Nonparametric Functional Estimation and Related Topics

Edited by

George Roussas

NATO ASI Series

Series C: Mathematical and Physical Sciences - Vol. 335

ON A PROBLEM IN SEMIPARAMETRIC ESTIMATION

S. R. JAMMALAMADAKA and X. WAN
Department of Statistics and Applied Probability
University of California
Santa Barbara, CA 93106 U.S.A.

ABSTRACT. The estimation problem in a semiparametric model , namely, the generalized Lehmann alternative model, is considered here. Suppose that two independent samples $X_1,...,X_m$ and $Y_1,...,Y_n$ with d.f.'s F and G, respectively, are observed. Assume that $G(\cdot)=H(F(\cdot);\theta)$, where the form of the function H is known, but F and the parameter θ are unknown. The problem is to estimate θ in the presence of the nuisance function F. We give two methods of estimating the parametric component of the model based on sample quantiles and the Mann-Whitney statistic. The asymptotic variances of these estimators are compared.

1. Introduction

Semiparametric models have become an active research topic in recent years. In such a model, there exist both parametric and nonparametric components. For example, Begun, Hall, Huang and Wellner(1983) study the model in which $X_1,...,X_n$ are i.i.d. with density $f=f(\cdot;\theta,g)$ with respect to Lebesgue measure μ on the real line, where θ is a real number and g belongs to a class of densities sufficiently small that θ is identifiable. There is considerable literature on the well known Lehmann alternatives(Lehmann(1953)). See, for example, Young(1973), Brooks(1974) and Savage(1980).

Fukui and Miura(1988) consider the estimation problem in the following two sample semiparametric model. Let $X_1,...,X_m$ be i.i.d. random variables(r.v.'s) with distribution function(d.f.) F and let $Y_1,...,Y_n$ be i.i.d. r.v.'s with d.f. $G(\cdot) = H(F(\cdot);\theta)$, where $\{H(x;\theta)\colon \theta\in(a,b)\}$ is a known family of d.f.'s on [0,1], while the true value of the parameter θ as well as the function F are unknown. It is desired to estimate θ in the presence of the nuisance function F. This semiparametric model is known as the generalized Lehmann alternative model since Lehmann alternative is a special case of this corresponding to $H(x;\theta) = x^\theta$. We now briefly review the work of Fukui and Miura (1988):

Since $G(x) = H(F(x);\theta)$, then $F(x) = H^{-1}(G(x);\theta)$. Let $D_{mn}(\theta)$ denote the Kolmogorov-Smirnov distance

G. Roussas (ed.), Nonparametric Functional Estimation and Related Topics, 647–656.

$$D_{mn}(\theta) = \sup_x \left|F_m(x) - H^{-1}(G_n(x);\theta)\right|$$

where $F_m(x)$ and $G_n(x)$ are the empirical d.f.'s of F and G, respectively. The minimum-distance estimator of θ is defined as the value $\hat{\theta}$ of θ, which minimizes $D_{mn}(\theta)$. Fukui and Miura give the asymptotic distribution of this $\hat{\theta}$.

Theorem (Fukui and Miura). If m+n = N tends to infinity such that $m/N \to \lambda$, where $0< \lambda< 1$, then

$$P\{ N^{1/2}(\hat{\theta} - \theta_0)\leq y \}$$

$$\to P\{ \sup_t [\frac{B_1(t)}{\lambda^{1/2}} - \frac{B_2(H(t;\theta_0))}{(1-\lambda)^{1/2} h_t(t;\theta_0)} + \frac{h_\theta(t;\theta_0)}{h_t(t;\theta_0)} \times y]$$

$$+ \inf_t [\frac{B_1(t)}{\lambda^{1/2}} - \frac{B_2(H(t;\theta_0))}{(1-\lambda)^{1/2} h_t(t;\theta_0)} + \frac{h_\theta(t;\theta_0)}{h_t(t;\theta_0)} \times y] \geq 0 \}$$

where

$$m^{1/2}[F_m(x) - F(x)] \to B_1(F(x)), \qquad n^{1/2}[G_n(x) - G(x)] \to B_2(G(x)),$$

$B_1(t)$ and $B_2(t)$ are independent Brownian bridges, $h_t(t;\theta)$ and $h_\theta(t;\theta)$ are the partial derivatives of $H(t;\theta)$ with respect to t and θ, respectively. #

This limiting distribution is very complicated and does not seem to have a normal approximation. Although Fukui and Miura give an algorithm to compute this $\hat{\theta}$, it seems that a great deal of computation needs to be done before $\hat{\theta}$ can finally be found.

