Local Conditional Sufficiency

By D. A. S. Fraser

University of Toronto

[Received February 1963. Revised November 1963]

Summary

The concept of sufficiency can be examined locally for a neighbourhood of a parameter point θ_0 and it leads to an estimate that has minimum variance among locally-unbiased estimates (Fraser, 1964). This local estimate generates the natural global estimate for the case of the simple exponential model, but it does not do so for the important case of the general translation model. The global estimate, the Pitman estimate, for this translation model is rooted in ancillarity and conditional sufficiency. In this paper the concept of local conditional sufficiency is developed; the asymptotic distribution of the conditionally-sufficient statistic is investigated for large samples; and for the maximum-likelihood estimate a conditional large-sample variance is obtained which differs from that usually used.

1. Introduction

The use of a minimal sufficient (exhaustive) statistic as a means of reduction for a statistical problem is well established in theoretical statistics. So also are the ensuing methods of inference, provided the minimal statistic has the same dimension as the parameter (Fisher Sufficiency). But, where the minimal statistic has larger dimension than the parameter, new factors seem to enter and little unanimity is found as to the appropriate method of inference. One promising method is to be found in the writings of R. A. Fisher; it has, however, received rather little attention—perhaps because his partial formulations are anathema to most mathematical statisticians. Fisher recommends the use of an *ancillary* statistic, which has a fixed distribution, followed by analysis in the conditional framework given the ancillary statistic. If the conditioned variable has the same dimension as the parameter then the established methods of inference appropriate to the case of Fisher sufficiency can be used and the conditioned variable can conveniently be called *conditionally sufficient*. Fisher (1956, p. 85) comments on this approach to estimation: "If we must think in terms of random sampling it is only that *selection* of random samples which agree exactly with our own in regard to the value of A (the ancillary statistic) that is relevant to assessing its real precision". Fisher (1956, pp. 84, 132, 161, 165) also gives a variety of examples illustrating the method.

In applying Fisher's method one encounters the problem of determining the appropriate ancillary statistic. Fisher (1956, pp. 119, 163) recognized this problem as being of fundamental importance and entitled it "The problem of the Nile". He did not give a general solution, but in many examples he presented on an *ad hoc* basis some very natural-looking ancillary statistics. A careful survey of his examples reveals that in all the continuous cases the ancillary statistic is in fact the maximal invariant statistic under a natural group of transformations on the sample space.

Perhaps, then, an essential but unspecified ingredient of what Fisher had in mind for an ancillary statistic was the position structure that derives from the transformation model. But however this may be, it can be noted that *the problem of the Nile is solved for the general transformation-parameter model* and that the solution specializes to give the ancillaries presented in all of Fisher's continuous examples.

Transformation-parameter models have, however, further advantages for purposes of inference. In particular they provide a position relationship between variable and parameter and this seems to lead to stronger inference statements. Ordinary sufficiency seems strangely out-of-step with this. For, in the regular case, a one-dimensional sufficient statistic is available only for samples from an exponential distribution and, further, the results of Lindley (1958) show that such a statistic has position relationship only for the normal and gamma cases. Thus to obtain position relationship between statistic and parameter *all translation models can be analysed by conditional sufficiency, but only the normal and gamma can be so analysed by sufficiency*. Perhaps the importance of sufficiency has been overestimated.

A similar disparity seems to hold for estimation examined locally on the parameter space. Local sufficiency and local minimum-variance unbiased estimation are developed in Fraser (1964). It is found that the translation-form of analysis is available only if the distribution is locally of normal or gamma form and that for translation models the local estimate does not generate the Pitman global estimate except for the normal and gamma cases. This seeming inadequacy of sufficiency has prompted the development of the local conditional sufficiency in this paper.

