Strong Consistency of the Non-parametric Maximum Likelihood Estimator of Correlated Normal Random Variables

Xiangjie Xue, Yong Wang

Department of Statistics, University of Auckland, New Zealand.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

SCIENCE
DEPARTMENT OF STATISTICS

Background

Maximum likelihood estimation has its fundamental role in the field of statistics because of its general utility. Based on the proofs in Wald (1949), Kiefer and Wolfowitz (1956) was able to extend the consistency result from parametric MLE to non-parametric MLE (NPMLE) under the independence assumption. It is shown that under regularity conditions, the maximum likelihood estimator of a structural parameter is strongly consistent, when the (infinitely many) incidental parameters are independently distributed variables with a common unknown distribution function, and that the latter is also consistently estimated although it is nor assumed to belong to a parametric class. We take it a little bit further by considering the special yet important problem of estimating the distribution function of location parameters of weakly correlated normal random variables. It is shown that under certain conditions, the consistency of NPMLE can be established if the likelihood function for the independent variables is used as the likelihood function for the weakly correlated variables.

Assumptions

Let $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{\mu}, \sigma^2\boldsymbol{R})$ where $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_p)^T$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_p)^T$ and $\boldsymbol{R}$ is the correlation matrix of size $p \times p$. Let $G_p(z) = p^{-1}\sum_{i=1}^{p} \mathbb{1}(\mu_i \le z)$ and G be the limit of G_p as $p \to \infty$.

Assumption 1: G has finite second moment. That is,

$$\int \mu^2 dG(\mu) = \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} \mu_i^2 < \infty.$$

Assumption 2: There exists $\epsilon > 0$ such that

$$\lim_{p\to\infty} \frac{1}{p^{2-\epsilon}} \sum_{i,j=1}^{p} |\mathrm{Cor}(Z_i, Z_j)| = \lim_{p\to\infty} \frac{1}{p^{2-\epsilon}} \sum_{i,j=1}^{p} |r_{ij}| = 0.$$

Assumption 2 is very similar to the definition of weak correlation in Azriel and Schwartzman (2015). The introduction of ϵ is to make sure that the Strong Law of Large Numbers (SLLN) in Lyons (1988) holds when we make use of it.

Consistency

Let $\mathcal{G}$ be the set of all CDFs with finite second moment. We use the metric defined by Kiefer and Wolfowitz (1956), that is, for $G_a, G_b \in \mathcal{G}$,

$$d(G_a, G_b) = \int |G_a(\theta) - G_b(\theta)| e^{-|\theta|} d\tau(\theta).$$

Proposition: If Assumptions 1 and 2 hold, then the NPMLE $\hat{G}_p$ that satisfies

$$\hat{G}_p = \arg\max_{G \in \bar{\mathcal{G}}} \sum_{i=1}^{p} \log \int \phi\left(\frac{z_i - \theta}{\sigma}\right) dG(\theta)$$

converges weakly to G almost surely. $\bar{\mathcal{G}}$ is the complete space of $\mathcal{G}$ (the space together with the limits of its Cauchy sequences in the sense of the metric defined above). Note that the log-likelihood function above is the log-likelihood function for independent random variables.

Reference

Azriel, D. and Schwartzman, A. (2015). The empirical distribution of a large number of correlated normal variables. *Journal of the American Statistical Association*, 110(511):1217–1228.

Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, 27(4):887–906.

Lyons, R. (1988). Strong laws of large numbers for weakly correlated random variables. *Michigan Math. J.*, 35.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *The Annals of mathematical statistics*, 20:595 – 601.

Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 69(2):185–198.

Wang, Y. (2017). *nspmix: Nonparametric and Semiparametric Mixture Estimation*. R package version 1.4-0. https://CRAN.R-project.org/package=nspmix.

Address: Level 3, Building 303, 38 Princess St, Auckland, New Zealand 1010.
Email: x.xue@auckland.ac.nz; yongwang@auckland.ac.nz;

Simulation Results

For each p value specified, we generated $\boldsymbol{z} = (z_1, z_2, \cdots, z_p)^T$ from a p-variate normal random variables with mean $\boldsymbol{\mu} = (-1, 1, \cdots, -1, 1)^T$ and covariance $\boldsymbol{\Sigma}$ with $\sigma_{ij} = \rho^{|i-j|}$, for $\rho = 0.5$ (AR1). Let $\epsilon = 0.5$. Then

$$\lim_{p\to\infty} \frac{1}{p^{3/2}} \sum_{i,j=1}^{p} |r_{ij}| \le \lim_{p\to\infty} \frac{2p}{p^{3/2}} \sum_{i=0}^{\infty} \rho^i = \lim_{p\to\infty} \frac{4p}{p^{3/2}} = 0,$$

