Ann. Inst. Statist. Math.
32 (1980), Part A, 95-97

NONEXISTENCE OF COMPLETE SUFFICIENT STATISTICS FOR STATIONARY k-STATE MARKOV CHAINS, $k \geqq 3$

A. L. WRIGHT

(Received Mar. 27, 1979)

Abstract

It is shown that there does not exist a complete sufficient statistic for the class of all k-state Markov chains, $k \geqq 3$. When $k=2$ there is a complete sufficient statistic.

1. Introduction

Consider the set of all Markov chains with a finite or countable state space S. Assume that either $S=\{1,\cdots,k\}$, $k \geqq 2$, or else S is the positive integers. The purpose of this note is to point out that there exists a complete sufficient statistic for the set of all *stationary* chains with state space S if and only if $S=\{1,2\}$. We will not try to describe the well-known role of complete sufficient statistics in the theory of uniformly most powerful unbiased tests and in minimum variance unbiased estimation. The books of Lehmann [4] and Zacks [6], among others, contain information and background. The concept completeness of a sufficient statistic is due to Lehmann and Scheffé [5].

2. Results

By definition, the parameter space for the set of Markov chains with state space S is $\Theta=\left\{(p_{ij}):\ p_{ij} \geqq 0,\ \sum_{j \in S} p_{ij}=1 \text{ for each } i \in S\right\}$. Fix integer $n \geqq 2$ and let $(x_1,\cdots,x_n)$ denote a sample from the chain. It turns out that if the initial state x_1 is fixed then there *does* exist a complete sufficient statistic for Θ, as shown by Denny and Yakowitz [3], and Denny and Wright [2]. This implies that there is a complete sufficient statistic for Θ if the distribution of x_1 is known to the statistician, that is, if the initial distribution is the same for all $\theta \in \Theta$.

AMS 1970 subject classifications: Primary 62M05; Secondary 62B05, 62J10.
Key words and phrases: Markov chains, complete sufficient statistics.

By definition, the parameter space for the set of all stationary Markov chains with state space S is $\Theta_S=\{(p_{ij}): (p_{ij})\in\Theta$, there exists at least one stationary initial distribution (p_i) for $(p_{ij})\}$. Recall the definition of the transition count statistic $\{f_{ij}: (i, j)\in S\times S\}$: f_{ij} is the number of m, $1\leqq m\leqq n-1$, for which $x_m=i$ and $x_{m+1}=j$ (see, for example, Billingsley [1]).

Suppose $S=\{1, 2\}$. It is mentioned on p. 338 of Denny and Wright [2] that the following function is a complete sufficient statistic for Θ_S:

$$(x_1, \cdots, x_n)\to(x_1, f_{11}, f_{22}, 0) \qquad \text{if } f_{12}=f_{21},$$

and

$$(x_1, \cdots, x_n)\to(0, f_{11}, f_{22}, f_{12}+f_{21}) \qquad \text{if } f_{12}\neq f_{21}.$$

THEOREM. *If $S\neq\{1, 2\}$, $(x_1, \{f_{ij}\})$ is a minimal sufficient statistic for Θ_S which is not complete. Consequently there does not exist a complete sufficient statistic for Θ_S.*

PROOF. To simplify notation we assume that S is finite. The factorization theorem shows sufficiency: $\mathrm{P}_\theta(x_1, \cdots, x_n)=p_{x_1}\prod_{ij} p_{ij}{}^{f_{ij}}h(x_1, \cdots, x_n)$. To prove minimality we use Theorem 6.3 of Lehmann and Scheffé [5]. Let $(x_1, \cdots, x_n)$ and $(y_1, \cdots, y_n)$ denote two samples and let $\{f_{ij}\}$ and $\{g_{ij}\}$ denote their respective transition counts. By the above mentioned theorem it is enough to prove: if $(x_1, \{f_{ij}\})\neq(y_1, \{g_{ij}\})$ then there exist $\theta_i\in\Theta_S$, $i=1, 2$, so that

$$\frac{\mathrm{P}_{\theta_1}(x_1, \cdots, x_n)}{\mathrm{P}_{\theta_1}(y_1, \cdots, y_n)}\neq\frac{\mathrm{P}_{\theta_2}(x_1, \cdots, x_n)}{\mathrm{P}_{\theta_2}(y_1, \cdots, y_n)}. \tag{1}$$

Assume first that $\{f_{ij}\}=\{g_{ij}\}$ and $x_1\neq y_1$. If θ_1 is doubly stochastic then the left side of (1) is one, and since the right side of (1) is p_{x_1}/p_{y_1}, there is clearly θ_2 so that (1) holds. Assume now that $\{f_{ij}\}\neq\{g_{ij}\}$, and with no loss of generality assume $f_{12}>g_{12}$. For each $\varepsilon>0$ we define doubly stochastic $\theta_1(\varepsilon)$ as follows: $p_{12}(\varepsilon)=\varepsilon$, $p_{ij}(\varepsilon)=(k-2+\varepsilon)/(k-1)^2$ for $2\leqq i\leqq k$ and $j\neq 2$, and $p_{ij}(\varepsilon)=(1-\varepsilon)/(k-1)$ otherwise. As ε tends to zero the left side of (1) tends to zero, so there is θ_2 so that (1) holds. This proves minimality. As pointed out in Billingsley [1], x_n is a function of $(x_1, \{f_{ij}\})$. Let F be the function of the sufficient statistic defined by $F(x_1, \{f_{ij}\})=x_1-x_n$. To see that $(x_1, \{f_{ij}\})$ is not complete, we use the stationarity assumption and conclude that $\mathrm{E}_\theta(F)=0$, $\theta\in\Theta_S$. The last statement of the theorem is a consequence of the preceding facts, and the theorem is proved.

Sometimes incomplete sufficient statistics occur because the parameter space is "restricted." The above proof shows that $(x_1, \{f_{ij}\})$ will not be a complete sufficient statistic for the class of all distributions for which it is sufficient and for which x_1 and x_n have the same mean.

UNIVERSITY OF ARIZONA

REFERENCES

[1] Billingsley, P. (1961). Statistical methods in Markov chains, *Ann. Math. Statist.*, **32**, 12-40.
[2] Denny, J. L. and Wright, A. L. (1978). On tests for Markov dependence, *Zeit. Wahrscheinlichkeitsth.*, **43**, 331-338.
[3] Denny, J. L. and Yakowitz, S. J. (1978). Admissible run-contingency type tests for independence and Markov dependence, *J. Amer. Statist. Ass.*, **73**, 177-181.
[4] Lehmann, E. L. (1959). *Testing Statistical Hypotheses*, Wiley, New York.
[5] Lehmann, E. L. and Scheffé, H. (1950, 1955). Completeness, similar regions, and unbiased estimation, *Sankhyā*, **10**, 305-340 and *Sankhyā*, **15**, 219-236.
[6] Zacks, S. (1971). *The Theory of Statistical Inference*, Wiley, New York.