Canad. Math. Bull. Vol. **18** (3), 1975

A PROPERTY OF MAXIMUM LIKELIHOOD ESTIMATORS FOR INVARIANT STATISTICAL MODELS

BY

PETER TAN AND CONSTANTIN DROSSOS

ABSTRACT. This paper generalizes some results on pivotal functions of maximum likelihood estimators of location and scale parameters and the related ancillary statistics obtained by Antle and Bain, and Fisher. It shows that the maximum likelihood estimator of the parameter in an invariant statistical model is an essentially equivariant estimator or a transformation variable in a structural model. In the latter case, ancillary statistics in the sense of Fisher used in conjunction with the maximum likelihood estimators can be easily recognized. It is also remarked that the values of maximum likelihood estimators from samples having the same "complexion" are simply related to those of other, perhaps simpler, transformation variables. In the development it also points out the importance of using the correct definition of the likelihood function originally proposed by Fisher.

Introduction. Recently, Antle and Bain (1969) show that if α and β are the location and scale parameters in the probability density function

$$(1)\quad f(x;\alpha,\beta)=\frac{1}{\beta}g\left(\frac{x-\alpha}{\beta}\right),\qquad -\infty<x<\infty,-\infty<\alpha<\infty, 0<\beta<\infty,$$

then the maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$ have a property described in the following theorem.

THEOREM 1. *$\hat{\beta}_s=\hat{\beta}/\beta$, $\hat{\alpha}_s=(\hat{\alpha}-\alpha)/\beta$ and $\hat{\alpha}_p=(\hat{\alpha}-\alpha)/\hat{\beta}$ are each distributed independent of α and β.*

Applying Theorem 1, Antle, Klimko and Harkness (1970) obtain confidence intervals for the parameters of the logistic distribution.

Similarly, Thoman, Bain and Antle (1969) prove that if α and β are the shape and scale parameters of a two-parameter Weibull distribution whose probability density function is given by

$$(2)\qquad f(x;\alpha,\beta)=(\beta/\alpha)(x/\alpha)^{\beta-1}\exp\{-(x/\alpha)^{\beta}\}, x>0, \alpha>0, \beta>0,$$

then the following theorem they give makes it possible to obtain confidence intervals for both α and β.

Received by the editors December 17, 1973 and, in revised form, April 29, 1974.

https://doi.org/10.4153/CMB-1975-074-2 Published online by Cambridge University Press

THEOREM 2. *If $\hat{\alpha}$ and $\hat{\beta}$ are the maximum likelihood estimators of the Weibull parameters in* (2), *then*

$$\hat{\beta}/\beta \quad \text{and} \quad \hat{\beta}\ln(\hat{\alpha}/\alpha)$$

are each distributed independent of α and β.

The proofs of both theorems are simple results of a group property of the maximum likelihood estimates inherited from the group structure of the parent models.

The results in Theorem 1 have actually been given earlier by the late Sir R. A. Fisher (1959, p. 163) who observes also that

$$\mu_i = (x_i - \hat{\alpha})/\hat{\beta}, \qquad i = 1, \ldots, N,$$

provide a set of $N-2$ functionally independent ancillary statistics from a sample of size N, whose probability distributions are independent of the parameters and whose values may be said "to specify the complexion of the sample" because the values μ_i will be unchanged if the values x_i we observed had instead the values $X_i = a + cx_i$ for $-\infty < a < \infty$ and $c > 0$. Fisher proposes that in the theory of estimation the precision of the estimates $\hat{\alpha}$ and $\hat{\beta}$ "should be judged solely by reference to the variation of estimates among samples having the same complexion", and that exact probability statements be made about the parameters.

