Ann. Inst. Statist. Math.
37 (1985), Part A, 235-240

ON ESTIMATING OF THE NUMBER OF CONSTITUENTS OF A FINITE MIXTURE OF CONTINUOUS DISTRIBUTIONS

JŌGI HENNA

(Received Nov. 11, 1982; revised Sept. 21, 1984)

Summary

Suppose that H is a mixture of distributions for a given family $\mathcal{F}$. A necessary and sufficient condition is obtained under which H is, in fact, a finite mixture. An estimator of the number of distributions constituting the mixture is proposed assuming that the mixture is finite and its asymptotic properties are investigated.

1. Introduction

Let $\mathcal{F}=\{F_{\boldsymbol{\theta}}(x): \boldsymbol{\theta} \in R_1^k\}$ be a family of known one-dimensional cumulative distribution functions (cdf's) and $G^\circ(\boldsymbol{\theta})$ any cdf such that $P_{G^\circ}(R_1^k)=1$, where P_{G° is the probability measure induced by G° and R_1^k a compact subset of the k-dimensional Euclidean space R^k. Let $F_{\boldsymbol{\theta}}(x)$ be continuous in x for each $\boldsymbol{\theta}$ and continuous in $\boldsymbol{\theta}$ for each x. Let $H_{G^\circ}(x)$ be the continuous cdf defined by

$$H_{G^\circ}(x)=\int_{R_1^k} F_{\boldsymbol{\theta}}(x)dG^\circ(\boldsymbol{\theta}) . \tag{1.1}$$

H_{G° will be called a mixture of $\mathcal{F}$ and, especially, a finite mixture when the support of P_{G° is a finite subset of R_1^k.

Suppose that H_{G° is known to be a finite mixture and let $\boldsymbol{X}=(X_1, X_2,\cdots, X_n)$ be a random sample from H_{G°. Our problem is to construct an estimator $\hat{m}_n$ of the number $m_\circ$ of cdf's constituting H_{G° and to investigate its asymptotic property. The method employed here consists of examining the number of cdf's constituting the finite mixture which is "closest" to the empirical distribution function of $\boldsymbol{X}$. Still another important problem discussed in this paper is a criterion as to whether or not a given mixture is finite.

To the author's knowledge ([2], [3] and others), no estimator of $m_\circ$

Keywords: Empirical distribution function, finite mixture, identifiability, mixing cdf.

has been proposed so far. What is then the purpose of estimating m_0? For instance, the estimator $\hat{m}_n$ is useful in classification problems. More precisely, we may regard it as the number of populations when we wish to classify an observation into one of several populations.

In Section 2, we give some notation, an estimator $\hat{m}_n$ and a preliminary lemma. In Section 3, we give a necessary and sufficient condition for H_{G° to be a finite mixture. In Section 4, we show that $\hat{m}_n = m_0$ holds with probability one for all n sufficiently large. In Section 5, we give some examples.

2. Notations and a preliminary lemma

We define a subfamily of mixing cdf's by

$$\mathcal{G}_m = \Big\{ G_m:\ G_m = (g_1, g_2, \cdots, g_m;\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m),$$

$$\sum_{j=1}^{m} g_j = 1,\ 0 \leqq g_j \leqq 1,\ \boldsymbol{\theta}_j \in R_1^k,\ j = 1, 2, \cdots, m \Big\},$$

$$(m = 1, 2 \cdots),$$

where $G_m = (g_1, g_2, \cdots, g_m;\ \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m)$ is the discrete cdf with jump g_j at $\boldsymbol{\theta}_j$ $(j = 1, 2, \cdots, m)$.

Let $\hat{G}_{m,n}$ be any G_m in $\mathcal{G}_m$ which minimizes

$$S_n(G_m) = \int \{H_{G_m}(x) - F_n(x)\}^2 dF_n(x) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{m} g_j F_{\boldsymbol{\theta}_j}(X_{(i)}) - \frac{i}{n} \right\}^2,$$

where $F_n(x)$ and $X_{(i)}$ are the empirical distribution function and the ith order statistic of $\boldsymbol{X}$ respectively. The existence of $\hat{G}_{m,n}$ is guaranteed since $S_n(G_m)$ is a continuous function of G_m on a compact subset of $R^{m(k+1)}$.

For the case $G^\circ = G^\circ_{m_0}$ with finite but unknown m_0, where $0 < g_j^\circ < 1$ $(j = 1, 2, \cdots, m_0)$, we propose an estimator $\hat{m}_n$ of m_0 defined as follows:

$$\hat{m}_n = \text{the minimal integer } m \text{ such that } S_n(\hat{G}_{m,n}) < \lambda^2(n)/n,$$

where $\lambda(n) \uparrow \infty$, $\lambda^2(n)/n \to 0$ as $n \to \infty$ and $\sum \{\lambda^2(n)/n\} e^{-2\lambda^2(n)} < \infty$.

