ROBUST BAYES FORECASTING FOR GROUPED BINARY DATA WITH KNOWN DISTORTION INTERVAL

Maxim Pashkevich

Department of Mathematical Modelling and Data Analysis, Faculty of Applied Mathematics and Computer Science, Belarusian State University, 4 F. Skaryny Ave., Minsk, Belarus, e-mail: *pma@omegasoftware.com*

Abstract. The paper is devoted the problem of robust forecasting for the beta-mixed hierarchical models of grouped binary data in the case of stochastic additive distortions of binary observations. In the case of known lower and upper bounds of the distortion intervals, a new robust minimax Bayes predictor is developed. The performance of the proposed forecasting technique is validated by computer simulation.

1. Introduction

Grouped binary data frequently arises in longitudinal studies that are carried out over a group of similar objects [4]. A natural way to describe this kind of data is using the binomial model [2]. However, it was noticed that this simple model often leads to inaccurate statistical inference due to so called "over-dispersion" effects [1]. These effects may occur for two main reasons: intergroup correlation (violation of the independence assumption of the experiment outcomes for a particular object) and intragroup correlation that is caused by heterogeneity among objects [7]. For these reasons, special "random effects" models are used to describe the heterogeneity and correlated outcomes [3].

Beta-mixed hierarchical models of grouped binary data are widely used in practical applications when the information about the experiments conditions is not available [4]. The most popular models of this class are the beta-binomial and beta-logistic models. To forecast the response probabilities in the beta-mixed hierarchical model, the Bayes approach is traditionally used.

In real life, the observed binary outcomes are often misclassified, and the classical statistical procedures that are optimal for the hypothetical model may loose their "good" properties under distortions [6]. In our previous work [8], it was proved that the classical Bayes predictor is not optimal under distortions and a new robust predictor for the case of known distortion levels was proposed. In this paper, we consider the case when the distortion levels interval is known and propose a new robust minimax predictor. The performance of the developed forecasting technique is illustrated by computer simulation results.

2. Mathematical Models and Problem Statement

Let us consider k objects with properties $Z_i \in R^m$, $i = 1, 2, \dots, k$, and a random event A, and let $B_i = (B_{i1}, B_{i2}, \dots, B_{in_i}) \in \{0,1\}^{n_i}$ be results of n_i Bernoulli experiments with the event A over the object i. Let us make the following three assumptions.

A$_1$. The probability properties of the objects are stable in time.

A$_2$. For the object i, the probability of success p_i is a random variable that has the beta distribution with true unknown parameters $\alpha_i^0 = f_\alpha(Z_i),\ \beta_i^0 = f_\beta(Z_i)$.

A$_3$. The random variables $p_1, p_2, \dots, p_k$ are independent in total.

Let us call the defined above set of models the family of the beta-mixed hierarchical models for grouped binary data. A concrete model is determined by a pair of functions $f_\alpha(.) : R^m \to R^+$, $f_\beta(.) : R^m \to R^+$.

Suppose that the data $B = \{B_1, B_2, ..., B_k\}$ is distorted by random errors $\{\eta_{ij}\}$ as

$$\tilde{b}_{ij} = b_{ij} \oplus \eta_{ij}, \quad P\{\eta_{ij} = 1 \mid b_{ij} = r\} = \varepsilon_r^0, \quad \varepsilon_r^0 \in [\varepsilon_r^{min}, \varepsilon_r^{max}], \quad r = 0, 1, \tag{1}$$

where $\oplus$ is the exclusive logical XOR operator, $\{\eta_{ij}\}$ are the independent Bernoulli random variables ($i = 1, 2, \dots, k,\ j = 1, 2, \dots, n_i$), ε_r^0 are the true unknown values of the distortion levels, and the $\varepsilon_r^{min}, \varepsilon_r^{max}$ are the known lower and upper bounds for ε_r^0.

The problem is to forecast the unknown success probabilities $p_1, p_2, \dots, p_k$ having the distorted sample $X = \{x_1, x_2, \dots, x_n\}$ that is calculated using the contaminated matrix $\tilde{B}$ as $x_i = \sum_{j=1}^{n} \tilde{b}_{ij}$.

