PARTLY SUFFICIENT STATISTICS AND COMPLETE CLASS THEOREMS FOR STATISTICAL DECISION PROBLEMS WITH NUISANCE PARAMETERS

Okuma, Akimichi
Department of Mathematics, Kyushu University

https://doi.org/10.5109/13094

出版情報：統計数理研究. 16 (3/4), pp.47-53, 1975-03. Research Association of Statistical Sciences

PARTLY SUFFICIENT STATISTICS AND COMPLETE CLASS THEOREMS FOR STATISTICAL DECISION PROBLEMS WITH NUISANCE PARAMETERS

By

Akimichi Okuma*

(Received October 4, 1974)

§ 1. Introduction.

In most of statistical inferences we propose a sufficient statistic for the family of distributions of the samples. The reason why we use the sufficient statistic is that the set of all procedures for the inference based on the sufficient statistic may consist of an essentially complete class for it.

For some statistical problems we need not to infer all of the parameters which are in the probability model of the samples. Some of these, say θ, are interesting for us and are the object of our statistical inference. We are in the case that the samples to be used in the inference are depending on some other parameters, ξ, as well as θ. In some of these cases it may be enough to find out a statistic which does not lose the infornation of θ in the samples instead of to find out a sufficient statistic of (θ, ξ).

In this point of view, we introduce a partly sufficient statistic for θ, which is required weaker conditions than the sufficiency for (θ, ξ), that is, if T is sufficient for (θ, ξ) then it is partly sufficient.

The main purposes of the present paper are to find the relation between a sufficient statistic and a partly sufficient statistic, and by using the partly sufficient statistic to find an essentially complete class for a statistical decision problem.

In section 2 we give the definitions of the partly sufficient statistic and complementary sufficiency, and we prove some relation between them. For minimal sufficiency there have been discussed in many papers, such as Halmos & Savage [5], Bahadur [1] and Dynkin [2]. As in these papers we define the necessity of a statistic and prove the result similar to Theorem 1 in Dynkin [2]. Linnik introduced the partial sufficiency (see Kagan [6] or Kagan et al. [7]), but it is diffierent from the partly sufficient statistic in the present paper.

In section 3 we introduce a statistical decision problem with nuisance parameters as in Okuma [9] and we generalize the complete class theorem in Ferguson [3], Okuma [8] etc., that is, if there exists a partly sufficient statistic then the class of all decision functions based on it is an essentially complete class. The corresponding result for a decision problem with mixed parmeters is given in Theorem 5.

* Department of Mathematics, Kyushu University, Fukuoka.

§ 2. Partly sufficient statistics.

As in Okuma [9] the class of all decision functions based on a sufficient statistic in the sense of Fraser [4], which we call *F-sufficiency* in the sequel, is an essentially complete class for a statistical decision problem with nuisance parameters. But the definition of F-sufficiency requires two conditions, which are

(i) *the induced probability measure of the statistic, T, is independent of the nuisance parameters, say ξ, and*

(ii) *for any fixed ξ, the conditional probability measure of the samples given T does not depend on the parameter of interesting, say θ.*

The existence of F-sufficient statistic may not be guaranteed in general. As a matter of fact, trivial statistic can not be F-sufficient since its probability measure depends on ξ. In usual sense the trivial statistic should be sufficient. According to the above aspects, we give the definitions of the *partly sufficient statistic* and the *complementary sufficient statistic.*

Let X be a random variable or a k-dimmentional random vector on $(\mathcal{X}, \mathfrak{B})$ where $\mathfrak{B}$ is a σ-field on $\mathcal{X}$, and $P_{\theta,\xi}$ be its probability measure where θ and ξ are points in some sets Θ and Ξ respectively.