The existence of ancillary statistics which are functions of maximum likelihood estimators for the Weibull distribution (2) can also be easily recognized if we observe that for all constants $a > 0$, and $c > 0$, the maximum likelihood estimates $\hat{\alpha}(X)$ and $\hat{\beta}(X)$ based on a sample $X = (x_1, \ldots, x_N)$ are related to $\hat{\alpha}([a, c]X)$ and $\hat{\beta}([a, c]X)$ based on another sample $[a, c]X$ obtained by a transformation $[a, c]$ of X, where

$$[a, c]X = ([a, c]X_1, \ldots, [a, c]x_N)$$

and

$$[a, c]x_i = ax_i^{1/c}, \qquad i = 1, \ldots, N.$$

The relation is the following:

$$\hat{\alpha}([a, c]X) = a(\hat{\alpha}(X))^{1/c}, \qquad \hat{\beta}([a, c]X) = c\hat{\beta}(X).$$

This is shown by Tan and Sherif (1973).

In this paper we shall give the general result regarding the maximum likelihood estimates for invariant statistical models which contain Theorem 1 and Theorem 2 as special cases or examples.

2. **Main results.** Let A be a vector random variable defined on an open sample space $S \subset R^N$, θ be a vector parameter with open parameter space Ω which may be taken here as an interval in a Euclidean space R^L, and $f(X; \theta)$ the probability density function of the probability measure P_θ on (S, β) with respect to the Euclidean volume, μ. Let $G = \{g\}$ be a group of one-to-one transformations on S such that $gB \in \beta$ whenever $B \in \beta$, and let $\bar{G} = \{\bar{g}\}$ be the induced group of transformations on Ω defined by the identity

$$P_\theta[B] = P_{\bar{g}\theta}[gB]$$

for every B in β and g in G. We assume that $\bar{g}$ is one-to-one and $\bar{g}\Omega=\Omega$ for all $\bar{g}\in\bar{G}$. The family $\{f(X;\theta):\theta\in\Omega\}$ of densities is then said to specify a statistical model invariant under G.

Given $f(x;\theta)$ we shall define the likelihood function of θ for $X=x$ to be

$$L(\cdot\,;x)=\{kf(x;\,\cdot\,):k\in R^{+}\}, \tag{3}$$

which is an equivalence class of similarly shaped functions on Ω involving an arbitrary multiplicative positive constant k. It follows that for each $g\in G$, we have

$$L(\theta;x)=L(\bar{g}\theta;gx), \tag{4}$$

which has been proved by Fraser (1966). We remark that the definition (3) was originally proposed by Fisher in 1922 and is also the current definition of the likelihood function widely used in the literature, but it is not the usual definition found in many introductory texts in mathematical statistics which omits the multiplicative constant k. This incorrect omission of the constant k in the definition of likelihood function would immediately invalidate the result (4) and may have hindered the development of some other results in the theory of statistical inference.

Given the likelihood function $L(\theta;X)$, the maximum likelihood estimator of θ is a (β-measurable) function $\hat{\theta}:S\to\Omega$ satisfying

$$L(\hat{\theta};X)=\sup_{\theta\in\Omega}L(\theta;X)\quad a.s.\ [\mu].$$

THEOREM 3. *For a statistical model* $\{f(x;\theta):\theta\in\Omega\}$ *invariant under a group* G *of transformations, the maximum likelihood estimator* $\hat{\theta}$ *has the following property:*

$$\hat{\theta}(gX)=\bar{g}\hat{\theta}(X) \tag{5}$$

for all g *in* G *and almost all* X *in* $s[\mu]$.

Zacks (1971, p. 319) calls such an estimator satisfying condition (5) an essentially equivariant estimator.

Proof.

$$\begin{aligned}\sup_{\theta\in\Omega}L(\theta;gX)&=L(\hat{\theta}(gX);gX)\\&=L(g^{-1}\hat{\theta}(gX);X)\le\sup_{\theta\in\Omega}L(\theta;X)\\&=L(\hat{\theta}(X);X)=L(\bar{g}\hat{\theta}(X);gX)\\&\le\sup_{\theta\in\Omega}L(\theta;gX).\end{aligned}$$

Hence (5) follows.

