

A subsufficient algebra related to the structure of UMVUEs

Abram Kagan

Dept. of Mathematics
University of Maryland, College Park

Statistics Seminar
September 22, 2011

- Let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a family of probability distributions on a measurable space $(\mathcal{X}, \mathcal{A})$, $\mathcal{X} = \{x\}$, \mathcal{A} - a σ -algebra of subsets of \mathcal{X} , parameterized by a general parameter θ . The notation $X \sim P_\theta$ means that an observation $X \in \mathcal{X}$ is distributed according to P_θ .
- The talk introduces and studies a new statistical structure, a σ -subalgebra (in what follows σ - before algebra and subalgebra will be omitted) $\mathcal{B} \subseteq \mathcal{A}$ defined by the following property. If $h(X)$ is a statistic with $E_\theta(h(X)) = 0$, $E_\theta(|h(X)|^2) < \infty$, $\forall \theta \in \Theta$ (a so called unbiased estimator of zero with finite variance), then

$$E_\theta(h(X)|\mathcal{B}) = 0, \forall \theta \in \Theta. \quad (1)$$

If the class of unbiased estimators of zero consists only of $h(X) \equiv 0$, then $\mathcal{B} = \mathcal{A}$ satisfies (1). Plainly, in this case any subalgebra of \mathcal{A} satisfies (1).

Sufficiency and subsufficiency

- Let now $\tilde{\mathcal{A}}$ be a sufficient algebra for θ . Take a \mathcal{B} -measurable statistic $\gamma(X)$ and set $\tilde{\gamma}(X) = E_{\theta}(\gamma(X)|\tilde{\mathcal{A}})$. Plainly,

$$\text{var}_{\theta}(\gamma(X)) \leq \text{var}_{\theta}(\tilde{\gamma}(X)), \quad (2)$$

the equality sign holding iff $\gamma(X) = \tilde{\gamma}(X)$ with P_{θ} -probability one. Since $\tilde{\gamma}(X)$ is a statistic, $h(X) = \gamma(X) - \tilde{\gamma}(X)$ is an unbiased estimator of zero and if \mathcal{B} satisfies (1), then

$$E_{\theta}[\gamma(X) - \tilde{\gamma}(X)|\mathcal{B}] = \gamma(X) - E_{\theta}(\tilde{\gamma}(X)|\mathcal{B}) = 0$$

whence

$$\text{var}_{\theta}(\gamma(X)) \geq \text{var}_{\theta}(\tilde{\gamma}(X)). \quad (3)$$

From (2) and (3), $\gamma(X) = \tilde{\gamma}(X)$ with P_{θ} -probability one.

- Since this holds for any \mathcal{B} -measurable $\gamma(X)$, \mathcal{B} is a subalgebra of any sufficient algebra and hence of the minimal sufficient algebra when the latter exists. In general, for an arbitrary statistic $\gamma(X)$ (not an unbiased estimator of zero), $E_\theta[\gamma(X)|\mathcal{B}]$ is not a statistic so that \mathcal{B} is a proper subalgebra of the minimal sufficient algebra. We call an algebra \mathcal{B} satisfying (1) *subsufficient*.
- The following simple lemma is well known.

Lemma

A statistic $\gamma(X)$ with $E_\theta(|\gamma(X)|^2) < \infty, \forall \theta \in \Theta$ is a UMVUE (of $g(\theta) = E_\theta(\gamma(X))$) if it is uncorrelated with any unbiased estimator of zero $h(X)$ with $E_\theta(|h(X)|^2) < \infty, \forall \theta \in \Theta$,