Our aim in this paper is to develop some methods which will provide estimators of θ involving less computational work while providing consistent estimation with limiting normal distributions.

These results can also be applied to test the hypothesis H_0: $\theta=\theta_0$. Notice that in many cases there exists a $\theta_0 \epsilon \Theta$, such that $H(x;\theta_0) = x$. Thus $G(x) = H(F(x);\theta_0) = F(x)$, so that it covers the usual two sample problem of testing H_0': F=G as a special case of testing H_0: $\theta=\theta_0$.

Although our results focus on a scalar θ, some results can be readily extended to the case when θ is a vector. The following examples illustrate some applications of the generalized Lehmann alternative models.

Example 1.1. The idea of proportional hazards was introduced by Cox(1972). If

$X_1,\ldots,X_n$ are independent r.v.'s with X_i having d.f. $G(x)=1-[1-F(x)]^{\theta_i}$, $i=1,\ldots,n$, then the ratio $r(x,\theta_i)/\ r(x,\theta_j)$ does not depend on x, where $r(x,\theta) = G'(x)/\ [1-G(x)]$ is the hazard function. This condition often appears to be at least approximately satisfied in many biological applications.

Example 1.2. Consider a mixture problem in which the d.f. of the observations is a linear combination of two d.f.'s F(x) and G(F(x)), i.e.,

$$H(F(x);\theta) = (1-\theta)\ G(F(x)) + \theta F(x), \qquad 0 \leq \theta \leq 1.$$

Assume that F is unknown but $G(\cdot)$ is known. It is often important to estimate the mixing parameter θ. If G also depends on an unknown parameter, then this is an example of a model with vector parameter.

This paper is arranged as follows. Section 2 contains results about an estimator $\hat{\theta}$ of θ obtained by matching the pth quantiles of the two samples. Section 3 deals with an estimator θ^* based on the Mann-Whitney statistic. In section 4, comparison is made between $\hat{\theta}$ and θ^* through examples. We will make the following assumptions throughout:

Assumption A. F(x) is an absolutely continuous d.f. on R^1. $H(x;\theta)$ is a d.f. on [0,1] for every $\theta=(\theta_1,\ldots,\theta_k)\in\Theta \subset R^k$ such that

$$\frac{\partial H}{\partial x} > 0 \text{ and } \frac{\partial H}{\partial \theta_i} \neq 0.$$

In the case when θ is a real number, we may assume without loss of generality that $\frac{\partial H}{\partial \theta} < 0$ so that $h(x;\theta)$ is a strictly decreasing function of θ. Furthermore, assume that $\Theta = (a,b)$, where a and/or b may be infinity. Assume that $H(x;a+)=1$ and $H(x;b-)=0$ for any x, $0< x <1$. Assume that there exists a λ, $0< \lambda< 1$, such that $m/N \rightarrow\lambda$ as $m\rightarrow\infty$, $n\rightarrow\infty$, where N=m+n is the total of the two sample sizes.

