Теорія Ймовір. та Матем. Статист.
Вип. 69, 2003

Theor. Probability and Math. Statist.
No. 69, 2004, Pages 95–102
S 0094-9000(05)00617-4
Article electronically published on February 8, 2005

ASYMPTOTIC NORMALITY OF IMPROVED WEIGHTED EMPIRICAL DISTRIBUTION FUNCTIONS

UDC 519.21

R. MAĬBORODA AND O. KUBAĬCHUK

Abstract. Weighted empirical distribution functions are often used to estimate the distributions of components in a mixture. However, weighted empirical distribution functions do not possess some properties of probability distribution functions in the case of negative weight coefficients. We consider a method allowing one to improve weighted empirical distribution functions and obtain an estimator that is a distribution function. We prove that this estimator is asymptotically normal. The limit distribution of the improved weighted empirical distribution function coincides with that of the initial estimator.

1. Introduction

Let $\Xi_N = \{\xi_{1:N}, \dots, \xi_{N:N}\}$ be a sample from a mixture with varying concentrations, that is, $\xi_{j:N}$, $j = 1, \dots, N$, are jointly independent random variables, and

$$\mathsf{P}\{\xi_{j:N} < x\} = \sum_{m=1}^{M} w_{j:N}^m H_m(x)$$

where M is the total number of components in the mixture, H_m is the distribution function of the component m, and $w_{j:N}^m$ is the concentration of the component m for the observation j, that is, the probability that an object of the component m occurs. (It is clear that $w_{j:N}^m \ge 0$ and $\sum_{m=1}^M w_{j:N}^m = 1$.) We assume in this paper that the concentrations of components are known. Thus the problem is to estimate the distribution functions H_k.

The weighted empirical distribution functions

$$\hat{F}_N(x, a) = \frac{1}{N} \sum_{j=1}^{N} a_{j:N} \mathbf{1}\{\xi_{j:N} < x\} \tag{1}$$

are proposed in [1] as estimators for H_k, where

$$a = (a_{1:N}, \dots, a_{N:N})$$

is a nonrandom vector of weight coefficients. It is shown in [1] that these estimators are unbiased, consistent, asymptotically normal, and minimax for appropriate weight coefficients. If, however, some coefficients $a_{j:N}$ are negative, then the function $\hat{F}_N(x, a)$ is not nondecreasing, and therefore it is not a probability distribution function. This circumstance does not play an important role for some applications but for some others

2000 *Mathematics Subject Classification.* Primary 62G30; Secondary 62G20.

it does. For example, this is the case for the bootstrap method where a crucial assumption is that the functions are probability distributions (otherwise one cannot simulate a bootstrap sample).

One can improve the weighted empirical distribution functions $\hat{F}_N(x,a)$ by putting

$$F_N^+(x,a) = \sup_{y<x} \hat{F}_N(y,a). \tag{2}$$

The function $F_N^+(x,a)$ assumes only positive values and is nondecreasing. However, it may assume values greater than 1. Thus we consider the function

$$F_N^*(x,a) = \min\left(1, F^+(x,a)\right). \tag{3}$$

In what follows we describe an effective procedure to evaluate improved weighted empirical distribution functions and study their asymptotic behavior. Under certain conditions we show that they are asymptotically normal estimators and their limit distribution is the same as that of the weighted empirical distribution functions defined by (1). Thus the asymptotic behavior of the empirical process

$$B_N^+(x) = \sqrt{N}\left(F_N^+(x,a) - H_k(x)\right) \tag{4}$$

is the same as that of the empirical process

$$B_N(x) = \sqrt{N}\left(\hat{F}_N(x,a) - H_k(x)\right). \tag{5}$$

2. The procedure for evaluating improved weighted empirical distribution functions

First we assume that all members of the sample $\Xi_N = (\xi_{1:N}, \dots, \xi_{N:N})$ are distinct. Denote by σ the permutation of numbers $1, 2, \dots, N$ for which the members of the sample are arranged in ascending order: $\xi_{\sigma(1):N} < \xi_{\sigma(2):N} < \cdots < \xi_{\sigma(N):N}$. (The numbers $\sigma(j)$, $j = 1, \dots, N$, are called the "inverse ranks", since $\sigma^{-1}(j)$ is the rank of the observation j in the sample.) Since the function $\hat{F}_N(x,a)$ is constant on the intervals $(\xi_{\sigma(j):N}, \xi_{\sigma(j+1):N})$, so is the function $F_N^+(x,a)$ defined by (2). Thus

