4^{th} Day. May 11, 2011 (= 3hours) [Non-parametric Statistics]

Nonparametric Inference for Generalized Lehmann's Alternatives

(Transformation Models)

(Hinted by "A Family of Distributions with Monotone Likelihood")

: (i) a principle of Hodges-Lehmann type estimation.

: **(ii). One Sample Problem.** Location Parameter.

: **(iii).Two Sample Problem.**

: (iv). Remarks.

(iv-a). Possible applications/relations of **Transformation Models** to statistical problems in Quantitative Finance.

e.g. Skew-Normal Distributions.

(iv-b). Some remarks for extensions of the above results in iid cases to **weakly dependent cases:** Brief Descriptions.

References

R.Miura (7). "A Note on the Principle of Hodges-Lehmann Type Estimation"

Keiei Kenkyu, Vol.37, No.5.6, pp. 185-192, January 1987

R.Miura (9). "Rank Estimates in a Class of Semiparametric Two-Sample Models"(with D.M. Dabrowska and K.A. Doksum), Annals of Institute of Statistical Mathematics, No.41, pp.63-79, 1989

R.Miura (10). "One-Sample Estimation for generalized Lehmann's Alternative Models" (co-author Tsukahara)

Statistica Sinica Vol. 3, No.1, January 1993.

R.Miura(1985). Special lecture at the meeting of Japanese association of Mathematics.

"Hodges-Lehmann type estimates and Generalized Lehmann's Alternative models"(in Japanese).

Remarks. Following

: Sana Louhichi (2000). "Weak Convergence for Empirical Processes of Associated Sequences." ***Ann. Inst. Henri Poincaré, Probabilités et Statistiques*** **36, 5 (2000) 547–567**

: Azzalini? Skew symmetry

Generalized Lehmann's Alternative Models (GLA models)

: **Ideas and Models.** Transformation of Distributions

: **Statistics.** Rank statistics.

: **Estimation under** H-L type principles.

: **Asymptotic normality.** Every-path arguement

: **Remarks.** Skew Symmetry.

From iid sequence to associated sequence.

Section 1. May 11. 2011

: (1-1). Generalized Lehmann's Alternative Model. (GLA model).

: (1-2).Remark.

Skew symmetry in GLA model.

: (1-3). One sample problem.

Statistics.

On the Model.

: Generalized Lehmann's Alternative Models.

:Recall Lehmann's Alternative in the book

: Hajek & Sidak (1967) [Theory of Rank Test].

§2. 一般化された Lehmann 対立仮説モデル

特徴：順位の確率の計算がFの知識がなくても行える。

例．
$$\begin{cases} X_i \overset{d.}{\sim} F, & 1 \le i \le m, \quad i.i.d. \\ Y_j \overset{d.}{\sim} G, & 1 \le j \le n, \quad i.i.d. \end{cases} \qquad N = m + n$$

$$F = \psi(G), \quad f = \psi'(G)\cdot g, \quad \psi : [0,1] \xrightarrow{1:1} [0,1], \nearrow$$

$$\Pr\{R_1 = r_1, \cdots, R_m = r_m\} = \int \cdots \int_S f(z_1)\cdots f(z_N)\, dz_1 \cdots dz_N$$

$S = \{(z_1, \cdots, z_N) : z_i$ の $(z_1,\cdots,z_N)$ における Rank が r_i, $1 \le i \le m\}$

$$= \binom{N}{m}^{-1} E\left\{\prod_{i=1}^{m} \frac{f(V(r_i))}{g(V(r_i))}\right\}, \quad \underset{1\le i\le N}{V(i)} \overset{d.}{\sim} G, \text{ order stat}$$

$$= \binom{N}{m}^{-1} E\left\{\prod_{i=1}^{m} \psi'(G(V(r_i)))\right\}$$

$$= \binom{N}{m}^{-1} E\left\{\prod_{i=1}^{m} \psi'(U(r_i))\right\}, \quad \underset{1\le i\le N}{U(i)} \sim [0,1]\text{一様からの order stat.}$$

確率は $\mathcal{F}\times\mathcal{F}$ 上のグラフ上で一定である。

$\psi^{-1}(t)$ の例

$$\psi^{-1}(t) = t^{\theta}, \quad 1-(1-t)^{\theta}, \quad (1-\theta)t + \theta t^2, \quad \frac{e^{\theta t}-1}{e^{t}-1},$$

$$\overset{\shortparallel}{h(t:\theta)} \qquad E(E^{-1}(t)+\theta), \quad : t \in [0,1]$$

Definition. Let Θ be an interval in the real line. A function $h(t;\theta)$ for $t \in (0,1)$ and $\theta \in \Theta$ which satisfies the following (1) and (2) is called the generalized Lehmann's alternative model;

(1) $h(0;\theta) = 0$ and $h(1;\theta) = 1$ for any $\theta \in \Theta$. $h(t;\theta)$ is a strictly monotone function of t.

(2) There exists $\theta^* \in \Theta$ such that $h(t;\theta^*) = t$ for $t \in (0,1)$. And for $\theta > \theta'$, $h(t;\theta) < h(t;\theta')$ for all t (or $<$ may be reversed for all t and $\theta > \theta'$).

We shall also call $h(F(\cdot);\theta)$ a generalized Lehmann's alternative model. In

Examples. Let F and G be d.f.'s which are connected through the generalized Lehmann's alternative model $G = h(F;\theta)$.
(i) If $h(t;\theta) = 1 - (1-t)^\theta$ for $\theta \in (0,\infty)$, then

$$\log \Lambda_G = \theta \log \Lambda_F,$$

where Λ_F and Λ_G are cumulative hazard functions corresponding to F and G respectively. This model is the well-known proportional hazards model proposed by Cox (1972).
(ii) Taking $h(t;\theta) = t[(1-t)\theta + t]^{-1}$ for $\theta \in (0,\infty)$ yields the proportional odds model:

$$\frac{G}{1-G} = \theta^{-1}\frac{F}{1-F}.$$

This model has been considered by Ferguson (1967) and in more general regression setting by Pettitt (1984), among others.
The above two models have useful and important

... and important applications in survival analysis. Other examples of our model include

(iii) $h(t;\theta) = (1-\theta)t + \theta t^2$ for $\theta \in [0,1)$ (Contamination),

(iv) $h(t;\theta) = (e^{\theta t} - 1)/(e^{\theta} - 1)$ for $\theta \in (0,\infty)$.

(iii) was considered in Lehmann (1953) and (iv) was found in Ferguson (1967). Both of these are Lehmann alternatives for which the locally most powerful rank test is Wilcoxon.

(v) $h(t;\theta) = t^{\theta}$ for $\theta \in (0,\infty)$ (Lehmann (1953)),

(vi) $h(t;\theta) = \sum_i c_i(\theta)t^i$ with $\sum_i c_i(\theta) = 1$ and $c_i(\theta) \geq 0$ for $\theta \in \Theta$ (Mixture of extremals by a discrete distribution).

(vii) $h(t;\theta) = E(E^{-1}(t) - \log\theta)$ for $\theta \in (0,\infty)$ where E is a known distribution function over the real line. This model can be rewritten as $\psi(X) = \log\theta + \epsilon$ where $X \sim G, \epsilon \sim E$ and $\psi = E^{-1} \circ F$, and includes (i) and (ii).

See Dabrowska, Doksum and Miura (1989) for other examples and Tsukahara (1991) for interesting relations among such models.

Estimation principle. Miura (1987)

A Note on the Principle of Hodges-Lehmann Type Estimation

Ryozo Miura
Department of Business, Osaka City University

ABSTRACT: The principle of Hodges-Lehmann type estimation is recognized, in this paper, to find the parameter value with which the transformed observations are uniformly distributed on the unit interval. The statistical models in this paper are expressed in a way that they allow such transformations. The models include those for location, scale and the generalized Lehmann's alternatives. The asymptotic behaviour of the maximum likelihood type estimates for the models are studied in general. The nonparametric estimability for the models is also considered. The procedures for the usual rank estimates and the maximum likelihood type estimates for a location parameter will be examined along with this principle.

2. Models and the Principle

(a) Let D and Θ be intervals in the real line. A function $\varphi(x\,;\theta)$ for $x \in D$, $\theta \in \Theta$ is called a parametrized transformation if

(i) for each $\theta \in \Theta$, $\varphi(x\,;\theta)$ is a continuous monotone function from D onto D.

(ii) for $\theta_1 > \theta_2$, $\varphi(x\,;\theta_1) < \varphi(x\,;\theta_2)$ for any $x \in D$.

A statistical model is the explicit expression of the "understanding" for the observation. Here our understanding is that the observation when transformed by a parametrized transformation coincides with a basic distribution; X is distributed with $G(\cdot) = F(\varphi(\cdot\,;\theta))$, equivalently $\varphi(X\,;\theta)$ is distributed with F, or $F(\varphi(X\,;\theta))$ is uniformly distributed on the (0, 1) interval.