2. Estimation Based On Sample Quantiles

Let $X_1' <\cdots< X_m'$ and $Y_1' < \cdots < Y_n'$ be the order statistics of the X and Y samples, respectively. For a fixed positive number p, $0< p< 1$, the sample pth quantile of the Y sample is defined as $Y'_{[np]+1}$. Let ξ_p denote the population pth quantile of Y, then $G(\xi_p) = H(F(\xi_p);\theta) = p$. Substituting estimators $Y'_{[np]+1}$ for ξ_p, and the empirical d.f. F_m for F, we may write, approximately, $H(F_m(Y'_{[np]+1});\theta) = p$. Let X_k' be the largest observation in the X sample such that $X'_k \le Y'_{[np]+1}$. When N=m+n is large, we would expect $F_m(Y'_{[np]+1})$ to be very close to $F_m(X'_k) = k/m$. Hence, our estimator $\hat{\theta}$ is the solution to the equation

$$H(k/m;\theta) = p \tag{2.1}$$

By Assumption A, $\frac{\partial H}{\partial \theta} < 0$ and $H(x;a+)=1$, $H(x;b-)=0$. This guarantees the existence and uniqueness of $\hat{\theta}$. For brevity we write Y^*_n for $Y'_{[np]+1}$. Let $\zeta_p = F(\xi_p)$ and let $H_2^{-1}(z,p)$ be the solution to the equation $H(z;\theta)=p$ with respect to θ for all z and p, such that $0< z< 1$, $0< p< 1$. Let $h_1(z;\theta)$ and $h_2(z;\theta)$ denote the partial derivatives of $H(z;\theta)$ with respect to z and θ, respectively. Thus $\hat{\theta} = H_2^{-1}(F_m(Y^*_n),p)$ and

$$N^{1/2}(\hat{\theta}-\theta_o) = N^{1/2}[\, H_2^{-1}(F_m(Y^*_n),p) - H_2^{-1}(\zeta_p,p)\,]$$

$$= N^{1/2} \frac{\partial}{\partial x} H_2^{-1}(x,p)\Big|_{x=\zeta_p} \cdot [\, F_m(Y^*_n) - \zeta_p] + o_p(1)$$

$$= -\left(\frac{N}{m}\right)^{1/2} \frac{h_1(\zeta_p;\theta_o)}{h_2(\zeta_p;\theta_o)} \cdot m^{-1/2} \sum_{i=1}^{m} [\, I(X_i \le Y^*_n) - \zeta_p] + o_p(1). \tag{2.2}$$

The last equality indicates that it suffices to find the limiting distribution of

$$A_N = m^{-1/2} \sum_{i=1}^{m} [\, I(X_i \le Y^*_n) - \zeta_p].$$

Applying the probability integral transformation F on both samples, $X_1,\ldots,X_m$ become $U_1,\ldots,U_m$, an i.i.d. sample from the uniform distribution on [0,1], and $Y_1,\ldots,Y_n$ become $Z_1,\ldots,Z_n$, an i.i.d. sample with d.f. $H(z;\theta)$. Although the values of $U_1,\ldots,U_m$ and $Z_1,\ldots,Z_n$ are unknown because F is unknown, the order relations in the combined samples remain the same since F is strictly increasing. Thus A_N can also be written as

$$A_N = m^{-1/2} \sum_{i=1}^{m} [\, I(U_i \le Z^*_n) - \zeta_p\,]$$

where $Z^*_n = Z'_{[np]+1}$.

Lemma 1. Let $\{X_i, i=1,...,n\}$ be a sequence of continuous r.v.'s with continuous probability density function f(x). Let μ_p denote the population pth quantile and X_n^* the sample pth quantile. Suppose that μ_p is unique with $f(\mu_p)$ positive. Then the sequence of the random variables $n^{1/2}(X_n^*-\mu_p)$ has asymptotically a normal distribution with mean zero and variance τ^2, where

$$\tau^2 = \frac{p(1-p)}{[f(\mu_p)]^2} . \qquad \#$$

Lemma 2. Under Assumption A, if θ is the true value of the parameter, then A_N has a limiting normal distribution with mean zero and variance τ_o^2, where

$$\tau_o^2 = \zeta_p - \zeta_p^2 + \frac{\lambda\, p(1-p)}{(1-\lambda)[h_1(\zeta_p;\theta)]^2} .$$

Proof: Define

$$B_N = m^{-1/2} \sum_{i=1}^{m} [I(U_i \leq \zeta_p) - \zeta_p] + m^{1/2}(Z_n^* - \zeta_p).$$

