DISCRIMINATION AND ESTIMATION OF THE DIFFERENCE OF LOCATION PARAMETERS

Rep. Fac. Sci. Kagoshima Univ. (Math. Phys. Chem.) No. 5, p. 7–14, 1972

DISCRIMINATION AND ESTIMATION OF THE DIFFERENCE OF LOCATION PARAMETERS

By

Hajime YAMATO

(Received September 30, 1972)

1. Summary and Introduction.

The application of an estimator of a distribution function to a univariate discriminant problem is presented. The estimator also is utilized to estimate the difference of location parameters of two distributions which are the same symmetric distributions except for the locations.

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population with an unknown distribution function $F(x)$. The empirical distribution function $F_n^*(x)$ can be expressed as follows,

$$F_n^*(x)=\frac{1}{n}\sum_{j=1}^{n} e_0(x-X_j)\,,$$

where e_0 is the unit distribution function, i.e.,

$$e_0(x)=\begin{cases}0 \text{ if } x<0\\ 1 \text{ if } x\geqq 0\,.\end{cases}$$

Now we consider an estimator of the distribution function $F(x)$ given by

$$F_n(x)=\frac{1}{n}\sum_{j=1}^{n} W_n(x-X_j)\,, \tag{1.1}$$

where W_n is a given distribution function. In case of the estimation of an absolutely continuous distribution function, by taking an absolutely continuous distribution function as W_n, $F_n(x)$ is again absolutely continuous, which provides an estimator of the density function $f(x)$ as follows,

$$f_n(x)=\frac{1}{n}\sum_{j=1}^{n} w_n(x-X_j)\,, \tag{1.2}$$

where w_n is the derivative of W_n.

Density estimators can be used to approximate the discriminant function and Glick [2] discussed this problem based on general density estimators including the multivariate analogue to the estimator of the form (1.2). To estimate the derivatives of a density function, the derivatives of the above density estimator $f_n(x)$ was utilized by Schuster [9]. The density estimator $f_n(x)$ and its bivariate analogue were used to estimate a conditional density and a regression curve by Rosenblatt [8], Nadaraya [6], Watson [9] etc.. An

estimator of a mode, which is determined by the density estimator $f_n(x)$, was discussed by Parzen [7] and Nadaraya [5]. The applications to compound decision of the density estimator $f_n(x)$ were discussed by Johns and Ryzin [3] etc.. On the other hand with the density estimator $f_n(x)$ its integration over $(-\infty, x]$, i,e., the estimator $F_n(x)$ of the form (1.1), was used to estimate a hazard rate by Watson and Leadbetter [11], [12] and Leadbetter [4]. To estimate a hazard rate Murthy [5] utilized the density estimator $f_n(x)$ and its integration over $[x, +\infty)$, which is an estimator of $1-F(x)$ and is denoted by the similar form to (1.1). Furthermore he construct two asymptotic equivalent estimators of the jump S_i corresponding to the saltus $x=x_i$ of the distribution function $F(x)$, which are denoted as follows: the one is $2[F_n^*(x_i)-F_n(x_i)]$ in our notation and the other is $f_n(x_i)/K(0)B_n$, where $f_n(x)$ is given by (1.2) with $w_n(x)=B_nK(B_nx)$, B_n is a positive constant and K is a symmetric density. We shall propose another applications of the estimator $F_n(x)$ of the form (1.1).

In section 2, we utilize the estimator $F_n(x)$ for a univariate discriminant problem in case two underlying density functions are the same symmetric unimodal function except for the locations.

In section 3, we present an estimator of the difference of the location parameters of two distributions, which are the same symmetric distributions except for the locations.

2. Discrimination.

Suppose that there are two populations π_1 and π_2, which have densities $f(x)$ and $f(x-\theta)$ respectively, where f is symmetric and unimodal, and θ is a location parameter. Suppose that we have an observation z and we know a priori that it has come from either of two populations π_1 and π_2. We assume that the losses due to two kinds of misclassification are same, where one misclassification is that if the observation is actually from π_1 we classify it as coming from π_2 and the other is that if the observation is actually from π_2 we classify it as coming from π_1. If we know a priori $f(x)$ and θ, and no a priori probabilities are known, then according to the minimax procedure we decide as follows: if $f(z)\geqq f(z-\theta)$ then we decide that the observation z has come from π_1 and otherwise we decide that the observation z has come from π_2. In case $\theta>0$, it is equivalent to decide z from π_2 if $z\geqq\theta/2$, and z from π_1 otherwise. In case $\theta<0$, we decide conversely.