$$F_N^+(x,a) = \frac{1}{N}\sum_{j=1}^N b_j^+ \mathbf{1}\left\{\xi_{j:N} < x\right\} = \frac{1}{N}\sum_{j=1}^N b_{\sigma(j):N}^+ \mathbf{1}\left\{\xi_{\sigma(j):N} < x\right\}$$

where b_j^+ are some coefficients that depend (in contrast to $a_{j:N}$) on the sample Ξ_N.

The idea of the procedure is as follows. Moving from left to right along the sequence of order statistics, we consecutively improve the coefficients $a_{\sigma(j):N}$ so that the sum

$$\hat{S}_j^N = N\hat{F}_N\left(\xi_{\sigma(j):N}, a\right) = \sum_{i\colon \xi_{i:N} \le \xi_{\sigma(j):N}} a_{i:N}$$

become "lower" than all its predecessors.

Thus the procedure is as follows:

(1) evaluate the inverse ranks $\sigma(j)$, $j = 1, \dots, N$, in the sample Ξ_N.
(2) Put $b_{\sigma(1)}^+ = \max(a_{\sigma(1):N}, 0)$, $\hat{S}_1 = a_{\sigma(1):N}$, $S_1^+ = b_{\sigma(1)}^+$.
(3) For j from 2 to N perform
  (a) $\hat{S}_j = \hat{S}_{j-1} + a_{\sigma(j):N}$;
  (b) $b_{\sigma(j)}^+ = \max\left(\hat{S}_j - S_{j-1}^+, 0\right)$;
  (c) $S_j^+ = S_{j-1}^+ + b_{\sigma(j)}^+$.

To evaluate the coefficients b_j^* for the function F_N^* defined by equality (3), Step (3) of this procedure must contain the following algorithm: until $\hat{S}_j < N$ put $b^*_{\sigma(j)} = b^+_{\sigma(j)}$; otherwise if $\hat{S}_{j_0} \le N$ for some j_0, put $b^*_{\sigma(j_0)} = N - \hat{S}_{j_0}$ and $b^*_{\sigma(j)} = 0$ for all $j > j_0$.

Note that the procedure of evaluating inverse ranks is similar to the sorting algorithm for a sample. The number of operations required by fast sorting algorithms is of order $CN \ln N$. Steps (2) and (3) require CN operations. Thus the total number of operations needed to evaluate the coefficients b^+ and b^* is of order $CN \ln N$. Algorithms that require a total number of operations of this order are called fast.

If there are several equal members in a sample, say $\xi_{j_1:N} = \xi_{j_2:N} = \cdots = \xi_{j_l:N}$, then it is reasonable to remove all of them from the sample except for $\xi_{j_1:N}$ and to change its weight coefficient to $a^*_{j_1:N} = a_{j_1:N} + \cdots + a_{j_l:N}$. After this change, the coefficients of improved weighted empirical distribution functions can be evaluated according to the described procedure.

3. The asymptotic behavior of weighted empirical distribution functions

Before we start the study of the asymptotic behavior of improved weighted empirical distribution functions, we recall some results concerning usual weighted empirical distribution functions defined by (1). Let $\hat{F}_N(x,a)$ be regarded as an estimator of $H_k(x)$ and let the weight coefficients $a = a^k$ be such that $\hat{F}_N(x,a)$ is an unbiased estimator. It is known that the following conditions are sufficient for $\hat{F}_N(x,a)$ to be unbiased:

$$\left\langle a^k w^m \right\rangle_N = \mathbf{1}\{k=m\} \quad \text{for all } m = 1,\dots,M, \tag{6}$$

where $\langle \cdot \rangle_N$ is the average over the whole sample: $\langle a \rangle_N = N^{-1}\sum_{j=1}^N a_{j:N}$. We denote by $\langle a \rangle$ the limit $\langle a \rangle = \lim_{N\to\infty} \langle a \rangle_N$ if it exists.