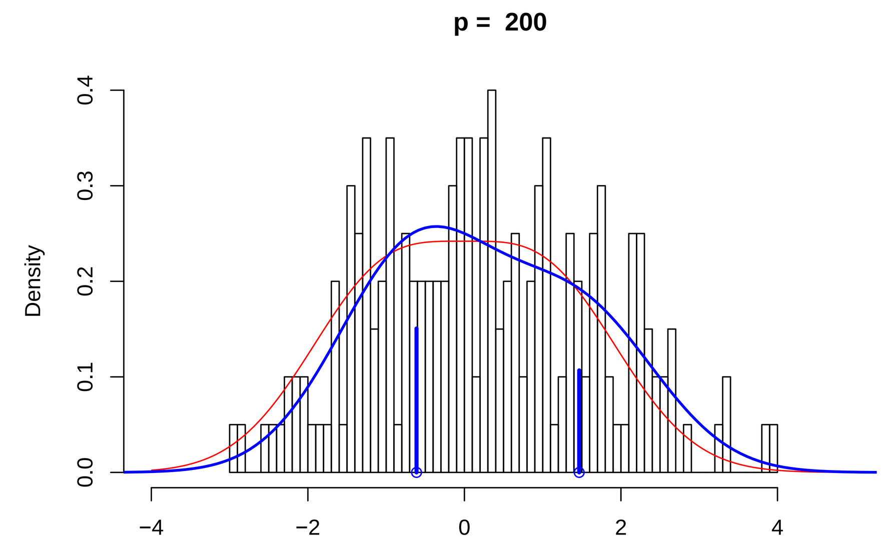
that is, Assumption 2 holds. Assumption 1 also holds since

$$\lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} \mu_i^2 \le \lim_{p\to\infty} \frac{1}{p}\sum_{i=1}^{p} 1 = 1 < \infty.$$

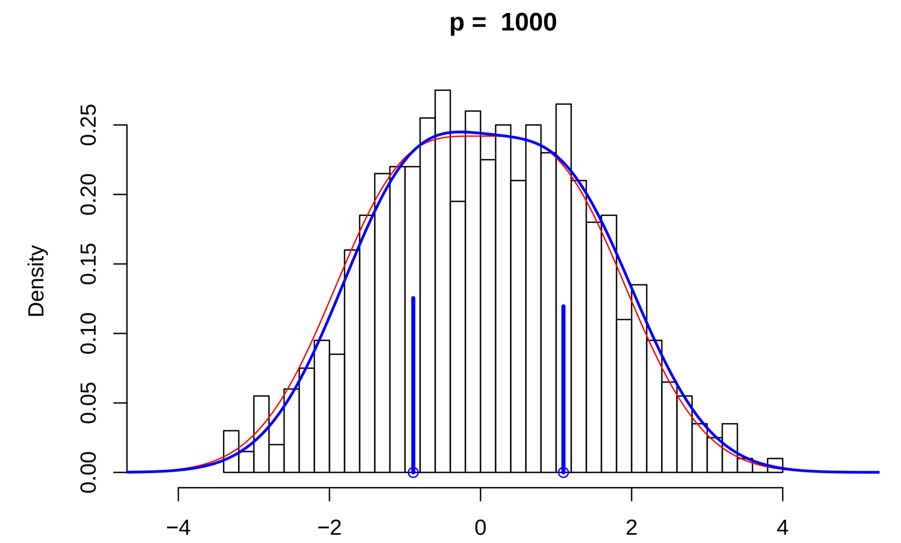
Wang (2007) studied the computation of the NPMLE of a mixing distribution and came up with a quadratically convergent method. Since we are dealing with finite sample, the estimator may be rewritten as

$$\hat{G}_p = \arg\max_{G\in\bar{\Theta}} \sum_{i=1}^{p} \log \int \phi\left(\frac{z_i - \theta}{\sigma}\right) dG(\theta) = \arg\max_{G\in\bar{\Theta}} \sum_{i=1}^{p} \log \sum_{j=1}^{m} \pi_j \phi\left(\frac{z_i - \theta_j}{\sigma}\right),$$

where $\pi_j > 0$ is the weight put on each component density with mean θ_j for $j = 1, \ldots, m$, $\sum_{j=1}^{m} \pi_j = 1$. Hence the NPMLE $\hat{G}_p$ of G_p can be computed using this method which is implemented in the R package `nspmix` (Wang, 2017), available on CRAN. Note that in the process of calculating the NPMLE, we do not specify the number of true groups, nor calculate $\hat{G}$ conditioned on the number of groups and use an information criterion to find the "best" number of groups. This is essentially different from the maximum likelihood estimate for parametric mixtures.



