Уральский государственный архитектурно-художественный университет
Национальный исследовательский Томский государственный университет
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина

НОВЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ СЛОЖНЫХ СТРУКТУР

МАТЕРИАЛЫ
ОДИННАДЦАТОЙ МЕЖДУНАРОДНОЙ КОНФЕРЕНЦИИ
6–10 июня 2016 г.

Томск
Издательский Дом Томского государственного университета
2016

Второе слагаемое в (3) показывает величину уменьшения дисперсии главной части оценки (2) при использовании совместно некомплектных наблюдений объемов n_1 и n_2 соответственно. В случаях $n_1 = 0$, ($\lambda_1 = 0$) и $n_2 = 0$ ($\lambda_2 = 0$) из (3) получаем дисперсии главных частей оценок (1) соответственно.

Положим $n_1 = k_1 n$, $k_1 \geq 0$ и $n_2 = k_2 n$, $k_2 \geq 0$. Тогда при $n \to \infty$ закон распределения $L(\sqrt{n}(P_n^{ab}(AB) - P(AB)) \to N(0, \sigma^2)$ (стремится к нормальному закону) с нулевым средним и дисперсией σ^2, равной

$$\sigma^2 = P(AB)P(\overline{AB}) - P^2(AB)\left[\frac{k_1}{1+k_1+k_2} \times \frac{P(\overline{A})}{P(A)} + \frac{k_2}{1+k_1+k_2} \times \frac{P(\overline{B})}{P(B)} - \frac{2k_1k_2}{(1+k_1+k_2)^2}\frac{P(AB) - P(A)P(B)}{P(A)P(B)}\right]. \quad (4)$$

Формула (4) позволяет анализировать влияние доли некомплектных наблюдений на точность оценивания.

Выводы. Предложен подход к обработке данных с пропусками, основанный на использовании некомплектных наблюдений при оценивании неизвестных вероятностей событий. Полученные оценки улучшают (не ухудшают) точность оценивания по СКО при любых объемах комплектных и некомплектных наблюдений. Этот факт подтверждает целесообразность включения некомплектных наблюдений в анализ данных. При увеличении доли некомплектных наблюдений точность оценивания повышается за счет привлечения информации, содержащейся в них. Методы заполнения пропусков и переход к «полным» данным к такому свойству не приводят. Предложенные оценки позволяют пересчитывать таблицы сопряженности признаков с учетом некомплектных наблюдений и на основании этого делать более надежные статистические выводы.

Литература

1. *Дмитриев Ю.Г., Устинов Ю.К.* Статистическое оценивание распределений вероятности с использованием дополнительной информации. Томск : Изд-во Том. ун-та, 1988. 194 с.

2. *Дмитриев Ю.Г., Князева А.А.* Оценки вероятностей событий при наличии данных с пропусками // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2009. № 4 (9). С. 43–51.

ОЦЕНКИ ВЕРОЯТНОСТИ С УЧЕТОМ АПРИОРНЫХ ДОГАДОК

Ю.Г. Дмитриев, Т.О. Кошевая

Национальный исследовательский Томский государственный университет, Томск, Россия
dmit@mail.tsu.ru, tomakoshevaya@gmail.com

Введение. Проблема сокращения объема дорогостоящих экспериментальных данных при сохранении требуемой точности оценивания приводит к необходимости привлечения дополнительной информации в структуру оценок искомых характеристик. В работах [1–3] рассматривается ситуация, когда исследователь на основании своего опыта и знаний может высказать предположение о значении интересующей его характеристики (такое значение принято называть априорной догадкой), выраженной линейным функционалом от неизвестного распределения вероятностей. Там же предложены оценки с учетом *одной* априорной догадки и проанализировано ее влияние на СКО оценок. В данной работе предлагаются оценки для вероятности некоторого события с учетом *многих* априорных догадок, анализируется их влияние на СКО оценок при конечном объеме наблюдений, даются рекомендации по их применению.

Постановка задачи. Пусть $X_1, \ldots, X_n$ – независимые результаты эксперимента объема n, а $P = P(B)$ – вероятность некоторого события B, которое может наступить в данном эксперименте. Пусть имеется m априорных догадок p_i, $i = \overline{1,m}$, которые выступают в качестве возможных значений неизвестной вероятности P. Требуется построить оценку P, учитывающую совместно непараметрическую оценку $\hat{P} = \hat{P}(B) = n^{-1}\sum_{z=1}^{n} I_B(X_z)$ и все p_i, здесь $I_B(.)$ – индикаторная функция. Рассмотрим комбинированную оценку $\hat{P}_\lambda = \hat{P} - \sum_{i=1}^{m}\lambda_i\left(\hat{P} - p_i\right)$, где весовые коэффициенты λ_i вектора $\lambda = (\lambda_1, \ldots, \lambda_m)$ выбраны из условия минимума среднеквадратической ошибки (СКО) $S^2(\lambda_i) = M[\hat{P} - \lambda_i\left(\hat{P} - p_i\right) - P]^2$ и определяются выражением $\lambda_i = 1/\left(1 + n\Delta_i^2/\sigma^2\right)$, в котором $\sigma^2 = P(1-P)$, а $\Delta_i = P - p_i$.

