9-1993

A Note on Distributed Estimation and Sufficiency

R. Viswanathan
Southern Illinois University Carbondale, viswa@engr.siu.edu

Published in Viswanathan, R. (1993). A note on distributed estimation and sufficiency. IEEE Transactions on Information Theory, 39(5), 1765 - 1767. doi: 10.1109/18.259674

so the BCH code is better than the RS code. Note however that the increase in the number of potential users per slot does have a price: The increase in the code length $n' = w$ (and, therefore, in the frame length N and in the number $\sigma = w/2$ of successful packets per slot) increases the decoding complexity per slot, since the erasure-decoding algorithm of the other shortened RS code is super linear in w (it is quadratic in w for most standard erasure-decoding algorithms).

Example: Consider the numerical Example 5 in [1]. Take $p = 13$, $k = 4$. The following table illustrates the parameters of the protocol sequence sets for RS and BCH codes, resp.:

	RS ($r = 1$)	BCH ($r = 2$)
T	169	28561
M	3	3
N	156	2184
σ	6	84
R_{sum}	3/26	3/26
T/N	1.08	13.08

References

[1] N. Q. A., L. Györfi, J. L. Massey, "Constructions of binary constant-weight cyclic codes and cyclically permutable codes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 940–948, May 1992.

[2] E. N. Gilbert, "Cyclically permutable error-correcting codes," *IEEE Trans. Inform. Theory*, vol. 9, pp. 175–182, July 1963.

[3] R. E. Blahut, *Theory and Practice of Error Control Codes.* Reading, MA: Addison-Wesley, 1984.

[4] N. Zierler, "Linear recurring sequences," *J. SIAM*, vol. 7, no. 1, pp. 31–48, Mar. 1959.

[5] J. L. Massey, "The capacity of the collision channel without feedback," in *Abstracts of Papers, IEEE Int. Symp. Inform. Theory*, 1982, p. 101.

[6] J. L. Massey and P. Mathys, "The collision channel without feedback," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 192–204, Mar. 1985.

[7] B. S. Tsybakov and N. B. Likhanov, "Packet communication on a channel without feedback," *Probl. Inform. Transm.*, (in Russian), vol. XIX, no. 2, pp. 64–84, 1983.

[8] L. A. Bassalygo and M. S. Pinsker "Limited multiple-access of a nonsynchronous channel," *Probl. Inform. Transm.* (in Russian), vol. XIX, no. 4, pp. 92–96, 1983.

A Note on Distributed Estimation and Sufficiency

R. Viswanathan

***Abstract*—In relation to distributed parameter estimation, the notion of local and global sufficient statistics is introduced. It is shown that when a sufficiency condition is satisfied by the probability distribution of a random sample, a global sufficient statistic is obtainable as a function of local sufficient statistics. Several standard distributions satisfy the said sufficiency condition.**

***Index Terms*—Estimation, sufficiency, fusion of estimates.**

Manuscript received April 15, 1992; revised December 10, 1992. This work was supported by ONR under Grant N00014-93-1-1092.

The author is with the Department of Electrical Engineering, Southern Illinois University at Carbondale, Carbondale, IL 62901-6603.

IEEE Log Number 9210707.

I. Introduction

In distributed estimation, inferences about a parameter are to be made based on partial information. A typical situation is the following. Several partial (local) data sets are available and separate inferences based on local data sets, such as local estimates or sufficient statistics pertaining to local data, are used to obtain an overall assessment of the parameter. If all the local data are available together, then the problem is one of classical estimation [1]. In distributed detection and estimation, some interesting and counter intuitive results are possible [2], [6]. Here, only those probability distributions that admit sufficient statistics are considered [1]. The utility of the sufficient statistic is that it is of reduced dimension as compared to the dimension of the data and that it achieves this reduction without any loss of information, because it carries all the relevant information that the data has, regarding the parameter. (The whole data is trivially sufficient, but this has no dimensionality reduction. Hence, it is assumed that those distributions that admit only the trivial sufficient statistic do not possess any sufficient statistic). In this discussion it is assumed that conditioned on the parameter, the data samples are statistically independent. In the next section we have some preliminaries that define the terminologies. In Section III, we pose the question: Given several local sufficient statistics and a global sufficient statistic pertaining to the whole data, does a function of local sufficient statistics exist such that this function is the global sufficient statistic? A sufficient condition on the probability distribution assures the existence of such a function. Also, several standard distributions are shown to satisfy this condition.