Now, on the basis of the two random samples of size n_1 and n_2, $X_1^1, X_2^1, \cdots, X_{n_1}^1$ and $X_1^2, X_2^2, \cdots, X_{n_2}^2$, drawn from populations π_1 and π_2 respectively, we consider an approximation to the above minimax procedure in case we know nothing about f and θ except that f is symmetric with respect to the axis of ordinates and unimodal. From the previous discussion we know that the estimation of the parameter $\theta/2$ presents its approximation. On the other hand, whether θ is positive or negative, $x=\theta/2$ satisfies the equation

$$F_1(x)=1-F_2(x)\,, \tag{2.1}$$

where F_1 and F_2 are distribution functions of populations π_1 and π_2 respectively. If the carriers of $f(x)$ and $f(x-\theta)$ overlap, then the equation (2.1) has a unique solution $x=\theta/2$. If the carriers of $f(x)$ and $f(x-\theta)$ do not overlap, then the solution of (2.1) consists of a

finite interval, whose middle point is $\theta/2$. It stead of setimating $\theta/2$ directly, we shall estimate $\theta/2$ by solving the equation

$$F_{1,n_1}(x)=1-F_{2,n_2}(x) \tag{2.2}$$

where $F_{1,n_1}(x)$ and $F_{2,n_2}(x)$ are estimators of $F_1(x)$ and $F_2(x)$ respectively, whose form is denoted by (1.1) with given distribution functions W_{n_1} and W_{n_2} on the basis of the two samples, i.e.,

$$F_{1,n_1}(x)=\frac{1}{n_1}\sum_{j=1}^{n_1} W_{n_1}(x-X_j^1)$$

and

$$F_{2,n_2}(x)=\frac{1}{n_2}\sum_{j=1}^{n_2} W_{n_2}(x-X_j^2)\,.$$

The solution of the above equation, in general, can not be expressed in a explicit form. Especially if n_1 and n_2 are equal, which we shall denote by n, and if we take $F_{1,n}(x)$ and $F_{2,n}(x)$ with $W_n=e_0$, then the n-th smallest value among $X_1^1,\cdots,X_n^1,X_1^2,\cdots,X_n^2$ is a solution of (2.2) with $n_1=n_2=n$ and $W_n=e_0$. The solution of (2.2) is easily obtained by drawing curves $y=F_{1,n_1}(x)$ and $y=1-F_{2,n_2}(x)$ on the (x, y) plane and reading the x-axis of a intersection of the two curves. If the above curves overlap then we shall take an arbitrary point of the x-axis of the overlapping part as a solution of (2.2). Let x_{n_1,n_2} denote such a solution of (2.2). We propose to decide as follows: if

$$x_{n_1,n_2}>0 \text{ and } z\geqq x_{n_1,n_2}$$

then we decide that the new observation z has come from π_2 and if

$$x_{n_1,n_2}>0 \text{ and } z<x_{n_1,n_2}$$

then we decide that the ovservation z has come from π_1. In case of $x_{n_1,n_2}<0$ we decide conversely. Then the probability of misclassification if the observation z is from π_1 is

$$\begin{aligned} P(2\,|\,1,A)=&I_{(0,\infty)}(x_{n_1,n_2})[1-F_1(x_{n_1,n_2})] \\ &+I_{(-\infty,0)}(x_{n_1,n_2})F_1(x_{n_1,n_2}) \end{aligned} \tag{2.3}$$

and the probability of misclassification if the observation z is from π_2 is

$$\begin{aligned} P(1\,|\,2,A)=&I_{(1,\infty)}(x_{n_1,n_2})F_2(x_{n_1,n_2}) \\ &+I_{(-\infty,0)}(x_{n_1,n_2})[1-F_2(x_{n_1,n_2})]\,, \end{aligned} \tag{2.4}$$

where I_S denotes the indicator function of a set S.

