THE BEHAVIOR OF MAXIMUM LIKELIHOOD ESTIMATES UNDER NONSTANDARD CONDITIONS

PETER J. HUBER
SWISS FEDERAL INSTITUTE OF TECHNOLOGY

1. Introduction and summary

This paper proves consistency and asymptotic normality of maximum likelihood (ML) estimators under weaker conditions than usual.

In particular, (i) it is not assumed that the true distribution underlying the observations belongs to the parametric family defining the ML estimator, and (ii) the regularity conditions do not involve the second and higher derivatives of the likelihood function.

The need for theorems on asymptotic normality of ML estimators subject to (i) and (ii) becomes apparent in connection with robust estimation problems; for instance, if one tries to extend the author's results on robust estimation of a location parameter [4] to multivariate and other more general estimation problems.

Wald's classical consistency proof [6] satisfies (ii) and can easily be modified to show that the ML estimator is consistent also in case (i), that is, it converges to the θ_0 characterized by the property $E(\log f(x, \theta) - \log f(x, \theta_0)) < 0$ for $\theta \neq \theta_0$, where the expectation is taken with respect to the true underlying distribution.

Asymptotic normality is more troublesome. Daniels [1] proved asymptotic normality subject to (ii), but unfortunately he overlooked that a crucial step in his proof (the use of the central limit theorem in (4.4)) is incorrect without condition (2.2) of Linnik [5]; this condition seems to be too restrictive for many purposes.

In section 4 we shall prove asymptotic normality, assuming that the ML estimator is consistent. For the sake of completeness, sections 2 and 3 contain, therefore, two different sets of sufficient conditions for consistency. Otherwise, these sections are independent of each other. Section 5 presents two examples.

2. Consistency: case A

Throughout this section, which rephrases Wald's results on consistency of the ML estimator in a slightly more general setup, the parameter set Θ is a locally compact space with a countable base, $(\mathfrak{X}, \mathfrak{A}, P)$ is a probability space, and $\rho(x, \theta)$ is some real-valued function on $\mathfrak{X} \times \Theta$.