3. Robust Forecasting

In our previous work [8], it was proved that for the case of known distortion levels $\varepsilon_0^0, \varepsilon_1^0$, the Bayes forecast function for the i-th object can be expressed as

$$\widehat{p}_\varepsilon^i(s;\varepsilon_0,\varepsilon_1) = \sum\nolimits_{l=0}^{n} \omega_{sl}^i(\varepsilon_0,\varepsilon_1)\cdot(\alpha_0^i + l)\big/(\alpha_0^i+\beta_0^i+n_i), \tag{2}$$

where $s = x_i$ is the number of successes observed for the object i, and

$$\omega_{sl}^i(\varepsilon_0,\varepsilon_1) = C_{n_i}^l\, w_{sl}^i(\varepsilon_0,\varepsilon_1)\, \alpha_0^{i\,[l+]}\beta_0^{i\,[(n_i-l)+]}\Big/\sum\nolimits_{j=0}^{n_i}(C_{n_i}^j\, w_{sj}(\varepsilon_0,\varepsilon_1)\, \alpha_0^{i\,[j+]}\beta_0^{i\,[(n_i-j)+]}),$$

$$w_{sl}(\varepsilon_0,\varepsilon_1) = \sum_{t=\max(l,s)}^{\min(n,l+s)} C_l^{t-s} C_{n-l}^{t-l}\varepsilon_0^{t-l}(1-\varepsilon_0)^{n-t}\varepsilon_1^{t-s}(1-\varepsilon_1)^{l+s-t}, \quad s,l = 0,1,\ldots,n.$$

Let us consider the problem of construction the minimax forecast of the response probability p_i that provides the minimum mean square error for the “worst” values of the distortion levels. Since the Bayes forecast (2) is mean square optimal when $\varepsilon_0 = \varepsilon_0^0$, $\varepsilon_1 = \varepsilon_1^0$, the minimax forecast function can be calculated as $\widehat{p}_\varepsilon^i(s;\varepsilon^*)$, where the tuning parameter $\varepsilon^* = (\varepsilon_0^*,\varepsilon_1^*)^T$ is defined as the solution of the following optimization problem

$$\max_{\varepsilon^0\in\mathcal{E}} r^2(\widehat{p}_\varepsilon^i(s;\varepsilon)\,|\,\varepsilon^0) \to \min_{\varepsilon\in\mathcal{E}}. \tag{3}$$

Here $r^2(\widehat{p}_\varepsilon^i(s;\varepsilon)\,|\,\varepsilon^0)$ is the mean square error of the forecast $\widehat{p}_\varepsilon^i(s;\varepsilon)$ on condition that the true values of the distortion levels were ε^0, $\varepsilon = [\varepsilon_0^{min},\varepsilon_0^{max}]\times[\varepsilon_1^{min},\varepsilon_1^{max}]$. The problem (3) is solved in the following Theorem that is based on the given below lemmas.

Lemma 1. *Let $\widehat{p}^i(s)$ be some forecast function of the response probability p_i, then the mean square error of $\widehat{p}^i(s)$ can be expressed as*

$$r^2(\widehat{p}^i) = \alpha_0^{i\,[2+]}\big/(\alpha_0^i+\beta_0^i)^{[2+]} + \sum\nolimits_{l=0}^{n_i}\left(\widehat{p}^{i\,2}(s) - 2\widehat{p}^i(l)\cdot\widehat{p}_\varepsilon^i(l;\varepsilon^0)\right)\cdot\pi_s^{\varepsilon,i}(\varepsilon^0),$$

where $\pi_r^{\varepsilon,i}(\varepsilon_0,\varepsilon_1) = \sum\nolimits_{j=0}^{n_i} w_{rj}^i(\varepsilon_0,\varepsilon_1)\cdot\pi_j^{0,i}, \quad \pi_j^{0,i} = C_{n_i}^j\, B(\alpha_i^0+j,\beta_i^0+n_i-j)\big/B(\alpha_i^0,\beta_i^0)$.

