An Alternative Regularity Condition for Hajek's Representation Theorem

by

Luke Tierney
School of Statistics
University of Minnesota

Technical Report No. 465

Abstract

Hajek's representation theorem states that under certain regularity conditions the limiting distribution of an estimator can be written as the convolution of a certain normal distribution with some other distribution. This result, originally developed for finite dimensional problems, has been extended to a number of infinite dimensional settings where it has been used, for example, to establish the asymptotic efficiency of the Kaplan-Meier estimator. The purpose of this note is to show that the somewhat unintuitive regularity condition on the estimators that is usually used can be replaced by a simple one: It is sufficient for the asymptotic information and the limiting distribution of the estimator to vary continuously with the parameter being estimated.

Introduction

Consider the problem of estimating a real valued parameter θ using a sequence of estimators $\{T_n\}$ based on data from a distribution with a well behaved likelihood. Hajek's representation theorem (Hajek (1970), and Roussas (1972) with a characteristic function proof due to Bickel) states that under certain regularity conditions on the sequence $\{T_n\}$ the limiting distribution

$$\mathcal{L}(\theta) = \lim_{n \to \infty} \mathcal{L}(\sqrt{n}(T_n - \theta)|\theta)$$

can be written as

$$\mathcal{L}(\theta) = N(0, i(\theta)^{-1}) * \mathcal{L}_1(\theta)$$

for some distribution $\mathcal{L}_1$. Here $\mathcal{L}(Y|\theta)$ denotes the distribution of the random variable Y when the true parameter is θ, $i(\theta)$ denotes the asymptotic information and * represents the convolution operator. Convergence of distributions is in the sense of weak convergence.

Hajek's representation theorem is useful for studying the asymptotic efficiency of estimators. Recently it has been extended to nonparametric settings where it has been used to show that the empirical distribution function (Beran (1977b)), the Kaplan-Meier estimator (Wellner (1982)) and Cox' partial likelihood estimators for the proportional hazards model (Begun et. al. (1983)) are asymptotically efficient. All these extensions use as their regularity condition on their estimators a variation of Hajek's original condition which states that the representation theorem holds at any θ where $\mathcal{L}(\sqrt{n}(T_n-\theta_n)|\theta_n) \to \mathcal{L}(\theta)$ for any sequence θ_n of the form $\theta_n = \theta + O(n^{-1/2})$. An estimator satisfying this condition at a particular θ will be called <u>Hajek regular</u> at θ (see Wong (1985)).

A regularity condition on the sequence $\{T_n\}$ is needed to rule out superefficiency. The local condition of Hajek regularity is rather natural from a mathematical point of view since it fits readily into the proof. On the other hand, by taking a more global point of view (and at the expense of adding a layer to the proof) it is possible to show that an alternative condition that may be easier to interpret and to verify is also sufficient: If the parameter space is an open set, the limit $\mathcal{L}(\theta) = \lim_{n\to\infty} \mathcal{L}(\sqrt{n}(T_n-\theta)|\theta)$ exists for all θ and $\mathcal{L}(\theta)$ is continuous in θ (in the weak convergence topology), then Hajek's representation is valid for all θ. A proof of this result in this one dimensional setting is given in the next two sections; by choosing a suitable one dimensional subfamily the proof can be extended to the nonparametric settings mentioned above. Simple examples given in the final section show that this alternative regularity condition is neither implied by nor does it imply Hajek regularity. Before stating the theorem we formulate our regularity condition on the likelihood.

<u>A Well Behaved Likelihood</u>

Rather than state explicit sufficient conditions on the likelihood we adopt the following convention: The likelihood will be called well-behaved at θ if there exists a number $i(\theta)$ such that for any $\{T_n\}$ that is Hajek-regular at θ we have

$$\mathcal{L}(\sqrt{n}(T_n-\theta)|\theta) \longrightarrow N(0,i(\theta)^{-1})*\mathcal{L}_1(\theta)$$

for some distribution $\mathcal{L}_1(\theta)$. The likelihood will be called well behaved on an open subset Ω of $\mathbb{R}$ if it is well behaved at every $\theta\varepsilon\Omega$ and $i(\theta)$ is continuous on Ω.