Since the first term of B_N is the sums of i.i.d. r.v.'s, by the Lindeberg-Levy central limit theorem, it has an asymptotic normal distribution with mean zero and variance $\tau_1^2=\zeta_p(1-\zeta_p)$. By Lemma 1, the second term of B_N has a limiting normal distribution with mean zero and variance

$$\tau_2^2 = \frac{\lambda\, p(1-p)}{(1-\lambda)[h_1(\zeta_p;\theta)]^2} .$$

Since U_i and Z_n^* are independent, the two limiting normal distributions are also independent of each other. Hence B_N has a limiting normal distribution with mean zero and variance τ_0^2. It can be verified that $E(A_N-B_N)^2 \rightarrow 0$. So that A_N and B_N have the same asymptotic distribution. #

Now from Lemma 2 and (2.2), we obtain

Theorem 1. Assume θ is the true value of the parameter, then under Assumption A, the asymptotic distribution of the sequence of random variables $N^{1/2}(\hat{\theta}-\theta)$ has a normal distribution with mean zero and variance σ_o^2, where

$$\sigma_o^2 = [h_2(\zeta_p;\theta)]^{-2} \left[\frac{h_1^2(\zeta_p;\theta)\,\zeta_p(1-\zeta_p)}{\lambda} + \frac{p(1-p)}{1-\lambda} \right]. \tag{2.3}$$

If $\hat{\zeta}_p$ is the pth quantile of H with this $\hat{\theta}$ substituted for θ, i.e., $\hat{\zeta}_p$ is the solution of

$$p = H(\hat{\zeta}_p;\hat{\theta})$$

and

$$\hat{\sigma}^2 = [\, h_2(\hat{\zeta}_p;\hat{\theta})\,]^{-2} \left[\frac{h_1^2(\hat{\zeta}_p;\hat{\theta})\,\hat{\zeta}_p(1-\hat{\zeta}_p)}{\lambda} + \frac{p(1-p)}{1-\lambda} \right].$$

then it is easily verified that $N^{1/2}(\hat{\theta}-\theta)/\hat{\sigma}$ is asymptotically N(0,1). One can also use this result to obtain asymptotic confidence limits for the unknown θ.

this result can also be used to test the hypothesis H: $\theta=\theta_o$. To improve the performance of the test statistic, one should minimize σ_o^2. (2.3) indicates that σ_o^2 depends on the quantities p, θ_o, λ and the function $H(x;\theta_o)$. All those quantities except p have been specified at the time of testing. However, we can choose an optimal value p_o so that σ_o^2 is minimized for the given values θ_o, λ and the function $H(x;\theta_o)$. Thus the optimal estimator and the optimal test is to use this p_o in the quantile matching method.

Example 2.1. Suppose that $H(x;\theta) = x^\theta$. Solving the equation

$$[F_m(Y_n^*)]^\theta = p$$

we get

$$\hat{\theta} = [\ln p]/ \ln[F_m(Y_n^*)].$$

According to Theorem 1, $\mathrm{Var}(\hat{\theta}) = N^{-1}\sigma_o^2 + o(N^{-1})$. To compute σ_o^2, we find

$$h_1(x;\theta) = \theta x^{\theta-1}, \qquad h_2(x;\theta) = x^\theta \ln x, \qquad \mu_p = p^{1/\theta}.$$

Substituting these into (2.3), we have after some simplification

$$\sigma_o^2 = \frac{\theta^4}{\lambda(\ln p)^2} \left[p^{-1/\theta} - 1 + \frac{\lambda(1-p)}{(1-\lambda)p\theta^2} \right]. \tag{2.4}$$

3. Estimation Based on the Mann-Whitney Statistic

In section 2, the estimator $\hat{\theta}$ is developed based on the statistic involving $F_m(Y_n^*)$, the value of the empirical d.f. F_m at Y_n^*. This motivates us to consider the statistic

$$W_{xy} = \frac{1}{n} \sum_{i=1}^{n} F_m(Y_i),$$

the average of the empirical d.f. F_m at all the n Y observations. By the definition of the empirical d.f. F_m, W_{xy} can be written as

$$W_{xy} = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} I(X_j < Y_i)$$

where $I(\cdot)$ is the indicator function. Notice that $W_{xy} = 1 - W_{yx}$, where

$$W_{yx} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I(Y_i < X_j). \tag{3.1}$$