The existence of $\hat{m}_n$ for all n sufficiently large is guaranteed with probability one by Lemma 4.3 which will be shown later.

The following lemma is a simple consequence of Polya's theorem (see Rao [4], p. 120) and the Glivenko-Cantelli theorem. We omit the proof.

LEMMA 2.1. *Let* $\{G_{m,n}\}_{n=1}^{\infty}$ *be any sequence of cdf's in* $\mathcal{G}_m$ *such that* $G_{m,n} \to G_m^*$ *as* $n \to \infty$. *Then*

$$\int \{H_{G_{m,n}}(x)-H_{G^\circ}(x)\}^2 dF_n(x) \to \int \{H_{G_m^*}(x)-H_{G^\circ}(x)\}^2 dH_{G^\circ}(x)$$

with probability one (with respect to $P_{H_{G^\circ}}^{(\infty)}$*).*

For a sequence $\{G_{m,n}\}=\{(g_{1,n},\cdots, g_{m,n};\ \boldsymbol{\theta}_{1,n},\cdots,\boldsymbol{\theta}_{m,n})\}$, the convergence $G_{m,n}\to G_m^*=(g_1^*,\cdots, g_m^*;\ \boldsymbol{\theta}_1^*,\cdots,\boldsymbol{\theta}_m^*)$ means $g_{j,n}\to g_j^*$ and $\boldsymbol{\theta}_{j,n}\to\boldsymbol{\theta}_j^*$ $(j=1, 2, \cdots, m)$ as $n\to\infty$.

3. A necessary and sufficient condition for H_{G° to be a finite mixture

In order to give a necessary and sufficient condition for H_{G° to be a finite mixture on the basis of $\boldsymbol{X}$, we need the following identifiability condition.

(A–3.1) For any finite mixture H_G, the relationship $H_G=H_{G^*}$ implies that $G=G^*$.

LEMMA 3.1. *Under* (A–3.1), *suppose that* H_{G° *is not a finite mixture. Then*

$$P_{H_{G^\circ}}^{(\infty)}\{\liminf_{n\to\infty} S_n(\hat{G}_{m,n})>0\}=1$$

for any finite m.

PROOF. Assume that the conclusion does not hold. Then there exists a Borel subset A of R^∞ such that $P_{H_{G^\circ}}^{(\infty)}(A)>0$ and, if $(X_1, X_2,\cdots)\in A$, then $\|H_{G^\circ}-F_n\|\to 0$ as $n\to\infty$ (by the Glivenko-Cantelli theorem) and $\liminf_{n\to\infty} S_n(\hat{G}_{m,n})=0$ for a finite m, where $\|\ \|$ denotes the sup norm. Then there exists a subsequence $\{\hat{G}_{m,r}\}$ of $\{\hat{G}_{m,n}\}$ such that

$$S_r(\hat{G}_{m,r})=\int \{H_{\hat{G}_{m,r}}(x)-F_r(x)\}^2 dF_r(x)\to 0 \tag{3.1}$$

as $r\to\infty$. Let $\{\hat{G}_{m,s}\}$ be any subsequence of $\{\hat{G}_{m,r}\}$ such that $\hat{G}_{m,s}\to G_m^*$ as $s\to\infty$, where G_m^* is a member of $\mathcal{G}_m$. We have

$$\begin{aligned}\int &\{H_{\hat{G}_{m,s}}(x)-F_s(x)\}^2 dF_s(x)\\ &=\int \{H_{\hat{G}_{m,s}}(x)-H_{G^\circ}(x)\}^2 dF_s(x)\\ &\quad+2\int \{H_{\hat{G}_{m,s}}(x)-H_{G^\circ}(x)\}\{H_{G^\circ}(x)-F_s(x)\}dF_s(x)\\ &\quad+\int \{H_{G^\circ}(x)-F_s(x)\}^2 dF_s(x)\ .\end{aligned} \tag{3.2}$$

The second and third terms on the right hand side of (3.2) converges to 0 as $s\to\infty$ by $\|H_{G^\circ}-F_s\|\to 0$. Hence, by Lemma 2.1, we have

$$\int\{H_{\hat{G}_{m,s}}(x)-F_s(x)\}^2dF_s(x)\to\int\{H_{G_m^*}(x)-H_{G^\circ}(x)\}^2dH_{G^\circ}(x)\ .$$

It follows from (3.1) and the assumption (A-3.1) that $G_m^*=G^\circ$, i.e., that H_{G° is a finite mixture, contradicting to the assumption of the lemma.