As a motivating example for the conditional form of analysis, consider the Welch (1939) example of a sample (x_1, x_2) from the uniform distribution $(\theta \pm \frac{1}{2})$. This statistical model has translation invariance: under the transformation $y = x + k$ the sample (x_1, x_2) produces a sample (y_1, y_2) from the uniform distribution $(\theta + k \pm \frac{1}{2})$ and, in fact, any distribution in the model can be obtained from any other by a single appropriate transformation. The statistic $(x_1 - x_2)$ is maximal invariant; it is the natural ancillary and has the triangular distribution on $(-1, +1)$. The conditional distribution of $\frac{1}{2}(x_1 + x_2)$, given this ancillary, is uniform on

$$\{\theta \pm \tfrac{1}{2}(1 - |x_1 - x_2|)\}.$$

The ancillary approach can be described in terms of a reduction—not of the observable variable but of the statistical problem: the reduction is to the conditional model given the ancillary. The new sample space is a slice of the old sample space (along the orbit determined by the observed value of $x_1 - x_2$) rather than a projection to a lower dimensional space, as is the case with a sufficient-statistic reduction.

2. The Linearizing Transformation

Consider a real variable x having a distribution function $F(x \mid \theta)$ that is stochastically increasing in the neighbourhood of θ_0:

$$F_\theta(x \mid \theta_0) = \frac{\partial}{\partial \theta_0} F(x \mid \theta) < 0;$$

an increase in θ at θ_0 causes the distribution to shift in the positive direction on the axis. In this section a transformation of x will be constructed which puts the problem into translation-invariant form in the neighbourhood of θ_0. In later sections this transformation provides the basis for local ancillarity and local conditional sufficiency.

Consider an increment in the parameter from θ_0 to $\theta_0+\delta$. The distribution function at x correspondingly decreases by an amount $-F_\theta(x|\theta_0)\,\delta$. Consider now an increase in the variable from x to $x+h$. The distribution function correspondingly increases by an amount $F_x(x|\theta_0)\,h$ when the parameter value is θ_0. If δ and h are in the ratio determined by

$$-F_\theta(x|\theta_0)\,\delta = F_x(x|\theta_0)\,h,$$

then $F(x|\theta_0)$ will be approximately equal to $F(x+h|\theta_0+\delta)$. Thus a change from θ_0 to $\theta_0+\delta$ can be viewed as a topological shifting of the distribution, the amount of the shift at the point x being given by

$$-\frac{F_\theta(x|\theta_0)}{F_x(x|\theta_0)}\,\delta.$$

Or the distribution can be said to shift at the rate

$$-\frac{F_\theta(x|\theta_0)}{F_x(x|\theta_0)}$$

under parameter change at θ_0.

The preceding rate function provides the basis for defining a new variable l whose rate of shift is everywhere equal to one. Let $l = l(x)$ be a new variable satisfying

$$\frac{dx(l)}{dl} = -\frac{F_\theta(x|\theta_0)}{F_x(x|\theta_0)};$$

$$l(x) = -\int^x \frac{F_x(x|\theta_0)}{F_\theta(x|\theta_0)}\,dx.$$

The variable $l(x)$ in the preceding formula is indeterminate to an additive constant of integration. The constant can be chosen naturally in a variety of ways; for example, *it can be chosen so that* $l = \theta_0$ *when* x *is equal to the median of the* θ_0 *distribution.* It should be noted that the transformation is produced on the basis of properties at θ_0 and in certain contexts might advantageously be denoted by $l = l(x|\theta_0)$.

Let $G(l|\theta)$ be the distribution function for the new variable:

$$G(l|\theta) = F\{x(l)|\theta\}.$$

The probability density function can be obtained from this as

$$\begin{aligned} G_l(l|\theta) &= -F_x\{x(l)|\theta\}\frac{F_\theta\{x(l)|\theta_0\}}{F_x\{x(l)|\theta_0\}} \\ &= F_x\{x(l)|\theta\}\,\omega\{x(l)\}, \end{aligned}$$

wherein $\omega(x)$ is a function of x only. At $\theta = \theta_0$ this simplifies to

$$G_l(l|\theta_0) = -F_\theta\{x(l)|\theta_0\}$$

and is seen to be the fiducial density function $-F_\theta$, only treated as a function of l rather than θ.