Theorem 3.1. *Let*

(1) *for some $A < \infty$,*

$$\sup_{j,N} |a_{j:N}| < A;$$

(2) *the limits $\langle w^l w^m (a^k)^2 \rangle$ exist for all $l, m = 1,\dots,M$;*
(3) *H_m are continuous functions on $\mathbb{R}$ for all $m = 1,\dots,M$;*
(4) *condition* (6) *holds.*

Then the processes $\tilde{B}_N(x)$ and $B(x)$ can be defined on a common probability space such that

(1) *the processes $\tilde{B}_N(x)$ have the same distribution as $B_N(x)$;*
(2) *$B(x)$ is a Gaussian stochastic process with almost sure continuous paths and zero mean, and with covariance function given by*

$$\begin{aligned} \mathsf{E}B(x)B(y) = \sum_{m=1}^M \left\langle w^m \left(a^k\right)^2 \right\rangle H_m(\min(x,y)) \\ - \sum_{i,m=1}^M \left\langle w^m w^i \left(a^k\right)^2 \right\rangle H_m(x) H_i(y); \end{aligned}$$

(3) $\sup_{x\in\mathbb{R}} |\tilde{B}_N(x) - B(x)| \to 0$ *almost surely as $N \to \infty$.*

Proof. The assumptions of Theorem 3.1 imply the weak convergence of B_N to B in the space $D(\mathbb{R})$ of functions without discontinuities of the second kind (see Theorem 2 in [2]). The sample continuity of $B(x)$ can be proved in a standard way by using the Dudley condition [3]. Now a theorem by Skorokhod [4] completes the proof of Theorem 3.1. □

4. The asymptotic behavior of improved weighted empirical distribution functions

The process B_N^+ defined by equality (4) can be represented in terms of the process $B_N(x)$ as follows:

$$B_N^+(x) = \sqrt{N}\left(\sup_{y<x}\left(F(y) + B_N(y)/\sqrt{N}\right) - F(x)\right). \tag{7}$$

In what follows we identify the process $B_N(x)$ with the process $\tilde{B}_N(x)$ constructed in Theorem 3.1, since their distributions coincide and we are interested in the weak convergence. The process B_N^+ defined by equality (7) is understood as $B_N^+(x)$.

A point x is called a point of increase of the function H_k if for all $\delta > 0$

$$H_k(x) - H_k(x-\delta) > 0.$$

The set of all points of increase of the function H_k is denoted by $\operatorname{supp} H_k$.

Theorem 4.2. *Let the assumptions of Theorem* 3.1 *hold and*

$$\operatorname{supp} H_m \subseteq \operatorname{supp} H_k$$

for all $m = 1, \dots, M$. *Then*

$$\sup_{x\in\mathbb{R}} \left|B_N^+(x) - B_N(x)\right| \to 0 \tag{8}$$

in probability as $N \to \infty$.

Corollary 4.1. *Let the assumptions of Theorem* 4.2 *hold. Then the process* B_N^+ *weakly converges to the process* B *in the uniform metric in the space* $D(\mathbb{R})$. *Moreover, the process* B_N^* *weakly converges to* B *in the uniform metric in* $D((-\infty, s])$ *for all interior points of increase* $s \in \operatorname{supp} H_k$.

Proof. There are two steps in the proof. First we prove the pointwise convergence in probability, that is, we prove that

$$\mathsf{P}\left\{|B_N^+(x) - B_N(x)| > \varepsilon\right\} \to 0, \qquad N \to \infty, \tag{9}$$

for all $x \in \mathbb{R}$. Then we prove (8) using (9).

Consider the transformation $\xi_{j:N} \to \tilde{\xi}_{j:N} = (2/\pi)\arctan \xi_{j:N}$. Let $\tilde{H}_m$ be the distribution function of the random variable $(2/\pi)\arctan \eta_m$ where η_m is a random variable with the distribution H_m. Then $(\tilde{\xi}_{j:N}, j = 1, \dots, N)$ is a sample from a mixture with varying concentrations for which the distributions of components are $\tilde{H}_m$ and the concentrations are $w_{j:N}^m$. Note that if $\tilde{B}_N$ and $\tilde{B}_N^+$ are the corresponding empirical processes constructed from $(\tilde{\xi}_{j:N}, j = 1, \dots, N)$, then $\tilde{B}_N((2/\pi)\arctan(x)) = B_N(x)$ and

$$\sup_x \left|\tilde{B}_N(x) - \tilde{B}_N^+(x)\right| = \sup_x \left|B_N(x) - B_N^+(x)\right|.$$

Since $\operatorname{supp} \tilde{H}_m \subseteq [-1, 1]$, the latter means that we can restrict our consideration to the case of samples such that $\operatorname{supp} H_m \subseteq [-1, 1]$.

