International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 6 (Special Issue), 2007, pp 55-61

Minimal Sufficient Statistics Emerge from the Observed Likelihood Functions

D.A.S Fraser
Dept. of Statistics
University of Toronto
Toronto, Canada

Amir Naderi
Dept. of Mathematical Sciences
Isfahan University of Technalogy
Isfahan, Iran

[Received December 11, 2005; Accepted July 8, 2007]

Abstract

The likelihood statistic guides statistical analysis in almost all areas of application. A precise definition of likelihood statistic is given, and simple and easy to use criteria are proposed to establish under weak conditions that minimal sufficiency in statistics emerges from observed likelihood functions. Some examples are presented.

Keywords and Phrases: Probability measures, derivatives of probability measures, likelihood, likelihood statistic, minimal sufficiency.

AMS Classification: 60A10, 62B05.

1 Introduction

With model $f_\theta(x)$, the observed likelihood function from x_0 is $cf_\theta(x_0)$ where c is an arbitrary positive constant; this can be formalized in terms of an equivalence class $R(f_\theta(x_0)) = \{cf_\theta(x_0), c \in R^+\}$ of functions on the parameter space, say Ω. The likelihood map $L(\theta|x)$ is defined as the mapping that carries a point x in the sample space X to $R(f_\theta(x))$, see Naderi [14] or Fraser et al [8]. Fisher [4]-[7] noted that the general dependence of $L(\theta|x)$ on x is that of minimal sufficiency but did this without detailed formulation. Neyman [15], Halmos and Savage [10], Lehmann and Scheffé

[12], Dynkin [3] and Bahadur [1] extended Fisher's results. The theory of minimal sufficiency was initiated by Lehmann and Scheffé [12] and Dynkin [3]. Pitcher [16] constructed a family of probability measures for which a minimal sufficient statistic did not exist. Landers and Rogge [11] showed that even for a dominated family of probability measures a minimal sufficient statistic did not exist. Barndorff-Nielsen et al [2] and Fraser et al [8], [9] however showed that the minimal sufficient statistic exists under some regularity conditions. This paper establishes that minimal sufficiency in statistics emerges from the observed likelihood functions under weak conditions.

2 Likelihood Partition

For each θ in Ω, let $(X, \mathcal{B}, P_\theta)$ be a probability space, P_θ be absolutely continuous with respect to a σ-finite measure μ, and $f_\theta(x)$ be the probability density of P_θ with respect to μ. Suppose $P = \{P_\theta : \theta \in \Omega\}$ and $\mathcal{F} = \{f_\theta(x) : \theta \in \Omega\}$. For each $\theta, f_\theta(x)$ is unique up to a set of μ-measure zero. The general uniqueness of $f_\theta(x)$ however, is required for the definition of our mappings. This may be achived by defining the probability density $f_\theta(x)$ as the $\lim_{n\to\infty} \frac{P_\theta(E_n)}{\mu(E_n)}$ for all sequences $\{E_n\}$ converging regularly to x and possessing some covering properties. For a detailed treatment see for example Naderi [14] or Fraser et al [8]. The question of the existence, in general, of a minimal sufficient partition or the associated statistic involves measure theoretic difficulties and some regularity conditions need to be imposed on $\mathcal{F}$. Two such regularity conditions are:

Separability Condition. $\mathcal{F}$ is separable if there exists a fixed countable subset $\mathcal{F}_0 = \{f_\theta(x) : \theta \in \Omega_0\}$ of $\mathcal{F}$ such that for each $f_\theta(x)$ in $\mathcal{F}$ there is a sequence $\{f_{\theta_n}(x)\}$ in $\mathcal{F}_0$ for which

$$\int_X |f_\theta(x) - f_{\theta_n}(x)| d\mu(x) \to 0$$

as $n \to \infty$. In this case $\mathcal{F}_0$ is said to be dense in $\mathcal{F}$. When X is a Euclidean space, the class of probability densities is a separable metric space with distance defined by $d(g,h) = \int_X |g(x) - h(x)| d\mu(x)$.