We will study the statistic W_{yx} which is known as the Mann-Whitney statistic, a special case of a U-statistic. If we apply the probability integral transformation F on both X and Y samples, W_{yx} can be expressed as

$$W_{yx} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} I(Z_i < U_j). \tag{3.2}$$

Let

$$M_1(\theta) = E_\theta[\, I(Z_1 < U_1)\,] = \int_0^1 H(u;\theta)\, du \tag{3.3}$$

be the expectation of W_{yx}. Under Assumption A, it is clear that $M_1(\theta)$ is a differentiable function. Since W_{yx} is a consistent and asymptotically normal(CAN) estimator for $M_1(\theta)$, we propose an estimator θ^* of θ as

$$\theta^* = M_1^{-1}(W_{yx}). \tag{3.4}$$

As in section 2, we will study the asymptotic behavior of the random variables $N^{1/2}(\theta^* - \theta)$ when the true value of the parameter is θ. Using the mean value theorem, we have, from (3.4),

$$N^{1/2}(\theta^{*}-\theta) = [M_1'(\theta)]^{-1} N^{1/2}[W_{yx}- M_1(\theta)] + o_p(1). \qquad (3.5)$$

Define $M_2(\theta) = P_\theta(Z_1< U_1, Z_1< U_2)$, $M_3(\theta) = P_\theta(Z_1< U_1, Z_2< U_1)$. Using conditional probabilities, $M_2(\theta)$ and $M_3(\theta)$ can be expressed as

$$M_2(\theta) = 2\int_0^1 (1-u)\, H(u;\theta)\, du\ , \qquad (3.6)$$

$$M_3(\theta) = \int_0^1 H^2(u;\theta)\, du\ . \qquad (3.7)$$

Theorem 2. Suppose that Assumption A holds. If θ is the true value of the parameter, then the random sequence $N^{1/2}(\theta^{*}-\theta)$ has an asymptotic normal distribution with mean zero and variance σ_1^2, where

$$\sigma_1^2 = [M_1'(\theta)]^{-2}\, \tau_1^2$$

$$\tau_1^2 = \frac{1}{\lambda} M_3(\theta) + \frac{1}{1-\lambda} M_2(\theta) - \frac{1}{\lambda(1-\lambda)} M_1^2(\theta). \qquad (3.8)$$

Proof: From the theory of U-statistics (see for example, Lehmann(1975)), using expressions (3.3), (3.6) and (3.7) we have

$$\mathrm{Var}(W_{yx}) = \frac{1}{nm}\{\, M_1(\theta)[1-M_1(\theta)] + (n-1)[M_3(\theta)-M_1^2(\theta)]$$
$$+ (m-1)[M_2(\theta)-M_1^2(\theta)]\,\}.$$

Notice that

$$\frac{m}{N} \to \lambda,\ \frac{n}{N} \to 1-\lambda, \text{ as } N\to\infty,$$

so

$$\mathrm{Var}\,[N^{1/2}(W_{yx}- M_1(\theta)] \to \tau_1^2\ .$$

It is well known that the limiting distribution of the statistic $N^{1/2}[W_{yx}- M_1(\theta)]$ is normal with mean zero and variance τ_1^2. Hence by (3.5), the limiting distribution of $N^{1/2}(\theta^{*}-\theta_0)$ is $N(0,\sigma_1^2)$. #

Example 3.1. Suppose that $H(x;\theta) = x^{\theta}$. Then

$$M_1(\theta)= \frac{1}{\theta+1}, \quad M_2(\theta)= \frac{2}{(\theta+1)(\theta+2)}, \quad M_3(\theta)= \frac{1}{2\theta+1}.$$