$$\text{cov}_\theta(\gamma(X), h(X)) = 0.$$

- Due to Lemma 1, for a subsufficient \mathcal{B} any \mathcal{B} -measurable statistic is a UMVUE. This is what makes subsufficiency a topic of interest, at least for the foundations of statistical inference. Suppose that a subsufficient \mathcal{B} is such that for every estimable parametric function $g(\theta)$ (i. e., such that there exists a statistic $\gamma(X)$ with $E_\theta(\gamma(X)) = g(\theta)$, $E_\theta(|\gamma(X)|^2) < \infty$) there exists a \mathcal{B} -measurable $\tilde{\gamma}(X)$ with $E_\theta(\tilde{\gamma}(X)) = g(\theta)$, $E_\theta(|\tilde{\gamma}(X)|^2) < \infty$. Then \mathcal{B} is a complete sufficient subalgebra.
- Indeed, for any statistic $\gamma(X)$ with finite variance there exists such a \mathcal{B} -measurable statistic $\tilde{\gamma}(X)$ also with finite variance that $h(X) = \gamma(X) - \tilde{\gamma}(X)$ is an unbiased estimator of zero and, due to subsufficiency of \mathcal{B} ,

$$E_\theta[\gamma(X) - \tilde{\gamma}(X)|\mathcal{B}] = E_\theta[\gamma(X)|\mathcal{B}] - \tilde{\gamma}(X) = 0.$$

Thus, \mathcal{B} is sufficient. The completeness is trivial.

Sufficiency vs subsufficiency

- Note in passing that if a subalgebra \mathcal{C} is such that every estimable parametric function is also estimable by a \mathcal{C} -measurable statistic, this does not imply sufficiency of \mathcal{C} . Thus, the assumed subsufficiency is essential in the previous paragraph. For a general family \mathcal{P} , some (estimable) parametric functions admit UMVUEs while the others do not.
- In the next section it is shown that even when every UMVUE is \mathcal{B} -measurable, \mathcal{B} is still a proper subalgebra of the minimal sufficient algebra. The structure of \mathcal{B} seems more complicated than that of the minimal sufficient algebra. Namely, let $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ be a dominated family, $P_\theta \ll \mu$ with $dP_\theta/d\mu = p(x; \theta)$. If $\tilde{\mathcal{A}}$ is the minimal sufficient algebra, then $x_1, x_2 \in \mathcal{X}$ simultaneously belong or not to any set $A \in \tilde{\mathcal{A}}$ iff

$$p(x_1; \theta) = c(x_1, x_2)p(x_2; \theta), \forall \theta \in \Theta \quad (4)$$

for some $c(x_1, x_2)$ (in this case, x_1 and x_2 belong to the same element of the minimal sufficient partition of \mathcal{X}). In other words, if T is the minimal sufficient statistic, $T(x_1) = T(x_2)$ iff (4) holds.

Subsufficiency for categorical data

- Thus, in order to decide if x_1, x_2 belong to the same element of the minimal partition, one needs to deal only with $p(x_1; \theta)$ and $p(x_2; \theta)$. This is not the case with subsufficient \mathcal{B} .
- Turning to categorical data, let X be a categorical random variable taking values in $\mathcal{X} = x_1, x_2, \dots, x_N$ with probabilities

$$P_\theta(X = x_i) = p_i(\theta), \theta \in \Theta.$$

Plainly, a parametric function $g(\theta)$ is estimable iff $g(\theta) \in \text{span}\{p_1(\theta), \dots, p_N(\theta)\}$.

- It is convenient to treat functions on Θ as elements of a linear space $V = \text{span}\{v_1, \dots, v_N\}$. Its null element is denoted $\mathbf{0}$. One may always assume that $\mathbf{0} \notin \{v_1, \dots, v_N\}$. Set $p_i(\theta) = v_i$. If

$$\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots \cup \mathcal{X}_m$$

is the minimal sufficient partition of \mathcal{X} , then x_i, x_j belong to the same element of the partition iff

$$p_i(\theta) = c_{ij} p_j(\theta), \theta \in \Theta$$

or, equivalently, if v_i and v_j are collinear, $v_i \parallel v_j$.