In what follows, we consider the asymptotic properties of the above approximate procedure. We assume the same conditions as stated in constructing our approximation to the minimax procedure and suppose that $W_{n_1}\to e_0$ as $n_1\to\infty$ and $W_{n_2}\to e_0$ as $n_2\to\infty$.

LEMMA 1.

$$\lim_{n_1,n_2\to\infty}\{F_1(x_{n_1,n_2})-[1-F_2(x_{n_1n_2})]\}=0 \quad \textit{with probability one.} \tag{2.5}$$

PROOF. We can reduce (2.2) to

$$[F_{1,n_1}(x_{n_1,n_2})-F_1(x_{n_1,n_2})]+F_1(x_{n_1,n_2})$$
$$=1-F_2(x_{n_1,n_2})-[F_{2,n_2}(x_{n_1,n_2})-F_2(x_{n_1,n_2})]\,.$$

By Theorem 2 in Yamato [13] and the assumption we have

$$\lim_{n_1,n_2\to\infty}|F_{i,n_i}(x_{n_1,n_2})-F_i(x_{n_1,n_2})|\leqq\lim_{n_i\to\infty}\sup_{-\infty<x<\infty}|F_{i,n_i}(x)-F_i(x)|=0$$

with probability one for $i=1, 2$. Therefore we have (2.5).

From the above lemma the solution of (2.2) is asymptotically a solution of (2.1) with probability one.

LEMMA 2. *If the carriers of $f(x)$ and $f(x-\theta)$ overlap, then*

(2.6) $$\lim_{n_1,n_2\to\infty}x_{n_1,n_2}=\theta/2 \quad \textit{with probability one.}$$

PROOF. Under the assumption the equation (2.1) has a unique solution $x=\theta/2$ and in a neighborhood of the point $x=\theta/2$ the continuous function $F_1(x)+F_2(x)$ is strictly increasing. Consequently the continuous function $y=F_1(x)-[1-F_2(x)]$ is strictly increasing in a neighborhood of the point $x=\theta/2$ and therefore for any $\varepsilon>0$ there exists a $\delta>0$ such that

$$|y|<\delta \text{ implies } |x-\theta/2|<\varepsilon\,,$$

i. e., $|F_1(x)-[1-F_2(x)]|<\delta$ implies $|x-\theta/2|<\varepsilon$.

On the other hand from (2.5) for the above $\delta>0$ there exists a positive integer N_0 such that

$$|F_1(x_{n_1,n_2})-[1-F_2(x_{n_1,n_2})]|<\delta \text{ for } n_1,n_2>N_0 \text{ with probability one.}$$

Henceforth it holds that

$$|x_{n_1,n_2}-\theta/2|<\varepsilon \text{ for } n_1, n_2>N_0 \text{ with probability one,}$$

which implies (2.6).

LEMMA 3. *Suppose that the carriers of $f(x)$ and $f(x-\theta)$ do not overlap. Let $[a, b]$ denote the closure of the interval which is sandwiched between the above carriers. Then*

$$a\leqq\liminf_{n_1,n_2\to\infty}x_{n_1,n_2}\leqq\limsup_{n_1,n_2\to\infty}x_{n_1,n_2}\leqq b \textit{ with probability one.}$$

We should note that under the conditions of Lemma 3 the solution of (2.1) consists of the interval $[a, b]$.

PROOF. Suppose that there exists a set C with $P(C)>0$ such that

$$\liminf_{n_1,n_2\to\infty}x_{n_1,n_2}<a \text{ on } C.$$

By the continuity of F_1 and F_2 and by Lemma 1, the equation (2.1) has a solution $\liminf_{n_1,n_2\to\infty}$ $x_{n_1,n_2}<a$ on C, which contradicts to the assumption. Similarly we can show that $\limsup_{n_1,n_2\to\infty}$ $x_{n_1,n_2}\leqq b$ with probability one.

