

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

ПРОВЕРКА ГИПОТЕЗ СОГЛАСИЯ ДЛЯ РЕГРЕССИОННОЙ МОДЕЛИ ПРИ НАЛИЧИИ КЛАССИФИКАЦИИ ЗАВИСИМОЙ ПЕРЕМЕННОЙ

Е.С. Агеева

Белгосуниверситет, факультет прикладной математики и информатики, Минск, Беларусь
helenaageeva@gmail.com

Регрессионной моделью описываются многие процессы, однако на практике зачастую данные наблюдаются с некоторыми отклонениями от классической модели [1]. Эта работа посвящена такому искажению как группирование [2], а точнее одному из его частных случаев — классификации.

Пусть задана регрессионная модель

$$Y_t = F^0(X_t, \theta^0) + u_t, \quad t = 1, \dots, n,$$

где $X_t \in \mathbf{X} \subseteq \mathbb{R}^N$, $t = 1, \dots, n$, — независимые регрессоры, $F^0(\cdot, \cdot) : \mathbf{X} \times \Theta \rightarrow \mathbb{R}^1$ — некоторая функция, заданная с точностью до параметра $\theta^0 \in \Theta \subseteq \mathbb{R}^m$, $\{u_t\}_{t=1}^n$ — случайная ошибка наблюдения, распределенная по нормальному закону с нулевым математическим ожиданием и неизвестной дисперсией: $\mathcal{L}\{u_t\} = \mathcal{N}(0, (\sigma^0)^2)$.

Пусть задано разбиение числовой прямой на K непересекающихся интервалов ($2 \leq K < +\infty$):

$$A_1 = (a_0, a_1], \quad A_2 = (a_1, a_2], \quad \dots, \quad A_{K-1} = (a_{K-2}, a_{K-1}], \quad A_K = (a_{K-1}, a_K),$$

где $-\infty = a_0 < a_1 < \dots < a_K = +\infty$. Истинные значения $\{Y_t\}_{t=1}^n$ не наблюдаются. Вместо этого мы знаем только номера интервалов $\{\nu_t\}_{t=1}^n$, в которые эти значения попали:

$$Y_t \in A_{\nu_t}, \quad \nu_t \in \{1, 2, \dots, K\}.$$

В [3] были доказаны сильная состоятельность и асимптотическая нормальность оценки максимального правдоподобия параметров $((\theta^0)', (\sigma^0)^2)'$ при условии, что регрессионная функция $F^0(\cdot, \cdot)$ известна с точностью до параметра. В данной работе рассмотрена задача проверки гипотез согласия:

$$H_0 : F^0(\cdot, \cdot) \in \mathcal{F}, \quad H_1 : F^0(\cdot, \cdot) \notin \mathcal{F},$$

где $\mathcal{F} = \{F(X; \theta), \theta \in \Theta, X \in \mathbf{X}\}$ — предполагаемое параметрическое семейство функции регрессии. Например, требуется установить можно ли использовать линейную регрессию для описания данных.

Литература

1. Kharin Yu. *Robustness in statistical forecasting*. Springer, 2013.
2. Heitjan D. F. *Inference from Grouped Continuous Data: A Review* // Statistical Science. 1989. Vol. 4, no. 2. P. 164–183.
3. Ageeva H., Kharin Yu. *ML estimation of multiple regression parameters under classification of the dependent variable* // Lithuanian Mathematical Journal. 2015. Vol. 55, no. 1, P. 48–60.