*Proceedings of the 4th International Conference on Statistics: Theory and Applications (ICSTA'22)*
*Prague, Czech Republic – July 28- 30, 2022*
*Paper No. 149*
*DOI: 10.11159/icsta22.149*

# Jackknife Estimator Consistency for Nonlinear Mixture

**Rostyslav Maiboroda[1], Vitaliy MIroshnychenko[2]**
[1]Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine
mre@univ.kiev.ua; vitaliy.miroshnychenko@gmail.com
[2]Taras Shevchenko National University of Kyiv
Volodymyrska str., 60, Kyiv, Ukraine

## Extended Abstract

This paper continues our studies of the jackknife (JK) technique application for estimation of estimators' covariance matrices in models of mixture with varying concentrations (MVC) [2, 3]. On JK applications for homogeneous samples, see [1]. In MVC models one deals with a non-homogeneous sample, which consists of subjects belonging to $M$ different sub-populations (mixture components). One knows the probabilities with which a subject belongs to the mixture components and these probabilities are different for different subjects. Therefore, the considered observations are independent but not identically distributed.

We consider objects from a mixture with various concentrations. All objects from the sample $\Xi_n$ belongs to one of M different mixture components. Each object from the sample $\Xi_n = (\xi_j)_{j=1}^{n}$has observed characteristics $\xi_j = (X_j, Y_j) \in \mathbb{R}^D$ and one hidden $\kappa_j$. $\kappa_j = m$ if $j$-th objects belongs to the $m$-th component. These numbers are unknown, but we know the mixing probabilities $p_{j;n}^{m} = P\{\kappa_j = m\}$. The $X_j$ is a vector of regressors and $Y_j$ is a response in the regression model

$$Y_j = g\left(X_j, b^{(\kappa_j)}\right) + \varepsilon_j. \tag{1}$$

Here $b^{(m)} \in \Theta \subseteq \mathbb{R}^d$ is a vector of unknown regression parameters for the $m$-th component, the $g: \mathbb{R}^{D-1} \times \Theta \to \mathbb{R}$ is a known regression function, $\varepsilon_j$ is a regression error term. Random variables $X_j$ and $\varepsilon_j$ are independent and their distribution is different for different components.

The estimator $\hat{b}_n^{(k)}$ for the regression parameter $b^{(k)}$ is a measurable solution to the GEE equation

$$S_n^k(\gamma) = \sum_{j=1}^{n} a_{j;n}^k \left(Y_j - g\left(X_j, \gamma\right)\right) \dot{g}\left(X_j, \gamma\right) = 0. \tag{2}$$

This equation might have more than one solution. Any of these solutions could be taken as $\hat{b}_n^{(k)}$ estimator. The symbol $\dot{g}(x, \gamma)$ means the gradient of function $g$ by the $\gamma$ term. $a_{j;n}^k$ are the minimax weights defined in [4]. The minimax weights matrix $A_{;n}$ defined using the mixing matrix $P_{;n} = \left(p_{j;n}^k\right)_{j=1,k=1}^{n,M}$:

$$A_{;n} = \left(a_{j;n}^k\right)_{j=1,k=1}^{n,M} = P_{;n}\left(P_{;n}^T P_{;n}\right)^{-1}. \tag{3}$$

In [3] it is shown that under suitable conditions $\hat{b}_n^{(k)}$ are asymptotically normal, i.e.

$$\sqrt{n}\hat{V}_n^{-\frac{1}{2}}\left(\hat{b}_n^{(k)} - b^{(k)}\right) \to^W N_d(0, E), \text{as n} \to \infty. \tag{4}$$

$$\hat{V}_n = {M^{(k)}}^{-1} Var\, S_n^k\left(b^{(k)}\right)\left({M^{(k)}}^T\right)^{-1};\; M^{(k)} = -E\dot{S}_n^k\left(b^{(k)}\right).$$

Let us denote $V^{(k)} = \lim_{n\to\infty} \hat{V}_n$ as a limit covariance matrix. For the nonlinear functions $g$, the matrix $V^{(k)}$ could be estimated by the following JK-estimator

$$^{JK}\hat{V}_n^{(k)} = \sum_{j=1}^{n}\left(\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right)\left(\hat{b}_{i-;n}^{(k)} - \hat{b}_n^{(k)}\right)^T. \tag{5}$$

Here $\hat{b}_{i-;n}^{(k)}$ are the regression parameter estimators build by the samples $\Xi_{i-;n}$, which are the sample $\Xi_n$ without $j$-th object and matrices $A_{i-;n} = P_{i-;n}\left(P_{i-;n}^T P_{i-;n}\right)^{-1}$. The matrix $P_{i-;n}$ is the matrix $P_{;n}$ without the $j$-th object.

Let us denote $s(\xi_j,\gamma) = \left(Y_j - g(X_j,\gamma)\right)\nabla g(X_j,\gamma)$.

**Theorem** Assume that following assumptions hold:

1.$b^{(k)}$ is an inner point of $\Theta$.

2.$M^{(k)}$ is a finite and non-singular for all $k = 1,\dots,M$.

3.The limits $\lim_{n\to\infty} n\sum_{j=1}^{n} a_{j;n}^{k} a_{j;n}^{m} p_{j;n}^{i} p_{j;n}^{l}$ exist for all $k, m, l, i = 1,\dots,M$.

4.Matrix $\Gamma_\infty = \frac{1}{n}\lim_{n\to\infty} P_{;n}^{T} P_{;n}$ exists and non-singular.

5.There exists a function $h: \mathbb{R}^D \to \mathbb{R}$ such that $\sup_{\gamma\in\Theta}|s(x,\gamma)| \le h(x)$, $\sup_{\gamma\in\Theta}|\dot{s}(x,\gamma)| \le h(x)$, $\sup_{\gamma\in\Theta}|\ddot{s}(x,\gamma)| \le h(x)$, and moreover, $\exists E\left(h(\xi_{(m)})\right)^{\alpha} < \infty$, for $\alpha > 4$ and $m = 1,\dots,M$. $\xi_{(m)}$ is a r.v. related to the $m$-th component.

6.$\hat{b}_n^{(k)}$ is a $\sqrt{n}$-consistent estimator of $b^{(k)}$.

7. $\sup_{i=1,\dots,n}\left|\hat{b}_{i-;n}^{(k)} - b^{(k)}\right| \to^{P} 0$ as $n \to \infty$.

Than

$$^{JK}\hat{V}_n^{(k)} - V^{(k)} \to^{P} 0, \text{as } n \to \infty. \quad (6)$$

## References


[1] J. Shao and D. Tu, *The Jackknife and Bootstrap*, Springer-Verlag, 2012.

[2] R. Maiboroda, O. Sugakova, *Jackknife covariance matrix estimation for observations from mixture*. Modern Stochastics: Theory and Applications, Vol.6, Iss.4 pp. 495 - 513, - 2019.

[3] V. Miroshnychenko and R. Maiboroda, *Asymptotic normality of modified LS estimator for mixture of nonlinear regressions*, Modern Stochastics: Theory and Applications, Vol.7, Iss.7 pp. 435 - 448, - 2020.

[4] R. Maiboroda and O. Sugakova, *Statistics of mixtures with varying concentrations with application to DNA microarray data analysis*. Journal of nonparametric statistics. 24 , No 1 201-205 (2012).