II. Preliminaries

Consider the problem of estimating a parameter Θ using the observations $Z_1, Z_2, \cdots, Z_N$. In the context of distributed processing, the whole data can be called global data and any proper subset of the global set local data. Whenever several local data sets are considered, we assume them to be mutually exclusive and collectively exhaustive. Hence, the conditional distribution of the global data given the parameter Θ and the prior distribution of Θ provide a complete characterization of the estimation problem. Any sufficient statistic [1] that pertains to the whole data will be called a global sufficient statistic. A sufficient statistic that pertains to local data will be called a local sufficient statistic. One could similarly define local and global likelihood functions, local estimates, and global estimates. The definition of local and global sufficiency given here is different from the one used by Barankin and Katz [7]. In [7], the variation of the dimensionality of a sufficient statistic as the sample $(z_1, z_2, \cdots, z_N)$ ranges over Euclidean N-space, leads to the definition of local and global sufficient statistics.

III. Local and Global Sufficiency

Using the terminology of distributed sensor processing [2], [3], let us consider a group of n sensors, with the ith sensor receiving observations $\{Z_{i1}, Z_{i2}, \cdots, Z_{in_i}\}$, for $i = 1, 2, \cdots, n$. Let G denote the global data set $\{Z_{ij}, i = 1, 2, \cdots, n, j = 1, 2, \cdots, n_i\}$ and let each Z_{ij} be independent and identically distributed with either a probability mass function $f(z \mid \Theta)$, when the observations are discrete, or with a probability density function $f(z \mid \Theta)$, when the observations are continuous. Here Θ denotes a one-dimensional parameter defined on an appropriate parametric space. For the sake of notational convenience, $f(\cdot)$ is used to denote both the marginal

0018-9448/93$03.00 © 1993 IEEE

and the joint distributions. From the context it will be clear as to which distribution is being referred.

Lemma 1: If $f(z \mid \Theta)$ is such that $\prod_{k=1}^{n_i} f(z_{ik} \mid \Theta)$, for any positive integer n_i, possesses a single dimensional sufficient statistic Y_i, then $\{Y_1, Y_2, \cdots, Y_n\}$ are jointly sufficient for Θ, corresponding to the global function

$$f(G \mid \Theta) = \prod_{i=1}^{n} \prod_{k=1}^{n_i} f(z_{ik} \mid \Theta). \quad (1)$$

The proof is obtained from the following observation. Since local sufficient statistics $\{Y_i\}$ exist by assumption, the application of the factorization theorem [1], [4] implies that $f(G \mid \Theta)$ is a product of a function independent of Θ, and a function depending on $\{Y_i, i = 1, 2, \cdots, n\}$ and Θ. Hence, $\{Y_i\}$ are jointly sufficient for Θ.

Remark: The assumption of the existence of a one-dimensional sufficient statistic for any sample size implies the existence of a sufficient statistic S, for Θ, pertaining to the global data G. Since the function $f(G \mid \{Y_i\})$ is independent of Θ, given $\{Y_i, i = 1, 2, \cdots, n\}$, an equivalent data set $\{\hat{Z}_{ij}\}$, equivalent to G, and hence an equivalent global sufficient statistic $\hat{S}$ can be obtained. $\hat{S}$ is in general not same as S, for any given global data G, but the distributions of $\hat{S}$ and S are equal a.e. Though an equivalent $\hat{S}$ can be obtained from the $\{Y_i\}$, it would be nice if S itself can be obtained from the $\{Y_i\}$. As shown by the next lemma, with an additional restriction on the nature of factorization admitted by $f(z \mid \Theta)$, there exists a function $q(\cdot)$ such that S equals $q(Y_1, Y_2, \cdots, Y_n)$.