COROLLARY 3.1. *Under* (A-3.1), *a necessary and sufficient condition for* H_{G° *to be a finite mixture is that there exists a finite* m *such that* $S_n(\hat{G}_{m,n})\to 0$ *with probability one as* $n\to\infty$ (*with respect to* $P_{H_{G^\circ}}^{(\infty)}$).

PROOF. Assume that $G^\circ=G_{m_\circ}^\circ$ with finite $m_\circ$. Then, by the definition of $\hat{G}_{m_\circ,n}$, we have

$$0\leqq S_n(\hat{G}_{m_\circ,n})\leqq\int\{H_{G_{m_\circ}^\circ}(x)-F_n(x)\}^2dF_n(x)\leqq\|H_{G_{m_\circ}^\circ}-F_n\|^2\ .$$

Accordingly, $S_n(\hat{G}_{m_\circ,n})\to 0$ with probability one as $n\to\infty$ by the Glivenko-Cantelli theorem.

If, conversely, H_{G° is not a finite mixture, then the probability that $S_n(\hat{G}_{m,n})\to 0$ as $n\to\infty$ is 0 for any finite m by Lemma 3.1.

4. The asymptotic behavior of $\hat{m}_n$ as $n\to\infty$ in the case of a finite mixture

In order to prove that $\hat{m}_n=m_\circ$ holds with probability one for all n sufficiently large in case $G^\circ=G_{m_\circ}^\circ$ with finite $m_\circ$, where $0<g_j^\circ<1$ ($j=1,2,\cdots,m_\circ$), we assume the following identifiability condition which is weaker than (A-3.1).

(A-4.1) For any two finite mixture H_G and H_{G^*}, the relationship $H_G=H_{G^*}$ implies that $G=G^*$.

To derive the asymptotic behavior of $\hat{m}_n$, we need to investigate some asymptotic properties of $S_n(\hat{G}_{m,n})$ as $n\to\infty$. We can show the following lemma in the same way to the proof of Lemma 3.1.

LEMMA 4.1. *Under* (A-4.1), *suppose that* $m<m_\circ$. *Then*

$$P_{H_{G_{m_\circ}^\circ}}^{(\infty)}\{\liminf_{n\to\infty}S_n(\hat{G}_{m,n})>0\}=1\ .$$

LEMMA 4.2. $S_n(\hat{G}_{m,n})\geqq S_n(\hat{G}_{m+1,n})$ *for any* n.

PROOF.

$$S_n(\hat{G}_{m,n})=\min S_n(g_1,g_2,\cdots,g_m,0;\ \boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\cdots,\boldsymbol{\theta}_m,\boldsymbol{\theta}_{m+1})$$

$$\geqq \min S_n(g_1, g_2, \cdots, g_m, g_{m+1}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_m, \boldsymbol{\theta}_{m+1})$$
$$= S_n(\hat{G}_{m+1,n}) .$$

LEMMA 4.3. *Suppose that* $m_\circ \leqq m$. *Then*

$$P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{S_n(\hat{G}_{m,n}) < \lambda^2(n)/n \text{ for all } n \text{ sufficiently large}\} = 1 .$$

PROOF. Suppose that $m_\circ \leqq m$. Then, by the last lemma, we have

$$0 \leqq S_n(\hat{G}_{m,n}) \leqq S_n(\hat{G}_{m_\circ,n}) \leqq \| H_{G^\cdot_{m_\circ}} - F_n \|^2 .$$

Accordingly, we have the conclusion by Theorem 2 of Chung [1].

THEOREM 4.1. *Under* (A–4.1), *we have*

$$P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{\hat{m}_n = m_\circ \text{ for all } n \text{ sufficiently large}\} = 1 .$$

PROOF. By Lemmas 4.1 and 4.2, we have

$$P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{\hat{m}_n \leqq m_\circ - 1 \text{ for an infinite number of } n\text{'s}\}$$
$$\leqq P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{S_n(\hat{G}_{m_\circ-1,n}) < \lambda^2(n)/n \text{ for an infinite number of } n\text{'s}\}$$
$$= 0 .$$

By Lemmas 4.2 and 4.3, we have

$$P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{\hat{m}_n \geqq m_\circ + 1 \text{ for an infinite number of } n\text{'s}\}$$
$$\leqq P^{(\infty)}_{H_{G^\cdot_{m_\circ}}}\{S_n(\hat{G}_{m_\circ,n}) \geqq \lambda^2(n)/n \text{ for an infinite number of } n\text{'s}\}$$
$$= 0 .$$

5. Examples

The following propositions are derived by a modification of the proof of Theorem 2 of Teicher [5].