Now we are going to prove relation (9). Since $B_N^+(x) \ge B_N(x)$, we only need to show that for all $\varepsilon > 0$

$$\mathsf{P}\left\{B_N^+ \ge B_N(x) + \varepsilon\right\} \to 0$$

as $N \to \infty$.

If for some $\delta > 0$ and $x \in \mathbb{R}$

$$(x - \delta, x) \cap \operatorname{supp} H_k = \varnothing$$

(therefore $(x-\delta,x)\cap\operatorname{supp} H_m=\varnothing$ for all $m=1,\dots,M$ by assumptions of the theorem), then

$$\mathsf{P}\{\xi_{j:N}\in(x-\delta,x)\}=0,$$

whence

$$\hat F_N(x,a)=\hat F_N(x-\delta,a),\qquad H_k(x)=H_k(x-\delta),\qquad F_N^+(x,a)=F_N^+(x-\delta,a),$$
$$B_N(x)=B_N(x-\delta),\qquad B_N^+(x)=B_N(x-\delta).$$

Put $s(x)=\sup\{x'\in\operatorname{supp} H_k\colon x'<x\}$. Then $s(x)\in\operatorname{supp} H_k$ and

$$B_N(x)=B_N(s(x)),\qquad B_N^+(x)=B_N^+(s(x)).$$

Hence it is sufficient to prove relation (9) only for $x\in\operatorname{supp} H_k$.

Let δ be a number such that $0<\delta<\varepsilon$, $t_0\in\mathbb{R}$, and $r\in\mathbb{N}$. Put

$$t_j=t_0+\delta j,\qquad A_N=\left\{B_N^+(x)\ge B_N(x)+\varepsilon\right\},\qquad A_N^-=\{B_N(x)<t_0\},$$
$$A_N^+=\left\{B_N^+\ge t_r+\varepsilon\right\},\qquad A_N^j=\left\{B_N(x)\in[t_j,t_j+\delta],B_N^+>t_j+\varepsilon\right\}.$$

Then $A_N\subseteq A_N^+\cup A_N^-\bigcup_{j=0}^{r-1}A_N^j$.

Fix $z>0$ and $\varepsilon>0$ and set $\delta=\varepsilon/2$. Now we estimate the probabilities of the events A_N^+, A_N^-, and A_N^j. Since for all fixed $\lambda>0$

$$\mathsf{P}\{|B_N(x)-B(x)|>\lambda\}\to 0$$

as $N\to\infty$ and

$$\mathsf{P}\{|B(x)|>\lambda\}\to 0,\qquad \lambda\to\infty,$$

there is t_0 such that $p_N^-=\mathsf{P}(A_N^-)<\varepsilon/3$ for all sufficiently large N. Since

$$B_N^+(x)>B_N(x),$$

there is a sufficiently large r (and thus there is t_r) such that $p_N^+=\mathsf{P}(A_N^+)<\varepsilon/3$. Fix t_0 and r.

Now we estimate

$$p_N^j=\mathsf{P}\left(A_N^j\right)\le\mathsf{P}\left\{B_N(x)<t_{j+1},B_N^+(x)>t_{j+1}+\delta\right\}$$

(recall that $t_{j+1}=t_j+\delta$ and $\varepsilon=2\delta$). We have

$$\begin{aligned}\left\{B_N^+(x)>t_{j+1}+\delta\right\}&=\left\{\sup_{y<x}(H_k(y)+B_N(y))/\sqrt N>H_k(x)+(t_{j+1}+\delta)/\sqrt N\right\}\\&=\left\{\text{there exists } y\le x\colon B_N(y)>t_{j+1}+\delta+\sqrt N(H_k(x)-H_k(y))\right\}\end{aligned}$$

and hence

$$\begin{aligned}A_N^j\subseteq\Big\{&B_N(x)<t_{j+1}\text{ and there exists } y\le x\colon\\&\qquad B_N(y)>t_{j+1}+\delta+\sqrt N(H_k(x)-H_k(y))\Big\}.\end{aligned}$$

