Rose-Hulman Institute of Technology

Rose-Hulman Scholar

Mathematical Sciences Technical Reports (MSTR)

Mathematics

9-1990

Sets of Typical Subsamples

Joel Atkins
Rose-Hulman Institute of Technology

G.J Sherman
Rose-Hulman Institute of Technology

Follow this and additional works at: https://scholar.rose-hulman.edu/math_mstr

Part of the Algebra Commons, Applied Mathematics Commons, and the Statistical Methodology Commons

SETS OF TYPICAL SUBSAMPLES

Joel E. Atkins and Gary J. Sherman

MS TR 90-04

September 1990

Department of Mathematics
Rose-Hulman Institute of Technology
Terre Haute, IN 47803

FAX(812) 877-3198   Phone: (812) 877-8391

SETS OF TYPICAL SUBSAMPLES

JOEL E. ATKINS and GARY J. SHERMAN*

Department of Mathematics, Rose-Hulman Institute of Technology, Terre Haute, IN 47803

Abstract: A group theoretic condition on a set of subsamples of a random sample from a continuous random variable symmetric about 0 is shown to be sufficient to provide typical values for 0.

Key words: typical values, group, permutation.

1. Introduction

Hartigan's "typical value theorem" (1969) is the basis for random subsampling, a resampling plan which uses group theory to construct confidence intervals for the center of a symmetric distribution on the real line. Here are notation and terminology required to state the theorem. Let $X_{(1)}, X_{(2)}, \ldots X_{(N)}$ denote the ordered values $(X_{(1)} \leq X_{(2)}, \leq \ldots \leq X_{(N)})$ of a set of random variables $X = \{X_1, X_2, \ldots, X_N\}$. If it is equally probable that the parameter θ is in any one of the intervals $(-\infty, X_{(1)}), (X_{(1)}, X_{(2)}), \ldots (X_{(N)}, \infty)$, then the X_i's are called **typical values** for θ. Let $\mathcal{P}(Y)$ denote the set of all subsamples of a set of random variables $Y = \{Y_1, Y_2, \ldots, Y_n\}$. $\mathcal{P}(Y)$ is an abelian group with respect to symmetric difference ($g \circ h = g \cup h - g \cap h$ for $g, h \in \mathcal{P}(Y)$). For $g \in \mathcal{P}(Y)$, we denote $\sum_{Y_i \in g} Y_i$ by X_g. If $g = \emptyset$, then $X_\emptyset = 0$ by convention.

---

* Will handle correspondence concerning the manuscript.

1

Hartigan's Typical Value Theorem. Let Y be a set of independent, continuous random variables symmetric about 0 and let G be a set of subsamples of Y which includes the empty subsample. Then $\{X_g | g \in G, g \neq \emptyset\}$ is a set of typical values for 0, for all Y, if, and only if, G is a subgroup of $\mathcal{P}(Y)$.

When reference is made to Hartigan's theorem it is usually in the context of random samples and only the sufficiency of the group theoretic condition is used (e.g. Efron, 1982, p. 70). Does Hartigan's theorem hold if Y is a random sample? No, consider the following example. Let $Y = \{Y_1, Y_2, Y_3\}$ where the Y_i's are iid continuous random variables symmetric about 0 and let $G = \{\emptyset, \{Y_1, Y_2\}, \{Y_1, Y_3\}\}$. G is not a group, because $\{Y_1, Y_2\} \circ \{Y_1, Y_3\} = \{Y_2, Y_3\} \notin G$, and yet $X_1 = Y_1 + Y_2$ and $X_2 = Y_1 + Y_3$ are typical values for 0. To see this consider the eight possibilities for the signs of Y_1, Y_2 and Y_3

$SGN(Y_1)$	+	+	+	+	$-$	$-$	$-$	$-$
$SGN(Y_2)$	+	+	$-$	$-$	+	+	$-$	$-$
$SGN(Y_3)$	+	$-$	+	$-$	+	$-$	+	$-$
$Prob(X_1, X_2 < 0)$	0	0	0	1/3	1/3	1/2	1/2	1

All eight possibilities occur with equal probability because the distributions of the Y_i's are symmetric. The probability that 0 occurs in the interval $(X_{(2)}, \infty)$, which is the probability that both X_1 and X_2 are negative, is $(1/8)(0 + 1/3 + 1/3 + 1/2 + 1/2 + 1) = 1/3$. A similar argument shows the probability that 0 occurs in the interval $(-\infty, X_{(1)})$ is also 1/3.

Can those sets of subsamples which produce typical values for a continuous random variable symmetric about 0 be characterized? The purpose of this paper is to provide a sufficient condition which encompasses Hartigan's group theoretic condition, explains the previous example, and suggests a promising research problem.