Continuity Condition. $\mathcal{F}$ is continuous on Ω if Ω is a separable metric space and if, for each x in X, $f_\theta(x)$ is continuous in θ. The Continuity Condition is stronger than the Separability Condition.

Let R_+^Ω be the class of all functions from Ω to $R_+ = [0, \infty)$ and let $R_*^\Omega = \{R(g) : g \in R_+^\Omega\}$ where $R(g) = \{cg : c > 0\}$ is the same class reduced modulo the scale group. An element $R(f_\theta(x))$ of R_*^Ω is called the likelihood function from x. The likelihood map L is the function from X to R_*^Ω that maps x in X to $R(f_\theta(x))$. For countable Ω the likelihood map is always minimal sufficient. For a general Ω, however, the likelihood map is minimal sufficient under the Continuity Condition, and the restricted likelihood map to Ω_0 say $L|\Omega_0$, is minimal sufficient under the Separability Condition, see Naderi [14] or Fraser et al [8]. The partition of X induced by both L and $L|\Omega_0$ coincides with the minimal sufficient partition proposed by Lehmann and Scheffé [12] $a.e.[P]$, see

McDunnough et al [13].
The standardized likelihood map is defined as follows. There is a countable subfamily $\{P_{\theta_n^*}, n \geq 1\}$ of P which is equivalent to P (Halmos and Savage [10]), and a Probability measure λ on $\mathcal{B}$ defined by

$$\lambda(E) = \sum_{n=1}^{\infty} \alpha_n P_{\theta_n^*}(E) \ \ , \ \ \sum_{n=1}^{\infty} \alpha_n = 1 \ \ , \ \ \alpha_n > 0,$$

which is equivalent to P. Let f be the probability density of λ with respect to μ. Then

$$f(x) = \sum_{n=1}^{\infty} \alpha_n f_{\theta_n^*}(x)$$

except on a set N_1 of μ measure zero. Let $N_2 = \{x : f(x) = 0\}$, then $\lambda(N_2) = 0$. Since λ is equivalent to P, $P_\theta(N_2) = 0$ for every θ in Ω. If $N = N_1 \cup N_2$, since μ dominates P and N_2 is a null set for P, then N is a null set for P. The standardized likelihood map is a mapping r from X to R^Ω that carries x in $X - N$ to $q_\theta = dP_\theta/d\lambda = f_\theta(x)/f(x)$. The values of r on N are immaterial and may be defined arbitrarily.

Barndorff-Nielsen et al [2] have shown that r is minimal sufficient under the Continuity Condition in the Bahadur sense. Their result may also be obtained via the following theorem and the fact that the likelihood map is minimal sufficient under the Continuity Condition; or more directly by showing that the partition of X induced by r coincides $a.e.[P]$ with the minimal sufficient partition proposed by Lehmann and Scheffé [12], see McDunnough et al [13]. As $L|\Omega_0$ is a minimal sufficient statistic under the Separability Condition this theorem also implies that $r|\Omega_0$, the restriction of r to Ω_0, is minimal sufficient statistic under the Separability Condition.
Theorem 1. L and r induce the same partition of $X, a.e.[P]$.
Proof. For each x and y in $X - N$ let $r(x) = r(y)$. Then $q_\theta(x) = q_\theta(y)$ or $f_\theta(x)/f(x) = f_\theta(y)/f(y)$. This may be written as

$$f_\theta(x) = \frac{f(x)}{f(y)} f_\theta(y) \ \text{ or } \ f_\theta(y) = \frac{f(y)}{f(x)} f_\theta(x)\cdot$$

Hence $R(f_\theta(x)) = R(f_\theta(y))$ and thus $L(x) = L(y)$. Now for each x and y in $X - N$ let $L(x) = L(y)$. Then by definition $R(f_\theta(x)) = R(f_\theta(y))$.This implies that there is a nonnegative function $h(x, y)$ such that for each θ, $f_\theta(x) = h(x, y) f_\theta(y)$. But

$$\begin{aligned} f(x) &= \sum_{n=1}^{\infty} \alpha_n \ \ f_{\theta_n^*}(x) = \sum_{n=1}^{\infty} \alpha_n h(x,y) f_{\theta_n^*}(y) \\ &= h(x,y) f(y)\cdot \end{aligned}$$

Hence $h(x, y) = f(x)/f(y)$ and thus $f_\theta(x) = (f(x)/f(y)) f_\theta(y)$. This implies that $f_\theta(x)/f(x) = f_\theta(y)/f(y)$ or $r(x) = r(y)$. Hence for each x and y in $X - N$, $L(x) = L(y)$ if and only if $r(x) = r(y)$ and thus L and r induce the same partition of $X, a.e.[P]$.