Fix $l>0$. The latter event occurs if on the interval $[x-l,x]$, the process $B_N(y)$ exceeds either the level

$$\sqrt N(H_k(x)-H_k(y))+t_{j+1}+\delta$$

for some $y<x-l$ or the level $t_{j+1}+\delta$. Therefore

$$p_N^j\le\mathsf{P}(C_N)+\mathsf{P}(D_N)$$

where

$$C_N = \left\{ \sup_y B_N(y) > t_{j+1} + \delta + \sqrt{N}(H_k(x) - H_k(x-l)) \right\},$$

$$D_N = \left\{ B_N(x) < t_{j+1} \text{ and there exists } y \in [x-l, x] \text{ such that } B_N(y) > t_{j+1} + \delta \right\}.$$

Now we estimate

$$\mathsf{P}\{D_N\} \le \mathsf{P}\left\{ \sup_{|y-x|<l} |B_N(x) - B_N(y)| > \delta \right\}.$$

Since $B(x)$ is a sample continuous process, there is a sufficiently small l such that

$$\mathsf{P}\left\{ \sup_{|x-y|<l} |B(x) - B(y)| > \frac{\delta}{3} \right\} \le \frac{\varepsilon}{18r}. \tag{10}$$

According to Theorem 3.1

$$\mathsf{P}\left\{ \sup_x |B_N(x) - B(x)| > \frac{\delta}{3} \right\} \le \frac{\varepsilon}{18r} \tag{11}$$

for sufficiently large N. Since

$$|B_N(x) - B_N(y)| \le |B_N(x) - B(x)| + |B(x) - B(y)| + |B(y) - B_N(y)|,$$

conditions (10)–(11) imply

$$\mathsf{P}\{D_N\} \le \frac{\varepsilon}{6r}. \tag{12}$$

Fix l and estimate $\mathsf{P}(C_N)$. If $l > 0$, then

$$H_k(x) > H_k(x-l)$$

and

$$\sup_y B_N(y) \to \sup_y B(y) < \infty, \qquad N \to \infty.$$

Thus Theorem 3.1 implies that $\mathsf{P}(C_N) \to 0$ as $N \to \infty$. Hence

$$\mathsf{P}(C_N) \le \frac{\varepsilon}{6r} \tag{13}$$

for sufficiently large N. Combining inequalities (12) and (13) we get $p_N^j \le \varepsilon/(3r)$.

Finally

$$\mathsf{P}(A_N) \le p_N^+ + p_N^- + \sum_{j=1}^r p_N^j \le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \sum_{j=1}^r \frac{\varepsilon}{3r} \le \varepsilon$$

for sufficiently large N. Thus relation (9) is proved.

Now we prove (8). Let $[a, b]$ be an arbitrary interval. Since $B_N^+(x) \ge B_N(x)$, we have

$$\inf_{x\in[a,b]} B_N^+(x) \ge \inf_{x\in[a,b]} B_N(x).$$

Now we obtain an upper estimate for $\sup_{x\in[a,b]} B_N^+(x)$. Since B_N^+ is decreasing on intervals between jumps of $\hat{F}_N(x, a)$, the supremum is attained either at a point of jump of the function $\hat{F}_N(x, a)$ or at the left endpoint of the interval $[a, b]$. In the first case, the supremum is equal to $\sup_{x\in[a,b]} B_N(x)$. Thus

$$\sup_{x\in[a,b]} B_N^+(x) \le \max\left(B_N^+(a), \sup_{x\in[a,b]} B_N(x) \right). \tag{14}$$

Since $B(x)$ is sample continuous, for all $\lambda > 0$ and $\varepsilon > 0$ there is δ such that

$$\mathsf{P}\left\{\sup_j |B(t_j) - B(t_{j-1})| > \varepsilon\right\} < \lambda$$

for $t_j = -1 + \delta j$. Relation (9) implies that for sufficiently large N

$$\mathsf{P}\left\{\sup_j \left|B_N^+(t_j) - B_N(t_j)\right| > \varepsilon\right\} < \lambda. \tag{15}$$