Let $P(2|1, M)$ and $P(1|2, M)$ denote the probabilities of misclassifications due to the minimax procedure if the ovservation is from π_1 and π_2 respectively. Then we have

THEOREM 1. *According to our discriminant procedure, we have*

$$\lim_{n_1,n_2\to\infty} P(2|1, A)=P(2|1, M) \text{ with probability one}$$

and

$$\lim_{n_1,n_2\to\infty} P(1|2,A)=P(1|2,M) \quad \text{with probability one.}$$

PROOF. We prove in case of $\theta>0$ and we can prove similarly in case of $\theta<0$. At first we consider the case where the carriers of $f(x)$ and $f(x-\theta)$ overlap. Then we have

$$P(2|1, M)=1-F_1(\theta/2) \text{ and } P(1|2,M)=F_2(\theta/2).$$

By applying Lemma 2 and using the continuity of F_1 and F_2 on (2.3) and (2.4) we have

$$\lim_{n_1,n_2\to\infty} P(2|1,A)=1-F_1(\theta/2) \text{ with probability one}$$

and

$$\lim_{n_1,n_2\to\infty} P(1|2,A)=F_2(\theta/2) \text{ with probability one.}$$

Thus the theorem is proved in this case. Next we consider the case where the carriers of $f(x)$ and $f(x-\theta)$ do not overlap. Then we have

$$P(2|1,M)=0 \quad \text{and} \quad P(1|2,M)=0,$$

and we have $a, b>0$, where a and b are constants defined in Lemma 3. By applying Lemma 3 and using the continuity of F_1 and F_2 on (2.3) and (2.4), we have with probability one

$$\begin{aligned}\limsup_{n_1,n_2\to\infty} P(2|1,A) &= \limsup_{n_1,n_2\to\infty}[1-F_1(x_{n_1,n_2})] \\ &= 1-\liminf_{n_1,n_2\to\infty} F_1(x_{n_1,n_2}) \\ &\leqq 1-F_1(a)=0\end{aligned}$$

and

$$\begin{aligned}\limsup_{n_1,n_2\to\infty} P(1|2,A) &= \limsup_{n_1,n_2\to\infty} F_2(x_{n_1,n_2}) \\ &\leqq F_2(b)=0.\end{aligned}$$

Thus the theorem is proved.

3. Estimation of the difference of location parameters.

Suppose there are two populations whose distribution functions are given by $F(x)$ and $F(x-\theta)$ respectively, where F is continuous and symmetric and θ is a unknown location parameter. Let $X_1^1, X_1^2, \cdots, X_{n_1}^1$ and $X_1^2, X_2^2, \cdots, X_{n_2}^2$ be random samples from the two populations with distribution functions $F(x)$ and $F(x-\theta)$ respectively. Our aim in this section is to present an estimator of θ based on the two samples. $x=\theta/2$ satisfies always

$$F(x)=1-F(x-\theta)\,.$$

Similarly to section 2, we can obtain an estimator of $\theta/2$ by solving the equation

$$F_{1,n_1}(x)=1-F_{2,n_2}(x)\,, \tag{3.1}$$

where $F_{1,n_1}(x)$ and $F_{2,n_2}(x)$ are estimators of $F(x)$ and $F(x-\theta)$ constructed by the two samples respectively, whose form is denoted by (1.1) with given distribution functions W_{n_1} and W_{n_2}. Let x_{n_1,n_2} denote a solution of (3.1). Then Lemma 2 yields the following

PROPOSITION 1. *Suppose that the distribution function $F(x)$ is strictly increasing on $\{x;\ 0<F(x)<1\}$ and that the closures of sets $\{x;\ 0<F(x)<1\}$ and $\{x;\ 0<F(x-1)<1\}$ overlap. Let $W_{n_1}\to e_0$ as $n_1\to\infty$ and $W_{n_2}\to e_0$ as $n_2\to\infty$. Then we have*

$$\lim_{n_1,\,n_2\to\infty} 2x_{n_1,n_2}=\theta$$

with probability one.