The rate of shift of the new variable l can be easily checked. At θ_0 the rate is

$$-\frac{G_\theta(l\,|\,\theta_0)}{G_l(l\,|\,\theta_0)} = -\frac{F_\theta\{x(l)\,|\,\theta_0\}}{F_x\{x(l)\,|\,\theta_0\}\{dx(l)/dl\}}$$
$$= \frac{F_\theta\{x(l)\,|\,\theta_0\}}{F_x\{x(l)\,|\,\theta_0\}}\frac{F_x\{x(l)\,|\,\theta_0\}}{F_\theta\{x(l)\,|\,\theta_0\}}$$
$$= 1;$$

this is the result sought for in the choice of transformation.

The above analysis provides, incidentally, a natural context for a simple necessary and sufficient condition for global translation invariance. For the local translation invariance to extend to global translation invariance, the rate function must satisfy

$$-\frac{G_\theta(l\,|\,\theta)}{G_l(l\,|\,\theta)} = \psi(\theta);$$

this is necessary and sufficient. In terms of the original variable this becomes

$$\frac{F_x(x\,|\,\theta)}{F_\theta(x\,|\,\theta)} = \psi(\theta)\,\omega(x),$$

which states that the ratio of partial derivatives must factorize so that the variables separate.

The locally translation-invariant distribution can be propagated in both directions from $\theta = \theta_0$: the resulting distribution function is

$$G\{l-(\theta-\theta_0)\,|\,\theta_0\} = F[x\{l-(\theta-\theta_0)\}\,|\,\theta_0];$$

and the density function is

$$G_l\{l-(\theta-\theta_0)\,|\,\theta_0\} = -G_\theta\{l-(\theta-\theta_0)\,|\,\theta_0\}$$
$$= -F_\theta[x\{l-(\theta-\theta_0)\}\,|\,\theta_0],$$

which also has fiducial density function form, but with the parameter tied additively to the transformed variable l. This globally translation-invariant distribution agrees with the original distribution to a first derivative approximation at θ_0, and it provides thereby a transparent means for examining some local properties at θ_0.

The local translation invariance suggests that the quantity $z = l-\theta$ would be locally pivotal. Let $H(z\,|\,\theta)$ designate the distribution function of z:

$$H(z\,|\,\theta) = G(z+\theta\,|\,\theta).$$

The derivative of this at θ_0 is

$$H_\theta(z\,|\,\theta_0) = G_l(z+\theta_0\,|\,\theta_0)+G_\theta(z+\theta_0\,|\,\theta_0)$$
$$= 0.$$

Thus, to first-derivative change, the quantity $z = l-\theta$ is a *pivotal quantity locally at* θ_0. In the neighbourhood of θ_0 the relationship of l to θ is additive or *linear*; accordingly, the transformation $l(x)$ will be called the linearizing transformation at θ_0.

3. Local Ancillarity

In the preceding Section a transformation was developed that carried a variable x into a new variable l in such a way as to produce translation invariance at θ_0; this local translation invariance is characterized by $G_l(l\,|\,\theta_0) = -G_\theta(l\,|\,\theta_0)$, where $G(l\,|\,\theta)$

is the distribution function of the new variable l. In this Section a sample of n will be considered and an ancillary statistic developed for θ near to θ_0.

Consider a sample $(x_1, ..., x_n)$ from the distribution $F(x \mid \theta)$ and a corresponding sample $(l_1, ..., l_n)$ from the distribution $G(l \mid \theta)$. A change in θ at θ_0 can be interpreted as a uniform shift in the distribution for l and hence a uniform shift for each l_i: from $(l_1, ..., l_n)$ to $(l_1+h, ..., l_n+h)$. These local displacements in n-space correspond to the differential equation

$$dl_1 = ... = dl_n,$$

which has a general-solution orbit

$$(l_1^0+t, ..., l_n^0+t),$$

where $(l_1^0, ..., l_n^0)$ is a reference point on an orbit and t is a parameter giving position on the orbit. For the original variables, the orbits are given by

$$\{x(l_1^0+t), ..., x(l_n^0+t)\}.$$

The flow of probability in the one-dimensional case suggests that for a sample of n the probability would flow along the preceding orbits as θ changes near θ_0. The remainder of this Section presents a demonstration of this ancillarity property.

