Литература

1. *Breuer L., Dudin A. N., Klimenok V. I.* A retrial BMAP|PH|N system // Queueing Systems, 2002. Vol. 40. P. 433–457.
2. *Dudin A. N., Klimenok V. I.* A retrial BMAP|SM|1 system with linear repeated requests // Queueing Systems, 2000. Vol. 34. No 1–4 P. 47–66.

ВЛИЯНИЕ АДДИТИВНЫХ ИСКАЖЕНИЙ НА ОЦЕНКИ ПАРАМЕТРОВ БЕТА-ЛОГИСТИЧЕСКОЙ МОДЕЛИ ДАННЫХ

М. А. Пашкевич

Введение

Бета-логистическая модель (БЛМ) традиционно используется для моделирования группированных бинарных данных в случае наличия априорной информации о свойствах объектов, над которыми производятся испытания. Впервые эта модель была введена Хекманом [2], и с тех пор широко используется в политологии [5], маркетинге [6] и других областях [1, 4]. При этом для оценивания параметров БЛМ обычно применяется метод максимального правдоподобия (ММП) [2]. Однако на практике гипотетическая вероятностная модель наблюдений оказывается, как правило, неадекватной в силу искажений различных типов [3]. Поэтому возникает необходимость исследования влияния этих искажений на свойства классических оценок. В данной работе исследуется влияние аддитивных стохастических искажений бинарных данных на оценки максимального правдоподобия (ММП-оценки) параметров БЛМ.

Математические модели и постановка задачи

Пусть определена некоторая совокупность из k объектов и некоторое случайное событие A. Над каждым объектом i этой совокупности производится серия из n_i испытаний. Результаты испытаний описываются набором k бинарных векторов-строк $B = (B_1, B_2, \ldots, B_k)$, $B_i \in \{0,1\}^{n_i}$, $i = 1, 2, \ldots, k$, где $B_{ij} = 1$, если в испытании j для объекта i случайное событие A имело место, и $B_{ij} = 0$ в противном случае. Объекту номер i поставлен в соответствие вектор факторов $Z_i \in R^m$, описывающий свойства этого объекта. Предполагается, что:

П$_1$. Вероятностные свойства объектов в процессе испытаний не меняются.

П$_2$. Вероятность p_i наступления случайного события A для i-ого объекта является случайной величиной, которая имеет бета распределение с параметрами α_i^0, β_i^0, причем $p_1, p_2, \ldots, p_k$ независимы в совокупности.

П3. Параметры указанных бета-распределений α_i^0, β_i^0 связаны с векторами факторов Z_i выражениями $\alpha_i^0 = \exp(a_0^T Z_i)$, $\beta_i^0 = \exp(b_0^T Z_i)$, где a_0, $b_0 \in R^m$ – векторы параметров модели.

Пусть на результаты испытаний $B = (B_1, B_2, \ldots, B_k)$ воздействуют случайные ошибки $\{\eta_{ij}\}$, и наблюдаются искаженные бинарные векторы-строки $\widetilde{B} = \left(\widetilde{B}_1, \widetilde{B}_2, \ldots, \widetilde{B}_k\right)$:

$$\widetilde{B}_{ij} = B_{ij} \oplus \eta_{ij}, \quad i = 1,2,\ldots k, \quad j = 1,2,\ldots,n_i, \tag{1}$$

где $\oplus$ – операция сложения по модулю два, а $\{\eta_{ij}\}$, $i = 1, 2, \ldots, n_i$, $j = 1, 2, \ldots, k$, – независимые случайные величины Бернулли. При этом для каждого i, j имеет место следующая зависимость случайной величины η_{ij} от случайной величины B_{ij}:

$$P\left\{\eta_{ij} = 1 \middle| B_{ij} = 0\right\} = \varepsilon_0, \quad P\left\{\eta_{ij} = 1 \middle| B_{ij} = 1\right\} = \varepsilon_1. \tag{2}$$

Рассматривается задача статистического оценивания параметров a_0, b_0 по косвенным данным – искаженной выборке $X = \{x_1, x_2, \ldots, x_k\}$ объема k: $x_i = \sum_{j=1}^{n_i} \widetilde{b}_{ij}$, $i = 1,2,\ldots,k$, где x_i – число наступлений события A для объекта i. Необходимо исследовать влияние искажений (1), (2) на свойства ММП-оценок параметров БЛМ.

