The Annals of Statistics
1987, Vol. 15, No. 1, 427–431

AN ALTERNATIVE REGULARITY CONDITION FOR HÁJEK'S REPRESENTATION THEOREM[1]

By Luke Tierney

University of Minnesota

Hájek'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. Hájek's representation theorem [Hájek (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,

$$\mathscr{L}(\theta) = \lim_{n \to \infty} \mathscr{L}\left(\sqrt{n}\,(T_n - \theta)|\theta\right),$$

can be written as

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

for some distribution $\mathscr{L}_1$. Here $\mathscr{L}(Y|\theta)$ 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.

Hájek'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, Hall, Huang and Wellner (1983)] are asymptotically efficient. All these extensions use as their regularity condition on their estimators a variation of Hájek's original condition which states that the representation theorem holds at any θ, where $\mathscr{L}(\sqrt{n}\,(T_n - \theta_n)|\theta_n) \to \mathscr{L}(\theta)$ 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 *Hájek regular* at θ [see Wong (1986)].

Received May 1986; revised July 1986.
[1]Research supported in part by National Science Foundation Grant DMS-8503019.
AMS 1980 *subject classifications*. 62G20, 62G05.
Key words and phrases. Asymptotic efficiency, regular estimators.

A regularity condition on the sequence $\{T_n\}$ is needed to rule out superefficiency. The local condition of Hájek 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 $\mathscr{L}(\theta) = \lim_{n\to\infty}\mathscr{L}(\sqrt{n}(T_n - \theta)|\theta)$ exists for all θ and $\mathscr{L}(\theta)$ is continuous in θ (in the weak convergence topology), then Hájek's representation is valid for all θ. A proof of this result in this one-dimensional setting is given in the next two sections. Simple examples given in the final section show that this alternative regularity condition is neither implied by nor does it imply Hájek regularity. Before stating the theorem we formulate our regularity condition on the likelihood.

A well-behaved likelihood. 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 Hájek regular at θ, we have

$$\mathscr{L}\left(\sqrt{n}(T_n - \theta)|\theta\right) \to N\left(0, i(\theta)^{-1}\right) * \mathscr{L}_1(\theta),$$

for some distribution $\mathscr{L}_1(\theta)$. The likelihood will be called well behaved on an open subset Ω of $\mathbb{R}$ if it is well behaved at every $\theta \in \Omega$ 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 $\mathscr{L}(\sqrt{n}(T_n - \theta)|\theta) \to \mathscr{L}(\theta)$ for all $\theta \in \Omega$ and $\mathscr{L}(\theta)$ is continuous in θ. Then for each θ there exists a distribution $\mathscr{L}_1(\theta)$ such that*

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

The proof is based on the fact that for continuous $\mathscr{L}(\theta)$ Hájek's representation theorem can only fail for a set of θ values with measure zero, whereas the assumed continuity of $\mathscr{L}(\theta)$ 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 Hájek regularity is used in Bickel's proof to show that $\mathscr{L}(\sqrt{n}(T_n - \theta_n)|\theta_n) \to \mathscr{L}(\theta)$ for $\theta_n = \theta + h(1/\sqrt{n})$, 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 $\{T_n\}$ *weakly regular at* θ 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

$$\mathscr{L}\big(\sqrt{n}\,(k)(T_{n(k)} - \theta_{n(k)})|\theta_{n(k)}\big) \to \mathscr{L}(\theta) \quad \text{if } \theta_n = \theta + h/\sqrt{n}\,.$$

Then the representation will hold at θ if $\{T_n\}$ is weakly regular at θ. We state this as a lemma.