If the invariant statistical model $\{f(x;\theta):\theta\in\Omega\}$ is also a structural model in the sense of Fraser (1968, p. 49) so that the random variable X and the parameter θ

are related to an error variable E with a known distribution the sample space S by the structural equation

$$X = \theta E,$$

where θ takes values in a unitary group G of transformations of S onto S, then the maximum likelihood estimator $\hat{\theta}(X)$, denoted by $[X]$ for convenience, is a transformation variable [Fraser, 1968, p. 51] taking values in G and

$$[gX] = g[X] \quad \text{for all } g \text{ in } G. \tag{6}$$

COROLLARY. *If $f(X; \theta)$ is a structural model, then the function $\theta^{-1}\hat{\theta}(X)$ is a pivotal quantity and $D(X)=[x]^{-1}X$ is a set of ancillary statistics whose probability distributions are independent of the parameter θ.*

Proof.

(*i*) $\theta^{-1}\hat{\theta}(X) = [\theta^{-1}X] = [E] = \hat{\theta}(E)$;

(*ii*) $D(X) = [X]^{-1}X = [E]^{-1}E$.

Since $D(gX)=[X]^{-1}X$ for all g in G, D is the "complexion" of the sample conditional on which, according to Fisher, the precision of $\hat{\theta}(X)$ should be judged.

REMARKS. While the above results show that for a structural model the maximum likelihood estimate $\hat{\theta}(X)$, if it exists, is a transformation variable $[X]$, satisfying the relation (6), often there are other transformation variables satisfying the same relation. Suppose $[X]_1=T(X)$ is a statistic taking values in G and satisfying (6). Then $D_1(X)=[X]_1^{-1}X$ is a set of ancillary statistics. We see then there exists a unique g in G such that

$$D(X) = gD_1(X), \tag{7a}$$

$$[X] = [X]_1 g^{-1} \quad \text{or} \quad \hat{\theta}(X) = T(X)g^{-1}, \tag{7b}$$

where $g=T(D(X))$. Thus in cases where it is difficult to obtain numerical values of maximum likelihood estimates $\hat{\theta}(X)$ but relatively easy to calculate $T(X)$, for repeated samples having the same "complexion" $D(X)$, we may obtain $\hat{\theta}(X)$ from $T(X)$ by (7b). Simple examples of this are given by Tan and Drossos (1973).

In structural inference, the structural distribution of θ obtained from the conditional distribution of $\hat{\theta}(E)$ given $D(X)$ and that obtained from the conditional distribution of $T(E)$ given $D_1(X)$ are equivalent [Fraser, 1968, p. 32].

ACKNOWLEDGEMENT. This research is partially supported with a research grant from Carleton University. The authors wish to thank the referee for a remark on the definition of the likelihood function.

REFERENCES

1. C. E. Antle and L. J. Bain, *A property of maximum likelihood estimates of location and scale parameters*. SIAM Review **11** (1969), 251–253.

2. C. E. Antle, L. Klimko and W. Harkness, *Confidence intervals for the parameters of the locistic distribution*. Biometrika **57** (1970), 397–402.

https://doi.org/10.4153/CMB-1975-074-2 Published online by Cambridge University Press

3. R. A. Fisher, *Statistical Methods and Scientific Inference*, Second Edition, London: Oliver and Boyd, 1959.

4. R. A. Fisher, *On the mathematical foundations of theoretical statistics*. Phil. Trans. Roy. Soc. of London, Series A, **22** (1922), 309–368.

5. D. A. S. Fraser, *On sufficiency and conditional sufficiency*, Sankhya, Series A, **28** (1966), 145–150.

6. D. A. S. Fraser, *The Structure of Inference*. New York: John Wiley & Sons, 1968.

7. P. Tan and A. Sherif, *Some remarks on structural inference applied to Weibull distributions*. Statistische Hefte, Vol. 15, **4** (1974), 335–341.

8. P. Tan and C. Drossos, *Invariance properties of maximum likelihood estimators*. Mathematics Magazine, Vol. 48, **1** (1975), 37–41.

9. D. R. Thoman, L. J. Bain and C. E. Antle, *Inference on the parameters of the Weibull distribution*. Technometrics **11** (1969), 445–460.

10. S. Zacks, *The Theory of Statistical Inference*. New York: John Wiley & Sons, 1971.

CARLETON UNIVERSITY
OTTAWA, ONTARIO,
CANADA