Explicit sufficient conditions to insure that the likelihood is well-behaved can be found in the references cited above.

The Representation Theorem

Let Ω be an open subset of $\mathbb{R}$ and assume that the likelihood is well behaved on Ω. Then we have the following result.

Theorem

Suppose $\{T_n\}$ is such that $\mathcal{L}(\sqrt{n}(T_n-\theta)|\theta) \dashrightarrow \mathcal{L}(\theta)$ for all $\theta\varepsilon\Omega$ and $\mathcal{L}(\theta)$ is continuous in θ. Then for each θ there exists a distribution $\mathcal{L}_1(\theta)$ such that

$$\mathcal{L}(\theta) = N(0,i(\theta)^{-1})*\mathcal{L}_1(\theta).$$

The proof is based on the fact that for continuous $\mathcal{L}(\theta)$ Hajek's representation theorem can only fail for a set of θ values with measure zero, whereas the assumed continuity of $\mathcal{L}(\theta)$ and $i(\theta)$ implies that if the representation fails to hold for some θ it must fail on an interval of θ's, thus producing a contradiction. To begin the proof, note that Hajek regularity is used in Bickel's proof to show that $\mathcal{L}(\sqrt{n}(T_n-\theta_n)|\theta_n) \dashrightarrow \mathcal{L}(\theta)$ for $\theta_n = \theta + h\frac{1}{\sqrt{n}}$, and any h which, in turn, is used to derive a characteristic function identity that implies the representation theorem. The basis of the derivation is an analytic continuation argument for a function of h. To use this argument it is sufficient to prove that the functional identity holds for a bounded infinite set of h values. We call $\{T_n\}$ <u>weakly regular at θ</u> if there exists a bounded infinite subset H of $\mathbb{R}$ such that for any $h \in H$ there exists a subsequence n(k) of integers for which $\mathcal{L}(\sqrt{n}(k)(T_{n(k)}-\theta_{n(k)})|\theta_{n(k)}) \dashrightarrow \mathcal{L}(\theta)$ if $\theta_n = \theta + h/\sqrt{n}$. Then the representation will hold at θ if $\{T_n\}$ is weakly regular at θ. We state this as a lemma.

<u>Lemma 1</u>

If $\{T_n\}$ is weakly regular at θ then there exists a distribution $\mathcal{L}_1(\theta)$ such that

$$\mathcal{L}(\theta) = N(0, i(\theta)^{-1}) * \mathcal{L}_1(\theta)$$

The proof is a straightforward modification of Bickel's proof of the Hajek representation theorem as given in Roussas (1972) or Beran (1977a) and is therefore omitted.

Next note that any $\{T_n\}$ satisfying the continuous limit hypothesis of the theorem is weakly regular at Lebesque-almost all θ:

<u>Lemma 2</u>

If $\{T_n\}$ is such that $\mathcal{L}(\sqrt{n}(T_n-\theta)|\theta) \dashrightarrow \mathcal{L}(\theta)$ and $\mathcal{L}(\theta)$ is continuous in θ then $\{T_n\}$ is weakly regular for Lebesque-almost all θ in Ω.

Proof

This proof is a modification of Bahadur's (1964) proof of his Lemma 4. Let $\rho(F,G)$ be a bounded metric inducing weak convergence and let

$$f_n(\theta_1,\theta_2) = \begin{cases} \rho(\mathcal{L}(\sqrt{n}(T_n-\theta_1)|\theta_1),\ \mathcal{L}(\theta_2)) & \text{for } \theta_1,\ \theta_2 \varepsilon \Omega \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f_n(\theta+h/\sqrt{n},\theta) \leq f(\theta+h/\sqrt{n},\theta+h/\sqrt{n}) + \rho(\mathcal{L}(\theta+\frac{h}{\sqrt{n}}),\ \mathcal{L}(\theta)).$$

Since $\mathcal{L}(\cdot)$ is continuous the second term tends to zero for all θ and h. On the other hand, if

$$g_n(\theta) = f_n(\theta,\theta),$$

and Φ is the standard normal distribution then

$$\int g_n(\theta+h/\sqrt{n})d\Phi(\theta)$$

$$= \int g_n(\theta)\exp\{-\frac{h^2}{2n} + \frac{h}{\sqrt{n}}\theta\}\ d\Phi(\theta)$$
$$\rightarrow 0$$

by dominated convergence. Thus for any h we have $g_n(\theta) \rightarrow 0$ in Φ measure.