For the transformed variables an orbit is determined by specifying the differences

$$t_i = l_i - l_1 \quad (i = 2, ..., n).$$

These new variables $(t_2, ..., t_n)$ can be used as coordinates for the orbits. On an orbit there is one degree of freedom and a simple choice of coordinate is $t_1 = l_1$ (an alternative choice with some advantages is $t_1 = \bar{l}$). The Jacobian of the transformation from $(l_1, ..., l_n)$ to $(t_1, ..., t_n)$ is equal to 1. Simple change of variable and integration then give the following expression for the marginal density function for $(t_2, ..., t_n)$:

$$\int_{-\infty}^{\infty} G_l(t_1 \mid \theta)\, G_l(t_2+t_1 \mid \theta) \ldots G_l(t_n+t_1 \mid \theta)\, dt_1$$
$$= \int_{-\infty}^{\infty} \prod_{i=1}^{n} G_l(l_i - l_1 + t \mid \theta)\, dt.$$

In the second expression the probability density function for the differences $(l_2-l_1), ..., (l_n-l_1)$ is expressed in terms of these differences.

The local ancillarity property is established by showing that the derivative at θ_0 of the preceding expression is zero:

$$\begin{aligned}
&\frac{d}{d\theta_0} \int_{-\infty}^{\infty} \prod_{i=1}^{n} G_l(l_i - l_1 + t \mid \theta)\, dt \\
&= \int_{-\infty}^{\infty} \sum_{j=1}^{n} G_{l\theta}(l_j - l_1 + t \mid \theta_0) \prod_{i \neq j} G_l(l_i - l_1 + t \mid \theta_0)\, dt \\
&= -\int_{-\infty}^{\infty} \sum_{j=1}^{n} G_{ll}(l_j - l_1 + t \mid \theta_0) \prod_{i \neq j} G_l(l_i - l_1 + t \mid \theta_0)\, dt \\
&= -\int_{-\infty}^{\infty} \frac{d}{dt} \prod_{i=1}^{n} G_l(l_i - l_1 + t \mid \theta_0)\, dt \\
&= -\left[\prod_{i=1}^{n} G_l(l_i - l_1 + t \mid \theta_0)\right]_{t=-\infty}^{\infty} \\
&= 0;
\end{aligned}$$

the second step uses the identity $G_{l\theta} = -G_{ll}$ which derives from $G_\theta = -G_l$ by differentiation with respect to l. It follows then that the probability for any event for $l_2 - l_1, \ldots, l_n - l_1$ has derivative equal to zero at θ. Accordingly, the statistic $(l_2 - l_1, \ldots, l_n - l_1)$ is called *locally ancillary* at θ_0. In terms of the original variables this statistic is $\{l(x_2) - l(x_1), \ldots, l(x_n) - l(x_1)\}$.

4. Local Conditional Sufficiency

Consider a sample $(x_1, \ldots, x_n)$ from the distribution $F(x \mid \theta)$ and a corresponding sample $(l_1, \ldots, l_n)$ from the distribution $G(l \mid \theta)$. The statistic

$$\{l(x_2) - l(x_1), \ldots, l(x_n) - l(x_1)\}$$

is locally ancillary at θ_0. In this Section the conditional distribution, given the ancillary, will be examined.