Let T be a statistic defined on $(\mathcal{X}, \mathfrak{B})$ taking values in $(\mathcal{T}, \mathcal{B})$ i.e., T is a function of only X and is $\mathfrak{B}$-measenable, and let $P^T_{\theta,\xi}$ be the induced probability measure on $(\mathcal{T}, \mathcal{B})$, that is, for each $\theta \in \Theta$ and $\xi \in \Xi$,

$$P^T_{\theta,\xi}(B) = P_{\theta,\xi}(T^{-1}B) \tag{2.1}$$

for any $B \in \mathcal{B}$.

Let $P_{\Theta,\Xi}$ be the family of probability measures, $P_{\theta,\xi}$, $\theta \in \Theta$, $\xi \in \Xi$, and also let $P^T_{\Theta,\Xi}$ be the family of $P^T_{\theta,\xi}$, $\theta \in \Theta$ and $\xi \in \Xi$.

DEFINITION 1. *A statistic T is said to be a partly sufficient statistic of θ for $P_{\Theta,\Xi}$ if for each $A \in \mathfrak{B}$ there exists a $\mathcal{B}$-measurable function $P_\xi(A\mid t)$ which does not depend on θ such that for all $B \in \mathcal{B}$ and for all $P_{\theta,\xi} \in P_{\Theta,\Xi}$,*

$$P_{\theta,\xi}(A \cap T^{-1}B) = \int_B P_\xi(A\mid t)\,dP^T_{\theta,\xi}(t)\,. \tag{2.2}$$

In this definition we require the condition (ii) but do not (i). Therefore if T is a F-sufficient statistic of θ for $P_{\Theta,\Xi}$ then it is a partly sufficient statistic of θ for $P_{\Theta,\Xi}$. And also if T is a sufficient statistic of (θ, ξ) for $P_{\Theta,\Xi}$ in the usual sense then it is a partly sufficient statistic of θ as well as ξ for $P_{\Theta,\Xi}$, but it does not imply F-sufficiency of T. We may abbreviate the class of probability distributions, $P_{\Theta,\Xi}$, in the sequel when it is clear.

EXAMPLE 1. Let $X_1, \cdots, X_n$ be iid random variables whose common distribution is normal with mean μ, $-\infty < \mu < \infty$, and variance σ^2, $0 < \sigma^2 < \infty$. Then it is well known that $(\bar{X}_n, S_n^2)$, where $\bar{X}_n = \Sigma X_i/n$ and $S_n^2 = \Sigma(X_i - \bar{X}_n)^2$, is a sufficient statistic of (μ, σ^2). It is easy to show that $\bar{X}_n$ is a partly sufficient statistic of μ but S_n^2 is not a partly sufficient statistic of σ^2. Apparently both $\bar{X}_n$ and S_n^2 are not F-sufficient statistic of μ and σ^2 respectively. However, as we know, S_n^2 is the best statistic for a testing hypothesis about σ^2 when μ is unknown. Therefore we introduce the following new notion.

DEFINITION 2. *A statistic T is said to be a complementary sufficient statistic of θ for $P_{\theta,\Xi}$ if there exists a partly sufficient statistic S of θ for $P_{\theta,\Xi}$ such that (T,S) is a sufficient statistic of (θ,ξ) for $P_{\theta,\Xi}$ but S itself is not.*

In example 1, clearly S_n^2 is a complementary sufficient statistic of σ^2.

EXAMPLE 2. Let $X_1,\cdots,X_n$ be iid random variables and the common distribution is Gamma with parameters α and β, i.e.,

$$f(x\mid\alpha,\beta)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^\alpha e^{-\frac{x}{\beta}}I_{(0,\infty)}(x).$$

Then ΣX_i and ΠX_i are complementary sufficient as well as partly sufficient statistics of β and α respectively, but none of them are F-sufficient.

EXAMPLE 3. Let $X_1,\cdots,X_n$ be iid randon variables from an uniform distribution on (α,β), then $\mathrm{Min}\,X_i$ is a complementary sufficient and partly sufficient statistic of α, and $\mathrm{Max}\,X_i$ is also of β.