- If

$$p_1(\theta) = \theta/2, p_2(\theta) = \theta/3, p_3(\theta) = \theta^2, p_4 = 1 - 5\theta/6 - \theta^2, \theta < 2/3,$$

the minimal sufficient partition is

$$\mathcal{X}_1 = \{x_1, x_2\}, \mathcal{X}_2 = \{x_3\}, \mathcal{X}_4 = \{x_4\}.$$

For

$$p_1(\theta) = \theta/2, p_2(\theta) = \theta/6, p_3(\theta) = \theta/6 + \theta^2/2, p_4(\theta) = \theta^2/2, p_5(\theta) = 1 - 5\theta/6 - \theta^2, \theta \in (0$$

the minimal sufficient partition is

$$\mathcal{X}_1 = \{x_1, x_2\}, \mathcal{X}_2 = \{x_3\}, \mathcal{X}_3 = \{x_4\}, \mathcal{X}_4 = \{x_5\}.$$

It will be shown that in the first example the subsufficient partition coincides with the minimal sufficient but in the second they are different.

- The key to the construction the algebra \mathcal{B} lies in a special partition of V . Subsets V_1, \dots, V_m of V are called linearly independent if the subspaces $\text{span}\{V_1\}, \dots, \text{span}\{V_m\}$ are linearly independent, i. e., the relation

$$u_1 + \dots + u_m = \mathbf{0}$$

holding for $u_1 \in \text{span}\{V_1\}, \dots, u_m \in \text{span}\{V_m\}$ implies $u_1 = \mathbf{0}, \dots, u_m = \mathbf{0}$.

Lemma

There exists a partition

$$V_1 \cup \dots \cup V_m \quad (5)$$

of the set $\{v_1, \dots, v_N\}$ such that (i) V_1, \dots, V_m are linearly independent, (ii) (5) is irreducible, i. e., if for some $j \in \{1, \dots, m\}$

$$V_j = V_{j1} \cup V_{j2}, \quad V_{j1} \cap V_{j2} = \emptyset$$

and $V_1, \dots, V_{j-1}, V_{j1}, V_{j2}, V_{j+1}, \dots, V_m$ are linearly independent, then either $V_{j1} = V_j$ or $V_{j2} = V_j$.

Such a partition is unique up to the order in (5) and within the elements V_1, \dots, V_m of the partition.



- *Proof.* For $N = 1$ Lemma 2 is trivial. Proceeding by induction, assume that

$$V'_1 \cup \dots \cup V'_k \quad (6)$$

is the partition from Lemma 2 of the set $\{v_1, \dots, v_{N-1}\}$. If $V'_1, \dots, V'_k, \{v_N\}$ are linearly independent, the partition

$$V'_1 \cup \dots \cup V'_k \cup V'_{k+1}, \quad V'_{k+1} = \{v_N\}$$

satisfies (i) and (ii).

Suppose now that $V'_1, \dots, V'_k, \{v_N\}$ are linearly dependent. Due to the assumed linear independence of V'_1, \dots, V'_k , there is exactly one collection $(V'_{i_1}, \dots, V'_{i_l}), i_1, \dots, i_l \in \{1, \dots, k\}$ such that $v_N \in \text{span}(V'_{i_1}, \dots, V'_{i_l})$. Without loss in generality, one may assume that the collection consists of the first l elements of (6),

$$V'_{i_1} = V'_1, \dots, V'_{i_l} = V'_l, \quad l \leq k.$$

Then for $m = k - l + 1$ the partition

$$V_1 \cup \dots \cup V_m \text{ with } V_1 = V'_1 \cup \dots \cup V'_l \cup \{v_N\}, V_2 = V'_{l+1} \cup \dots \cup V'_k$$

satisfies (i) and (ii) from Lemma 2.

Proof of Lemma (contd.)

- To prove (iii), assume that besides (5), there is another partition of $\{v_1, \dots, v_N\}$ with properties (i) and (ii) of Lemma 2,

$$\tilde{V}_1 \cup \dots \cup \tilde{V}_n \quad (7)$$

for some n . Take the intersections

$$V_{i1} = V_i \cap \tilde{V}_1, \dots, V_{in} = V_i \cap \tilde{V}_n, \quad i = 1, \dots, k. \quad (8)$$

They are linearly independent since the new partition satisfies (i). But this contradicts irreducibility of (5) unless only one of the sets (8) is nonempty. The same arguments prove that only one of the intersections

$$\tilde{V}_{j1} = \tilde{V}_j \cap V_1, \dots, \tilde{V}_{jk} = \tilde{V}_j \cap V_k, \quad j = 1, \dots, n$$

is nonempty. This proves that for the partitions (5) and (7), $m = n$ and each \tilde{V}_j is one of V_i . \square

- As always, a proof by induction is non-constructive. To find an algorithm of effective partitioning a set of vectors satisfying (i) and (ii) from Lemma 2 seems to be an interesting problem that may be useful not only in describing UMVUEs.