Lemma 2: Let $f(z \mid \Theta)$ be factorizable as

$$f(z \mid \Theta) = g(z; \Theta)h(z)m(\Theta) \quad (2)$$

such that $h(z) > 0$,

$$g(z_1; \Theta)g(z_2; \Theta) = g(\psi(z_1, z_2); \Theta)k(z_1, z_2). \quad (3)$$

Then there exist local sufficient statistics $\{Y_i, i = 1, 2, \cdots, n\}$ and a function $q(\cdot)$ such that a global sufficient statistic S equals $q(Y_1, Y_2, \cdots, Y_n)$.

Proof: First let us observe the type of factorization. The same $g(\cdot)$ function appears on both sides of (3). Using (2) and (3), it is observed that

$$f(z_1, z_2 \mid \Theta) = g(\psi(z_1, z_2); \Theta)k(z_1, z_2)h(z_1)h(z_2)m^2(\Theta). \quad (4)$$

Hence, a one-dimensional sufficient statistic $\psi(z_1, z_2)$ is obtained from z_1 and z_2. Notice that ψ must be a symmetric function of its variables. Similarly, consider

$$f(z_1, z_2, z_3 \mid \Theta) = g(\psi(z_1, z_2); \Theta)g(z_3; \Theta) \cdot k(z_1, z_2)\prod_{i=1}^{3} h(z_i)m^3(\Theta). \quad (5)$$

Application of (3) yields a sufficient statistic pertaining to $\{z_1, z_2, z_3\}$ as $\psi(\psi(z_1, z_2), z_3)$. By induction to any sample size, we observe that there exists a local sufficient statistic Y_i, as a function of $\{Z_{ij}, j = 1, 2, \cdots, n_i\}$, for every i, and that there exists a function $q(\cdot)$ such that a global sufficient statistic S equals $q(Y_1, Y_2, \cdots, Y_n)$. □

Remark 1: Several standard distributions satisfy the factorization condition of lemma 2.

Example 1: The distribution of Z_{ij} belongs to an exponential family [1]. In this case, $f(z \mid \Theta) = a(\Theta)b(z)\exp(c(\Theta)d(z))$; $g(z; \Theta) = \exp(c(\Theta)d(z))$. Clearly $g(\cdot)$ satisfies (3). A local sufficient statistic Y_i is of the form $\sum_j d(Z_{ij})$ and a global sufficient statistic S is $\sum_i Y_i$.

Example 2: Consider Z_{ij} to be i.i.d. uniform on $(0, \Theta)$. $g(z; \Theta) = I_{(0,\Theta)}(z)$; $\psi(z_1, z_2) = \max(z_1, z_2)$; $k(z_1, z_2) = I_{(0, \max\{z_1, z_2\})}(\min\{z_1, z_2\})$. $Y_i = \max\{Z_{ij}, \text{ all } j\}$ and $S = \max\{Y_i, \text{ all } i\}$.

Example 3: Consider the density of the type [5]:

$$f(z \mid \Theta) = h(z)m(\Theta)I_{(a(\Theta), b(\Theta))}(z), \quad (6)$$

where $a(\Theta)$ and $b(\Theta)$ are monotonic functions in the opposite senses. For example, if $a(\Theta)$ is monotone decreasing and $b(\Theta)$ is monotone increasing, then a sufficient statistic, for all samples of sizes ≥ 2 is given by

$$t_N = \max\{a_1^{-1}(z_{(1)}), b_1^{-1}(z_{(N)})\}, \quad (7)$$

where $z_{(1)}$ and $z_{(N)}$ are the minimum and the maximum order statistics of the data $\{z_i, i = 1, 2, \cdots, N\}$, and $a_1^{-1}(\alpha)$ and $b_1^{-1}(\alpha)$ are the smallest roots of $a(x) = \alpha$ and $b(x) = \alpha$, respectively.