For the one sample case where $X_1, X_2, \ldots, X_n$ are independent and indentically distributed (abbreviated to i.i.d.) with $G(\cdot)=F(\varphi(\cdot\ ;\theta))$, we are to find the parameter value $\hat{\theta}$ which makes the set of transformed observations

$$\{\varphi(X_i\,;\hat{\theta}) : i=1, 2, \ldots n\}$$

distributed as likely with F as possible, or

$$\{F(\varphi(X_i\,;\hat{\theta})) : i=1, 2, \ldots n\}$$

distributed as uniformly on (0, 1) as possible. When the hypothesis "$\theta=\theta_1$" is tested, we are to judge how uniformly the above set, with the hypothetical parameter value θ_1, is distributed. For the judgement of the uniformity of the set of the variables, several types of statistics may be available.

Remark

Skew symmetric distribution as a specific case of Generalized Lehmann's Alternative model(GLAM)

Skew symmetric distribution.

Let F be a symmetric distribution function,
say standard normal N(0,1). F(x)+F(-x)=1.
F(x) be skewed to G(x)=F(x){2F(δx)}.
When δ=0, we have G(x)=F(x).

Note that F in {2F(δx)} can be replaced
with any other symmetric distribution function.

From my viewpoint, $\{2F(\delta x)\}$ is acting to make F skewed.

Then, take $h(t:\delta) = t\{2F(\delta F^{-1}(t))\}$.

$G(x)=h(F(x):\delta) = F(x)\{2F(\delta x)\}$ can be regarded as an example of GLA model.

This will be a semi-nonparametric approach for skew symmetry problems.

In order to make a skewed shape, a family of $\{h(t:\theta), \theta \in \Theta\}$ can be used.

We may fit the data to see what functional form for $h(t:\theta)$ be suitable.

: Estimation of $h(t:\theta)$ when F is fixed and a known function.

$\{(F(X_{(i)}), G_n(X_{(i)})), i=1,2,...,n\}$ for observatons X_i, $i=1,2,...,n$,

where

$G_n(.)$ is an empirical distribution function of these observations.

So,

when the skew generating part $h(t:\delta) = t\{2F(\delta F^{-1}(t))\}$

satisfy our conditions (Homework 4),

then,

the asymptotic normality of estimators in skew symmetry models

had been given in the framework of GLA model,

at least mathematically.

One sample Problem.

Rank Statistics

and

Estimation based on Rank statistics.

Statistica Sinica 3(1993), 83-101

ONE-SAMPLE ESTIMATION FOR GENERALIZED LEHMANN'S ALTERNATIVE MODELS

Ryozo Miura and Hideatsu Tsukahara

Hitotsubashi University and University of Illinois at Urbana-Champaign

Abstract: This paper shows that nonparametric estimation of θ for generalized Lehmann's alternative models $h(F;\theta)$ is possible, even in the one-sample problem, when symmetry of the basic distribution function F about zero, $F(x) = 1 - F(-x)$, is assumed. Simultaneous nonparametric estimators of μ and θ for the model $h(F(\cdot - \mu);\theta)$ are also provided under the symmetry of F. The asymptotic normality of these estimators is proved under certain regularity conditions.

1. Introduction

In this paper we consider the following model: the observations $X_1, X_2, \ldots, X_n$ are independent and identically distributed (i.i.d.) with a distribution function (d.f.) $G(x;\mu,\theta) = h(F(x-\mu);\theta)$, where $h(t;\theta)$ is a known transformation on (0,1) which satisfies the conditions (1) and (2) below, and F is an unknown d.f. Then the observations are said to follow a distribution called Lehmann's alternative (Lehmann (1953)). The Lehmann's alternative is in general a transformation on the space of distributions, but in our model we parametrize this transformation and define as follows (Miura (1985)).

Definition. Let Θ be an interval in the real line. A function $h(t;\theta)$ for $t \in (0,1)$ and $\theta \in \Theta$ which satisfies the following (1) and (2) is called the generalized Lehmann's alternative model;
(1) $h(0;\theta) = 0$ and $h(1;\theta) = 1$ for any $\theta \in \Theta$. $h(t;\theta)$ is a strictly monotone function of t.
(2) There exists $\theta^* \in \Theta$ such that $h(t;\theta^*) = t$ for $t \in (0,1)$. And for $\theta > \theta'$, $h(t;\theta) < h(t;\theta')$ for all t (or $<$ may be reversed for all t and $\theta > \theta'$).

We shall also call $h(F(\cdot);\theta)$ a generalized Lehmann's alternative model. In terms of random variables, the observations following a generalized Lehmann's alternative model $h(F;\theta)$ are somehow the transformed values of the basic random variables whose d.f. is F. The d.f. F is treated as a nuisance parameter and we consider the problems of estimating θ when μ is known to be zero and of estimating θ and μ simultaneously. This model includes many useful models as follows.

In the one-sample problem, it is not possible to estimate θ for generalized Lehmann's alternative models $h(F;\theta)$, when F is unknown and no restrictions are made on the shape of F. The parameter θ is not even identifiable in that case. Throughout this paper we assume:

$$F \text{ is continuous and } F(x) = 1 - F(-x). \tag{1.1}$$

Also note that (2) in the definition of the generalized Lehmann's alternative model implies

$$h(t;\theta) + h(1-t;\theta) \neq 1 \text{ for } t \in (0,1) \text{ and } \theta \in \Theta - \{\theta^*\}. \tag{1.2}$$

Under (1.1) and (1.2), θ is identifiable and can be estimated.

In Section 2, X_i's are i.i.d. with d.f. $G(x;\theta) = h(F(x);\theta)$ and we introduce a statistic based on ranks of transformed X_i's. We then define our estimator of θ by a generalization of the method of Hodges and Lehmann (1963), and prove its asymptotic normality under certain mild regularity conditions. In Section 3, the observations X_i's are i.i.d. with d.f. $G(x;\mu,\theta) = h(F(x-\mu);\theta)$ and simultaneous nonparametric estimators for μ and θ are defined using rank statistics similar to the one in Section 2. We show joint asymptotic normality of the simultaneous estimators assuming some conditions in addition to those for the case of Section 2. See also Miura (1987) for the principle of these estimation procedures.

2. Estimation of θ

In this section, $X_1, X_2, \ldots, X_n$ are i.i.d. with d.f. $G(x) = h(F(x); \theta_0)$ and θ_0 is to be estimated.

Let $G_n(\cdot)$ be the empirical distribution function of X_i's, that is,

$$G_n(x) \triangleq n^{-1} \sum_{i=1}^{n} I_{[X_i \leq x]},$$

where I_A is an indicator function of a set A and let $\tilde{G}_n(x)$ be a linearized version of G_n: let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ be the order statistics of X_i's and define $\tilde{G}_n(x)$ by

$$\tilde{G}_n(x) \triangleq \frac{x + iX_{(i+1)} - (i+1)X_{(i)}}{(n+1)(X_{(i+1)} - X_{(i)})}, \quad x \in \left[X_{(i)}, X_{(i+1)}\right],$$

for $i = 0, 1, \ldots, n$ with $X_{(0)} = X_{(1)} - 1/n$ and $X_{(n+1)} = X_{(n)} + 1/n$. For $i = 1, 2, \ldots, n$, let

$$Z_i(r) \triangleq \tilde{G}_n^{-1}\left(h\left(\frac{i}{n+1}; r\right)\right),$$

and define

$$R_i^+(r) = \text{ the rank of } |Z_i(r)| \text{ among } \{|Z_j(r)| : j = 1, 2, \ldots, n\}.$$

Note that $\tilde{G}_n^{-1}(h(\cdot;\theta_0))$ may be viewed as an estimator of F^{-1} and so $Z_i(\theta_0)$'s can be regarded as an approximation of the ordered sample from F. Also, by

virtue of the smoothness of $\tilde{G}_n$, we can cope with the problem of ties among the $Z_i(r)$'s. Let $J(t)$ be a score function which is monotone increasing in $t \in (0,1)$ and assume that $J(t)$ has a continuous derivative $J'(t)$ and satisfies $\int_0^1 J(t)dt = 0$. Then the statistic we shall use for inference concerning θ_0 is

$$S_n(r) \triangleq \frac{1}{n} \sum_{i:Z_i(r)>0} J\left(\left(1+\frac{R_i^+(r)}{n+1}\right)/2\right) + \frac{1}{n} \sum_{i:Z_i(r)<0} J\left(\left(1-\frac{R_i^+(r)}{n+1}\right)/2\right). \quad (2.1)$$

If J is symmetric about $\frac{1}{2}$ in the sense that $J(t) = -J(1-t)$, $0 \le t < 1$, then it is easy to see that

$$S_n(r) = \frac{1}{n} \sum_{i=1}^{n} J^*\left(\frac{R_i^+(r)}{n+1}\right) \operatorname{sign} Z_i(r),$$

where $J^*(t) = J((1+t)/2)$, $0 \le t \le 1$. So that the statistic $S_n(r)$ may be regarded as a signed linear rank statistic. The point is that under (1.1) and (1.2) $Z_1(r), Z_2(r), \ldots, Z_n(r)$ are thought of as a sample from a symmetric distribution *only when* $r = \theta_0$, and $S_n(r)$ gives the strongest support to $r = \theta_0$ when it is closest to zero. This makes it possible to estimate θ even in the one-sample situation. Then our estimator $\hat{\theta}_n$ of θ_0 is defined as the value of r which makes $|S_n(r)|$ closest to zero. Such r exists since $S_n(r)$ is nonincreasing in r.