Hence there is a subsequence $g_{n(k)}$ that converges to zero Φ -a.e. and hence Lebesque -a.e.. Now consider a bounded sequence h_n of distinct real numbers and take the union of all the corresponding null sets. At all θ in the complement of that union, and thus at Lebesque almost all θ, the sequence $\{T_n\}$ is weakly regular.

These two lemmas produce the proof of the theorem.

Proof of Theorem

Suppose the representation fails to hold for some θ_o. Since the mapping $\theta \dashrightarrow (\mathcal{L}(\theta),\theta)$ is continuous and the set

$$A = \{(N(0,i(\theta)^{-1})*\mathcal{L},\theta): \mathcal{L} \text{ a distribution}, \theta\varepsilon\Omega\}$$

is closed in the product topology, if $(\mathcal{L}(\theta_o),\theta_o) \notin A$ then we must have $(\mathcal{L}(\theta),\theta)\notin A$ for all θ in some neighborhood of θ_o. But, by Lemma 2, $\{T_n\}$ is weakly regular at almost all points in that neighborhood, which, by Lemma 1, provides a contradiction.

Comparison of Regularity and the Continuous Limit Condition

It is easy to construct examples of estimators that satisfy one of these conditions but not the other. Thus neither condition implies the other. Let $X_1,\ldots X_n$ be iid $N(\theta,1)$, let $T_n = \bar{X}$ if $|\bar{X}| \geq 1/\log n$ and

$T_n = \bar{X}*(1-\frac{1}{\sqrt{n}}) + X_1(\frac{1}{\sqrt{n}})$ if $|\bar{X}| < 1/\log n$. Then $\mathcal{L}(\theta) = N(0,1)$ for $\theta \neq 0$ but $\mathcal{L}(0) = N(0,2)$. So $\mathcal{L}(\theta)$ is not continuous at zero. On the other hand, if $\theta_n = O(n^{-1/2})$ then $\mathcal{L}(\sqrt{n}(T_n-\theta)|\theta_n) \dashrightarrow \mathcal{L}(0) = N(0,2)$ since

$P\{T_n \neq \bar{X}(1-\frac{1}{\sqrt{n}}) + X_1 \frac{1}{\sqrt{n}} \mid \theta_n\} \dashrightarrow 0$. So $\{T_n\}$ is Hajek regular at θ.

To find an example where T_n is continuous but not regular let f be some infinitely differentiable function such that $f(x) = 1$ if $x \leq 0$ or $x \geq 2$ and $f(1) = 2$, and for each n let $X_{1,n},\ldots,X_{n,n}$ be iid $N(\theta,f(\theta\sqrt{n}))$. Then $T_n = \bar{X}_n$ has $\mathcal{L}(\theta) = N(0,1)$ for all θ, ie $\mathcal{L}(\theta)$ is continuous, but $\mathcal{L}(\sqrt{n}(T_n-1/\sqrt{n})|\theta=1/\sqrt{n}) = N(0,2)$ for all n, so $\{T_n\}$ is not regular.

References

Bahadur, R.R. (1964), On Fisher's bound for asymptotic variances, Ann. Math. Statist., 35, 1545-1552.

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, 432-452.

Beran, R. (1977a), Estimating a distribution function, Ann. Statist., 5, 400-404.

Hajek, J. (1970), A characterization of limiting distributions of regular estimates, Z. Wahrs. Geb., 14, 323-330.

Roussas, G. (1972), Contiguity of Probability Measures, Cambridge University Press.

Wellner, J.A. (1982), Asymptotic optimality of the product limit estimator, Ann. Statist., 10, 595-602.

Wong, W.H. (1985), Theory of partial likelihood, Technical Report No. 155, Department of Statistics, University of Chicago.