MR0193707 (33 #1922) 62.20
Zehna, Peter W.
Invariance of maximum likelihood estimators.
Ann. Math. Statist. **37** 1966 744

Let $L(\theta)$ denote the likelihood function for a random variable whose possible distributions are indexed by θ. Assume that the maximum likelihood estimator $\hat{\theta}$ exists (then $L(\hat{\theta}) = \max_\theta L(\theta)$). The short note under review attempts to justify why $\hat{\lambda} = u(\hat{\theta})$ may then be called the maximum likelihood estimator of $\lambda = u(\theta)$, especially when u is not one-to-one. The author defines

$$M(\lambda) = \sup\{L(\theta) \colon u(\theta) = \lambda\},$$

referring to it as the likelihood function induced by u. It is easily seen that $L(\hat{\theta}) = M(\hat{\lambda}) = \max_\lambda M(\lambda)$; hence, $\hat{\lambda}$ is a maximum induced-likelihood estimator of λ.

The author indicates that when u is one-to-one, this "method" then coincides with the one usually employed. This means, essentially, that $M(\lambda)$ is then the likelihood function written in terms of the new parameter λ. {However, in the reviewer's opinion, consideration of $M(\lambda)$ misses the point, for when u is not one-to-one, $M(\lambda)$ in general appears not to be a likelihood function associated with any random variable. That $\hat{\lambda}$ then maximizes $M(\cdot)$ is perhaps interesting but irrelevant to maximum likelihood estimation.}

Justification for calling $\hat{\lambda}$ the maximum likelihood estimator of λ is implicit in the usual convention regarding simultaneous maximum likelihood estimation. If $\theta = (\theta_1, \theta_2)$ and $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)$ is its maximum likelihood estimator, then it seems generally agreed that the maximum likelihood estimator of $\theta_1 = u_1(\theta)$ is $\hat{\theta}_1 = u_1(\hat{\theta})$. (Cf. the simultaneous maximum likelihood estimation of the mean and variance of the normal distribution.) The result follows for any $u(\theta)$ if one simply adjoins to $u(\theta)$ another function $v(\theta)$ so that the mapping $\theta \to (u(\theta), v(\theta))$ is one-to-one. (E.g., if θ represents a location parameter and $u(\theta) = |\theta|$, one can choose $v(\theta) = \operatorname{sgn}\theta$.) The maximum likelihood estimator of the one-to-one function $w(\theta) = (u(\theta), v(\theta))$ is $w(\hat{\theta}) = (u(\hat{\theta}), v(\hat{\theta}))$, from which one obtains $u(\hat{\theta})$ as the maximum likelihood estimator of $u(\theta)$.

The reviewer has benefited from conversations with I. R. Savage. *R. H. Berk*