Адаптивные оценки. Поскольку коэффициенты λ_i неизвестны, то можно построить их оценки $\hat{\lambda}_i$ по имеющейся выборке и получить адаптивные комбинированные оценки вероятности P: $\hat{P}_{\hat{\lambda}} = \hat{P} - \sum_{i=1}^{m}\hat{\lambda}_i\left(\hat{P} - p_i\right)$. Рассмотрим оценки $\hat{\lambda}_{i,1} = 1/\left(1 + n\hat{\Delta}_i^2/\hat{\sigma}^2\right)$, где $\hat{\Delta}_i = \hat{P} - p_i$, $\hat{\sigma}^2 = \hat{P}(1-\hat{P})$, и $\hat{\lambda}_{i,2} = 1/\left(1 + n\hat{\Delta}_i^2/\sigma_i^2\right)$, где $\sigma_i^2 = p_i(1-p_i)$.

Соответствующие им адаптивные оценки $\hat{P}_1$ и $\hat{P}_2$ являются смещенными с СКО S_1^2 и S_2^2, которые можно вычислить для различных значений *n* и *P* на основании биномиального распределения. Возникает вопрос, при каких условиях такие адаптивные оценки оказываются предпочтительнее обычной оценки $\hat{P}$ по величине СКО $S^2(\hat{\lambda}) = M[\hat{P}_{\hat{\lambda}} - P]^2$. Отметим, что СКО $S^2(\lambda) = M[\hat{P}_{\lambda} - P]^2 = [(1 - \sum_{i=1}^{m} \lambda_i)^2 + (\sum_{i=1}^{m} \lambda_i \Delta_i)^2 n / \sigma^2]\sigma^2 / n$.

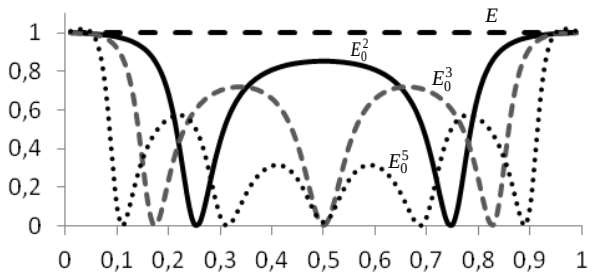

Рис. 1. Зависимость E_0^m от *P* и *m* при *n* = 100

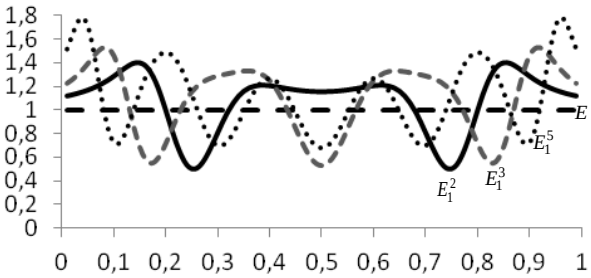

Рис. 2. Зависимость E_1^m от *P* и *m* при *n* = 100

Для сравнения СКО оценок рассмотрим отношения $E = D\hat{P} / D\hat{P} = 1$, $E_0^m = S^2(\lambda)$, $E_1^m = S_1^2 / D\hat{P}$, $E_2^m = S_1^2 / D\hat{P}$, где верхний индекс показывает количество априорных догадок. Зададим априорные догадки в виде: $p_i = (2i-1)/2m, i = 1, \ldots, m$. Графики отношений СКО в зависимости от *P, n, m* приводятся на рисунках.

Литература

1. *Dmitriev Yu., Tarassenko P., Ustinov Yu.* On Estimation of Linear Functional by Utilizing a Prior Guess // ITMM 2014, CCIS 487 / A. Dudin [et al.]. (eds.). 2014. P. 82–90.

2. *Дмитриев Ю.Г., Кошевая Т.О.* О комбинированных оценках вероятности // Известия вузов. Физика. 2015. Т. 58, № 11/2. С. 242–246.

3. *Dmitriev Yu., Tarassenko P.* On Adaptive Estimation Using a Prior Guess // Applied Methods of Statistical Analysis. Nonparametric Approach – AMSA-2015, Novosibirsk, Russia, 14–15 September, 2015 : Proceedings of the International Workshop. 2015. P. 49–55.

ИССЛЕДОВАНИЕ КАЧЕСТВА НЕПАРАМЕТРИЧЕСКИХ ОЦЕНОК ПРИ ИССЛЕДОВАНИИ МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ

Е.С. Катаева

Томский государственный университет систем управления и радиоэлектроники, Томск, Россия
kes@keva.tusur.ru

В данной работе на примере временных рядов среднемесячных температур, полученных на метеорологических станциях на острове Врангеля, в Томске, Уфе, на Оймяконе и в Никольском (Свердловская область) с 1955 по 2010 г., было исследовано качество оценки значений временного ряда непараметрической регрессией Надарая–Ватсона [1]

$$r(x,h) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)},$$

где *K*(*u*) – ядро функции, X_i – значения зависимой переменной (в нашем случае это моменты времени), Y_i – известные значения временного ряда, *h* – параметр размытости (определяет количество влияющих на оценку соседних значений ряда). В качестве ядра рассматривается функция гауссовского вида $K(u) = \frac{1}{\sqrt{2\pi}} exp\left\{-\frac{u^2}{2}\right\}$.

В качестве ошибки оценки использовалось эмпирическое абсолютное отклонение

$$E = \frac{1}{N} \sum_{t=1}^{N} \left| y_t - \hat{y}_t \right|,$$

где $\hat{y}_t$ – оценка значения ряда в момент *t* [2].