We can write

Graph of the statistics

as a function of r

Section 2.

: (2-1). Asymptotics (mathematics).

Outline of the proof for one sample problem.

: (2-2). Two sample problem.

Setting and statistics.

: (2-3). Remark.

Associated sequence of r.v.'s

asymptotics

mathematics

Next we shall state the assumptions which are necessary to prove the asymptotic normality of our estimator. Assume that $h(t;\theta)$ is continuously differentiable with respect to t and θ and let

$$h_1(t;\theta) \triangleq \frac{\partial}{\partial t} h(t;\theta), \qquad h_2(t;\theta) \triangleq \frac{\partial}{\partial \theta} h(t;\theta).$$

Let $u(t) = t(1-t)$. Assume, uniformly in θ in a neighborhood of θ_0,

(A.1) $$|J'(t)| \leq M\Big[u(h(t;\theta_0))\Big]^{-3/2+\delta}, \quad \text{for } \delta > 0$$

(A.2) $$\frac{1}{|h_1(t;\theta)|} \leq M < \infty$$

(A.3) $$|h_2(t;\theta)| \leq M\Big[u(h(t;\theta_0))\Big]^{1/2-\delta'}, \quad \text{for } \delta' > 0$$

where M is a universal constant. We require $\rho \triangleq \delta - \delta' > 0$. Further, assume

(A.4) $h_k(t;\theta) \sim h_k(t;\theta_0)$ uniformly in $t \in (0,1)$ as $\theta \to \theta_0$, $(k = 1, 2)$.

We can write

$$S_n(r) = \int_0^\infty J\left(\frac{1+H_{n,r}(x)}{2}\right) dL_{n,r}(x) + \int_{-\infty}^0 J\left(\frac{1-H_{n,r}(-x)}{2}\right) dL_{n,r}(x),$$

where

$$\begin{aligned}
u_n(t) &\triangleq \frac{1}{n}\left(\text{the number of } \left\{i : \frac{i}{n+1} \le t\right\}\right), \quad t \in (0,1), \\
L_{n,r}(x) &\triangleq \frac{1}{n}\left(\text{the number of } \{i : Z_i(r) \le x\}\right) \\
&= u_n\left(h^{-1}(\tilde{G}_n(x); r)\right), \quad x \in \mathbb{R}, \\
H_{n,r}(x) &\triangleq \frac{1}{n+1}\left(\text{the number of } \{i : |Z_i(r)| \le x\}\right), \quad x \in (0,\infty).
\end{aligned}$$

We set $H(x) \triangleq F(x) - F(-x)$ for $x \in (0,\infty)$.

For a function g on I ($I = [0,1]$ or $\mathbb{R}$), define $\|g\| = \sup_{t\in I}|g(t)|$. By Skorohod's representation theorem, there exists a probability space on which a sequence of i.i.d. uniform (0,1) random variables U_{ni}'s and a Brownian bridge U are defined and satisfy

$$\|U_n - U\| \xrightarrow{\text{a.s.}} 0, \quad n \to \infty, \tag{2.4}$$

where

$$\Gamma_n(t) \triangleq n^{-1}\sum_{i=1}^{n} I_{[U_{ni}\le t]}, \quad t \in (0,1),$$

$$U_n(t) \triangleq \sqrt{n}(\Gamma_n(t) - t), \quad t \in (0,1).$$

Using these U_{ni}'s, we shall represent the observation as $X_i = G^{-1}(U_{ni})$ for $i = 1, 2, \ldots, n$, which is called the special construction following Shorack and Wellner (1986). We shall then obtain convergence in probability of the estimator, but on the original probability space we can claim convergence in distribution only.

The following lemma is needed.

Lemma 2.1. *Let* $r = \theta_0 + b/\sqrt{n}$. *Then for the special construction* $X_i = G^{-1}(U_{ni})$ *and any given positive number* B, *we have, uniformly in* x *and* $|b| \leq B$,

$$\sqrt{n}\Big[L_{n,r}(x) - F(x)\Big] \xrightarrow{\text{a.s.}} A(F(x)), \quad n \to \infty, \tag{2.5}$$

where

$$A(t) \triangleq \frac{U(h(t;\theta_0))}{h_1(t;\theta_0)} - b \cdot \frac{h_2(t;\theta_0)}{h_1(t;\theta_0)}, \tag{2.6}$$

provided (A.2)-(A.4) *hold.*

Now set

$$\sigma^2(\theta) \triangleq \int_0^1 \alpha^2(t)dh(t;\theta) - \left[\int_0^1 \alpha(t)dh(t;\theta)\right]^2 + \int_0^{1'} \bar{\alpha}^2(t)dh(t;\theta) - \left[\int_0^1 \bar{\alpha}(t)dh(t;\theta)\right]^2 + 2\left[\int_0^1 \alpha(t)\bar{\alpha}(t)dh(t;\theta) - \int_0^1 \alpha(t)dh(t;\theta)\int_0^1 \bar{\alpha}(t)dh(t;\theta)\right],$$

and

$$\tau(\theta) \triangleq \int_0^1 h_2(t;\theta)d\{\alpha(t)+\bar{\alpha}(t)\},$$

where $\alpha(t)$ and $\bar{\alpha}(t)$ are defined by

$$\frac{d\alpha(t)}{dt} = \frac{J'(t)}{h_1(t;\theta)} \quad \text{and} \quad \frac{d\bar{\alpha}(t)}{dt} = \frac{J'(1-t)}{h_1(t;\theta)}$$

respectively.

Theorem 2.1. *Assume that $h(t;\theta)$ is continuously differentiable with respect to t and θ and $\tau(\theta_0) > 0$. Also let the assumptions* (1.1), (A.1)-(A.4) *hold. Then, as $n \to \infty$, we have*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(\theta_0)}{\tau^2(\theta_0)}\right).$$

Proof. Noting that $\int_0^1 J(t)dt = 0$, $\sqrt{n}S_n(r)$ can be expressed as

$$\begin{aligned} &\sqrt{n}\left[\int_0^\infty J\left(\frac{1+H_{n,r}(x)}{2}\right)dL_{n,r}(x) - \int_0^\infty J\left(\frac{1+H(x)}{2}\right)dF(x)\right] \\ +\ &\sqrt{n}\left[\int_{-\infty}^0 J\left(\frac{1-H_{n,r}(-x)}{2}\right)dL_{n,r}(x) - \int_{-\infty}^0 J\left(\frac{1-H(-x)}{2}\right)dF(x)\right]. \end{aligned} \quad (2.9)$$

Then the first term in (2.9) is decomposed to $\sum_{i=1}^2 B_{in} + \sum_{i=1}^3 C_{in}$ where

Then the first term in (2.9) is decomposed to $\sum_{i=1}^{2} B_{in} + \sum_{i=1}^{3} C_{in}$ where

$$\begin{aligned}
B_{1n} &\triangleq \int J\left(\frac{1+H}{2}\right) d\left\{\sqrt{n}(K_{n,r} - F)\right\}, \\
B_{2n} &\triangleq \frac{1}{2} \int \sqrt{n}(H_{n,r} - H) J'\left(\frac{1+H}{2}\right) dF, \\
C_{1n} &\triangleq \int J\left(\frac{1+H_{n,r}}{2}\right) d\left\{\sqrt{n}(L_{n,r} - K_{n,r})\right\}, \\
C_{2n} &\triangleq \frac{1}{2} \int \sqrt{n}\Big(H_{n,r} - H\Big) J'\left(\frac{1+H}{2}\right) d\Big(K_{n,r} - F\Big), \\
C_{3n} &\triangleq \sqrt{n} \int \left[J\left(\frac{1+H_{n,r}}{2}\right) - J\left(\frac{1+H}{2}\right) - \frac{1}{2}\Big(H_{n,r} - H\Big) J'\left(\frac{1+H}{2}\right)\right] dK_{n,r}.
\end{aligned}$$

Note that $(1+H)/2 = F$ due to the symmetry of F, which we shall use repeatedly without mention.

