〈Research Paper〉

Journal of the Korean
Statistical Society
Vol. 17, No. 2, 1988

Two Properties of Ancillary Statistics +

Yong Goo Lee*

ABSTRACT

Two properties of ancillary statistics are considered. One is to find a role of ancillary statistics in the statistical inference by showing that the ancillary statistic can recover the lost information and to give a criteria for comparing the conditional inference with unconditional inference. The other is to find an ancillary statistic of translation model and its relationship with observed Fisher information.

1. Introduction and Summary

In the parametric probability model, reliability of the statistical inference is largely depend on the precision of the estimator of the parameter. So improving the precision of the estimator is one of the most important topics in the statistical inference. Most of the cases, a function of sufficient statistic gives the best estimator, and an ancillary statistic is a statistic whose distribution is free of the parameter. So ancillary statistic by itself gives no information about the parameter. But some cases precision of the inference about the parameter can be improved by conditioning on the observed value of the ancillary statistic, if one exists.

But there are two difficulties with ancillary statistics. One is that there is no general method for constructing ancillary statistic and the other is that ancillary statistic may not be unique and can not be derived the best one such as minimal sufficient statistic. No one has suggested any solution for the first problem. But for the second problem, we can find some rules at Cox and Hinkley (1974, pp. 43-44) and Cox (1971).

In this research we have studied two properties of ancillary statistics. One is the usefulness of ancillary statistics based on the Fisher information (which is related with the second problem), and the other is to give a rule to find ancillary statistics of a translation model (which is related with the first problem). In section 2, we show that how the ancillary

+ This research was partially supported by Korea Science and Engineering Foundation Grant 864-0105-003-1

* Department of Applied Statistics, Chung Ang University, Seoul, Korea.

statistics can recover the lost information and give a criteria for comparing the conditional inference with unconditional inference. In section 3, we derive an ancillary statistic of a translation model and find relationship with derivatives of log likehood function at $\theta=\hat{\theta}$, which was proved to be ancillary statistic by Efron and Hinkley (1978).

2. Fisher Information and Ancillary Statistics

Fisher (1934) has shown that the maximum likelihood estimator (m.l.e.) which is a function of sufficient statistic, gives the best estimator for the parameter. But, in some cases, the m.l.e. can not contain all the information that the data has and ancillary statistic can recover the lost information.

In this section, we will study the relationship between ancillary statistic and Fisher information, show that how the ancillary statistic can recover the lost information and give a criteria for deciding what to condition on and whether to condition at all.

Let $\{f(X:\theta):\theta\varepsilon\Theta\}$ be a probability model and (T, U) be minimal sufficient statistic. Suppose the joint density of (T, U) is factored into conditional and marginal according to

$$f(t,u:\theta)= f_1(t \mid u:\theta)f_2(u:\theta) \tag{2.1}$$

where we deliberately allow θ in f_2. For f, f_1 and f_2, the regularity conditions of the likehood function are assumed to be satisfied. Let the logarithms of f, f_1, f_2 be l, m_1, m_2, so that if prime denotes derivative with respect to θ,

$$\begin{aligned} l &= m_1+ m_2 \\ l' &= m_1'+ m_2' \\ l'' &= m_1''+ m_2'' \end{aligned} \tag{2.2}$$

and by a simple calculation, we can get

$$\begin{aligned} &E(m_1' \mid u)= 0,\ El'=Em_1'= Em_2'= 0 \\ &E(m_1'^2 \mid u)= -E(m_1'' \mid u) \\ &E(l'^2)= -El'' \\ &E(m_2'^2)= -Em_2'' \end{aligned} \tag{2.3}$$

The Fisher information in (T,U) can be defined as

$$i_{t,u}(\theta)= -El''= El'^2, \tag{2.4}$$

and the information in U as

$$i_u(\theta) = -Em_2'' = Em_2'^2, \tag{2.5}$$

and the conditional information in t given U as

$$i_{t|u}(\theta) = -E(m_1'' \mid u) = E(m_1'^2 \mid u). \tag{2.6}$$

When U is distribution constant (which means that U is an ancillary statistic) then $m_2' = m_2'' = i_u(\theta) = 0$, and $i_{t|u}(\theta)$ could equivalently be defined as $-E(1'' \mid u)$.