Уклонения ММП-оценок в случае искаженной выборки

Теорема. Для описанной выше бета-логистической модели с искажениями (1), (2) имеют место следующие стохастические разложения для уклонений ММП-оценок параметров a_0, b_0 для некоторой выборки X:

$$\Delta a(X, \varepsilon_0, \varepsilon_1) = \widehat{a}(X, \varepsilon_0, \varepsilon_1) - a_0, \quad \Delta b(X, \varepsilon_0, \varepsilon_1) = \widehat{b}(X, \varepsilon_0, \varepsilon_1) - b_0,$$

$$\begin{pmatrix} \Delta a(X, \varepsilon_0, \varepsilon_1) \\ \Delta b(X, \varepsilon_0, \varepsilon_1) \end{pmatrix} = J^{-1}(a_0, b_0, X) \cdot g_e(a_0, b_0, X, \varepsilon_0, \varepsilon_1) + 1_{2m}\left(o(\varepsilon_0) + o(\varepsilon_1) + O_P\left(\frac{1}{\sqrt{k}}\right)\right), \tag{3}$$

где вектор $g_e(.)$ определяется как

$$\left(g_\varepsilon(a_0, b_0, X, \varepsilon_0, \varepsilon_1)\right)^T = \left(g^a(a_0, b_0, X, \varepsilon_0, \varepsilon_1), g^b(a_0, b_0, X, \varepsilon_0, \varepsilon_1)\right)^T,$$

$$g_l^a(a_0, b_0, X, \varepsilon_0, \varepsilon_1) = \sum_{i=1}^{k}\left(Z_{il}\alpha_i^0\left(\frac{-x_i\left(\beta_i^0 + n_i - x_i\right)}{\left(\alpha_i^0 + x_i - 1\right)^2}\varepsilon_0 + \frac{n_i - x_i}{\beta_i^0 + n_i - x_i - 1}\varepsilon_1\right)\right),$$

$$l = 1,2,\ldots,m,$$

$$g_l^b(a_0,b_0,X,\varepsilon_0,\varepsilon_1)=\sum_{i=1}^{k}\left(Z_{il}\beta_i^0\left(\frac{x_i}{\alpha_i^0+x_i-1}\varepsilon_0-\frac{(n_i-x_i)(\alpha_i^0+x_i)}{(\beta_i^0+n_i-x_i-1)^2}\varepsilon_1\right)\right),$$

$$l=1,2,\ldots,m,$$

а $J(.)$ есть следующая блочная матрица:

$$J(a_0,b_0,X)=\begin{pmatrix} J^{Aa}(a_0,b_0,X) & J^{Ab}(a_0,b_0,X) \\ J^{Ba}(a_0,b_0,X) & J^{Bb}(a_0,b_0,X)\end{pmatrix},$$

$$J_{ls}^{Aa}(a_0,b_0,X)=\sum_{i=1}^{k}\left(Z_{il}Z_{is}\alpha_i^0\left(\sum_{j=0}^{x_i-1}\frac{1}{\alpha_i^0+j}-\sum_{j=0}^{n_i-1}\frac{1}{\alpha_i^0+\beta_i^0+j}\right)\right)-$$

$$-\sum_{i=1}^{k}\left(Z_{il}Z_{is}\left(\alpha_i^0\right)^2\left(\sum_{j=0}^{x_i-1}\frac{1}{(\alpha_i^0+j)^2}-\sum_{j=0}^{n_i-1}\frac{1}{(\alpha_i^0+\beta_i^0+j)^2}\right)\right),\quad l,s=1,2,\ldots,m,$$

$$J_{ls}^{Ab}(a,b,X)=J_{ls}^{Ba}(a,b,X)=\sum_{i=1}^{k}\left(Z_{il}Z_{is}\alpha_i^0\beta_i^0\sum_{j=0}^{n_i-1}\frac{1}{(\alpha_i^0+\beta_i^0+j)^2}\right),\quad l,s=1,2,\ldots,m,$$

$$J_{ls}^{Bb}(a,b,X)=\sum_{i=1}^{k}\left(Z_{il}Z_{is}\beta_i^0\left(\sum_{j=0}^{n_i-x_i-1}\frac{1}{\beta_i^0+j}-\sum_{j=0}^{n_i-1}\frac{1}{\alpha_i^0+\beta_i^0+j}\right)\right)-$$

$$-\sum_{i=1}^{k}\left(Z_{il}Z_{is}\left(\beta_i^0\right)^2\left(\sum_{j=0}^{n_i-x_i-1}\frac{1}{(\beta_i^0+j)^2}-\sum_{j=0}^{n_i-1}\frac{1}{(\alpha_i^0+\beta_i^0+j)^2}\right)\right),\quad l,s=1,2,\ldots,m.$$