In this case a nontrivial sufficient statistic exists only for a sample size of greater than or equal to 2. The density of (6) does not satisfy the condition (3) of Lemma 2. However, a basic density formed with a sample size of 2 or greater satisfies (2) and (3). For example,

$$f(z_1, z_2 \mid \Theta) = h(z_1)h(z_2)m^2(\Theta)I_{(-\infty, \Theta)}(t_2), \quad (8)$$

where t_2 is the sufficient statistic given in (7) (with $N = 2$), satisfies (2) and (3). In this case,

$$S = \max\{Y_i, \text{ all } i\}, \quad \text{where } Y_i \text{ are the local sufficient statistics.}$$

Remark 2: Lemmas 1 and 2 can be extended to the case of multidimensional sufficient statistics.

Remark 3: Let us consider the Theorem 1 of [6]. In our notation, given Θ, Z_i's are independent Gaussian with mean Θ and variance σ_i^2. The variances are assumed known. Although Lemma 2 is in general stated for i.i.d. distributions, it happens to hold for the case of independent but nonidentical Gaussian with mean as the parameter (with a slight modification, $g(z_i/(\sigma_i^2); \Theta)$ replacing $g(z_i; \Theta)$ in (2)). A sufficient statistic is $S = \sum_{i=1}^{n} Z_i/(\sigma_i^2)$ and this equals the sum of local sufficient statistics. In [6], the authors consider the problem of whether the global minimum mean-square-error (mmse) estimate is obtainable as a function of local mmse estimates. In this regard, it is worth observing the following. Both the local mmse estimate and local sufficient statistic are real numbers. Therefore, from the communication constraint viewpoint of distributed estimation, there is no difference between transmitting an estimate and a sufficient statistic from a local site. When lemma 2 is satisfied, one can obtain a global sufficient statistic from the local sufficient statistics and hence obtain a global mmse estimate. In general, a Bayes' estimator is a function of a sufficient statistic, but the function need not be one to one. Under the conditions of Theorem 1 of [6], the authors have shown that the conditional mean estimate $E(\Theta \mid \text{all } Z_i)$ is a one to one function of S and hence, the estimate itself is sufficient. Since a Bayes' estimator is not always a sufficient statistic, it is preferable to "fuse" local sufficient statistics to obtain first a global sufficient statistic, when Lemma 2 is satisfied. When no sufficient statistic exists, one has no choice but to obtain local estimates and then fuse them in a best possible way.

IV. Conclusion

It is shown that a global sufficient statistic is obtainable as a function of local sufficient statistics, when a sufficiency condition is satisfied by the probability distribution of a random sample.

Acknowledgment

The author would like to thank Prof. S. Jeyaratnam and Prof. S. Panchapakesan for useful discussions.

References

[1] E. L. Lehmann, *Theory of Point Estimation.* New York: John Wiley, 1983.

[2] J. N. Tsitsiklis, "Decentralized detection," in *Advances in statistical signal processing* (vol. 2, Signal detection), H. V. Poor and J. B. Thomas, Eds. Greenwhich, CT: JAI Press, 1990.

[3] R. R. Tenny and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace and Electron. Syst.*, vol. 17, pp. 501–510, July 1981.

[4] T. S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach.* New York: Academic Press, 1967.

[5] V. S. Huzurbazar, *Sufficient Statistics.* New York: Marcel Dekker, 1976.

[6] E. B. Hall, A. E. Wessel, and G. L. Wise, "Some aspects of fusion in estimation theory," *IEEE Trans. Inform. Theory*, vol. 37, pp. 420–422, Mar. 1991.

[7] E. W. Barankin and M. Katz, "Sufficient statistics of minimal dimension," *Sankhya*, vol. 21, pp. 217–246, 1959.