Now consider the identity $1'' = m_1'' + m_1''$.

If we take first conditional and then marginal expectation we get

$$i_{t,u}(\theta) = Ei_{t|u}(\theta) + i_u(\theta) \tag{2.7}$$

In words, the total information equals the expectation of the conditional information plus the marginal information. When U is distribution constant so that $i_u(\theta) = 0$, this reduces to a well known result of Fisher which purports to explain how the ancillary statistic U recovers the lost information.

Equation (2.7) furnishes some possible criteria for deciding what to conditon on and whether to condition at all:

(a) U_1 is a better conditioning statistic than U_2 if

$$i_{u_1}(\theta) < i_{u_2}(\theta) \text{ for all } \theta. \tag{2.8}$$

(b) conditional inference using $t \mid U$ is prefered to unconditional inference using t if

$$Ei_{t|u}(\theta) > i_t(\theta) \text{ for all } \theta, \tag{2.9}$$

or equivaently if

$$i_t(\theta) + i_u(\theta) < i_{t,u}(\theta) \text{ for all } \theta. \tag{2.10}$$

Because of the dependence on θ, these two criteria give only partial results for using ancillary statistic as a conditioning statistic.

3. Translation Model and Ancillary Statistic

Fisher (1925, 1934) has addressed that the observed Fisher information is an ancillary statistic, and Efron and Hinkley (1978) has shown that the derivatives of the log likehood function at m.l.e. are ancillary statistics for translation model and (at least) approximate ancillary statistics for the other models.

In this section, we will show that the likehood shape statistic is an ancillary statistic for translation model and it is equivalent with the ancillary statistic defined by Efron and Hinkley (1978), and give an example that observed Fisher information is not an ancillary statistic for some nontranslation model. The m.l.e. of θ will be denoted by $\hat{\theta}$, and the regularty conditions for the likelihood function are assumed to be satisfied. Buehler (1982) has defined the translation model such that $f(X;\theta)$ is a tralslation model if X is one to one with $(\hat{\theta},U,V)$ where $(\hat{\theta},U)$ is sufficient for θ, U is distribution constant, V is a statistic which is needed to make X and $(\hat{\theta},U,V)$ have a same order, and the conditional density $f(\hat{\theta} \mid u;\theta)$ has the form of $f_0(\hat{\theta}-\theta \mid u)$.

Theorem 3.1. If $f(x:\theta)$ is a translation model, then $f(x:\theta) \propto f_0(\hat{\theta}-\theta \mid u)$, where $\propto$ means "proportional as a function of θ",

Proof. If J is the Jacobian of the transformation from x to $(\hat{\theta},u,v)$, then using sufficiency and ancillarity we have

$$\begin{aligned} f(x;\theta) &= f(\hat{\theta},u,v;\theta)J \\ &= f(\hat{\theta},u;\theta)f(v \mid \hat{\theta},u)J \\ &= f(\hat{\theta} \mid u:\theta)f(u)f(v \mid \hat{\theta},u)J \\ &\propto f_0(\hat{\theta}-\theta \mid u). \quad \blacksquare \end{aligned}$$

If x_1 and x_2 are values of X then we will denote $\hat{\theta}(x_j)$ by $\hat{\theta}_j$ and similarly for other statistics.