Доказательство теоремы основано на асимптотическом анализе функции правдоподобия с учетом результатов, полученных ранее в [7].

Результаты компьютерного моделирования

Для подтверждения полученных теоретических результатов была проведена серия компьютерных экспериментов. При этом в качестве номинальных параметров БЛМ были выбраны значения $a_0 = 1$, $b_0 = 2$, причем предполагалось, что $m = 1$, $n_1 = n_2 = \ldots = n_K = 10$, а факторы Z_i считались равномерно распределенными на отрезке [1,0; 1,1]. Уровни искажений ε_0, ε_1 совпадали и изменялись в пределах от 0,00 до 0,05 с шагом 0,01. Результаты компьютерного моделирования приводятся в таблице, в которой приняты следующие обозначения: $\varepsilon = \varepsilon_0 = \varepsilon_1$ – уровень искажений, Δa_{min}, Δa_{max} и Δb_{min}, Δb_{max} – границы 95-процентных экспериментальных доверительных интервалов для уклонений соответственно параметров a и b, Δa_{th}, Δb_{th} – теоретические значения уклонений, вычисленные с использованием выражения (3). Как следует из таблицы, проведенные компьютерные эксперименты подтверждают достоверность полученных в теореме выражений.

Таблица

Достоверность выражений для уклонений ММП-оценок БЛМ

ε	Δamin	Δamax	Δath	Δbmin	Δbmax	Δbth
0,00	–0,02	0,03	0,00	–0,01	0,02	0,00
0,01	0,08	0,12	0,85	0,04	0,08	0,05
0,02	0,14	0,18	0,15	0,09	0,13	0,10
0,03	0,21	0,26	0,23	0,15	0,19	0,15
0,04	0,28	0,32	0,31	0,18	0,23	0,20
0,04	0,35	0,40	0,39	0,24	0,28	0,25

Заключение

В работе получены выражения, позволяющие оценить уклонения ММП-оценок параметров БЛМ в случае аддитивных стохастических искажений бинарных данных. Теоретические результаты подтверждаются результатами компьютерного моделирования. Данные исследования были частично поддержаны грантом БГУ для молодых ученых 628/30.

Литература

1. *Dunn R., Wrigley, N.* Beta-Logistic Model of Urban Shopping Center Choice // Geographical Analysis. 1985. Vol. 17. P. 95–113.
2. *Heckman J. J.* A Beta-logistic Model for the Analysis of Sequential Labor Force Participation by Married Women // Journal of Political Economy. 1977. Vol. 85. P. 27–58.
3. *Kharin Yu*. Robustness in Statistical Pattern Recognition, Kluwer Academic Publishers, Dordrecht, 1996.
4. *Ming, K., Rosenbaum P. R.* Estimation and Testing with Overdispersed Proportions Using the Beta-Logistic Regression Model of Heckman and Willis // Biometrics**.** 2000. Vol. 56. P. 125–134.
5. *Dietz, N.* A Beta-Logistic Model of Presidential Influence on Voting on Civil Rights Issues in the House of Representatives, 1960-1988 // Midwest Annual Meeting Methodology Panel Working Papers*,* Jacksonvile, 1999. P. 265–273.
6. *Pfeifer P. E.* On Using the Beta-Logistic Model to Update Response Probabilities Given Nonresponse // Journal of Interactive Marketing. 1998. Vol. 12. P. 23–32.
7. *Харин Ю. С., Пашкевич М. А.* Статистическое оценивание бета-биномиального распределения при искажениях бинарных наблюдений // Весцi НАН Беларусi. Сер. фiз.-мат. навук. 2003. № 1. С. 11–17.