In the preceding Section the *marginal density* for the ancillary was obtained by integrating a *joint density function*. The ratio of these density functions gives the conditional density function. For notation let $(l_1, \ldots, l_n)$ determine an orbit, that is, determine the ancillary $(l_2 - l_1, \ldots, l_n - l_1)$, and let L_1 be a free variable giving position on the orbit. Then the conditional density for L_1 given the orbit described by $(l_1, \ldots, l_n)$ is

$$\frac{\Pi G_l(l_i - l_1 + L_1 \mid \theta)}{\int \Pi G_l(l_i - l_1 + t \mid \theta)\, dt}.$$

The conditional density for L_1 is of interest for θ near to θ_0. At θ_0 the derivative of the denominator with respect to θ is equal to zero. And at θ_0 the numerator has translation-invariant form:

$$\frac{\partial}{\partial \theta_0} \Pi G_l(l_i - l_1 + L_1 \mid \theta) = -\frac{\partial}{\partial L_1} \Pi G_l(l_i - l_1 + L_1 \;\; \theta_0),$$

which is obtained by use of the relation $G_l(l \mid \theta_0) = -G_\theta(l \mid \theta_0)$. It follows that the conditional density itself has translation-invariant form.

The variable L_1 has the same dimension as the parameter, and the statistic $(l_2 - l_1, \ldots, l_n - l_1)$ is ancillary for θ near θ_0; accordingly, it is natural to call L_1 *conditionally sufficient* (given the ancillary) *for θ near* θ_0. In addition L_1 has locally the translation invariance which is a quite general characteristic of Fisher's conditionally-sufficient statistics.

The density function for $X_1 = x(L_1)$ is obtained by simple change of variable:

$$\frac{\Pi(F_x[x\{l_i - l_1 + l(X_1)\} \mid \theta]\, \omega[x\{l_i - l_1 + l(X_1)\}])}{\int \Pi(F_x[x\{l_i - l_1 + t\} \mid \theta]\, \omega[x\{l_i - l_1 + t\}])\, dt}\, \omega^{-1}(X_1).$$

The locally translation-invariant distribution can be propagated in both directions from $\theta = \theta_0$; the resulting density function is

$$\frac{\Pi G_l\{l_i - l_1 + L_1 - (\theta - \theta_0) \mid \theta_0\}}{\int \Pi G_l(l_i - l_1 + t \mid \theta_0)\, dt}.$$

This globally translation-invariant distribution provides a first-derivative approximation at θ_0 and can be used to examine some local properties in a transparent manner.

5. Estimation Using a Conditionally-Sufficient Statistic

The logarithmic derivative of the likelihood function, called the *score* by Fisher, is prominent in the theory of estimation. Its variance, the *information*, appears in the Cramér–Rao inequality. The minimum-variance locally-unbiased estimate at θ_0 is based on the score and the information at θ_0; see Fraser (1964).

The log-likelihood for the conditioned variable x_1 is

$$k+\Sigma \ln F_x[x\{l_i-l_1+l(x_1)\}\,|\,\theta].$$

The derivative of this log-likelihood at θ_0 is

$$S(x_1)=\Sigma\frac{\partial}{\partial\theta_0}\ln F_x\{x(x_i)\,|\,\theta\},$$

which is the same as the score $\Sigma s(x_i)$ as calculated from the original observations $(x_1,\ldots,x_n)$.

Let $\operatorname{var}\{S(x_1)\,|\,l_i-l_1,\theta_0\}$ designate the conditional variance of $S(x_1)$ along the orbit defined by the ancillary statistic $\{l(x_2)-l(x_1),\ldots,l(x_n)-l(x_1)\}$. Then, by Fraser (1964), the locally-unbiased estimate having minimum variance is

$$\theta_0+\frac{S(x_1)}{\operatorname{var}\{S(x_1)\,|\,l_i-l_1,\theta_0\}}.$$

Conditionally on each orbit, this estimate is locally unbiased at θ_0. It follows that, unconditionally, it is locally unbiased at θ_0. As a consequence, its variance must be at least as large as the variance of

$$\theta_0+\frac{\Sigma s(x_i)}{\operatorname{var}\{\Sigma s(x_i)\,|\,\theta_0\}},$$