Definition 3.1. If $f(X;\theta)$ is the likehood function, $\hat{\theta}$ is the m.l.e., and $\theta+\hat{\theta}(X) \varepsilon \Theta$ (parameter space), x_1 and x_2 are any two points in the sample space, then we define that $W(X)$ is a likelihood shape statistic if

$$W(X_1)= W(X_2) \Leftrightarrow f[X_1;\theta+\hat{\theta}(X_1)] \propto f[X_2;\theta+\hat{\theta}(X_2)].$$

The likelihood shape statistic $W(X)$ specifies the shape of the likelihood function apart from its location, $\hat{\theta}$.

Theorem 3.2. If $f(X:\theta)$ is a translation model and $W(x)$ is a likelihood shape statistic. then

(i) $(\hat{\theta},W)$ is minimal sufficient

(ii) $u_1 = u_2$ implies $w_1 = w_2$

(iii) w is distribution constant.

Proof. (i) Using the known minimal sufficiency of the (normalized) likelihood it suffices to show $(\hat{\theta}_1,w_1) = (\hat{\theta}_2,w_2) \Leftrightarrow f(x_1,\theta) \propto f(x_2: \theta)$.
Assume the latter, then $\hat{\theta}_1 = \hat{\theta}_2$ and

$$f(x_1;\theta+\hat{\theta}_1) \propto f(x_2;\theta+\hat{\theta}_1) = f(x_2;\theta+\hat{\theta}_2),$$

and so $w_1 = w_2$. Next assume $(\hat{\theta}_1,w_1) = (\hat{\theta}_2,w_2)$.

$$\begin{aligned} f(x_1;\ \theta') &= f(x_1;\ \theta+\hat{\theta}_1), \text{ where } \theta' = \theta+\hat{\theta}_1 \\ &\propto f(x_2;\theta+\hat{\theta}_2), \text{ because } w_1 = w_2 \\ &= f(x_2;\theta+\hat{\theta}_1), \text{ because } \hat{\theta}_1 = \hat{\theta}_2 \\ &= f(x_2;\ \theta') \text{ for all } \theta' \end{aligned}$$

(ii) assume $u_1 = u_2$, then

$$\begin{aligned} f(x_1;\ \theta+\hat{\theta}_1) &\propto f_0(\hat{\theta}_1 - (\theta+\hat{\theta}_1);\ u_1) \text{ by Theorem 3.1} \\ &= f_0(-\theta;u_1) \\ &= f_0(-\theta;u_2) \\ &= f_0(\hat{\theta}_2 - (\theta+\hat{\theta}_2);\ u_2) \\ &\propto f(x_2;\theta+\hat{\theta}_2) \text{ by Theorem 3.1} \end{aligned}$$

which shows that $w_1 = w_2$.

(iii) this is a consequence of (ii). ■

Theorem 3.2. states that if we have a translation model, then the likelhood function determines an exact ancillary statistic W such that $(\hat{\theta},W)$ is minimal sufficient. The statistic W simply specifies the shape of the likelihood apart from its location $\hat{\theta}$. The shape W and location $\hat{\theta}$ together determine the likelihood function.

Examples 3.1. If $f(x;\theta)=\prod_{i=1}^{n} f(x_i-\theta)$ is a translation model and there is no sufficiency reduction beyond the order statistic, then W is the order statistic spacings $(X_{(2)}-X_{(1)}, \cdots, X_{(n)}-X_{(1)})$.

Example 3.2. (Efron and Hinkely, 1978) The parameter θ is measured by one of two measuring instruments whose errors are $N(0,\sigma_k^2)$, $K=0,1$, where σ_0^2 and σ_1^2 are known and unequal. Let $Pr(a_j=0)=Pr(a_j=1)=\frac{1}{2}$, then after n measurements we get $(a_1, x_1), \cdots, (a_n, x_n)$, $a_j=0$ or 1.
Assume that instrument 1 was used A times. Then $(\hat{\theta}, A)$ is the minimal sufficient statistic and W is equal to A.