3 Likelihood Statistic

We shall call a statistic T a likelihood statistic if the partition of X induced by T coincides with the partition of X induced by the likelihood map $L, a.e.[P]$. The likelihood statistic is essentially unique up to a one to one equivalence. Clearly L is a likelihood statistic and by Theorem 1, r is also a likelihood statistic. The definition of likelihood statistic implies that a likelihood statistic is minimal sufficient if and only if the likelihood map is minimal sufficient. Hence when Ω is countable, or more generally when $\mathcal{F}$ is continuous on Ω, a likelihood statistic is minimal sufficient. The Likelihood statistic guides statistical analysis in almost all areas of application. Some important likelihood statistics, for example, $(\bar{x}, s_x^2)$, arise from the likelihood function with a normal random sample. The next theorem gives a simple criteria to establish the minimal sufficiency of such statistics.

Theorem 2. Suppose θ_0 is an arbitrary but fixed point in Ω such that for each x in $X - N$, $f_{\theta_0}(x) > 0$. Suppose t is a statistic such that for each θ in Ω , $g_{\theta|\theta_0}(t(x)) = f_\theta(x)/f_{\theta_0}(x)$ is a function of $t(x)$ for x in $X - N$. If $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$, θ in Ω implies that $t(x) = t(y)$, then t is a likelihood statistic.

Proof. For each x and y in $X - N$, $t(x) = t(y)$ implies that $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ and hence that $f_\theta(x)/f_{\theta_0}(x) = f_\theta(y)/f_{\theta_0}(y)$. This in turn implies that $f_\theta(x) = (f_{\theta_0}(x)/f_{\theta_0}(y))f_\theta(y)$ and that $f_\theta(y) = (f_{\theta_0}(y)/f_{\theta_0}(x))f_\theta(x)$. Hence $R(f_\theta(x)) = R(f_\theta(y))$ and thus $L(x) = L(y)$. Now for each x and y in $X - N$, let $L(x) = L(y)$. Then by definition $R(f_\theta(x)) = R(f_\theta(y))$. Thus there is a nonnegative function $h(x, y)$, free of θ, such that for each θ in Ω, $f_\theta(x) = h(x, y)f_\theta(y)$. But, as in the proof of Theorem 1, $h(x, y) = f(x)/f(y)$. Hence for each θ in Ω, $f_\theta(x) = (f(x)/f(y))f_\theta(y)$. The last equation implies that for an arbitrary but fixed point θ_0 in Ω, $f_{\theta_0}(x) = (f(x)/f(y))f_{\theta_0}(y)$. Hence for each θ in Ω, $f_\theta(x)/f_{\theta_0}(x) = f_\theta(y)/f_{\theta_0}(y)$ or $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ and thus by the assumption $t(x) = t(y)$. Since for each x and y in $X - N$, $t(x) = t(y)$ if and only if $L(x) = L(y)$, the partition of X induced by t coincides with the partition of X induced by $L, a.e[P]$ and hence t is a likelihood statistic.

When Ω is Countable or more generally when $\mathcal{F}$ is continuous on Ω the likelihood statistic t in Theorem 2 is minimal sufficient. When the Continuity Condition is not fulfilled the criteria given in the next theorem may be used to establish the minimal sufficiency of the statistic t.

Theorem 3. Suppose $\mathcal{F}$ is separable and $\mathcal{F}_0 = \{f_\theta(x) : \theta \in \Omega_0\}$ is dense in $\mathcal{F}$. Define $g_{\theta|\theta_0}(t(x))$ as in Theorem 2. If $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y)), \theta$ in Ω_0, results in $t(x) = t(y)$ then t is a minimal sufficient statistic for P.