which is of minimum variance among locally-unbiased estimates. This inequality is easily checked; the variance of the conditional estimate is

$$E[1/\operatorname{var}\{S(x_1)\,|\,l_i-l_1,\theta_0\}]$$

whereas the variance of the unconditional estimate is

$$1/\operatorname{var}\{\Sigma s(x_i)\,|\,\theta_0\}.$$

The two scores are equal; the first expression can be viewed as the average of the reciprocal of a quantity and the second expression as the reciprocal of the average of the same quantity; the inequality then follows from the convexity of the reciprocal function.

The estimate based on conditional sufficiency can have *smaller* variance or it can have *larger* variance than the estimate based on sufficiency—*depending on the value of the ancillary statistic.* If viewed *unconditionally* its variance will be larger (unless its conditional variance is independent of the ancillary statistic). Thus minimum-variance and conditionality can be in conflict in the appraisal of locally-unbiased estimates.

6. The Large-Sample Conditional Distribution

The basic principle associated with conditionally-sufficient statistics is that they can be evaluated and interpreted entirely in terms of their conditional distribution given the ancillary statistic. In this Section the large-sample distribution form for the conditionally-sufficient statistic is examined, and found to be approximately normal under moderately general conditions.

Consider a sample $(x_1, \ldots, x_n)$ from the distribution $F(x \mid \theta)$, and the corresponding sample $(l_1, \ldots, l_n)$ from the distribution $G(l \mid \theta)$ obtained by linearizing at θ_0. The local ancillary statistic can conveniently be designated by $(l_1 - \bar{l}, \ldots, l_n - \bar{l})$ and the conditionally-sufficient statistic by $\bar{l}$.

The quantity $z = l - \theta$ is locally pivotal at θ_0. Its distribution function $H(z \mid \theta)$ satisfies

$$H(z \mid \theta) = G(z + \theta \mid \theta),$$

$$G(l \mid \theta) = H(l - \theta \mid \theta);$$

the local pivotality is expressed by $H_\theta(z \mid \theta_0) \equiv 0$. Let $h(z \mid \theta) = H_z(z \mid \theta)$ designate the density function for the quantity z; in terms of h the local pivotality becomes $h_\theta(z \mid \theta_0) \equiv 0$.

In the remainder of this Section the large-sample form of the conditional distribution will be investigated. The logarithm of the density function will be examined in terms of $\ln h(z \mid \theta)$. The implications for the original variables x will be considered in the next Section.

Consider the Taylor's expansion of $\ln h(z \mid \theta)$ about θ_0. For this, it will be convenient to write $\ln h(z \mid \theta) = \mathrm{lnh}\,(z \mid \theta)$ so that, for example, the second derivative of the log-density can be written

$$\frac{\partial^2}{\partial z^2} \ln h(z \mid \theta) = \mathrm{lnh}_{zz}(z \mid \theta).$$

Then

$$\mathrm{lnh}\,(z \mid \theta) = \mathrm{lnh}\,(z \mid \theta_0) + \tfrac{1}{2}(\theta - \theta_0)^2 \mathrm{lnh}_{\theta\theta}(z \mid \theta^*);$$

the linear term vanishes because $\mathrm{lnh}_\theta(z \mid \theta_0) = h_\theta(z \mid \theta_0)/h(z \mid \theta_0)$ which is zero by local pivotality, and θ^* is a value between θ_0 and θ that can depend on z.