Define the log likelihood function by full and abbreviated notations by

$$l(\theta;x)=l_\theta=\log f(x;\theta).$$

Derivatives with respect to θ will be denoted with varying degrees of abbreviation as

$$\begin{aligned} \partial l(\theta;x)/\partial\theta &= l'(\theta;x)=l_\theta'=l' \\ \partial^2 l(\theta;x)/\partial\theta^2 &= l''(\theta;x)=l_\theta'' \\ \partial^k l(\theta;x)/\partial\theta^k &= l^{(k)}(\theta;x)=l_\theta^{(k)}=l^{(k)} \end{aligned} \tag{3.1}$$

values at the maximum will be further abbreviated for example by

$$\hat{l}''=l_{\hat\theta}''=l''(\hat{\theta};x) \tag{3.2}$$

we will call

$$I(x)=-l'' \tag{3.3}$$

the observed Fisher information, and we will call

$$i(\theta)=E_\theta(l')^2=-E_\theta l''$$

the (expected) Fisher information.

Efron and Hinkley (1978) has proved that $\widetilde{W}=(\hat{1}^{''}, \hat{1}^{(3)}, \cdots)$ is an ancillary statistic for translation model. To the extend that $1(\theta;X)$ is an analytic function in θ with a convergent expansion about $\hat{\theta}$, the statistic W of Theorem 3.2 is quivalent to $\widetilde{W}$, for both statistics determine the shape of 1_θ about $\theta=\hat{\theta}$ and are determine by that shape.

While $(\hat{\theta},\widetilde{W})$ is in fact minimal sufficient, $\widetilde{W}$ nevertheless contains redundancies in that a finite vector will suffice to determine the rest. For Example 3.1, we could reasonably expect to use $(1^{''}, \cdots, 1^{(n)})$, and for Example 3.2, $1^{''}= I(x)$ suffices.

An interesting question is wheather ancillarity of $\widetilde{W}$ and in particular of $\hat{1}^{''}= I(x)$ carries over at least approximately to other models. These questions have been addressed by Fisher (1925, 1934) and by Efron and Hinkley (1978). Intuitively a large value of $I(x)$ corresponds to a sharply peaked likelihood and so large $I(x)$ seemingly indicates high precision. But ancillarity of $I(x)$ depends on the parameterization and in some cases $I(x)$ is not an ancillary statistic at all.

Example 3.3. Consider the autoregressive model

$$X_i = \theta X_{i-1} + \varepsilon_i, \; i=1,2, \cdots$$
$$\varepsilon_i \sim \text{iid } N(0, \sigma^2) \tag{3.4}$$

The log likelihood equals a constant plus $T\theta - V\theta^2/2$, where

$$T=\sum_{i=1}^{n} X_j X_{j-1} \quad \text{and} \quad V=\sum_{i=1}^{n} X_{i-1}^2 \tag{3.5}$$

then $\hat{\theta}= T/V$, $(\hat{\theta},V)$ is sufficient, and $I(x)= V$.
Specifically for $n=2$, $X_0=1$, $EV=E(X_0^2+ X_1^2)= 2+ \theta^2$, showing that V is not distribution constant.

References

(1) Buethler, R.J. (1982), Some Ancillary Statistics and Their Properties (with discussion and rejoinder), *Journal of the American Statistical Association,* 77, 581-594.

(2) Cox, D.R. (1971), The Choice between Alternative Ancillary Statistice, *Journal of the Royal Statistical Society,* Ser. B, 33, 251-255.

(3) Cox, D.R., and Hinkely, D.V. (1974), *Theoretical Statistics,* London: Chapman and Hall.

(4) Efron, B., and Hinkly, D.V. (1978), Assessing the Accuracy of the Maximum Likelihood Estimator: Observed vs Expected Fisher Information, *Biometrika,* 65, 457-487.

(5) Fisher, R.A. (1925), Theory of Statistical Estimation, *Proceedings of the Cambridge philosophical Society,*, 22, 700-725.

(6) Fisher, R.A. (1934), Two New Properties of Mathematical Likelihood, *Proceedings of the Royal Society,* Ser. A., 144, 285-307.