THEOREM 1. *If T and S are partly sufficient statistics of θ and ξ respectively, and let $(\mathcal{T},\mathcal{B})$ and $(\mathcal{S},\mathcal{C})$ be their range spaces. Let $W=(T,S)$ and let $\sigma(\mathcal{B}\times\mathcal{C})$ be the induced σ-field by $\mathcal{B}\times\mathcal{C}$. Then W is a sufficient statistic, taking values in $(\mathcal{T}\times\mathcal{S},\sigma(\mathcal{B}\times\mathcal{C}))$, of (θ,ξ).*

PROOF. By Radon-Nikodym Theorem, for each $A\in\mathfrak{B}$ there exsists a $\sigma(\mathcal{B}\times\mathcal{C})$-measurable function $P_{\theta,\xi}(\cdot\mid t,s)$ such that

$$P_{\theta,\xi}(A\cap T^{-1}B\cap S^{-1}C)=\int_{B\times C}P_{\theta,\xi}(A\mid t,s)\,dP_{\theta,\xi}^{T,S}(t,s) \tag{2.3}$$

for any $B\in\mathcal{B}$ and $C\in\mathcal{C}$. On the other hand, by the partical sufficiency of S, there exists $P_1(\cdot\mid s)$ such that

$$P_{\theta,\xi}(A\cap T^{-1}B\cap S^{-1}C)=\int_C P_1(A\cap T^{-1}B\mid s)\,dP_{\theta,\xi}^S(s) \tag{2.4}$$

where P_1 may depend on θ but not depend on ξ. The equation (2.3) may be written as

$$P_{\theta,\xi}(A\cap T^{-1}B\cap S^{-1}C)=\int_C\int_B P_{\theta,\xi}(A\mid t,s)\,dP_{\theta,\xi}^T(t\mid s)\,dP_{\theta,\xi}^S(s) \tag{2.5}$$

where $P_{\theta,\xi}^T(\cdot\mid s)$ is the conditional probability measure of T given $S=s$, so that $P_{\theta,\xi}^T(\cdot\mid s)$ does not depend on ξ, since $P_{\theta,\xi}^T(B\mid s)=P_{\theta,\xi}(T^{-1}B\mid s)$ and S is a partly sufficient statistic of ξ. From (2.4) and (2.5), for any $C\in\mathcal{C}$

$$\int_C P_1(A\cap T^{-1}B\mid s)\,dP_{\theta,\xi}^S(s)=\int_C\int_B P_{\theta,\xi}(A\mid t,s)\,dP_\theta^T(t\mid s)\,dP_{\theta,\xi}^S(s).$$

Therefore

$$P_1(A\cap T^{-1}B\mid s)=\int_B P_{\theta,\xi}(A\mid t,s)\,dP_\theta^T(t\mid s),\qquad a.s.\ \ P_{\theta,\xi}^S. \tag{2.6}$$

Let N be its exceptional null set in $\mathcal{S}$. Again by Radon-Nikodym Theorem, for any fixed $s\in N^c$ there exists a $\mathcal{B}$-measurable function $P_2(\cdot\mid t,s)$ such that

$$P_1(A\cap T^{-1}B\mid s)=\int_B P_2(A\mid t,s)\,dP_\theta^T(t\mid s) \tag{2.7}$$

where P_2 does not depend on θ since T is a partly sufficient statistic of θ, furthermore P_2 does not depend on ξ since P_1 does not. Then for any $B\in\mathscr{B}$, from (2.6) and (2.7),

$$\int_B P_2(A|t,s)dP_\theta^T(t|s)=\int_B P_{\theta,\xi}(A|t,s)dP_\theta^T(t|s)\,.$$

This implies that $P_{\theta,\xi}(A|t,s)$ depend neither on θ nor on ξ for almost all (t,s) and then for each $A\in\mathfrak{B}$, there exists a $\sigma(\mathscr{B}\times\mathcal{C})$-measurable $P(A|t,s)$ such that $P(A|t\,s)$, does not depend on (θ,ξ) and statisfies (2.3). Therefore, for any $D\in\sigma(\mathscr{B}\times\mathcal{C})$ we have

$$P_{\theta,\xi}(A\cap W^{-1}D)=\int_D P(A|w)dP_{\theta,\xi}^W(w)\,, \tag{2.8}$$

which completes the proof.