The conditional density function was derived in Section 4. Let $(l_1 - \bar{l}, \ldots, l_n - \bar{l})$ designate the orbit and let $\bar{L}$ be a free variable denoting the sample average for a possible sample among those having the same differences as the reference sample $(l_1, \ldots, l_n)$. Consider now the logarithm of the density function and ignore the normalizing constant; then

$$\begin{aligned}
&\ln \prod_{i=1}^{n} G_l(l_i - \bar{l} + \bar{L} \mid \theta) \\
&= \Sigma \ln g(l_i - \bar{l} + \bar{L} \mid \theta) \\
&= \Sigma \ln h(l_i - \bar{l} + \bar{L} - \theta \mid \theta) \\
&= \Sigma\, \mathrm{lnh}\,(l_i - \bar{l} + \bar{L} - \theta \mid \theta_0) + \tfrac{1}{2}(\theta - \theta_0)^2 \Sigma\, \mathrm{lnh}_{\theta\theta}(l_i - \bar{l} + \bar{L} - \theta \mid \theta^*).
\end{aligned}$$

Now consider a Taylor's expansion of each term in terms of the variable $\bar{L}-\theta$. For this, it is convenient to use $\bar{L}-(\theta-\theta_0)$ in deviation from a reference point a. The log-density function becomes

$$\begin{aligned}&\Sigma \ln h(l_i-\bar{l}+a-\theta_0 \mid \theta_0)+(\bar{L}-\theta+\theta_0-a)\Sigma \ln h_z(l_i-\bar{l}+a-\theta_0 \mid \theta_0)\\ &\quad+\tfrac{1}{2}(\bar{L}-\theta+\theta_0-a)^2\Sigma \ln h_{zz}(l_i-\bar{l}+a-\theta_0 \mid \theta_0)\\ &\quad+\tfrac{1}{6}(\bar{L}-\theta+\theta_0-a)^3\Sigma \ln h_{zzz}(l_i-\bar{l}+a'-\theta_0 \mid \theta_0)\\ &\quad+\tfrac{1}{2}(\theta-\theta_0)^2\Sigma \ln h_{\theta\theta}(l_i-\bar{l}+a-\theta_0 \mid \theta^*)\\ &\quad+\tfrac{1}{2}(\theta-\theta_0)^2(\bar{L}-\theta+\theta_0-a)\Sigma \ln h_{\theta\theta z}(l_i-\bar{l}+a''-\theta_0 \mid \theta^*).\end{aligned}$$

Now choose the reference point a so that

$$\Sigma \ln h_z(l_i-\bar{l}+a-\theta_0 \mid \theta_0)=0;$$

the log-density can then be rewritten as

$$\begin{aligned}&k+\tfrac{1}{2}\{n^{\frac{1}{2}}(\bar{L}-\theta+\theta_0-a)\}^2\Sigma \ln h_{zz}(l_i-\bar{l}+a-\theta_0 \mid \theta_0)/n\\ &\quad+\tfrac{1}{6}n^{-\frac{1}{2}}\{n^{\frac{1}{2}}(\bar{L}-\theta+\theta_0-a)\}^3\Sigma \ln h_{zzz}(l_i-\bar{l}+a'-\theta_0 \mid \theta_0)/n\\ &\quad+\text{similar terms.}\end{aligned}$$

Then, on the assumption that the expectations

$$E\{\ln h_z(z+\delta \mid \theta_0) \mid \theta_0+\epsilon\}, \ldots, E\{\ln h_{\theta\theta z}(z+\delta \mid \theta_0+\epsilon_1) \mid \theta_0+\epsilon_2\}$$

exist, so that the law of large numbers applies uniformly for small $\delta, \epsilon_1, \epsilon_2$, it follows that, *for n sufficiently large, the probability is close to one that the conditional distribution of*

$$n^{\frac{1}{2}}(\bar{L}-\theta+\theta_0-a)$$

is asymptotically normal with mean equal to zero and variance equal to the reciprocal of

$$-(1/n)\Sigma \ln h_{zz}(l_i-\bar{l}+a-\theta_0 \mid \theta_0),$$

where a is the solution of

$$\Sigma \ln h_z(l_i-\bar{l}+a-\theta_0 \mid \theta_0)=0.$$

7. Interpretation of the Large-Sample Distribution

In the preceding Section the large-sample conditional distribution of

$$n^{\frac{1}{2}}(\bar{L}-\theta+\theta_0-a)$$

was shown to be approximately normal. Consider briefly the interpretation of this in terms of the original variables.

