STATISTICAL RESEARCH REPORT No 9
Institute of Mathematics 1971
University of Oslo

ON THE OPTIMALITY OF STATISTICAL METHODS BASED UPON THE LIKELIHOOD FUNCTION IN A ONE-PARAMETER EXPONENTIAL FAMILY

by

Emil Spjøtvoll

ABSTRACT

The properties of statistical methods based upon the likelihood function for a one-parameter exponential family are studied. It is shown that the maximum likelihood estimate of a certain function of the parameter is the best unbiased estimate, for hypotheses with one-sided alternatives the likelihood ratio test is the uniformly most powerful test, and for hypotheses with two-sided alternatives the likelihood ratio test rejects when the values of the sufficient statistic are outside an interval. Under certain conditions it is also shown the test of hypotheses with two-sided alternatives is uniformly most powerful unbiased. The properties of the tests also carry over to the confidence intervals based upon the likelihood function.

1. INTRODUCTION

It is well known that in many cases statistical methods based upon the likelihood function are optimal in various ways. The use of the likelihood function has also been advocated on more intuitive grounds, see, for example, Barnard (1965), Kalbfleisch and Sprott (1970). Here we shall consider the one-parameter exponential family

$$f(x,\theta) = D^{-1}(\theta) \exp\{\theta T(x)\} \quad , \quad \theta \in \Omega , \qquad (1)$$

where $f(x,\theta)$ is a probability density with respect to a σ-finite measure μ over a Euclidean sample space, and Ω is a subset of the real line. We shall assume that Ω has interior points, and let Ω_0 be the interior of Ω. With the above assumptions, the statistic $T(X)$ is sufficient and complete. For examples and further properties of exponential families the reader is referred to Lehmann (1959, pp. 50-54).

At various point we shall assume that the exponential family we are considering satisfy some of the following conditions

A 1. The equation

$$\frac{D'(\theta)}{D(\theta)} = T(x)$$

has a unique solution θ with $\theta \in \Omega$ for almost all x.

A 2. $f(x,\theta)$ is continuous in θ for all x.

A 3. The family of densities $\{f(x,\theta) : \theta \in \Omega\}$ is invariant under a group G of measurable transformations of the sample space and μ is absolutely continuous with respect to μg^{-1} for all $g \in G$. Furthermore, the induced group $\bar{G}$ of transformations of Ω is transitive over Ω, and the transformations $\bar{g} \in \bar{G}$ are continuous.

Throughout the whole paper we shall assume that A 1 holds. A 3 is not satisfied for discrete exponential families, but is satisfied for all non-discrete exponential families that the author is aware of.

In Section 2 it is shown that for a certain function of the parameter the maximum likelihood estimate (MLE) is the best unbiased estimate. In Section 3 it is shown that for one-sided hypotheses the likelihood ratio test (LRT) is uniformly most powerful, and for two-sided hypotheses it is uniformly most powerful unbiased if all conditions A 1-3 are satisfied. Analogous results for confidence sets are given in Section 4.

2. MAXIMUM LIKELIHOOD ESTIMATES

Consider the integral

$$\int \exp\{\theta T(x)\}\, d\mu(x) \, . \qquad (2)$$

In Lehmann (1959, Theorem 9, pp. 52-53) it is proved that (1) considered as a function of the complex variable $\theta = \xi + i\eta$ is an analytic function in the region of parameter points for which $\xi \in \Omega_o$, and the derivatives of all orders with respect to θ of the integral (2) can be computed under the integral sign. Using this result and observing that

$$\int \exp\{\theta T(x)\}\, d\mu(x) = D(\theta) \, ,$$

we obtain

$$E\{T(X)^k\} = D^{(k)}(\theta)/D(\theta) \, , \quad \theta \in \Omega_o \, , \qquad (3)$$

where $D^{(k)}(\theta)$ denotes the k-th derivative of $D(\theta)$. Using (3) for $k = 1,2$ we also find

$$\mathrm{Var}\{T(X)\} = D''(\theta)/D(\theta) - \{D'(\theta)/D(\theta)\}^2 \, , \quad \theta \in \Omega_o \, . \qquad (4)$$

In the following we shall assume that (3) holds for all $\theta \in \Omega$ for $k = 1,2$.

In the proof of Theorem 1 we need the following

<u>Lemma 1.</u> $D'(\theta)/D(\theta)$ is a strictly increasing function of θ.

<u>Proof.</u> The derivative of $D'(\theta)/D(\theta)$ is found to be equal to $\mathrm{Var}\{T(x)\}$, see (4), which is > 0.

Theorem 1. (I) The MLE of θ is the unique solution $\hat{\theta}$ of the equation

$$D'(\theta)/D(\theta) = T(X) , \qquad (5)$$

and (II) the MLE of $D'(\theta)/D(\theta)$ is a best unbiased estimate of $D'(\theta)/D(\theta)$.