2

2. A theorem

Let the symmetric group on n symbols, S_n, act on $\mathcal{P}(Y)$ by defining

$$\sigma(g) = \begin{cases} \{Y_{\sigma(i)} | Y_i \in g\} & , \quad \text{if} \quad g \neq \emptyset \\ \emptyset & , \quad \text{if} \quad g = \emptyset \end{cases}$$

for each $\sigma \in S_n$.

Theorem. *Let Y be a set of n iid continuous random variables symmetric about 0 and let G be a set of subsamples of Y which includes the empty subsample. If for each $h \in G$ there exists $\sigma \in S_n$ such that $G \circ h = \{g \circ h | g \in G\} = \{\sigma(g) | g \in G\} = \sigma(G)$, then $\{X_g | g \in G, g \neq \emptyset\}$ is a set of typical values for 0.*

Proof: Notice that for each $h \in G$, $\{X_g - X_h | g \in G\}$ has the same distribution as $\{X_{g \circ h} | g \in G\}$ due to the symmetry of the Y_i's. By hypothesis there exists $\sigma \in S_n$ such that $\{X_{g \circ h} | g \in G\} = \{X_{\sigma(g)} | g \in G\}$. Since the Y_i's are iid and σ is a permutation $\{X_{\sigma(g)} | g \in G\}$ has the same distribution as $\{X_g | g \in G\}$. Thus the probability that X_h is the $(k+1)$th order statistic equals the probability that exactly k of the X_g's are negative. Since this observation is independent of h, we have that each X_h is the $(k+1)$th order statistic with equal probability. In particular, this is true for $X_\emptyset = 0$. Thus $\{X_g | g \in G, g \neq \emptyset\}$ is a set of typical values for 0.

3. Discussion

The sufficiency of Hartigan's group theoretic condition for random samples follows immediately from the Theorem: If G is a subgroup, then $G \circ h = G$ for each $h \in G$; i.e., we may choose σ to be the identity permutations for each $h \in G$. The example preceding the theorem satisfies our condition. Moreover, the construction in this example generalizes: If $G = \{\emptyset, \{Y_1, Y_2\}, \{Y_1, Y_3\}, \ldots, \{Y_1, Y_n\}\}$, then $G \circ \{Y_1, Y_j\} = \sigma(G)$ where σ is the transposition $(1, j)$. Indeed we have not been able to find a set of subsamples which produces typical values that does not

3

satisfy our condition.

Conjecture. *If G and Y are as in the Theorem, then $\{X_g | g \in G, \ g \neq \emptyset\}$ is a set of typical values for 0 if, and only if, G is a **set of typical subsamples**; i.e., for each $h \in G$ there exists $\sigma \in S_n$ such that $G \circ h = \sigma(G)$.*

Regardless of the resolution of this conjecture, the Theorem establishes the importance of sets of typical subsamples. We encourage the study of properties of such sets. This is essentially a combinatorial problem as can be seen by making use of a natural isomorphism from $\mathcal{P}(Y)$, under symmetric difference, to Z_2^n, the set of binary n-tuples under component-wise modulo 2 addition. The correspondence is defined by assigning $g \in \mathcal{P}(Y)$ to $b \in Z_2^n$ where the ith component of b is one if, and only if, $Y_i \in g$. The following table illustrates, in the context of Z_2^n, the ideas we have been discussing.

G		$G \circ h$	$\sigma(G)$
in $\mathcal{P}(Y)$	in Z_2^4	$h = 1000$	$\sigma = (2, 4)$
$\emptyset$	0000	1000	0000
$\{Y_1\}$	1000	0000	1000
$\{Y_1, Y_2\}$	1100	0100	1001
$\{Y_1, Y_2, Y_3\}$	1110	0110	1011
$\{Y_1, Y_2, Y_3, Y_4\}$	1111	0111	1111
$\{Y_2, Y_3, Y_4\}$	0111	1111	0111
$\{Y_3, Y_4\}$	0011	1011	0110
$\{Y_4\}$	0001	1001	0100

Here $n = 4$ and $G \circ h = \sigma(G)$ for $h = 1000$ and $\sigma = (2, 4)$. It is easy to verify that G is in fact a set of typical subsamples. This illustrates that *finding a set of typical*

subsamples is equivalent to finding a subset, G, of Z_2^n such that each translate of G by an element of G is in the orbit of G under the action of S_n.

We close by pointing out that if $G \subseteq Z_2^n$ is a set of typical subsamples because it is a subgroup, then G is an algebraic code (e.g. see MacWilliams and Sloan, 1977). Thus the study of sets of typical subsamples can be viewed as a generalization of algebraic coding theory.

References

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans*, (SIAM).

Hartigan, J.A. (1969), Using subsample values as typical values, Amer. Stat. Assoc. J. **64**, 1303-1317.

MacWilliams, F.J. and Sloane, N.J.A.(1977) *The Theory of Error-Correcting Codes*, (North Holland).