# INVARIANCE OF MAXIMUM LIKELIHOOD ESTIMATORS

By Peter W. Zehna

*U. S. Naval Postgraduate School*

One of the distinguishing features of the method of maximum likelihood in statistical estimation is the fact that it enjoys a certain invariance property. Briefly stated, if $\hat{\theta}$ is a maximum likelihood estimator for $\theta$, then $u(\hat{\theta})$ is a maximum likelihood estimator for $u(\theta)$ where $u$ is some function of $\theta$. Some textbooks on the subject avoid any explicit mention of properties that $u$ must possess in order for invariance to hold. When a proof of the property is given, it is at least assumed, either explicitly or implicitly that $u$ is 1-1 thereby defining a unique inverse.

Now if the assumption that $u$ be 1-1 is really necessary, then the invariance principle could not be invoked to find the maximum likelihood estimator for even as common a case as the variance, $p(1-p)$, of a Bernoulli random variable. Indeed, there may be some doubt as to the meaning of maximum likelihood in such a case. The purpose of this note is to point out that the notion of a maximum likelihood estimator for $u(\theta)$ when $u$ is not 1-1 can and should be made explicit. The method used for accomplishing this task has the desirable feature that it coincides with the usual method employed when $u$ is 1-1.

Suppose that parameter $\theta$ is restricted to lie in some set $\Theta$ and let $L(\theta)$ denote the likelihood function, a mapping from $\Theta$ to the real line. Assume that the maximum likelihood estimator $\hat{\theta}$ exists so that, $\hat{\theta} \varepsilon \Theta$ and $L(\hat{\theta}) \geqq L(\theta)$ for all $\theta \varepsilon \Theta$. Let $u$ be an arbitrary transformation from $\Theta$ to some set $\Lambda$. For convenience, we suppose that $\Lambda$ is the range of $u$ and we adopt the notation $\lambda = u(\theta)$.

Since $u$ is a function, $u(\hat{\theta})$ is a unique member, say $\hat{\lambda}$, of $\Lambda$. For each $\lambda \varepsilon \Lambda$, let $\Theta_\lambda = \{\theta; u(\theta) = \lambda\}$ and $M(\lambda) = \sup_{\theta \varepsilon \Theta_\lambda} L(\theta)$. Then $M$ is a real-valued function on $\Lambda$ to be called the *likelihood function induced by u*. Clearly, $M(\hat{\lambda}) = L(\hat{\theta})$ and the fact that $\hat{\lambda}$ maximizes $M$ is a trivial consequence of the inequality $M(\lambda) = \sup_{\theta \varepsilon \Theta_\lambda} L(\theta) \leqq \sup_{\theta \varepsilon \Theta} L(\theta) = L(\hat{\theta}) = M(\hat{\lambda})$ for all $\lambda \varepsilon \Lambda$. In this sense, it is reasonable to call $\hat{\lambda} = u(\hat{\theta})$ the maximum likelihood estimator for $u(\theta)$.

The author is indebted to Lt. Allen P. Fancher, USN, for his valuable assistance in surveying the available literature on the subject and his examination of several examples to amplify the result.

Received 23 November 1965; revised 21 January 1966.