- An element $\sum_{i=1}^N c_i v_i$ of span $\{v_1, \dots, v_N\}$ is called a *contrast* if

$$\sum_1^N c_i v_i = \mathbf{0}. \quad (9)$$

A contrast is called *decomposable* if

$$\sum_1^N c_i v_i = \sum_{v_i \in V'} c_i v_i + \sum_{v_i \in V''} c_i v_i \quad (10)$$

where $V' \cup V''$ is a partition of $\{v_1, \dots, v_N\}$ and the terms on the right hand side of (10) are nontrivial contrasts (i. e., not all the coefficients equal zero), and *indecomposable* otherwise.

Corollary

If $\sum_1^N c_i v_i$ is a contrast and

$$\sum_1^N c_i v_i = \sum_{v_i \in V_1} c_i v_i + \dots + \sum_{v_i \in V_m} c_i v_i,$$

where V_1, \dots, V_m are from Lemma 2, then

$$\sum_{v_i \in V_1} c_i v_i, \dots, \sum_{v_i \in V_m} c_i v_i \tag{11}$$

are also contrasts.



Lemma

Suppose that V_1 contains at least two vectors. For any two vectors $v_i, v_j \in V_1$, there exists an indecomposable contrast $\sum_{v_i \in V_1} c_i v_i$ with $c_i c_j \neq 0$.

- *Proof.* Re-enumerating the vectors v_1, \dots, v_N , if necessary, one may assume that $V_1 = (v_1, v_2, \dots, v_{k+1})$. Suffice to prove Lemma 3 for v_1, v_2 .

We shall proceed by induction in $|V_1|$, the number of elements in V_1 . For $V_1 = (v_1, v_2)$, the vectors v_1 and v_2 are linearly dependent (otherwise they would have belonged to different components of the partition) and neither is $\mathbf{0}$. Hence $c_1 v_1 + c_2 v_2 = \mathbf{0}$ for some c_1, c_2 with $c_1 c_2 \neq 0$.

Now assume that Lemma 3 holds for $|V_1| \leq k$ and prove it for $|V_1| = k + 1$.

If the set (v_1, v_2, \dots, v_k) is irreducible, the claim holds true due to the induction assumption.

Suppose now that v_1, v_2, \dots, v_k is reducible (though V_1 is irreducible). By virtue of Lemma 2, there is a partition

$$V_1 = V_{11} \cup V_{12} \cup \dots \cup V_{1l} \quad (12)$$

similar to (5). If v_1, v_2 belong to the same component of the partition (12), the claim of Lemma 3 again holds due to the induction assumption.

Proof of Lemma 3 (contd.)

- Consider the case when v_1, v_2 belong to different components of (12), $v_1 \in V_{11}, v_2 \in V_{12}$, say. There is no loss in generality in assuming that a basis of $\text{span}(V_{1i}), i = 1, \dots, l$ consists of elements of V_{1i} ,

$$v_{1i}, v_{2i}, \dots, v_{m_i i}, \quad v_{11} = v_1, \quad v_{21} = v_2.$$

Since

$$v_{11}, \dots, v_{1m_1}; \quad v_{21}, \dots, v_{2m_2}; \dots; \quad v_{l1}, \dots, v_{lm_l}$$

is a basis of $\text{span}(V_{11}, V_{12}, \dots, V_{1l})$ and V_1 is irreducible, there exists a unique representation

$$v_{k+1} = \sum_{i=1}^{k_1} c_{1i} v_{1i} + \sum_{i=1}^{k_2} c_{2i} v_{2i} + \dots + \sum_{i=1}^{k_l} c_{li} v_{li}. \quad (13)$$