LEMMA 1. *If $\{T_n\}$ is weakly regular at θ, then there exists a distribution $\mathscr{L}_1(\theta)$ such that*

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

The proof is a straightforward modification of Bickel's proof of the Hájek 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 Lebesgue almost all θ:

LEMMA 2. *If $\{T_n\}$ is such that $\mathscr{L}(\sqrt{n}(T_n - \theta)|\theta) \to \mathscr{L}(\theta)$ and $\mathscr{L}(\theta)$ is continuous in θ, then $\{T_n\}$ is weakly regular for Lebesgue 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

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

Then

$$f_n(\theta + \sqrt{n}/h, \theta) \le f\left(\theta + \frac{h}{\sqrt{n}}, \theta + \frac{h}{\sqrt{n}}\right) + \rho\left(\mathscr{L}\left(\theta + \frac{h}{\sqrt{n}}\right), \mathscr{L}(\theta)\right).$$

Since $\mathscr{L}(\cdot)$ 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

$$\begin{aligned} &\int g_n\left(\theta + \frac{h}{\sqrt{n}}\right) d\Phi(\theta) \\ &\quad = \int g_n(\theta) \exp\left\{-\frac{h^2}{2n} + \frac{h}{\sqrt{n}}\theta\right\} d\Phi(\theta) \\ &\quad \to 0, \end{aligned}$$

by dominated convergence. Thus for any h we have $f_n(\theta + h/\sqrt{n}, \theta) \to 0$ in Φ

measure. Hence there is a subsequence $f_{n(k)}(\theta + h/\sqrt{n(k)}, \theta)$ that converges to zero Φ a.e. and hence is Lebesgue 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 Lebesgue 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_0. Since the mapping $\theta \to (\mathscr{L}(\theta), \theta)$ is continuous and the set

$$A = \left\{\left(N\left(0, i(\theta)^{-1}\right) * \mathscr{L}, \theta\right): \mathscr{L} \text{ a distribution}, \theta \in \Omega\right\}$$

is closed in the product topology, if $(\mathscr{L}(\theta_0), \theta_0) \notin A$ then we must have $(\mathscr{L}(\theta), \theta) \notin A$ for all θ in some neighborhood of θ_0. 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 $X_1, \dots, X_n$ be iid $N(\theta, 1)$, let $T_n = \bar{X}$ if $|\bar{X}| \geq 1/\log n$ and

$$T_n = \bar{X}\left(1 - \frac{1}{\sqrt{n}}\right) + X_1\left(\frac{1}{\sqrt{n}}\right), \quad \text{if } |\bar{X}| < \frac{1}{\log n}.$$

Then $\mathscr{L}(\theta) = N(0,1)$ for $\theta \neq 0$ but $\mathscr{L}(0) = N(0,2)$. So $\mathscr{L}(\theta)$ is not continuous at zero. On the other hand, if $\theta_n = O(n^{-1/2})$, then $\mathscr{L}(\sqrt{n}(T_n - \theta)|\theta_n) \to \mathscr{L}(0) = N(0,2)$ since

$$P\left\{T_n \neq \bar{X}\left(1 - \frac{1}{\sqrt{n}}\right) + X_1\frac{1}{\sqrt{n}}\bigg|\theta_n\right\} \to 0.$$

So $\{T_n\}$ is Hájek regular at $\theta = 0$.

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 $X_{1,n}, \dots, X_{n,n}$ be iid $N(\theta, f(\theta\sqrt{n}))$. Then $T_n = \bar{X}_n$ has $\mathscr{L}(\theta) = N(0,1)$ for all θ, i.e., $\mathscr{L}(\theta)$ is continuous, but $\mathscr{L}(\sqrt{n}(T_n - 1/\sqrt{n})|\theta = 1\sqrt{n}) = N(0,2)$ for all n, so $\{T_n\}$ is not regular.

Acknowledgments. The author would like to thank D. Lambert and M. L. Eaton for helpful discussions.

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). Robust location estimates. *Ann. Statist.* **5** 431–444.

BERAN, R. (1977b). Estimating a distribution function. *Ann. Statist.* **5** 400–404.

HÁJEK, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. verw. Gebiete* **14** 323–330.

ROUSSAS, G. (1972). *Contiguity of Probability Measures.* Cambridge Univ. Press.

WELLNER, J. A. (1982). Asymptotic optimality of the product limit estimator. *Ann. Statist.* **10** 595–602.

WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14** 88–123.

DEPARTMENT OF THEORETICAL STATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455