The constant a is chosen so that

$$\Sigma \ln h_z(l_i-\bar{l}+a-\theta_0)=0.$$

But this can be simplified, for

$$\Sigma \ln h_z(l_i - \bar{l} + a - \theta_0 \mid \theta_0) = \Sigma \frac{\partial}{\partial l} \ln g(l_i - \bar{l} + a \mid \theta_0)$$

$$= \Sigma \frac{G_{ll}(l_i - \bar{l} + a \mid \theta_0)}{G_l(l_i - \bar{l} + a \mid \theta_0)}$$

$$= -\Sigma \frac{G_{\theta l}(l_i - \bar{l} + a \mid \theta_0)}{G_l(l_i - \bar{l} + a \mid \theta_0)}$$

$$= -\Sigma \frac{\partial}{\partial \theta} \ln g(l_i - \bar{l} + a \mid \theta_0)$$

$$\doteq -\Sigma \frac{\partial}{\partial \theta} \ln f\{x(l_i - \bar{l} + a) \mid \theta_0\}.$$

Thus the constant a should satisfy

$$\Sigma \frac{\partial}{\partial \theta} \ln f\{x(l_i - \bar{l} + a) \mid \theta_0\} = 0,$$

or equivalently satisfy

$$\Sigma s\{x(l_i - \bar{l} + a) \mid \theta_0\} = 0,$$

which uses the score at θ_0. In this latter form the equation is similar to a maximum-likelihood equation and in fact the constant a can be interpreted in terms of maximum likelihood. The linearized observations are $(l_1, \ldots, l_n)$; additively adjusted to have average equal to zero, they become $(l_1 - \bar{l}, \ldots, l_n - \bar{l})$; additively adjusted to have average value equal to a, they become $(l_1 - \bar{l} + a, \ldots, l_n - \bar{l} + a)$. Thus, *if the linearized observations are additively adjusted, a is that average value for them which makes the maximum-likelihood solution equal to the reference point* θ_0.

For large samples the conditional distribution of

$$n^{\frac{1}{2}}\{\bar{L} - (\theta - \theta_0) - a\}$$

is approximately normal with mean equal to zero and variance equal to the reciprocal of

$$-\frac{1}{n} \Sigma \ln h_{zz}(l_i - \bar{l} + a - \theta_0 \mid \theta_0) = -\frac{1}{n} \Sigma \frac{\partial^2}{\partial l^2} \ln g(l_i - \bar{l} + a \mid \theta_0)$$

$$= \frac{1}{n} \Sigma \frac{\partial^2}{\partial l \, \partial \theta} \ln g(l_i - \bar{l} + a \mid \theta_0)$$

$$= \frac{1}{n} \Sigma \frac{\partial^2}{\partial l \, \partial \theta} \ln f\{x(l_i - \bar{l} + a) \mid \theta_0\};$$

this is the sample average of the mixed logarithmic derivative at the linearly adjusted data point for which θ_0 is the maximum-likelihood solution.

REFERENCES

BASU, D. (1959), "The family of ancillary statistics", *Sankhyā*, **21**, 247–256.

FISHER, R. A. (1956), *Statistical Methods and Scientific Inference*. Edinburgh: Oliver & Boyd.

FRASER, D. A. S. (1961), "Invariance and the fiducial method", *Biometrika*, **48**, 261–280.

—— (1964), "On local unbiased estimation", *J. R. statist. Soc.* B, **26**, 46–51.

LINDLEY, D. V. (1958), "Fiducial distributions and Bayes' theorem", *J. R. statist. Soc.* B, **20**, 102–107.

OWEN, A. R. G. (1948), "Ancillary statistics and fiducial distributions", *Sankhyā*, **9**, 1–18.

WELCH, B. L. (1939), "On confidence limits and sufficiency with particular reference to parameters of location", *Ann. math. Statist.*, **10**, 58–69.