We now show that $\sum_{i=1}^{2} B_{in}$ converges in probability to a normal random variable. By (A.1)-(A.3) and the mean value theorem, Consequently

$$B_{1n} \xrightarrow{P} -\int_{\frac{1}{2}}^{1} \frac{U(h(t;\theta_0))}{h_1(t;\theta_0)} dJ(t) + b\int_{\frac{1}{2}}^{1} \frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} dJ(t) + \lambda(1/2),$$

where

$$\lambda(t) \triangleq \frac{U(h(1/2;\theta_0))J(1/2)}{h_1(1/2;\theta_0)} - bJ(1/2)\frac{h_2(1/2;\theta_0)}{h_1(1/2;\theta_0)}.$$

Concerning B_{2n}, note that Lemma 2.1 implies

$$\sqrt{n}\Big(H_{n,r}(x) - H(x)\Big) \xrightarrow{a.s.} A(F(x)) - A(1 - F(x)), \tag{2.12}$$

uniformly in $x > 0$ and $|b| \leq B$. Then, using argument as in B_{1n}, it is easy to see from (2.2), (2.12) and (A.1)-(A.4) that

$$\begin{aligned} B_{2n} \quad \xrightarrow{P} \quad & \frac{1}{2}\int_{\frac{1}{2}}^{1} \left[\frac{U(h(t;\theta_0))}{h_1(t;\theta_0)} - \frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)}\right] dJ(t) \\ & -\frac{b}{2}\int_{\frac{1}{2}}^{1} \left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} - \frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right] dJ(t). \end{aligned}$$

Thus

$$\sum_{i=1}^{2} B_{in} \xrightarrow{P} -\frac{1}{2}\int_{\frac{1}{2}}^{1}\left[\frac{U(h(t;\theta_0))}{h_1(t;\theta_0)}+\frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)}\right]dJ(t) + \frac{b}{2}\int_{\frac{1}{2}}^{1}\left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)}+\frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right]dJ(t)+\lambda(1/2). \tag{2.13}$$

Next we show that $\sum_{i=1}^{3} C_{in} \xrightarrow{P} 0$. For C_{1n}, note that $H_{n,r} \leq n/(n+1)$.

It can be seen in the similar way that the second term in (2.9) converges in probability to

$$-\frac{1}{2}\int_0^{\frac{1}{2}}\left[\frac{U(h(t;\theta_0))}{h_1(t;\theta_0)}+\frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)}\right]dJ(t)$$
$$+\frac{b}{2}\int_0^{\frac{1}{2}}\left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)}+\frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right]dJ(t)-\lambda(1/2).$$

Noting that

$$\int_0^1\left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)}+\frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right]dJ(t)=\tau(\theta_0),$$

we obtain asymptotic linearity: for any $B>0$

$$\sup_{|b|\le B}\left|\sqrt{n}S_n(r)+\frac{1}{2}T-\frac{1}{2}b\tau(\theta_0)\right|\xrightarrow{P}0, \tag{2.14}$$

where

$$T \triangleq \int_0^1 \left[\frac{U(h(t;\theta_0))}{h_1(t;\theta_0)} + \frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)}\right] dJ(t).$$

Now let $\epsilon > 0$ be a given number small enough to satisfy $\epsilon < \tau(\theta_0)/2$. Take $B_\epsilon > 1$ so large that

$$P\left\{|T| > \frac{B_\epsilon \tau(\theta_0)}{2}\right\} < \frac{\epsilon}{2}.$$

By asymptotic linearity (2.14), there exists an N_ϵ such that for all $n \geq N_\epsilon$,

$$P\left\{\sup_{|b| \leq B_\epsilon} \left|\sqrt{n} S_n(r) + \frac{1}{2}T - \frac{b}{2}\tau(\theta_0)\right| > \epsilon\right\} < \frac{\epsilon}{2}.$$

Thus for all $n \geq N_\epsilon$, any value b_n of b which minimizes $|\sqrt{n}S_n(r)| = |\sqrt{n}S_n(\theta_0 + b/\sqrt{n})|$ lies in $[-B_\epsilon, B_\epsilon]$ and it follows that

$$|b_n - T/\tau(\theta_0)| < \epsilon/\tau(\theta_0)$$

with probability exceeding $1-\epsilon$ (note that $T/\tau(\theta_0)$ minimizes $|-T/2+b\tau(\theta_0)/2|$). Noting that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is a value of b which minimizes $|\sqrt{n}S_n(r)|$, we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{P} \frac{T}{\tau(\theta_0)}.$$

An easy calculation shows that the random variable on the right-hand side has a normal distribution $N(0, \sigma^2(\theta_0)/\tau^2(\theta_0))$. Thus, as remarked above, we obtain the desired convergence in distribution of our estimator.

Remark. If $J(t) = -J(1-t)$, then $\alpha(t) = \bar{\alpha}(t)$, so that the asymptotic variance becomes simpler; in this case $\sigma^2(\theta)$ and $\tau(\theta)$ are given by

$$\sigma^2(\theta) = \int_0^1 \alpha^2(t)dh(t;\theta) - \left[\int_0^1 \alpha(t)dh(t;\theta)\right]^2,$$

and

$$\tau(\theta) = \int_0^1 h_2(t;\theta)d\alpha(t).$$

3. Simultaneous Estimation of μ and θ

In this section, let $X_1, X_2, \ldots, X_n$ be i.i.d. with d.f. $G(x) = h(F(x-\mu_0); \theta_0)$. The parameters μ_0 and θ_0 are both unknown and are to be estimated simultaneously.

Let $Z_i(r)$ be as in Section 2 and define

$$R_i^+(r,q) = \Big(\text{the number of } \{j : |Z_j(r) - q| \le |Z_i(r) - q|\}\Big).$$

In this section assume that F has a bounded continuous density f. Let $J_1(\cdot)$ and $J_2(\cdot)$ be the score function used for estimation of θ and μ respectively. $J_1(\cdot)$ and $J_2(\cdot)$ satisfy the conditions for the score functions in Section 2. In addition, $J_1(\cdot)$ and $J_2(\cdot)$ are assumed different enough to satisfy

$$\frac{\int_0^1 \left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} + \frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right] dJ_1(t)}{\int_0^1 \left[\frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} + \frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)}\right] dJ_2(t)} \neq \frac{\int_0^1 f(F^{-1}(t))dJ_1(t)}{\int_0^1 f(F^{-1}(t))dJ_2(t)}. \tag{3.1}$$

The rank statistics for the simultaneous inference of μ and θ are defined as follows:

$$S_{1n}(r,q) \triangleq \frac{1}{n}\sum_{i:Z_i(r)>q} J_1\left(\left(1+\frac{R_i^+(r,q)}{n+1}\right)/2\right)+\frac{1}{n}\sum_{i:Z_i(r)\leq q} J_1\left(\left(1-\frac{R_i^+(r,q)}{n+1}\right)/2\right), \tag{3.2}$$

and

$$S_{2n}(r,q) \triangleq \frac{1}{n}\sum_{i:Z_i(r)>q} J_2\left(\left(1+\frac{R_i^+(r,q)}{n+1}\right)/2\right)+\frac{1}{n}\sum_{i:Z_i(r)\leq q} J_2\left(\left(1-\frac{R_i^+(r,q)}{n+1}\right)/2\right). \tag{3.3}$$

(3.3)

Our estimators of μ and θ are derived from the simultaneous equations $S_{1n}(r,q) \approx 0$ and $S_{2n}(r,q) \approx 0$. Define

$$D_n \triangleq \left\{ (r,q) : \sum_{k=1}^{2} |S_{kn}(r,q)| = \min \right\}.$$

$D_n \subset \Theta \times \mathbf{R}$ is not empty for all $X_1, X_2, \ldots, X_n$ since $S_{kn}(r,q)$, as a function of r and q with fixed $X_1, X_2, \ldots, X_n$, takes on a finite number of different values. $S_{kn}(r,q)$, $(k = 1,2)$ are nonincreasing in each coordinate r and q separately, but it does not ensure the convexity of D_n, which may be used to determine the estimators uniquely. Our estimator $(\hat{\theta}_n, \hat{\mu}_n)$ is thus defined to be any point of D_n. Since $(\hat{\theta}_n, \hat{\mu}_n)$ may not be unique, there may be some arbitrariness in this definition. But, as will turn out in Theorem 3.2 below, all points in D_n are asymptotically equivalent; so, for large n, it will not matter much how $(\hat{\theta}_n, \hat{\mu}_n)$ chosen.

Define, for $x \geq 0$,

$$H_{n,r,q}(x) \triangleq \frac{1}{n+1}\Big(\text{the number of } \{i : |Z_i(r) - q| \leq x\}\Big).$$

Then we can write

$$R_i^+(r,q) = (n+1)H_{n,r,q}(|Z_i(r) - q|),$$

so that, for $k = 1, 2$, $S_{kn}(r,q)$ can be written as

$$\begin{aligned} S_{kn}(r,q) &= \int_q^\infty J_k\left(\frac{1 + H_{n,r,q}(x-q)}{2}\right) dL_{n,r}(x) \\ &\quad + \int_{-\infty}^q J_k\left(\frac{1 - H_{n,r,q}(-(x-q))}{2}\right) dL_{n,r}(x). \end{aligned}$$

To investigate the asymptotic behavior of S_{kn}, we assume, in addition to (A.1) with J replaced by J_k and (A.2)-(A.4),

$$|J_k'(t)| \leq M[u(t)]^{-1+\delta}, \quad \delta > 0. \tag{A.5}$$

We also introduce the following notation: let $r = \theta_0 + b_1/\sqrt{n}$, $q = \mu_0 + b_2/\sqrt{n}$ and

$$\boldsymbol{S}_n(r,q) \triangleq \Big(S_{1n}(r,q), S_{2n}(r,q)\Big)', \quad \boldsymbol{b} \triangleq (b_1, b_2)'.$$

Furthermore, for $k = 1, 2$

$$T_k \triangleq \int_0^1 \left\{ \frac{U(h(t;\theta_0))}{h_1(t;\theta_0)} + \frac{U(h(1-t;\theta_0))}{h_1(1-t;\theta_0)} \right\} dJ_k(t),$$

and set $\boldsymbol{T} \triangleq (T_1, T_2)'$. Let $D = [d_{kl}]$ denote a 2×2 matrix, where

$$\begin{aligned} d_{k1} &\triangleq \int_0^1 \left\{ \frac{h_2(t;\theta_0)}{h_1(t;\theta_0)} + \frac{h_2(1-t;\theta_0)}{h_1(1-t;\theta_0)} \right\} dJ_k(t), \\ d_{k2} &\triangleq -2\int_0^1 f(F^{-1}(t))\,dJ_k(t) \quad (k = 1, 2). \end{aligned}$$

Note that D is nonsingular because of (3.1). Then we have the following asymptotic linearity result.