PROPOSITION 5.1. *Let* $N_{(\theta,\sigma^2)}(x)$ *be the normal distribution function with mean* θ *and variance* σ^2. *Suppose that*

$$\int_D N_{(\theta,\sigma^2)}(x) dG_1(\theta, \sigma^2) = \int_D N_{(\theta,\sigma^2)}(x) dG_2(\theta, \sigma^2) \tag{5.1}$$

and either side of (5.1) *is a finite mixture, where* $P_{G_1}(D) = P_{G_2}(D) = 1$ *and* $D = (\theta_\circ, \infty) \times (0, \infty)$ *with finite* $\theta_\circ$. *Then* $G_1 = G_2$.

PROPOSITION 5.2. *Let* $F_{(\alpha,\beta)}(x) = \Gamma(\alpha)^{-1} \int_\beta^x (y-\beta)^{\alpha-1} e^{-(y-\beta)} dy$. *Suppose that*

$$\int_D F_{(\alpha,\beta)}(x) dG_1(\alpha, \beta) = \int_D F_{(\alpha,\beta)}(x) dG_2(\alpha, \beta) \tag{5.2}$$

and either side of (5.2) *is a finite mixture, where* $P_{G_1}(D)=P_{G_2}(D)=1$ *and* $D=(0, \infty)\times(-\infty, \infty)$. *Then* $G_1=G_2$.

Now we shall give some examples.

Example 5.1. Let $\mathcal{F}=\{N_{(\theta,\sigma^2)}(x):\ (\theta, \sigma^2)\in R_1^2\}$, where R_1^2 is a compact subset of $(-\infty, \infty)\times(0, \infty)$. The condition (A–3.1) (accordingly (A–4.1)) is satisfied by Proposition 5.1. So, Corollary 3.1 can be applied. Let

$$\hat{m}_n=\text{the minimal integer } m \text{ such that } S_n(\hat{G}_{m,n})<\frac{(\log n)^2}{n}\ .$$

Then Theorem 4.1 can be applied, that is,

$$P^{(\infty)}_{H_{G^*_{m_0}}}\{\hat{m}_n=m_0 \text{ for all } n \text{ sufficiently large}\}=1\ .$$

Example 5.2. Let $\mathcal{F}=\{F_{(\alpha,\beta)}(x);\ (\alpha, \beta)\in R_1^2\}$, where R_1^2 is a compact subset of $(0, \infty)\times(-\infty, \infty)$. The condition (A–3.1) (accordingly (A–4.1)) is satisfied by Proposition 5.2. So, Corollary 3.1 can be applied. Let $\hat{m}_n$ be that of the last example. Then Theorem 4.1 can be applied.

Example 5.3. Let $\mathcal{F}=\{F_{(\theta,\alpha)}(x): (\theta, \alpha)\in R_1^2\}$, where $F_{(\theta,\alpha)}(x)=\theta^\alpha[\Gamma(\alpha)]^{-1}\cdot\int_0^x y^{\alpha-1}e^{-\theta y}dy$ and R_1^2 is a compact subset of $(0, \infty)\times(0, \infty)$. The condition (A–4.1) is satisfied by Proposition 2 of Teicher [5]. Let $\hat{m}_n$ be that of the last example. Then Theorem 4.1 can be applied.

Acknowledgements

The author would like to acknowledge the continuing guidance and encouragement of Prof. M. Huzii. Also the author is very grateful to Dr. R. Shimizu, Dr. Y. Yajima, Dr. A. Nishio and Dr. Y. Isokawa for their valuable advice and discussions. The author also wishes to thank the referee for his helpful advice.

Ryukyu University

References

[1] Chung, K. L. (1949). An estimate concerning the Kolmogoroff limit distribution, *Tran. Amer. Math. Soc.*, **67**, 36-50.

[2] Gupta, S. S. and Huang, W. T. (1981). On mixtures of distributions: a survey and some new results on ranking and selection, *Sankhyā*, **43**, 245–290.

[3] Isaenko, O. K. and Urbakh, V. Y. (1977). Partitioning mixed probability distributions into their constituents, *J. Soviet Math.*, **7**, 148-160.

[4] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, John Wiley.

[5] Teicher, H. (1963). Identifiability of finite mixtures, *Ann. Math. Statist.*, **34**, 1265–1269.