Proof. The mean square error of the forecast $\widehat{p}^i(s)$ is defined as $E\{(p-\widehat{p}^i)^2\} = E\{p^2\} - 2E\{p\cdot\widehat{p}^i\} + E\{(\widehat{p}^i)^2\}$, where p is a Beta random variable with the parameters α_i^0, β_i^0. Using the properties of the Beta distribution, the first summand is expressed as $E\{p^2\} = \alpha_0^{i\,[2+]}\big/(\alpha_0^i+\beta_0^i)^{[2+]}$. Employing the conditional expectation formula, the second summand can be simplified to $E\{p\cdot\widehat{p}^i\} = \sum\nolimits_{l=0}^{n_i}\widehat{p}^i(l)\cdot\widehat{p}_\varepsilon^i(l;\varepsilon_0,\varepsilon_1)\cdot\pi_l^{\varepsilon,i}(\varepsilon_0,\varepsilon_1)$, and the third is calculated by definition as $E\{(\widehat{p}^i)^2\} = \sum\nolimits_{l=0}^{n_i}\widehat{p}^i(l))^2\pi_i^{\varepsilon,i}(\varepsilon_0,\varepsilon_1)$. The lemma is proved. ■

Lemma 2. *The mathematical expectation of the function $f(\xi) = (\alpha+\xi-1)^{-1}$ of the beta-binomial random variable ξ with the parameters n, α, β is calculated as*

$$E\{(\alpha+\xi-1)^{-1}\} = \frac{\alpha+\beta-1}{(\alpha-1)(\alpha+\beta+n-1)}.$$

Proof. By definition, the mathematical expectation of $f(\xi)$ is expressed as

$$E\{(\alpha+\xi-1)^{-1}\} = \sum_{r=0}^{n}\frac{1}{\alpha+r-1}\cdot C_n^r\,\frac{\prod_{j=0}^{r-1}(\alpha+j)\cdot\prod_{j=0}^{n-r-1}(\beta+j)}{\prod_{j=0}^{n-1}(\alpha+\beta+j)}.$$

Let us transform the latter expression in the following way

$$E\{(\alpha+\xi-1)^{-1}\} = \frac{1}{\alpha-1}\cdot\frac{\alpha+\beta-1}{\alpha+\beta+n-1}\cdot\sum_{r=0}^{n}C_n^r\,\frac{\prod_{j=0}^{r-1}(\alpha-1+j)\cdot\prod_{j=0}^{n-r-1}(\beta+j)}{\prod_{j=0}^{n-1}(\alpha-1+\beta+j)}.$$

Since the sum by r contains the elements of the probability row for the beta-binomial distribution with the parameters n, $\alpha - 1$, β, it is equal to one. It proves the lemma. ■

Lemma 3. *For the conditional mean square error of the forecast function $\widehat{p}_\varepsilon^i(s;\varepsilon^*)$, the following asymptotic expansion holds*:

$$r^2(\hat{p}^i_\varepsilon(s;\varepsilon^*)\,|\,\varepsilon^0)=r_0^2+\frac{n_i\beta_0^i}{(\alpha_0^i+\beta_0^i)(\alpha_0^i+\beta_0^i+n_i)^2}\cdot\varepsilon_0^0+\frac{n_i\alpha_0^i}{(\alpha_0^i+\beta_0^i)(\alpha_0^i+\beta_0^i+n_i)^2}\cdot\varepsilon_1^0+o(\varepsilon^0,\varepsilon^*),$$

where $r_0^2=\alpha_i^0\beta_i^0\big/((\alpha_i^0+\beta_i^0)(\alpha_i^0+\beta_i^0+1)(\alpha_i^0+\beta_i^0+n_i))$.

Proof. Using the results from our previous work [8], one can show that the following asymptotic expansions for the forecast functions $\hat{p}^i_\varepsilon(s;\varepsilon^*)$, $\hat{p}^i_\varepsilon(s;\varepsilon^0)$ hold

$$\hat{p}^i_\varepsilon(s\,|\,\varepsilon^*)=\hat{p}^i_0(s)\cdot\big(1+d_0^i(s)\varepsilon_0^*+d_1^i(s)\varepsilon_1^*\big)+o(\varepsilon_0^*,\varepsilon_1^*),$$

$$\hat{p}^i_\varepsilon(s\,|\,\varepsilon^0)=\hat{p}^i_0(s)\cdot\big(1+d_0^i(s)\varepsilon_0^0+d_1^i(s)\varepsilon_1^0\big)+o(\varepsilon_0^0,\varepsilon_1^0),$$