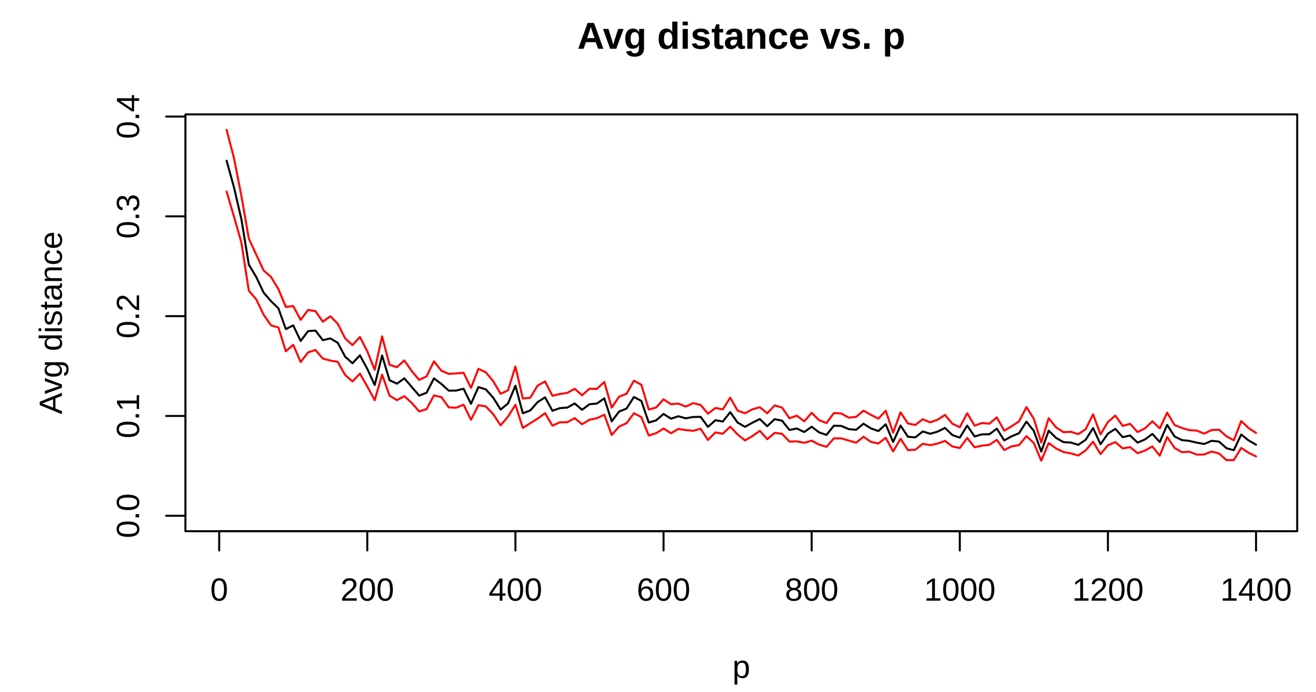
(a) $\hat{G}_{200}(z) = 0.585\mathbb{1}(z \ge -0.612) + 0.415\mathbb{1}(z \ge 1.467)$ and $d(\hat{G}_{200}, G) = 0.237$.



(b) $\hat{G}_{1000}(z) = 0.512\mathbb{1}(z \ge -0.89) + 0.488\mathbb{1}(z \ge 1.098)$ and $d(\hat{G}_{1000}, G) = 0.053$.

Histograms of independently generated observations for $p = 200, 1000$ respectively are shown above. The red line stands for the true density function while the blue line stands for the density function where NPMLE is used. The blue vertical lines show the locations and weights of the component means in $\hat{G}_p$.

The graph of mean distances between G_p and 100 iid $\hat{G}_p$ plotted against p is shown below. The distance is again the metric defined in Kiefer and Wolfowitz (1956). The black lines represents the mean distance for each p and the red dotted line shows the boundaries of the 95% confidence interval.



All the numerical studies above was performed in R (R Core Team, 2019), version 3.5.3.

LaTeX TikZposter