According to Theorem 3.1, δ can be chosen so that

$$\mathsf{P}\left\{\sup_{|x-y|<\delta} |B_N(x) - B_N(y)| > \varepsilon\right\} < \lambda,$$

whence

$$\mathsf{P}\left\{\max_j \left(\sup_{y\in[t_j,t_{j+1}]} B_N(x) - \inf_{y\in[t_j,t_{j+1}]} B_N(x)\right) > \varepsilon\right\} < \lambda. \tag{16}$$

For all $x \in [-1, 1]$ there is j such that $x \in [t_j, t_{j+1}]$ and

$$B_N^+(x) \geq \inf_{y\in[t_j,t_{j+1}]} B_N^+(y) \geq \inf_{y\in[t_j,t_{j+1}]} B_N(y).$$

It follows from (14) that

$$\begin{aligned} B_N^+(x) &\leq \sup_{y\in[t_j,t_{j+1}]} B_N^+(y) \leq \max\left(B_N^+(t_j), \sup_{y\in[t_j,t_{j+1}]} B_N(y)\right) \\ &\leq \max\left(B_N(t_j) + \varepsilon, \sup_{y\in[t_j,t_{j+1}]} B_N(y)\right) \leq \sup_{y\in[t_j,t_{j+1}]} B_N(y) + \varepsilon \end{aligned}$$

if the event under the probability sign in (15) occurs. Taking into account (15), we obtain

$$\mathsf{P}\left\{\text{for all } j, x \in [t_j, t_{j+1}],\ \inf_{y\in[t_j,t_{j+1}]} B_N(y) \leq B_N^+(x) \leq \sup_{y\in[t_j,t_{j+1}]} B_N(y) + \varepsilon\right\} < \lambda.$$

Combining this with (16) we get

$$\mathsf{P}\left\{\sup_x \left|B_N^+(x) - B_N(x)\right| > 2\varepsilon\right\} < 2\lambda.$$

Since ε and λ are arbitrary, the proof of the theorem is complete. □

Proof of the corollary. Note that $B_N^*(x) \leq B_N^+(x)$ by construction and

$$\left\{\sup_{x\in(-\infty,s]} \left|B_N^*(x) - B_N^+(x)\right| \neq 0\right\} \subseteq \left\{\sup_{x\in(-\infty,s]} \hat{F}_N(x,a) > 1\right\} = D_N.$$

Theorem 3.1 implies that

$$\sup_{x\in(-\infty,s]} \left|\hat{F}_N(x,a) - H_k(x)\right| \to 0$$

in probability. Since s is a point of increase,

$$H_k(x) < H_k(s) < 1.$$

Therefore $\mathsf{P}\{D_N\} \to 0$ as $N \to \infty$. □

Bibliography

1. R. E. Maĭboroda, *Estimation of the distributions of the components of mixtures having varying concentrations*, Ukrain. Matem. Zh. **48** (1996), no. 4, 562–566; English transl. in Ukrainian Math. J. **48** (1996), no. 4, 618–622. MR1417019 (97j:62055)
2. ———, *Correlation analysis of mixtures. I*, Teor. Imovir. Matem. Statist. **54** (1996), 99–108; English transl. in Theor. Probability Math. Statist. **54** (1997), 105–114. MR1644590 (99k:62102)
3. V. V. Buldygin and Yu. V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, TViMS, Kiev, 1998; English transl., AMS, Providence, RI, 2000. MR1743716 (2001g:60089)
4. A. V. Skorokhod, *Studies in the Theory of Random Processes*, Kiev University, Kiev, 1961; English transl., Addison-Wesley, New York, 1965. MR0185620 (32:3082b)

Department of Probability Theory and Mathematical Statistics, Faculty of Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine

E-mail address: mre@mechmat.univ.kiev.ua

Department of Probability Theory and Mathematical Statistics, Faculty of Mechanics and Mathematics, National Taras Shevchenko University, Academician Glushkov Avenue 6, Kyiv 03127, Ukraine

E-mail address: linsta@akcecc.kiev.ua

Received 26/SEP/2002

Translated by S. KVASKO