where $\hat{p}^i_0(s)$ is the classical Bayes forecast for the beta-mixed hierarchical models

$$\hat{p}^i_0(s)=E\{p_i\,|\,s\}=(\alpha_i^0+s)\big/(\alpha_i^0+\beta_i^0+n_i)$$

and the coefficients are calculated as

$$d_0^i(s)=-\frac{s(\beta_0^i+n_i-s)}{(\alpha_0^i+s)(\alpha_0^i+s-1)},\quad d_1^i(s)=\frac{n_i-s}{\beta_0^i+n_i-s-1}.$$

Besides, as follows from [6], $\pi_s^{\varepsilon,i}$ satisfies the following asymptotic expansion

$$\pi_s^{\varepsilon,i}=\pi_s^{0,i}\cdot\big(1+c_0^i(s)\cdot\varepsilon_0+c_1^i(s)\cdot\varepsilon_1\big)+o(\varepsilon_0,\varepsilon_1),$$

where

$$c_0^i(s)=\frac{s(\beta_0^i+n_i-s)}{\alpha_0^i+s-1}-(n_i-s),\quad c_1^i(s)=\frac{(n_i-s)(\alpha_0^i+s)}{\beta_0^i+n_i-s-1}-s.$$

Let us plug the given above expansions into the expression for the conditional mean square forecast error $r^2(\hat{p}^i_\varepsilon(s;\varepsilon)\,|\,\varepsilon^0)$ and collect the terms with ε_0^0, ε_1^0, ε_0^*, ε_1^*. The terms with ε_0^*, ε_1^* will be equal to zero; for the term with ε_0^0, the following expression holds

$$\mu_0^i=\sum_{s=0}^{n_i}\left(\frac{\alpha_0^i+s}{\alpha_0^i+\beta_0^i+n_i}\right)^2\cdot C_{n_i}^s\frac{B(\alpha_0^i+s,\beta_0^i+n_i-s)}{B(\alpha_0^i,\beta_0^i)}\cdot\left(\frac{2s(\beta_0^i+n_i-s)}{(\alpha_0^i+s)(\alpha_0^i+s-1)}-s\frac{\beta_0^i+n_i-s}{\alpha_0^i+s-1}+(n_i-s)\right).$$

The given sum can be considered the mathematical expectation of some function of the beta-binomial random variable ξ with the parameters n, α, β:

$$\mu_0^i=(\alpha_0^i+\beta_0^i+n_i)^{-2}\cdot E\left\{h(\xi)\cdot\frac{\alpha_0^i+\xi}{\alpha_0^i+\xi-1}\right\},$$

where $h(\xi)=(\alpha_0^i+\beta_0^i+1)\cdot\xi^2+\big({\alpha_0^i}^2+(-n_i+\beta_0^i-1)\alpha_0^i-2\beta_0^i-n_i\big)\cdot\xi+n_i\alpha_0^i(1-\alpha_0^i)$. Denote $\eta=\alpha_0+\xi-1$, then

$$\mu_0^i=(\alpha_0^i+\beta_0^i+n_i)^{-2}\cdot E\big\{h(\xi)+(v_2\cdot\eta+v_1+v_0\eta^{-1})\big\},\tag{4}$$

where

$$v_2=(\alpha_0^i+\beta_0^i+1)^2,\ v_1={\alpha_0^i}^2+(\beta_0^i-n_i-1)\alpha_0^i-(2\beta_0^i+n_i),\ v_0={\alpha_0^i}^2+(\beta_0^i+n_i-2)+(1-n_i-\beta_0^i).$$

Employing now the properties of the beta-binomial distribution and the result of Lemma 2:

$$E\{\xi\}=\frac{n_i\alpha_0^i}{\alpha_0^i+\beta_0^i},\ E\{\xi^2\}=\frac{n_i\alpha_0^i}{\alpha_0^i+\beta_0^i}+\frac{n_i^{[2-]}\cdot{\alpha_0^i}^{[2+]}}{(\alpha_0^i+\beta_0^i)^{[2+]}},\ E\{\eta^{-1}\}=\frac{\alpha_0^i+\beta_0^i-1}{(\alpha_0^i-1)(\alpha_0^i+\beta_0^i+n_i-1)},$$

to the expression (4) reduces μ_0^i to the coefficient for ε_0^0 given in the statement of this lemma. The coefficient for ε_0^1 is derived in the same way. ■

Theorem. *For the beta-mixed hierarchical model of grouped binary data under distortions* (1), *the minimax forecast* $\hat{p}^i_\varepsilon(s;\varepsilon^*)$ *is calculated on the upper bound of the distortion levels interval* $\varepsilon^*=\varepsilon^{max}$.