As an immediate consequence we have the following corollary.

COROLLARY. *Under the same assumptions as in Theorem* 1, *and if neither T nor S are sufficient statistics of (θ,ξ), then both T and S are complementary sufficient statistics of θ and ξ repectively.*

To discuss the minimality of a partly sufficient statistic, we will define a notion of partial necessity. The necessity of a statistic has been discussed in several papers (see Bahadur [1], Dynkin [2] etc.).

Let $T(x)$ and $S(x)$ be two functions defined on $\mathfrak{X}$. If $S(x_1)=S(x_2)$ implies $T(x_1)=T(x_2)$, then we say that T *is dependent on* S, and if T is dependent on S and also S is dependent on T then we call the functions T and S *equivalent.*

DEFINITION 3. *A statistic T is said to be a partly necessary statistic of θ for $P_{\theta,\Xi}$ if it is dependent on every partly sufficient statistic of θ for $P_{\theta,\Xi}$.*

DEFINITION 4. *A statistic T is said to be a minimal partly sufficient statistic of θ for $P_{\theta,\Xi}$ if it is a partly necessary and sufficient statistic of θ for $P_{\theta,\Xi}$.*

We assume that the family of probability measures $P_{\theta,\Xi}$ is dominated by some σ-finite measure μ then let $f(\cdot|\theta,\xi)$ be the density function corresponding to $P_{\theta,\xi}$ with respect to μ. And also we assume that $f(\cdot|\theta,\xi)$ be a positive continuous function on $\mathfrak{X}$ for each θ and ξ in the rest of this section.

By a similar proof of Corollary 1 in Halmos and Savage [5] we have the following factorization theorem for a partly sufficient statistic.

THEOREM 2. *T is a partly sufficient statistic of θ for $P_{\theta,\Xi}$ if there exist a function g of t which may depend on θ and ξ and a function h of x which may depend on ξ but does not on θ such that*

$$f(x|\theta,\xi)=g(T(x),\theta,\xi)\cdot h(x,\xi)\,.$$

The following theorem corresponds to Theorem 1 in Dynkin [2] where T is minimal sufficient. Let θ_0 be some fixed point in θ.

THEOREM 3. *Any statistic which is equivalent to $p_x(\theta,\xi)=\ln f(x|\theta,\xi)-\ln f(x|\theta_0,\xi)$ is a minimal partly sufficient statistic of θ for $P_{\theta,\Xi}$.*

PROOF. Let T be any equivalent statistic to $p_x(\theta,\xi)$. For any $x\in X$

$$f(x|\theta,\xi)=f(x|\theta_0,\xi)e^{p_x(\theta,\xi)}\,.$$

Since T is equivalent to $p_x(\theta,\xi)$, there exists a function l such that $p_x(\theta,\xi)=l(T(x),\theta,\xi)$, and then

$$f(x|\theta, \xi)=e^{l(T(x),\theta,\xi)}\cdot f(x|\theta_0, \xi).$$

Hence, from Theorem 2, T is a partly sufficient statistic of θ.

Let S be any partly sufficient statistic of θ. Then by Theorem 2 there exist g and h such that

$$f(x|\theta, \xi)=g(S(x)|\theta, \xi)h(x|\xi).$$

Therefore

$$\begin{aligned} p_x(\theta, \xi) &= \ln f(x|\theta, \xi)-\ln f(x|\theta_0, \xi) \\ &= \ln g(S(x)|\theta, \xi)-\ln g(S(x)|\theta_0, \xi) \end{aligned}$$

which depends on $S(x)$, and then $T(x)$ depends on $S(x)$, which shows that T is a partly necessary statistic of θ.

§ 3. Essentially complete class.

In this section we deal with a statistical decision problem with nuisance parameters.