Theorem 3.1. *Suppose that F has a bounded continuous density f and that* (A.1) *with J replaced by J_k and* (A.2)-(A.5) *all hold. Then*

$$\max_{k=1,2} \sup_{|b_k| \le B} \left| \sqrt{n} S_{kn}(r,q) + \frac{1}{2} T_k - \frac{1}{2}(d_{k1} b_1 + d_{k2} b_2) \right| \xrightarrow{P} 0, \quad n \to \infty, \tag{3.4}$$

for each $0 < B < \infty$.

Using matrix notation, express the relation (3.4) as

$$\sup_{|b_k| \le B} \left| \sqrt{n} \boldsymbol{S}_n(r,q) + \frac{1}{2} \boldsymbol{T} - \frac{1}{2} D \boldsymbol{b} \right| \xrightarrow{P} 0, \quad n \to \infty. \tag{3.5}$$

Proof of Theorem 3.1. Without loss of generality assume $\mu_0 = 0$. Note first that, as $n \to \infty$,

$$\begin{aligned} \left\| H_{n,r,q}(x) - H(x) \right\| &\xrightarrow{\text{a.s.}} 0, \quad x \geq 0, \\ \left\| L_{n,r}(x) - F(x) \right\| &\xrightarrow{\text{a.s.}} 0, \end{aligned}$$

and for the special construction $X_i = G^{-1}(U_{ni})$,

$$\begin{aligned} \sqrt{n}\Big[L_{n,r}(x) - F(x)\Big] &\xrightarrow{\text{a.s.}} A(F(x)), \\ \sqrt{n}\Big[H_{n,r,q}(x) - H(x)\Big] &\xrightarrow{\text{a.s.}} A(F(x)) - A(1 - F(x)), \quad x > 0, \end{aligned}$$

uniformly in x and $|b_k| \leq B$, where $A(t)$ is given by (2.6). Making use of

$$\int_{-\infty}^{\infty} J_k\left(\frac{1 + H(x)}{2}\right) dF(x) = 0,$$

we have

$$
\begin{aligned}
S_{kn}(r,q) \;=\; & \left[\int_q^\infty J_k\left(\frac{1+H_{n,r,q}(x-q)}{2}\right)dL_{n,r}(x) - \int_q^\infty J_k\left(\frac{1+H(x)}{2}\right)dF(x)\right] \\
& + \left[\int_{-\infty}^q J_k\left(\frac{1-H_{n,r,q}(-(x-q))}{2}\right)dL_{n,r}(x) \right. \\
& \left. - \int_{-\infty}^q J_k\left(\frac{1-H(-x)}{2}\right)dF(x)\right]. \qquad (3.6)
\end{aligned}
$$

We decompose the first term of the right-hand side of (3.6) to $\sum_{i=1}^{3} B_{in} + \sum_{i=1}^{3} C_{in}$, where

$$
\begin{aligned}
B_{1n} &\triangleq \int J_k\left(\frac{1+H(x-q)}{2}\right) d\left\{\sqrt{n}(K_{n,r}(x)-F(x))\right\},\\
B_{2n} &\triangleq \frac{1}{2}\int \sqrt{n}\Big(H_{n,r,q}(x-q)-H(x-q)\Big) J_k'\left(\frac{1+H(x-q)}{2}\right) dF(x),\\
B_{3n} &\triangleq \sqrt{n}\int \left[J_k\left(\frac{1+H(x-q)}{2}\right) - J_k\left(\frac{1+H(x)}{2}\right)\right] dF(x),\\
C_{1n} &\triangleq \int J_k\left(\frac{1+H_{n,r,q}(x-q)}{2}\right) d\left\{\sqrt{n}(L_{n,r}(x)-K_{n,r}(x))\right\},\\
C_{2n} &\triangleq \frac{1}{2}\int \sqrt{n}\Big(H_{n,r,q}(x-q)-H(x)\Big) J_k'\left(\frac{1+H(x-q)}{2}\right) d\Big(K_{n,r}(x)-F(x)\Big)\\
C_{3n} &\triangleq \sqrt{n}\int \left[J_k\left(\frac{1+H_{n,r,q}(x-q)}{2}\right) - J_k\left(\frac{1+H(x-q)}{2}\right)\right.\\
&\qquad \left. -\frac{1}{2}\Big(H_{n,r,q}(x-q)-H(x-q)\Big) J_k'\left(\frac{1+H(x-q)}{2}\right)\right] dK_{n,r}(x).
\end{aligned}
$$

Noting that the proof of Theorem 2.1 is valid uniformly in all continuous and symmetric F, one can use the same argument as in the proof of Theorem 2.1 to show the convergence of B_{1n} and B_{2n} and the asymptotic negligibility of $\sum_{i=1}^{3} C_{in}$.

Concerning B_{3n}, we have

$$\begin{aligned} B_{3n} &= b_2 \int \frac{J_k(F(x - b_2/\sqrt{n})) - J_k(F(x))}{b_2/\sqrt{n}} dF(x) \\ \longrightarrow \quad & -b_2 \int_0^\infty J_k'(F(x)) f(x) dF(x) = -b_2 \int_{\frac{1}{2}}^1 f(F^{-1}(t)) dJ_k(t), \end{aligned}$$

since f is bounded and continuous and (A.5) holds. This is verified by the dominated convergence theorem.

We can prove the convergence of the second term of the right-hand side of (3.6) quite similarly. We therefore obtain

$$\sqrt{n} S_{kn}(r, q) \xrightarrow{P} -\frac{1}{2} T_k + \frac{1}{2}(d_{k1} b_1 + d_{k2} b_2), \quad n \to \infty,$$

for $k = 1, 2$. Compactness of $[-B, B]$ and monotonicity of S_{kn} establishes the claimed uniformity in $|b_k| \leq B$ for each $0 < B < \infty$.

Once asymptotic linearity holds, one can see that each point of D_n has the same distribution as in Jurečková (1971). Let Σ denote the covariance matrix of $\boldsymbol{T}$. Then its k,l th entry σ_{kl} is given by

$$\begin{aligned}
\sigma_{kl} \triangleq\; & \int_0^1 \alpha_k(t)\alpha_l(t)dh(t;\theta_0) - \int_0^1 \alpha_k(t)dh(t;\theta_0) \int_0^1 \alpha_l(t)dh(t;\theta_0) \\
& + \int_0^1 \bar{\alpha}_k(t)\alpha_l(t)dh(t;\theta_0) - \int_0^1 \bar{\alpha}_k(t)dh(t;\theta_0) \int_0^1 \alpha_l(t)dh(t;\theta_0) \\
& + \int_0^1 \alpha_k(t)\bar{\alpha}_l(t)dh(t;\theta_0) - \int_0^1 \alpha_k(t)dh(t;\theta_0) \int_0^1 \bar{\alpha}_l(t)dh(t;\theta_0) \\
& + \int_0^1 \bar{\alpha}_k(t)\bar{\alpha}_l(t)dh(t;\theta_0) - \int_0^1 \bar{\alpha}_k(t)dh(t;\theta_0) \int_0^1 \bar{\alpha}_l(t)dh(t;\theta_0),
\end{aligned}$$

for $k,l = 1,2$, where $\alpha_k(t)$ and $\bar{\alpha}_k(t)$ are defined by

$$\frac{d\alpha_k(t)}{dt} = \frac{J_k'(t)}{h_1(t;\theta)} \quad \text{and} \quad \frac{d\bar{\alpha}_k(t)}{dt} = \frac{J_k'(1-t)}{h_1(t;\theta)}$$

respectively. Then d_{k1} becomes $\int_0^1 h_2(t;\theta_0)d\{\alpha_k(t) + \bar{\alpha}_k(t)\}$ for $k = 1,2$.

Theorem 3.2. *Suppose that all the conditions of Theorem 3.1 are satisfied. Then each point of D_n is asymptotically normal $N(0, D^{-1}\Sigma(D^{-1})')$, that is,*

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix} \xrightarrow{d} N\left(0, D^{-1}\Sigma(D^{-1})'\right),$$

as $n \to \infty$.

Proof of this theorem proceeds in a fashion quite similar to the last part of the proof of Theorem 2.1 and is not given here.