Proof. As follows from Lemmas 1, 3, the conditional mean square forecast error can be expressed as $r^2(\hat{p}^i_\varepsilon(s;\varepsilon^*)\,|\,\varepsilon^0)=r_0^2+\mu_0^i\cdot\varepsilon_0^0+\mu_1^i\cdot\varepsilon_1^0+o(.)$ where $\mu_0^i,\mu_1^i>0$. Hence, when ε^* is fixed, the maximum of $r^2(.)$ is achieved when $\varepsilon^0=\varepsilon^{max}$. On the other hand, from the properties of the Bayes forecast follows that when ε^0 is fixed, the minimum of $r^2(.)$ is reached when $\varepsilon^*=\varepsilon^0$. As a result, the solution of the problem (3) is $\varepsilon^*=\varepsilon^{max}$. ■

4. Computer simulation results

To demonstrate the performance of the developed robust forecasting technique, the following computer simulation was made. Assuming that the true parameter values of the beta distribution were $\alpha = 0.5$, $\beta = 9.5$, there were generated $k = 10000$ realizations of the corresponding beta random variable. Then, for each realization (object), a random Bernoulli sample of size $n = 10$ was generated. Every sample was distorted using the expression (1); it was assumed that $\varepsilon_0^0 = \varepsilon_1^0 \in [0, 0.05]$ and the true value of the distortion level ε belonged to the set $\{0, 0.01, \ldots, 0.05\}$. For each distortion level, the classical and proposed predictors of the response probabilities were calculated. Finally, 95%-confidence intervals of the mean square forecast error were computed for both predictors. The results of the computer simulation allow making a conclusion that the proposed robust minimax predictor ensures much lower mean square forecast error when compared to the classical approach. In particular, for the case $\varepsilon = 0.05$, the classical predictor leads to the error 0.070, while the error of the proposed predictor is only 0.054 that is 1.3 times better.

5. Conclusion

In the case of known lower and upper bounds of the additive stochastic distortion levels, the developed robust minimax Bayes predictor for the beta-mixed hierarchical models of grouped binary data ensures much lower mean square forecast error then the classical Bayes predictor.

References

[1]. Brooks S.P. "On Bayesian analysis and finite mixtures for proportions", *Statistics and Computing*, **11**, (2001), 179-190.

[2]. Collet, D. *Modelling Binary Data*. London: Champton and Hall/CRC, (2002).

[3]. Coull, B.A., and Agresti, A. "Random effects modelling of multiple binomial responses using the multivariate binomial logit-normal distribution", *Biometrics*, **56**, (2000), 73-80.

[4]. Diggle, P.J., Heagerty, P., Liang, K.-Y. and Zeger, S.L. *Analysis of Longitudinal Data*. Oxford: University Press, (2002).

[5]. Kharin Yu. *Robustness in Statistical Pattern Recognition*. Kluwer Academic Publishers, Dordrecht, (1996).

[6]. Kharin Yu.S., Pashkevich M.A. "Statistical Estimation of Parameters for the Beta-Binomial Distribution under Distortions of Binary Observations", *Journal of the National Academy of Sciences of Belarus (Series in Physics and Mathematics)*, **1**, (2003), 11-17.

[7]. Neuhaus, J.M. "Analysis of clustered and longitudinal binary data subject to response misclassification", *Biometrics*, **58**, (2002), 675-683.

[8]. Pashkevich M., Kharin Yu. "Robust Analysis of Clustered Binary Data: Beta-Binomial Model under Response Misclassification", *Proceeding of the 7th International Conference on Pattern Recognition and Information Processing*, Minsk, Belarus, May 21-23, (2003), 32-36.