Proof. Consider $L|\Omega_0$, the restriction of the likelihood map L to Ω_0. A similar argument as in the proof of Theorem 2 implies that $L|\Omega_0(x) = L|\Omega_0(y)$ if and only if $t(x) = t(y), a.e.[P]$. Thus both $L|\Omega_0$ and t induce the same partition of $X, a.e.[P]$. But $L|\Omega_0$ is minimal sufficient under the Separability Condition for P, hence t is minimal sufficient for P.

Note that the Separability Condition holds when X is a Euclidean space. Hence the criteria in Theorem 3 are widely applicable. In practice the conditions of Theorems

2 and 3 are verified by showing that $t(x) = t(y)$ is a θ-free solution to the equation $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$, for θ in Ω for the continuity case and for θ in Ω_0 for the separability case.

4 Examples

1. Consider a random sample from the Normal (θ, σ_0^2) distribution with σ_0^2 known and (θ, σ_0^2) in $R \times R^+$ (location normal model). If $\theta_0 = 0$ and $t(x) = \bar{x}$ then $g_{\theta|\theta_0}(t(x)) = \exp(-\frac{n}{2\sigma_0^2}(\theta^2 - 2\theta\bar{x}))$. For each θ in R, $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ if and only if for each θ in R, $\exp(\frac{n\theta}{\sigma_0^2}(\bar{x} - \bar{y})) = 1$. A θ-free solution to the equation is $\bar{x} = \bar{y}$. Hence $t(x) = t(y)$ and thus $t(x) = \bar{x}$ is a minimal sufficient statistic.

2. Consider a random sample from Normal (μ, σ^2) distribution with μ, σ^2 both unknown and $\theta = (\mu, \sigma^2)$ in $R \times R^+$ (location-scale normal model). If $\theta_0 = (0, 1)$ and $t(x) = (\bar{x}, s_x^2)$ then

$$g_{\theta|\theta_0}(t(x)) = \sigma^{-n} \exp(-\frac{n\mu^2}{2\sigma^2} - \frac{n}{2}(\frac{1}{\sigma^2} - 1)\bar{x}^2 + \frac{n\mu}{\sigma^2}\bar{x} - \frac{(n-1)}{2}(\frac{1}{\sigma^2} - 1)s_x^2)\cdot$$

For each θ in $R \times R^+$, $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ if and only if for each μ in R and each σ^2 in R^+,

$$\exp(-\frac{n}{2}(\frac{1}{\sigma^2} - 1)(\bar{x}^2 - \bar{y}^2) + \frac{n}{\sigma^2}\mu(\bar{x} - \bar{y}) - \frac{(n-1)}{2}(\frac{1}{\sigma^2} - 1)(s_x^2 - s_y^2)) = 1\cdot$$

A θ-free solution to the equation is $\bar{x} = \bar{y}$ and $s_x^2 = s_y^2$. Hence $t(x) = (\bar{x}, s_x^2) = (\bar{y}, s_y^2) = t(y)$ and thus $t(x) = (\bar{x}, s_x^2)$ is a minimal sufficient statistic.

3. Consider a random sample from an exponential model with probability density

$$f_\theta(x) = \gamma(\theta) \exp\{t_1(x)\varphi_1(\theta) + \cdots + t_r(x)\varphi_r(\theta)\}h(x)$$

where $1, \varphi_1(\theta), \cdots, \varphi_r(\theta)$ are linearly independent on Ω (exponential family). If $t(x) = (t_1(x), \cdots, t_r(x))$ then

$$g_{\theta|\theta_0}(t(x)) = \frac{\gamma(\theta)}{\gamma(\theta_0)} \exp\{\sum_{i=1}^{r} t_i(x)(\varphi_i(\theta) - \varphi_i(\theta_0))\}$$

For each θ, $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ if and only if for each θ,

$$\exp\{\sum_{i=1}^{r}(t_i(x) - t_i(y))(\varphi_i(\theta) - \varphi_i(\theta_0))\} = 1\cdot$$

A θ-free solution to the equation is $t_i(x) = t_i(y)$ for $i = 1, \cdots, r$. Hence $t(x) = (t_1(x), \cdots, t_r(x)) = (t_1(y), \cdots, t_r(y)) = t(y)$ and thus $t(x) = (t_1(x), \cdots, t_r(x))$ is a minimal sufficient statistic.