Remark 3.1. The above results are also simplified in the case that $J(t) = -J(1-t)$. The $k,1$ th entry of the matrix D becomes, for $k = 1, 2$,

$$d_{k1} = 2\int_0^1 h_2(t;\theta_0)d\alpha_k(t).$$

Also we get

$$\sigma_{kl} = 4\left[\int_0^1 \alpha_k(t)\alpha_l(t)dh(t;\theta_0) - \int_0^1 \alpha_k(t)dh(t;\theta_0)\int_0^1 \alpha_l(t)dh(t;\theta_0)\right].$$

Further, letting Λ denote a 2×2 matrix with k,l th entry λ_{kl} given by

$$\lambda_{11} \triangleq -\int_0^1 f(F^{-1}(t))dJ_2(t), \quad \lambda_{12} \triangleq \int_0^1 f(F^{-1}(t))dJ_1(t),$$

$$\lambda_{21} \triangleq -\int_0^1 h_2(t;\theta_0)d\alpha_2(t), \qquad \lambda_{22} \triangleq \int_0^1 h_2(t;\theta_0)d\alpha_1(t),$$

and

$$\gamma \triangleq \int_0^1 h_2(t;\theta_0)d\alpha_2(t)\int_0^1 f(F^{-1}(t))dJ_1(t) - \int_0^1 h_2(t;\theta_0)d\alpha_1(t)\int_0^1 f(F^{-1}(t))dJ_2(t),$$

we have

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\mu}_n - \mu_0 \end{pmatrix} \xrightarrow{d} N\left(0, \frac{1}{\gamma^2}\Lambda\Sigma\Lambda'\right).$$

Remark 3.2. The efficient scores for estimation are not yet known. But for testing, the locally most powerful rank test may be given with the following score functions:

$$J_1 : \left.\frac{\frac{\partial}{\partial\theta} g(x;\mu,\theta)}{g(x;\mu,\theta)}\right|_{x=G^{-1}(t;\mu,\theta)} = \frac{h_{12}(h^{-1}(t;\theta);\theta)}{h_1(h^{-1}(t;\theta);\theta)},$$

where

$$h_{12}(t;\theta) \triangleq \frac{\partial^2}{\partial t \partial \theta} h(t;\theta).$$

This is independent of μ and f, but depends on θ. For testing, θ may be the value for the null hypothesis.

$$J_2 : \left.\frac{\frac{\partial}{\partial\mu} g(x;\mu,\theta)}{g(x;\mu,\theta)}\right|_{x=G^{-1}(t;\mu,\theta)}$$
$$= -\frac{h_{11}(h^{-1}(t;\theta);\theta) f(F^{-1}(h^{-1}(t;\theta)))}{h_1(h^{-1}(t;\theta);\theta)} - \frac{f'(F^{-1}(h^{-1}(t;\theta)))}{f(F^{-1}(h^{-1}(t;\theta)))},$$

where

$$h_{11}(t;\theta) \triangleq \frac{\partial^2}{\partial t^2} h(t;\theta).$$

This is independent of μ, but depends on f and θ.

Two sample problem

•

1. Introduction

We consider the two sample problem where $X_1,\ldots,X_m$ and $Y_1,\ldots,Y_n$ are independent random samples from populations with continuous distribution functions F and G, respectively. Many of the models in which rank (partial, marginal) likelihood methods are useful can be put in the form

$$F(t) = D(H(t), \theta_1), \qquad G(t) = D(H(t), \theta_2)\,, \tag{1.1}$$

where $H(t)$ is an unknown continuous distribution function, $D(u, \theta)$ is a known continuous distribution function on $(0, 1)$, and θ_1 and θ_2 are in some parameter set Θ.

For inference based on rank likelihood, the above model is equivalent to the model obtained by using the distributions of $U_i = F(X_i) = D(H(X_i), \theta_1)$ and $V_j = F(Y_j) = D(H(Y_j), \theta_1)$. These distributions are

$$\tilde{F}(u) = u, \quad u \in (0,1), \qquad \tilde{G}(v) = D(D^{-1}(v, \theta_1), \theta_2), \quad v \in (0,1).$$

In the case where $\{D(u,\theta): \theta \in \Theta\}$ is a group under composition satisfying $D(u,1) = u$, $D(D^{-1}(u,\theta_1),\theta_2) = D(u,\theta)$, $\theta = \theta_2/\theta_1$, we can write

$$\tilde{F}(u) = u, \quad u \in (0,1), \qquad \tilde{G}(v) = D(v,\theta), \quad v \in (0,1). \tag{1.2}$$

From this point on we assume that (1.2) is satisfied. The distribution function F is treated as a nuisance parameter and we consider the problem of estimating θ. This model goes back to Lehmann (1953), and includes the following models that have important applications in survival analysis, reliability, and other areas.

Example 1.1. (*Proportional hazard model*) If F and G have proportional hazards, then $D(v, \theta) = 1 - [1 - v]^{1/\theta}$, $\theta > 0$. Lehmann (1953) and Savage (1956) considered testing in this model. Cox (1972, 1975) developed estimation procedures in a much more general regression problem with censored data.

Example 1.2. (*Proportional odds model*) For any continuous distribution H the odds rate is defined by $r_H = H/(1 - H)$. If F and G have proportional odds rates, in the sense that $r_G(t) = \theta^{-1} r_F(t)$, then $D(v, \theta) = v[(1 - v)\theta + v]^{-1}$. This model has been considered by Ferguson (1967), and Bickel (1986) in the two-sample case and in more general regression models by Bennett (1983) and Pettitt (1984), among others.

3. Rank-inversion estimates

In this section, we introduce rank-inversion estimates based on the ideas of Hodges-Lehmann (1963). Again, we start by assuming that F is known and let U_i, V_j and $D(u, \theta)$ be as in Section 2. In particular, we assume that $D(u, \theta)$ is monotone decreasing in θ. Note that $U_1, \ldots, U_m$, $D(V_1, \theta), \ldots, D(V_n, \theta)$ all have the same distribution when $\theta = \theta_0$, where θ_0 is the true value of the parameter. Let $R_i(\theta)$ denote the rank of U_i among $U_1, \ldots, U_m$, $D(V_1, \theta), \ldots, D(V_n, \theta)$, and let

$$T_N(\theta) = m^{-1} \sum_{i=1}^{m} J_N\left(\frac{R_i(\theta)}{N+1}\right),$$

denote a linear rank function with monotone increasing score function J_N. For F known, the Hodges-Lehmann estimate of θ is obtained by solving $T_N(\theta) = \int_0^1 J(u)\,du$ for θ, where $J(u)$ is the limit of $J_N(u)$. Without loss of generality, we assume $\int_0^1 J(u)\,du = 0$.

Suppose now that F is unknown. Let $X_{(1)} < \cdots < X_{(m)}$ be the vector of order statistics of X_i's. Let $\tilde{F}$ be defined by

$$\tilde{F}(u) = \frac{u + iX_{(i+1)} - (i+1)X_{(i)}}{(m+1)(X_{(i+1)} - X_{(i)})},$$

for $X_{(i)} \leq u \leq X_{(i+1)}$, $i = 1, \ldots, m-1$. Thus on the interval $[X_{(1)}, X_{(m)}]$, $\tilde{F}$ is a linearized version of the right-continuous distribution function $mF_m/(m+1)$, where $F_m(u) = m^{-1} \# \{i: X_i \leq u\}$. Further, let $Y_{(1)}$ and $Y_{(n)}$ be the first and the last order statistics of the Y_j's. If $Y_{(1)} < X_{(1)}$ or $Y_{(n)} > X_{(m)}$, then we extend $\tilde{F}$ to the interval $[\min(X_{(1)}, Y_{(1)}), \max(X_{(m)}, Y_{(n)})]$ linearly with $\tilde{F}(Y_{(1)}) = 1/(N+1)$ if $Y_{(1)} < X_{(1)}$ and $\tilde{F}(Y_{(n)}) = N/(N+1)$ if $Y_{(n)} > X_{(m)}$.

Further, let $\tilde{R}_i(\theta)$ be the rank of $\tilde{F}(X_i)$ among $\tilde{F}(X_1),\ldots,\tilde{F}(X_m)$, $D(\tilde{F}(Y_1),\theta),\ldots,D(\tilde{F}(Y_n),\theta)$. Let $\tilde{\theta}_R$ be any "solution" to

$$\tilde{T}_N(\theta) = m^{-1}\sum_{i=1}^{m} J_N\left(\frac{\tilde{R}_i(\theta)}{N+1}\right) = 0 .$$

More precisely, let $\tilde{\theta}_R$ be any point in $[\theta_R^*, \theta_R^{**}]$ where $\theta_R^* = \sup\{\theta: \tilde{T}_N(\theta) < 0\}$ and $\theta_R^{**} = \inf\{\theta: \tilde{T}_N(\theta) > 0\}$. Similar estimates have also been considered by Doksum and Nabeya (1984) and Miura (1985).