As in section 2 let $(\mathcal{X}, \mathfrak{B})$ be a sample space and $P_{\theta,\Xi}$ be a family of probability measures on $(\mathcal{X}, \mathfrak{B})$. Let Θ be a parameter space of interesting, and let $(\mathfrak{A}, \mathcal{A})$ be a space of actions, each element of which is prescribing a decision for θ. We restrict our attention to some convex set of behavioral decision functions, $\mathcal{D}$, that is, any $\delta \in \mathcal{D}$ is a function from the sample space $\mathcal{X}$ to the family of all distribution functions on $(\mathfrak{A}, \mathcal{A})$.

According to Okuma [9] we formulate a statistical decision problem with nuisance parameters as follows.

For each $\xi \in \Xi$

$$\mathcal{P}_\xi : (\Theta, \mathfrak{A}, L, P_{\theta,\xi}). \tag{3.1}$$

That is, we choose an action which prescribes some decision for $\theta \in \Theta$, according to the value of the sample X whose distribution is given by $P_{\theta,\xi}$, and its loss is given by $L(\theta, a)$.

To avoid the mathematical complication we assume the measurability and integrability if necessary.

Next, we suppose a prior distribution G on Ξ, and let P_θ be the marginal distribution of X, i. e.,

$$P_\theta=\int_\Xi P_{\theta,\xi}dG(\xi). \tag{3.2}$$

And let define a new statistical decision problem for mixed parameters model.

$$\mathcal{P}_G : (\Theta, \mathfrak{A}, L, P_\theta). \tag{3.3}$$

For any $\delta \in \mathcal{D}$ the risk functions of δ for $\mathcal{P}_\xi$ and $\mathcal{P}_G$ are denoted by

$$R(\theta, \xi, \delta)=\int_{\mathcal{X}} \hat{L}(\theta, \delta_x)dP_{\theta,\xi}(x) \tag{3.4}$$

and

$$\hat{R}(\theta, \delta)=\int_{\mathcal{X}} \hat{L}(\theta, \delta_x)dP_\theta(x) \tag{3.5}$$

respectively, where

$$\hat{L}(\theta, \delta_x) = \int_{\mathfrak{A}} \hat{L}(\theta, a) d\delta_x(a) . \tag{3.6}$$

For any statistic T, let $\mathcal{D}_T$ be the class of all $\delta \in \mathcal{D}$ such that δ is based on T, that is, if $T(x_1) = T(x_2)$ for $x_1, x_2 \in \mathcal{X}$ then δ_{x_1} and δ_{x_2} are essentially same distributions on $\mathfrak{A}$.

Then we have the following complete class theorem for $\mathcal{P}_\xi$, $\xi \in \Xi$.

THEOREM 4. *For any partly sufficient statistic T of θ for $P_{\theta,\Xi}$, $\mathcal{D}_T$ is an essentially complete class in $\mathcal{D}$ for $\mathcal{P}_\xi$.*

PROOF. By the definition of partial sufficiency of T, the conditional distribution of X given T, $P_\xi^{X|T}$, is independent of θ (it may depend on ξ), then taking the $P_\xi^{X|T}$-mixture of δ, δ^ξ, i. e.,

$$\delta_t^\xi = E[\delta_x | T = t] \tag{3.7}$$

for each $t \in \mathcal{T}$, $\delta^\xi \in \mathcal{D}_T$. By using Lemma 2 in Okuma [9]

$$\begin{aligned} R(\theta, \xi, \delta^\xi) &= \iint \hat{L}(\theta, \delta_x) dP_\xi^{X|T} dP_{\theta,\xi}^T \\ &= \int L(\theta, \delta_x) dP_{\theta,\xi} \\ &= R(\theta, \xi, \delta) . \end{aligned} \tag{3.8}$$

Then $\mathcal{D}_T$ is an essentially complete class for $\mathcal{P}_\xi$.