4. Consider a random sample from Uniform $(0, \theta)$ distribution, $\theta > 0$. If $\theta_0 = 1$ and $t(x) = x_{(n)} = \max(x_1, \cdots, x_n)$ then $g_{\theta|\theta_0}(t(x)) = \theta^{-n} I_{[x_{(n)}, \infty)}(\theta)$. For each $\theta > 0$, $g_{\theta|\theta_0}(t(x)) = g_{\theta|\theta_0}(t(y))$ if and only if for each $\theta > 0$, $I_{[x_{(n)}, \infty)}(\theta) = I_{[y_{(n)}, \infty)}(\theta)$. A θ-free solution to the equation is $x_{(n)} = y_{(n)}$. Hence $t(x) = t(y)$ and thus $t(x) = x_{(n)}$ is a minimal sufficient statistic.

Acknowledgments

This work was financially supported by Isfahan University of Technology, Isfahan, Iran and the Natural Sciences and Engineering Research Council of Canada.

References

1. Bahadur, R.R. (1954). Sufficiency and statistical decision functions, Ann. Math. Statist., **25**, 423-462.

2. Barndorff-Nielsen, O., Hoffmann-Jorgensen, J., and Pedersen, L. (1976). On the minimal sufficiency of the likelihood function, Scand. J. Statist., **3**, 37-38.

3. Dynkin, E.B. (1951). Necessary and sufficient statistics for a family of probability distributions, English translation in Select. Transl. Math. Statist. Prob., **1**, 1961, 23-41.

4. Fisher, R.A. (1921). On the "probable error" of a coefficient of correlation deduced from a small sample, Metron I, 3-32.

5. Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics, Philos. Trans. Roy. Sor. London Ser., **A 222**, 309-368.

6. Fisher, R.A.(1934). Two new properties of mathematical likelihood, Proc. Roy. Soc. Ser., **A 144**, 285-307.

7. Fisher, R.A. (1956). Statistical Methods and Scientific Inference, Oliver and Boyd, London.

8. Fraser, D.A.S., McDunnough, P., Naderi, A. and Plante, A. (1995). On the definition of probability densities and sufficiency of the likelihood map, Prob. Math. Statist., **15**, 301-310.

9. Fraser, D.A.S., McDunnough, P., Naderi, A., and Plante, A. (1997). From the likelihood map to Euclidean minimal sufficiency, Prob. Math. Statist., **17**, 223-230.

10. Halmos, P.R. and Savage, L.J. (1949). Application of the Radon-Nikodym theorem to the theory of sufficient statistics, Ann. Math. Statist., **20**, 225-241.

11. Landers, D. and Rogge, L. (1972). Minimal sufficient σ-fields and minimal sufficient statistics. Two counterexamples, Ann. Math. Stat., **43**, 2045-2049.

12. Lehmann, E.L. and Scheffé, H.(1950). Completeness, similar regions, and unbiased estimation, Sankhya, **10**, 305-340.

13. McDonnough, P. and Naderi, A. (2001). Likelihood map and minimal sufficiency: The intertwined concepts, Data Analysis from Statistical Foundations. A Festschrift in honor of D.A.S Fraser, Editor: A.K.MD.E. Saleh, Nova Science Publishers Inc., New York, 3-8.

14. Naderi, A. (1988). Measurability and other properties of likelihood map, Ph.D. Thesis, University of Toronto, 1988.

15. Neyman, J. (1935). Sur un teorema concernente le cosidette statistiche sufficienti, Giorn. Ist. Ital. Att., **6**, 320-334.

16. Pitcher, T.S. (1957). Sets of measures not admitting necessary and sufficient statistics or subfields., Ann. Math. Stat., **28**, 267-268.