko@gmail.com

Во многих стохастических моделях функция отклика Y зависит не от всех факторов $X = (X_1, \dots, X_n)$, а лишь от некоторого *значимого* набора $\{X_{k_1}, \dots, X_{k_r}\}$, где $1 \le k_1 < \dots < k_r \le n$. Описанная ситуация часто возникает в медико-биологических исследованиях. Для решения подобных задач в [2] был разработан метод, основанный на понижении размерностей. Основная идея заключается в том, что функции f, прогнозирующие отклик Y по различным наборам факторов, упорядочиваются в соответствии с некоторым функционалом ошибки $Err(f)$ для произвольных штрафных функций ψ. Распределение случайных элементов X и Y неизвестно, поэтому статистические выводы основаны на оценках $\widehat{Err}_K(f_{PA})$ функционала ошибки, вовлекающих предсказательный алгоритм f_{PA}, оценку штрафной функции $\widehat{\psi}$ и K-кросс валидацию. Рассматриваются результаты применения построенной процедуры к анализу сгенерированных данных.
В статье [1] описанная выше задача была обобщена на случай многозначной функции отклика. Была доказана строгая состоятельность предложенной оценки $\widehat{Err}_K(f_{PA})$, а также установлена центральная предельная теорема. Полученные результаты в области перестановочных случайных величин позволили продолжить данное исследование.
Пусть $\mathfrak{X} = \{X_{n,i},\ i = 1, \dots, n\}_{n \in \mathbb{N}}$ - треугольный массив построчно перестановочных случайных величин. С помощью метода Линдеберга установлены центральные предельные теоремы вида

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(X_{n,i} - L_{n,i}\right) \xrightarrow{\mathsf{d}} \mathsf{N}(0, \sigma^2), \quad n \to \infty, \tag{1}$$

где $L_{n,i},\ i \le n$, зависят от $X_{n,1}, \dots, X_{n,n}$. Введем обозначения $\widehat{\mu}_n := n^{-1} \sum_{i=1}^{n} X_{n,i}$ и $\widehat{\sigma}_n^2 := n^{-1} \sum_{i=1}^{n} \left(X_{n,i} - \widehat{\mu}_n\right)^2$. При естественных ограничениях на массив $\mathfrak{X}$ справедлива следующая

Теорема 1. Пусть задана возрастающая последовательность натуральных чисел m_n такая, что $m_n \le n$ и $m_n/n \to \alpha < 1$ при $n \to \infty$. Если $\mathbb{E} \sup_{n \in \mathbb{N}} \left((X_{n,1} - \widehat{\mu}_n)/\widehat{\sigma}_n\right)^4 < \infty$, то

$$\frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \left(\frac{X_{n,i} - \widehat{\mu}_{k_n}}{\widehat{\sigma}_{k_n}}\right) \to Z_{0,1-\alpha} \sim \mathsf{N}(0, 1-\alpha), \quad n \to \infty.$$

Используемая при доказательстве техника позволяет получать оценки на скорость сходимости.

Литература

1. А.В. Булинский, А.С. Ракитько, Оценивание небинарного случайного отклика, Доклады РАН, 2014, т. 455, №6, с. 1-5.

2. A. Bulinski, O. Butkovsky, V. Sadovnichy, A. Shashkin, P. Yaskov, A. Balatskiy, L. Samokhodskaya and V. Tkachuk, Statistical Methods of SNP Data Analysis and Applications, Open Journal of Statistics, Vol. 2 No. 1, 2012, pp. 73-87.

Слова благодарности

Автор выражает благодарность и признательность профессору, д.ф.-м.н. Булинскому Александру Вадимовичу за постановку задачи и руководство в работе.