REMARK 1. The above $\mathcal{D}_T$ does not depend on the volue of ξ. Thus Theorem 4 shows that if we have a partly sufficient statistic of θ then no matter what ξ is, the class of all decision functions based on the statistic constitutes an essentially complete class, and then we may restrict our decision class to the essentially complete class. It is important because interesting is only θ and then we can reduce the samples to a partly sufficient statistic of θ.

A similar result for $\mathcal{P}_G$ is in the following theorem, in which we require stronger assumptions than in Theorem 4.

THEOREM 5. *Let T be a partly sufficient statistic of θ for $P_{\theta,\Xi}$. If either*

(i) *the induced probability measure of T does not depend on ξ or*

(ii) *the postereor distribution of ξ given T, say $H(\xi | T)$, does not depend on θ, then $\mathcal{D}_T$ is an essentially complete class for $\mathcal{P}_G$.*

PROOF. Assume (i), for each $\delta \in \mathcal{D}$ let $\delta_t^* = \int_\Xi \delta_t^\xi \, dG(\xi)$ where δ^ξ is defined by (3.7) then $\delta^* \in \mathcal{D}_T$, furthemore by using Fubini's Theorem and Lemma 2 in Okuma [9] we have

$$\begin{aligned} \hat{R}(\theta, \delta^*) &= \iint_\Xi \hat{L}(\theta, \delta_t^\xi) dG(\xi) dP_\theta^T(t) \\ &= \int_\Xi R(\theta, \xi, \delta^\xi) dG(\xi) . \end{aligned}$$

From (3.8), $\hat{R}(\theta, \delta^*) = \hat{R}(\theta, \delta)$.

Next, we assume (ii), let

$$\delta_t^* = \int_\Xi \delta_t^\xi \, dH(\xi | t) \tag{3.9}$$

then $\delta^* \in \mathcal{D}_T$ and also by using Fubini's Theorem and Lemma 2 in Okuma [9], we have

$$R(\theta, \delta^*) = \iint_{\Xi} \hat{L}(\theta, \delta_t^{\xi}) dH(\xi \mid t) dQ_\theta^T$$

where Q_θ^T is the G-mixture of $P_{\theta,\xi}^T$ and δ^{ξ} is in (3.7). Then from (3.9) and Lemma 2 in Okuma [9] again,

$$\hat{R}(\theta, \delta^*) = \hat{R}(\theta, \delta),$$

which completes the proof.

REMARK 2. As we mentioned in Remark 1, for $\mathcal{P}_G$, $\mathcal{D}_T$ does not depend on G, hence similarly in Remark 1, no matter what the prior distribution G is, we can restrict our class of decision functions to $\mathcal{D}_T$.

References

[1] R.R. BAHADUR, *Sufficiency and statistical decision functions.* Ann. Math. Stat., 25 (1954), 423–462.

[2] E.B. DYNKIN, *Necesary and sufficient statistics for a family of probability distributions.* Select. Transl. Math. Stat. and Prob., 1, 23–41.

[3] T.S. FERGUSON, Mathematical Statistics. Academic Press, New York (1967).

[4] D.A.S. FRASER, *Sufficient statistic with nuisance parameters.* Ann. Math. Stat., 27 (1956), 836–842.

[5] P.R. HALMOS AND L.J. SAVAGE, *Application of the Radon-Nikodym theorem to the theory of sufficient statistics.* Ann. Math. Stat., 20 (1949), 225–241.

[6] A.M. KAGAN, *Partial sufficiency and unbiased estimation of polynomials in a translation parameter.* Soviet Math. Dokl., 8 (1967), 745–747.

[7] A.M. KAGAN, JU. V. LINNIK AND C.R. RAO, Characterization problems in mathematical statistics. Wiley, New York (1972).

[8] A. OKUMA, *On invariant sequential decision problems.* Mem. Fac. Sci., Kyushu Univ., 27 (1973), 121–134.

[9] A. OKUMA, *On statistical decision problems with nuisance parameters.* Bull. Math. Stat., 16, No. 3～4 (1975). to appear.