The above proposition shows that under the conditions of Proposition 1, $2x_{n_1,n_2}$ is a consistent estimator of θ. The above problem is equivalent to the estimation of a quantile. If we put

$$G(x)=\frac{1}{2}\{F(x)+F(x-\theta)\}$$

then the parameter $\theta/2$ is a median of the distribution $G(x)$. We estimate $G(x)$ by

$$G(x)=\frac{1}{2}\{F_{1,n_1}(x)+F_{2,n_2}(x)\}$$

and we construct an estimator of θ, $2x_{n_1,n_2}$, where $G(x_{n_1n_2})=1/2$, which is equivalent to (3.1).

For practical purposes it arises a question what W_n we should take. In case of absolutely continuous distributions, following Epanechnikov [1] we propose to take

$$W_n(x)=\begin{cases} 0 & (x\leqq-\sqrt{5}h_n) \\ \dfrac{1}{2}+\dfrac{3x}{4\sqrt{5}h_n}-\dfrac{x^3}{20\sqrt{5}h_n^{\,3}} & (-\sqrt{5h_n}<x<\sqrt{5h_n}\,)\,, \\ 1 & (x\geqq\sqrt{5h_n}\,) \end{cases}$$

where $\{h_n\}$ is a sequence of positive numbers. The optimum sequence $\{h_n\}$ depends on the unknown density and therefore it may be unavoidable to take $h_n=1/n^r$, where r is a suitable constant between 0 and 1.

The author wishes to thank Prof. A. Kudo of Kyushu University for his kind encouragement and suggestions.

References

[1] EPANECHNIKOV, V.A. (1969), Non-parametric estimation of a multivariate probability density, Theor. Prob. Appl., Vol. 14, pp. 153–158.

[2] GLICK, N (1972), Sample-based classification procedures derived from density estimators, Jour. Amer. Statist. Ass., Vol. 67, pp. 116–122.

[3] JOHNS, M.V. and RYZIN, J. VAN (1972), Convergence rate for empirical Bayes two-action problems II. Continuous case, Ann. Math. Statist., Vol. 43, pp. 934–947.

[4] LEADBETTER, M.R. (1963), On the non-parametric estimation of probability densities, Technical Report No. 11, Research Triangle Institute.

[5] MURTHY, V.K. (1965), Estimation of jumps, reliability and hazard rate, Ann. Math. Statist., Vol. 36, pp. 1032–1040.

[6] NADARAYA, E.A. (1965), On non-parametric estimates of density functions and regression curves, Theor. Prob. Appl., Vol. 10, pp. 186–190.

[7] PARZEN, E. (1962), On estimation of a probability density function and mode, Ann. Math. Statist., Vol. 33, pp. 1065–1076.

[8] ROSENBLATT, M. (1969) Conditional probability density and regression estimators, Multivariate Analysis-II (edited by Krishnaiah), pp. 25–31, Academic Press, New York.

[9] SCHUSTER, E.F. (1969), Estimation of a probability density function and its derivatives, Ann. Math. Statist., Vol. 40, pp. 1187–1195.

[10] WATSON, G.S. (1964), Smooth regression analysis, Sankhyā, Ser. A, Vol. 26, pp. 359–372.

[11] WATSON, G.S. and LEADBETTER, M.R. (1964), Hazard analysis. I, Biometrika, Vol. 51, pp. 175–184.

[12] WATSON, G.S. and LEADBETTER, M.R. (1964), Hazard analysis. II, Sankyā, Ser. A, Vol. 26, pp. 101–116.

[13] YAMATO, H. (1972), Uniform convergence of an estimator of a distribution function (to appear in Bull. Math. Statist.).

Errata

Page	Line	Wrong	Corrected
Back cover	↓10	Estimation of the Difference of	Estimating the Difference in
1	↓ 9	functinal	functional
〃	↓12	Chang	Chyug
2	↓ 7	continuos	continuous
7	↓ 1	ESTIMATION OF THE	ESTIMATING THE
〃	↓ 2	OF	IN
〃	↓ 8	of	in
11	↑ 6	**of location**	**in location**
12	↓17	$G(x) =$	$G_n(x) =$
〃	↓18	$G(x_{n_1 n,_2})$	$G_n(x_{n_1,n_2})$