Proof. (I) The first derivative of $f(x,\theta)$ w.r.t. θ is

$$D^{-1}(\theta) \exp\{\theta T(x)\} \; [T(x) - D'(\theta)/D(\theta)] . \qquad (6)$$

Hence the density has a stationary point for θ satisfying (6). Since by Lemma 1 $D'(\theta)/D(\theta)$ increases with θ , it is seen from (6) that it must be a maximum. (II) The MLE of $D'(\theta)/D(\theta)$ is $T(X)$ by (5), and by (3) it is unbiased. Since the distribution of the sufficient statistic $T(X)$ is complete, $T(X)$ is also the unbiased estimate with minimum variance.

Remark. The fact that $D'(\theta)/D(\theta)$ increases with θ (see Lemma 1) is useful in cases when we need to find the MLE of θ from (5) by numerical methods.

An example could be the truncated binomial distribution with truncation point a

$$P[X = x] = \frac{\binom{n}{x} p^x (1-p)^{n-x}}{1 - \sum_{i=0}^{a} \binom{n}{i} p^i (1-p)^{n-i}} \qquad x = a+1, \ldots, n ,$$

which can be written in the form (1) with

$$\theta = \log \frac{p}{1-p} \qquad (7)$$

$$D(\theta) = (1+e^{\theta})^n \{1 - \sum_{i=0}^{a} \binom{n}{i} e^{i\theta} (1+e^{\theta})^{-n}\}$$

$$T(x) = x$$

We find

$$\frac{D'(\theta)}{D(\theta)} = ne^{\theta} \frac{(1+e^{\theta})^{n-1} - \sum_{i=1}^{a} \binom{n-1}{i-1} e^{(i-1)\theta}}{(1+e^{\theta})^n - \sum_{i=0}^{a} \binom{n}{i} e^{i\theta}} . \qquad (8)$$

We know from Lemma 1 that $D'(\theta)/D(\theta)$ increases with θ, but this result is not easily obtained directly from (8). The MLE $\hat{\theta}$ of θ can now be found by numerical methods from the equation

$$\frac{D'(\theta)}{D(\theta)} = x$$

The MLE $\hat{p}$ of p is then obtained from

$$\hat{p} = \frac{e^{\hat{\theta}}}{e^{\hat{\theta}}+1}$$

Estimation of parameters in truncated discrete distributions is important in some application. For literature on this problem, see a recent paper by Selvin (1971).

3. LIKELIHOOD RATIO TESTS

We need the following

Lemma 2. The MLE $\hat{\theta}(T(x))$ of θ is a strictly increasing function of $T(x)$.

Proof. Equation (5) defines $\hat{\theta}(T(x))$ implicitly as a function of $T(x)$

$$D'(\hat{\theta}(T(x)) - T(x)D(\hat{\theta}(T(x)) = 0$$

The derivative of $\hat{\theta}(T(x))$ with respect to $T(x)$ exists and is given by

$$\hat{\theta}'(T(x)) = [D''(\hat{\theta}(T(x)))/D(\hat{\theta}(T(x)))-\{D'(\hat{\theta}(T(x)))/D(\hat{\theta}(T(x)))\}^2]^{-1}$$

which is > 0 by (4).

Theorem 2. The level α LRT of

$H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ is UMP if α is a possible level of the LRT.

Proof. The LRT rejects H if

$$L(x) = \frac{\sup_{\theta \leq \theta_0} f(x,\theta)}{\sup_{\theta} f(x,\theta)} < \text{constant}.$$

The set of possible levels for a LRT is

$$\{\gamma : \gamma = \sup_{\theta \leq \theta_0} P_\theta\{L(X)<c\} \text{ for some } c \geq 0\}$$

If X is a discrete random variable the above set will not usually contain all numbers between 0 and 1 .

We have that

$$\sup_\theta f(x,\theta) = D(\hat{\theta})^{-1} \exp\{\hat{\theta}T(x)\}$$

where $\hat{\theta}$ is the solution of (5). If $\hat{\theta} \leq \theta_0$, then also

$$\sup_{\theta \leq \theta_0} f(x,\theta) = D(\hat{\theta})^{-1} \exp\{\hat{\theta}T(x)\}$$

and hence $L(x) = 1$.

If $\hat{\theta} > \theta_0$, then

$$\sup_{\theta \leq \theta_0} f(x,\theta) = f(x,\theta_0)$$

since, by Lemma 1 $D'(\theta)/D(\theta)$ is an increasing function of θ , and then by (6) $f(x,\theta)$ increases when θ increases to $\hat{\theta}$. Hence in this case

$$L(x) = D(\hat{\theta})D(\theta_0)^{-1} \exp\{(\theta_0-\hat{\theta})T(x)\} , \tag{9}$$

and

$$\log L(x) = \log D(\hat{\theta})-\log D(\theta_0)+(\theta_0-\hat{\theta})T(x) . \tag{10}$$

The derivative of $\log L(x)$ w.r.t. $T(x)$ is

$$\frac{D'(\hat{\theta}(T(x)))}{D(\hat{\theta}(T(x)))} \theta'(T(x))+\theta_0-\hat{\theta}(T(x))-T(x)\hat{\theta}'(T(x)) . \tag{11}$$

By (5) $D'(\hat{\theta}(T(x)))/D(\hat{\theta}(T(x))) = T(x)$, and (11) reduces to

$$\theta_0 - \hat{\theta}(T(x)) . \tag{12}$$

Since we are considering the case $\hat{\theta}(T(x)) > \theta_0$, the derivative is negative and hence $L(x)$ is a decreasing function of $T(x)$. To reject when $L(x) <$ constant is therefore equivalent to rejecting when

$T(x) >$ constant.