Hence

$$\theta^{*} = [W_{yx}]^{-1} - 1 \text{ and } \operatorname{Var}(\theta^{*}) = N^{-1}\sigma_1^2 + o(N^{-1})$$

with

$$\sigma_1^2 = (\theta+1)^4 \left[\frac{1}{\lambda(2\theta+1)} + \frac{2}{(1-\lambda)(\theta+1)(\theta+2)} - \frac{1}{\lambda(1-\lambda)(\theta+1)^2} \right]. \tag{3.9}$$

4. The Comparison between $\hat{\theta}$ and θ^{*}.

Between the two estimators $\hat{\theta}$ and θ^{*}, we should prefer the one with smaller asymptotic variance. In Theorems 1 and 2, we have given the asymptotic variances of $\hat{\theta}$ and θ^{*}, denoted by σ_0^2 and σ_1^2, respectively. Furthermore, in Examples 2.1 and 3.1, we derived estimators $\hat{\theta}$ and θ^{*} when $H(x;\theta) = x^{\theta}$. The specific expressions of σ_0^2 and σ_1^2 are given in (2.4) and (3.9). For a selected number of values of θ and λ, we compute σ_0^2 and σ_1^2. Here σ_0^2 is the minimum variance corresponding to the optimal value of p. The results are listed in Table 4.1.

From Table 4.1, it is seen that for small values of θ, the estimator $\hat{\theta}$ is superior to θ^{*}, while θ^{*} is better for other values of θ. In other words, there is no overall winner. In particular, if we test the two sample problem H_0: F=G by using this model, i.e., testing H_0: θ=1, the estimator θ^{*} is preferred.

TABLE 4.1 Asymptotic variances of $\hat{\theta}$ and θ^* for the Lehmann alternative model $H(x;\theta) = x^\theta$.

	$\lambda = 1/3$	$\lambda = 1/2$	$\lambda = 2/3$
θ=0.1	σ_0^2=0.121	σ_0^2=0.120	σ_0^2=0.141
	σ_1^2=0.117	σ_1^2=0.135	σ_1^2=0.188
θ=0.5	σ_0^2=1.840	σ_0^2=1.655	σ_0^2=1.860
	σ_1^2=1.519	σ_1^2=1.463	σ_1^2=1.772
θ=1	σ_0^2=6.949	σ_0^2=6.177	σ_0^2=6.949
	σ_1^2=6.000	σ_1^2=5.333	σ_1^2=6.000
θ=1.5	σ_0^2=16.01	σ_0^2=14.24	σ_0^2=15.96
	σ_1^2=14.57	σ_1^2=12.39	σ_1^2=13.31

5. REFERENCES

Begun, J. M., Hall, W. J., Huang, W. M. and Wellner, J. A.(1983). 'Information and asymptotic efficiency in parametric-nonparametric models', Ann. Statist. 11, pp. 432 - 452.

Brooks, R. J.(1974). 'Bayesian analysis of the two-sample problem under the Lehmann alternatives', Biometrika, **61**, pp. 501 - 507.

Cox, D. R.(1972). 'Regression models and life tables', J. Roy. Statist. Soc. **B34**, pp. 187 - 220.

Fukui, M. and Miura, R.(1988). 'Kolmogrov-Smirnov estimation for the generalized Lehmann alternative models: two sample problem', in K. Matusita(ed.), Statistical Theory and Data Analysis II, Elsevier Science Publisher B. V. (North-Holland), pp. 101 - 110.

Lehmann, E. L.(1953). 'The power of rank tests', Ann. Math. Statist. **24**, pp. 23 - 43.

Lehmann, E. L.(1975). 'Nonparametrcs: statistical methods based on ranks', Holden-Day, San Francisco.

Savage, I. R.(1980). 'Lehmann alternatives', in B. V. Gnedenko, M. L. Puri and I. Vincze (eds.), Nonparametric Statistical inference II, North-Holland Publishing Company, pp. 795 - 821.

Young, D. H.(1973). 'Distributions of some censored rank statistics under Lehmann alternatives for the two sample case', Biometrika, **60**, pp. 543 - 549.