Example 3.1. Assuming F is known, the Hodges-Lehmann type rank estimate based on $J(u) = 2u - 1$ is asymptotically optimal for the proportional odds model. Let $L(x) = 1/[1 + e^{-x}]$ be the logistic distribution function and note that if we set $W_i = L^{-1}(F(X_i))$ and $Z_j = L^{-1}(F(Y_j))$, then W_i and Z_j follow a logistic shift model ($W_i \sim L(w), Z_j \sim L(z - \log\theta)$) with parameter $\log\theta$. Since the ranks are invariant under the increasing transformation L^{-1}, it follows that the Hodges-Lehmann estimate of θ is

$$\tilde{\theta}_{HL} = \exp\{\underset{i,j}{\text{median}}\,(Z_j - W_i)\}\,.$$

The corresponding $\tilde{\theta}_R$, which is appropriate when F is unknown, is

$$\tilde{\theta}_R = \exp\{\underset{i,j}{\text{median}}\,(L^{-1}(\tilde{F}(Y_j)) - L^{-1}(\tilde{F}(X_i)))\}\,.$$

We return to the general case and show the asymptotic normality of $\sqrt{N}(\tilde{\theta}_R - \theta_0)$. From (1.2), we have

$$D(D^{-1}(u,\theta),\delta) = D(u,\delta/\theta), \qquad D(u,1) = u\,. \tag{3.1}$$

Let

$$\dot{D}(u,\theta) = \frac{\partial}{\partial\theta} D(u,\theta) \qquad \eta = \int_0^1 J'(u)\dot{D}(u,1)\,du\,,$$

$$\tau^2(\theta) = \pi_1^{-1}\int_0^1 J^2(u)\,du + \pi_0^{-1}\left[\int_0^1 \alpha^2(u)\,du - \left(\int_0^1 \alpha(u)\,du\right)^2\right],$$

where α is defined by

$$\frac{d\alpha(u)}{du} = J'(D(u,\theta))[d(u,\theta)]^2\,.$$

Further, let

$$G_n^\theta(u) = n^{-1}\,\#\,[j\colon D(\tilde{F}(Y_j),\theta) \le u] = G_n\tilde{F}^{-1}D(u,\theta^{-1})\,,$$

$$H_N^\theta(u) = \{mF_m\tilde{F}^{-1}(u) + nG_n^\theta(u)\}/(N+1)\,,$$

$$G^\theta(u) = D(D^{-1}(u,\theta),\theta_0) = D(u,\theta_0/\theta)\,,$$

(by (3.1)) and $H^\theta(u) = \pi_0 u + \pi_1 G^\theta(u)$. In terms of these functions, we have

$$\tilde{T}_N(\theta) = m^{-1}\sum_{i=1}^{m} J_N\left(\frac{\tilde{R}_i(\theta)}{N+1}\right) = \int J_N(H_N^\theta(\tilde{F}))\,dF_m = \int J_N(H_N^\theta)\,dF_m\tilde{F}^{-1}\,.$$

Assume

(B.1) $r_{1N} = \sqrt{N}\int\{J_N(H_N^\theta) - J(H_N^\theta)\}dF_m\tilde{F}^{-1} \xrightarrow[P]{} 0$ as $N \to \infty$ uniformly

for θ in a neighbourhood of θ_0.

Moreover,

(B.2) J is a differentiable function with bounded continuous derivative J', and $0 < \int J^2(u)du < \infty$.

Finally, we assume that the limits $\pi_0 = \lim_{N\to\infty}(m/N)$ and $\pi_1 = \lim_{N\to\infty}(n/N)$ exist and are strictly between 0 and 1.

THEOREM 3.1. *If $D(u,\theta)$ is decreasing in θ, and if the preceeding conditions hold, then $\sqrt{N}(\tilde{\theta}_R - \theta_0)$ has asymptotically a normal distribution with mean zero and variance $\theta_0^2\tau^2(\theta_0)/\eta^2$.*

PROOF. As in the case of Hodges-Lehmann (1963), θ_R^*, θ_R^{**} and any point between them, such as $\tilde{\theta}_R$, will have the same asymptotic distribution. Further,

$$P(\sqrt{N}(\theta_R^*/\theta_0 - 1) \leq t) = P(\sqrt{N}\,\tilde{T}_N(\delta) > 0)\ ,$$

where $\delta = \theta_0(1 + t/\sqrt{N})$. We have

$$\begin{aligned}\sqrt{N}\,\tilde{T}_N(\delta) &= \sqrt{N}\int J_N(H_N^\delta(u))\,dF_m\tilde{F}^{-1}(u) \\ &= \sqrt{N}\int J(H_N^\delta(u))\,d[F_m\tilde{F}^{-1}(u) - u] + \sqrt{N}\int[J(H_N^\delta(u) - J(H^\delta(u))]\,du \\ &\quad + \sqrt{N}\int[J(H^\delta(u) - J(u)]\,du + r_{1N} = I_1 + I_2 + I_3 + r_{1N}\ .\end{aligned}$$

Remark.

Asymptotics.

Weak convergence of Empirical distribution of associated sequence of rv's has been proved By Louhichi(2000) and Yu(1993),Shao&Yu(1996).

Iid asymptotics could be extended to these weakly dependent cases.

Appendix 1.

Special construction where every-path argument is allowed.

Pyke and Shorack (1968)

The Annals of Mathematical Statistics
1968, Vol. 39, No. 3, 755–771

WEAK CONVERGENCE OF A TWO-SAMPLE EMPIRICAL PROCESS AND A NEW APPROACH TO CHERNOFF-SAVAGE THEOREMS[1]

BY RONALD PYKE AND GALEN R. SHORACK

University of Washington

0. Summary. An empirical stochastic process for two-sample problems is defined and its weak convergence studied. The results are based upon an identity which relates the two-sample empirical process to the more usual one-sample empirical process. Based on this identity a relatively simple proof of a Chernoff-Savage theorem is obtained. The c-sample analogues of these results are also included.

2. The 1-sample empirical process. For $m \geqq 0$ let $\{W_m(t): 0 \leqq t \leqq 1\}$ denote stochastic processes on a probability space $(\Omega, \mathfrak{A}, P)$ whose sample functions are points in some metric space $(\mathfrak{M}, \delta)$.

DEFINITION 2.1. We write $W_m \to_L W_0$ relative to $(\mathfrak{M}, \delta)$ if $\lim_{m\to\infty} E[\psi(W_m)] = E[\psi(W_0)]$ for all bounded real functionals ψ defined on $\mathfrak{M}$ which are continuous in the δ-metric and are such that $\psi(W_m)$, $m \geqq 0$, are measurable with respect to $\mathfrak{A}$. Such convergence is called *convergence in law* or *weak convergence*. (If the W_m process and W_0-process are measurable with respect to the Borel sets of $(\mathfrak{M}, \delta)$, so that their image laws on $\mathfrak{M}$ are well defined, then the above definition is equivalent to the usual definition of weak convergence as given in Prokhorov (1956) for example.)

Suppose $\mathfrak{M} = D$, the set of all right continuous real valued functions on [0, 1] having only jump discontinuities. In this case two possible metrics are $\delta = \rho$, the uniform metric defined by

$$\rho(f, g) = \sup_{0 \leq t \leq 1} |f(t) - g(t)|, \tag{2.1}$$

and $\delta = d$, the Prokhorov metric on D as defined by Prokhorov (1956). Prokhorov showed that (D, d) is a complete separable metric space and that $U_m \to_L U_0$ relative to (D, d). Actually, since all jumps of the U_m-process equal $m^{-\frac{1}{2}}$, it is possible to show that $U_m \to_L U_0$ relative to the stronger uniform topology of the non-separable metric space (D, ρ). We will obtain this result as Lemma 2.1. (It should be pointed out that Dudley (1966) gives a definition of weak convergence for non-separable spaces which is more general than Definition 2.1 above in that his use of upper and lower integrals enables him to place a less restrictive assumption of measurability upon the function ψ. Also, the statement "$U_m \to_L U_0$ on (D, ρ) in the sense of Prokhorov's definition" is false; see Chibisov (1965) for a statement of the measurability difficulties.)

Since $U_m \to_L U_0$ relative to (D, d) and (D, d) is a complete separable metric space it is possible, (see item 3.1.1 of Skorokhod (1956)), to construct processes $\{\tilde{U}_m(t): 0 \leqq t \leqq 1\}, m \geqq 0$, with sample functions in D and having the same finite dimensional df's as $\{U_m(t): 0 \leqq t \leqq 1\}$, $m \geqq 0$, but which in addition satisfy $d(\tilde{U}_m, \tilde{U}_0) \to_{a.s.} 0$. Let us make an independent construction for the V_n-processes so that

$$(2.2) \qquad d(\tilde{U}_m, \tilde{U}_0) \to_{a.s.} 0, \qquad d(\tilde{V}_n, \tilde{V}_0) \to_{a.s.} 0$$

where all processes are defined on a single probability space $(\tilde{\Omega}, \tilde{\mathfrak{A}}, \tilde{P})$. This is the probability space we shall work on in what follows. Note that if we set $\tilde{F}_m = m^{\frac{1}{2}}\tilde{U}_m(F) + F$, then $\tilde{F}_m$ is a.s. a df having exactly m discontinuities each of magnitude m^{-1}. (We shall henceforth drop the symbol $\sim$ from the notation.)

Based on the above construction, we shall prove a series of lemmas about the U_m-processes.

Warning. The results obtained below which involve convergence stronger than convergence in law may apply only to the specially constructed processes. Only the implied convergence in law should be assumed to hold for the original processes unless further checking is done.

Appendix 2.

: Associated sequence of random variables.