By Lehmann (1959, pp. 68-69) this is the UMP test of H against K.

Theorem 3. The LRT of

$H : \theta = \theta_o$ against $K : \theta \neq \theta_o$
is of the form: Reject when

$$T(x) < c_1 \quad \text{or} \quad T(x) > c_2 \tag{13}$$

where c_1 and c_2 are related by

$$\log \{D(\hat{\theta}(c_2))/D(\hat{\theta}(c_1))\} = c_2\hat{\theta}(c_2)-c_1\hat{\theta}(c_1) . \tag{14}$$

Proof. The LRT rejects H if

$$L(x) = \frac{f(x,\theta_o)}{\sup_\theta f(x,\theta)} < \text{constant} \tag{15}$$

Here $L(x)$ is equal to (9), and the derivative of $\log \{L(x)\}$ is given by (12). Let t_o be the number such that $\hat{\theta}(t_o) = \theta_o$. Then the derivative of $\log \{L(x)\}$ is positive when $T(x) < t_o$ and negative when $T(x) > t_o$. It follows that $L(x)$ as a function of $T(x)$ has a maximum at t_o and decreases when $T(x)$ decreases or increases from t_o. By (15) the form of the rejection region must be as given in the theorem. From (10) and (15) we get (14).

Theorem 4. Under assumptions A2-3 the LRT test of

$H : \theta = \theta_o$ against $K : \theta \neq \theta_o$

is UMP unbiased.

Proof. It has been shown by Spjøtvoll (1971) that the confidence sets

$$S(x) = \{\theta : \frac{f(x,\theta)}{\sup_\theta f(x,\theta)} \geq c\} \tag{16}$$

for θ, are unbiased. Here c is determined so that $P_\theta \{\theta \in S(T(X))\} = 1-\alpha$. Hence the test which rejects the null hypothesis when $\theta_o \notin S(TX))$ is also unbiased. That $\theta_o \notin S(T(X))$ is equivalent to

$$\frac{f(x,\theta_o)}{\sup_\theta f(x,\theta)} < c$$

which by Theorem 3 is equivalent to $T(X) < c_1$ and $T(X) > c_2$.

To prove that this test is UMP unbiased it is enough to show (see Lehmann (1959), p. 127) that the test is of the form reject when

$$D(\theta_0)^{-1}(k_1+k_2T(x))\exp\{\theta_0 T(x)\} < D(\theta')^{-1}\exp\{\theta' T(x)\} \qquad (17)$$

where $\theta' \neq \theta_0$ and k_1 and k_2 are suitably chosen constants. The equations

$$k_1+k_2c_i = e^{(\theta'-\theta_0)c_i} \qquad i = 1,2$$

have always a solution with respect to k_1 and k_2. With these k_1 and k_2 the region (13) is of the form (17). The theorem is proved.

4. LIKELIHOOD RATIO CONFIDENCE SETS

The author (1971) has defined a $1-\alpha$ likelihood ratio confidence set (LRCS) to be a confidence set of the form (16) where c is the largest number such that $P_\theta\{\theta \in S(T(X))\} \geq 1-\alpha$.

Corresponding to Theorem 3 and 4 we have

Theorem 5. The $1-\alpha$ LRCS for is of the form $[c_1(T(X)), c_2(T(X))]$ where c_1 and c_2 satisfy (14).

Theorem 6. If A2-3 hold, then the LRCS is UMP unbiased.

The LRCS in a family of distributions with one real parameter is not necessarily an interval when this family is not an exponential family. An example is the Cauchy distribution, see Spjøtvoll (1971, Example 3).

Hudson (1968) has studied the LRCS for a binomial p.

REFERENCES

BARNARD, G.E. (1965) The use of the likelihood function in statistical practice. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1, 17-40.

HUDSON, D.J. (1968). Interval estimation from the likelihood function. Bell Telephone Lab. Tech. Rep.

KALBFLEISCH, J.D. and SPROTT, D.A. (1970). Application of likelihood methods to models involving large numbers of parameters. J.R. Statist. Soc. B, 32, 175-208.

LEHMANN, E.L. (1959). Testing Statistical Hypotheses. Wiley, New York.

SELVIN, S. (1971). Maximum likelihood estimation in the truncated, single parameter, discrete, exponential family. The American Statistician, 25, 41-42.

SPJØTVOLL, E. (1971). Unbiasedness of likelihood ratio confidence sets in cases without nuisance parameters. To appear in J.R. Statist. Soc. B.