: Convergence of empirical distribution

: Sana Louhichi (2000) and H.Yu(1993),Q.M.Shao & H.Yu(1996)

[13] Shao Q.M., Yu H., Weak convergence for weighted empirical processes of dependent sequences, Ann. Probab. 24 (4) (1996) 2052–2078.

[15] Yu H., A Glivenko–Cantelli lemma and weak convergence for empirical processes of associated sequences, Probab. Theory Related Fields 95 (1993) 357–370.

Ann. Inst. Henri Poincaré, Probabilités et Statistiques 36, 5 (2000) 547–567

Weak convergence for empirical processes of associated sequences

by
Sana LOUHICHI 1
UMR C 8628 CNRS, Université de Paris-Sud, Bât. 425, Mathématiques, F-91405 Orsay Cedex, France

Manuscript received in 29 October 1998, revised 22 March 2000

1. INTRODUCTION, NOTATIONS AND PREVIOUS RESULTS

Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary sequence of random variables (r.v's) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let F be the common distribution function of $(X_n)_{n\in\mathbb{Z}}$. The empirical distribution function F_n of $X_1, \ldots, X_n$ is defined as:

$$F_n(x) := F_n(x, \omega) = \frac{1}{n} \sum_{1\leqslant i\leqslant n} \mathbf{1}_{X_i(\omega)\leqslant x}, \quad x \in \mathbb{R}.$$

The empirical process G_n based on the observations $X_1, \ldots, X_n$ is defined by:

$$G_n(x) := G_n(x, \omega) = \sqrt{n}\left[F_n(x, \omega) - F(x)\right]. \tag{1}$$

Let $D[-\infty, +\infty]$ be the space of *cadlag* functions on $[-\infty, +\infty]$ having finite limits at $\pm\infty$. Suppose that $D[-\infty, +\infty]$ is equipped with the Skorohod topology. The usual Empirical Central Limit Theorem (ECLT) gives conditions under which the empirical process $\{G_n(x), \ x \in \mathbb{R}\}$ converges in distribution, as a random element of $D[-\infty, +\infty]$, to a Gaussian process G with zero mean and covariance

$$\mathrm{Cov}(G(x), G(y)) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbf{1}_{X_0\leqslant x}, \mathbf{1}_{X_k\leqslant y}). \qquad (2)$$

The proof of such theorem requires two steps:

Step 1. Establish the convergence of finite-dimensional distributions.

Step 2. Establish the tightness property.

In general, it remains to prove step 2 since step 1 follows from a suitable central limit theorem, usually well known.

For the sake of simplicity, we suppose in the sequel that the marginal distribution function F is *continuous* on $\mathbb{R}$. This restriction allows to suppose that the marginal law is $\mathcal{U}([0, 1])$: the uniform law over $[0, 1]$ (cf. Billingsley [2]).

Definition.

$\{X_n$, n= 1,2,....$\}$ is a sequence of associate random variables
if for every finite subcollection $X_{i1}, ..., X_{in}$ and
for every pair of coordinatewise non decreasing functions
h(.) and k(.) : $\mathbb{R}^n \to \mathbb{R}$,

$$\mathrm{Cov}(h(X_{i1}, ..., X_{in}), k(X_{i1}, ..., X_{in})) \geq 0,$$

whenever the covariance is defined.

Memo: This paper mentions that Gaussian processes are associated
if and only if their covariance function is positive,
referring a paper(Pitt(1982)).

2. MAIN RESULT AND APPLICATION

THEOREM 1. – *Let $(X_n)_{n\in\mathbb{Z}}$ be a stationary associated sequence with continuous marginal distribution F. Assume that, for $n \in \mathbb{N}^*$,*

$$\operatorname{Cov}\big(F(X_1), F(X_n)\big) = \mathcal{O}\big(n^{-b}\big), \quad \textit{for } b > 4. \tag{10}$$

Then

$$G_n(.) \to G(.) \quad \textit{in } D[-\infty, +\infty],$$

where $G_n(.)$ is defined by (1) *and G is the zero-mean Gaussian process with covariance defined by* (2).

$\mathrm{Cov}(F(X_1), F(X_n))$ can be written as,

$$\mathrm{Cov}(F(X_1), F(X_n)) = \iint f(u)f(s)\mathrm{Cov}(1_{\{X_1 \le u\}}, 1_{\{X_n \le s\}})duds$$

$$\le \| f \|_{\infty}^{2} \mathrm{Cov}(X_1, X_n).$$

$(Memo: (Y1, Yn)$ be independent of (X1,Xn) and be distributed as the same as (X1,Xn).

Then, we have

$$\mathrm{Cov}(F(X1),F(Xn)) = \frac{1}{2} E[\{F(X1)-F(Y1)\}\{F(Xn)-F(Yn)\}]$$

Note: $F(X)=P\{V \le X\}=E_F(1\{V \le X\}) = \int 1\{v \le X\} f(v)dv$

, where V follows the distribution F. Also note;

$$F(x)-F(y)=\int [1\{v \le x\}-1\{v \le y\}]f(v)dv.$$

Representation of

Statistics as functionals of empirical distribution functions

and

Von Mises' Type Asymptotic Theory.

Basics made Simple.

$X_1, X_2, \ldots, X_{n,}$

$X_i \sim F$

An Empirical Distribution Function:

$\mathrm{F}_n(\mathrm{x}) = \sum_{i=1}^{n} 1_{\{\mathrm{X_i} < \mathrm{x}\}}$.

Von Mises Functional T is a functional of F_n.

For a precise definition, see Filippova(1962).

$\mathrm{T}[\mathrm{F}_n] = \int \psi(x) \mathrm{d}[\mathrm{F}_n(x) - \mathrm{F(x)}]$: Statistics

Rough sketch for Asymptotic Theory of Statistics.

$X \sim F$

$\varphi(x:\theta)$

θ: estimand; $\int \varphi(x:\theta)\mathrm{dF(x)} = 0$

Observe X_i, i=1,2,.... then we have F_n .

θ_n is an estimator ; $\int \varphi(x:\theta_n)\mathrm{d}F_n(x) = 0$

$$0 = \int \varphi(x:\theta_n)\mathrm{d}F_n(x) - \int \varphi(x:\theta)\mathrm{dF(x)}$$

$$\approx \int \varphi(x:\theta_n)\mathrm{d}[F_n(x) - F(x)] + \int \{\varphi(x:\theta_n) - \varphi(x:\theta)\}\mathrm{dF(x)}$$

$$\approx \int \varphi(x:\theta)\mathrm{d}[F_n(x) - F(x)] + (\theta_n - \theta)\int \{\frac{\partial}{\partial\theta}\varphi(x:\theta)\}\mathrm{dF(x)}$$

Simple Example

Mean

Variance

Covariance

[Classics: X_i , i=1,2,...n. iid with G(.).]

: Empirical Disribution Function $G_n(.)$.

Then, Estimand and Estimator are: for example, for means

: Estimand: θ such that $\int (x-\theta)dG(x)=0$

: Estimator: $\hat{\theta}$ such that $\int (x-\hat{\theta})dG_n(x)=0$

in Fillipova(1962.Theory of Probability and its applications)

Or, "substitution principle" says, $\int xdG(x)$ can be estimated by $\int xdG_n(x)$.

In that world, $\{G_n(.)-G(.)\}$ played a fundamental role for deriving the asymptotic distribution of the estimation erros (asymptotic normality).

Remark: This can be discussed, as an extention, for the case where observations X_i , i=1,2,...n, are "weakly dependent".

Variance σ^2=E[(X-μ)^2], μ=E[X].

Note : $\int (\text{x-}\mu)^2 f(x)dx = \sigma^2$.

So, $\int (\frac{x-\mu}{\sigma})^2 f(x)dx - 1 = 0,\ 1 = \int$ dF.

Then, $\int \{(\frac{x-\mu}{\sigma})^2 - 1\}$dF(x)=0.

$$\sigma^2{}_n = \int (x - \mu)^2 d\, F_n(x)$$

:*** φ(x: σ) = $(\frac{x-\mu}{\sigma})^2 - 1$.*****

Homework

: (1). Discuss if

$$h(t:\delta) = t\{2F(\delta F^{-1}(t))\}$$

satisfies the conditions (1) and (2) in slide page 7;

Definition. Let Θ be an interval in the real line. A function $h(t;\theta)$ for $t \in (0,1)$ and $\theta \in \Theta$ which satisfies the following (1) and (2) is called the generalized Lehmann's alternative model;
(1) $h(0;\theta) = 0$ and $h(1;\theta) = 1$ for any $\theta \in \Theta$. $h(t;\theta)$ is a strictly monotone function of t.
(2) There exists $\theta^* \in \Theta$ such that $h(t;\theta^*) = t$ for $t \in (0,1)$. And for $\theta > \theta'$, $h(t;\theta) < h(t;\theta')$ for all t (or $<$ may be reversed for all t and $\theta > \theta'$).

We shall also call $h(F(\cdot);\theta)$ a generalized Lehmann's alternative model. In

: (2). Have you been learning anything from my lectures?

Please describe briefly. Also, please give me an information of your specialized field of study; such as Mathematical Statistics, Applied Statistics and Probability Theory and etc.