One has $c_{11}c_{21} \neq 0$. Indeed, were $c_{11} = 0$ it would imply that $v_1 = v_{11}$ is linearly independent of v_2, \dots, v_{k+1} contradicting irreducibility of V_1 . Similarly, $c_{21} = 0$ implies the linear independence of v_2 and v_1, v_3, \dots, v_{k+1} . Thus,

$$\sum_{i=1}^{k_1} c_{1i} v_{1i} + \sum_{i=1}^{k_2} c_{2i} v_{2i} + \dots + \sum_{i=1}^{k_l} c_{li} v_{li} - v_{k+1}$$

is the required contrast. It is indecomposable due to the uniqueness of the representation (13). \square

- Let $c = (c_1, \dots, c_N)^T$, $\gamma = (\gamma(1), \dots, \gamma(N))^T$ be vectors in \mathbb{R}^N . The *Hadamard product* $h = \gamma \odot c$ of γ and c is the vector with components $h_1 = \gamma(1)c_1, \dots, h_N = \gamma(N)c_N$.

Lemma

Suppose γ is such that for any contrast $\sum_1^N c_i v_i$, $\sum_1^N h_i v_i$ is also a contrast. Then γ is constant on the index set $I_k = \{k_1, \dots, k_l\}$ of each V_k , $k = 1, \dots, m$, i. e., if $V_k = \{v_{k_1}, \dots, v_{k_l}\}$, then

$$\gamma(k_1) = \dots = \gamma(k_l) = \gamma_k. \quad (14)$$

We shall prove Lemma 4 for V_1 assuming that $v_1, v_2 \in V_1$ (if it contains only one vector, the claim is trivially true). Dividing $g(x)$ by $\max\{|\gamma(x)|, x \in I_1\}$, one may assume that $|\gamma(x)| \leq 1, x \in I_1$. Let $\sum_{i \in I_1} c_i v_i$ be an indecomposable contrast with $c_1 c_2 \neq 0$. Such a contrast exists due to Lemma 3.

Using the condition of Lemma 4 repeatedly, we get that the linear combinations

$$\sum_{i \in I_1} \gamma(i) c_i v_i, \sum_{i \in I_1} [\gamma(i)]^2 c_i v_i, \dots, \sum_{i \in I_1} [\gamma(i)]^n c_i v_i \quad (15)$$

are all contrasts.

Let $I_1 = I_1' \cup I_1''$ where

$$I_1' = \{i \in I_1 : |\gamma(i)| = 1\}, I_1'' = \{i \in I_1 : |\gamma(i)| < 1\}.$$

Letting $n = 2, 4, 6, \dots \rightarrow \infty$ in (15) leads to that $\sum_{i \in I_1'} c_i v_i$ is a contrast which contradicts indecomposability of $\sum_{i \in I_1} c_i v_i$ unless $I_1'' = \emptyset$.

- Let now $I_1^+ = \{i \in I_1 : \gamma(i) = +1\}$, $I_1^- = \{i \in I_1 : \gamma(i) = -1\}$. Since

$$\sum_{i \in I_1} \gamma(i) c_i v_i = \sum_{i \in I_1^+} c_i v_i - \sum_{i \in I_1^-} c_i v_i,$$

we have $\sum_{i \in I_1^+} c_i v_i = \sum_{i \in I_1^-} c_i v_i$ and thus, $\sum_{i \in I_1^+} c_i v_i$ and $\sum_{i \in I_1^-} c_i v_i$ are contrasts again contradicting indecomposability of the contrast $\sum_{i \in I_1} c_i v_i$ unless one of the sets I_1^+ , I_1^- is empty. Since $c_1 c_2 \neq 0$, $\gamma(1) = \gamma(2)$. \square

- Return to the statistical setup when an observation X takes values x_1, \dots, x_N with probabilities

$$P_\theta(X = x_i) = p_i(\theta), \quad i = 1, \dots, N.$$

A parametric function $g(\theta)$ is estimable from X iff

$$g(\theta) = \sum_1^N c_i p_i(\theta) \tag{16}$$

for some constants c_1, \dots, c_N .

Theorem

An estimable parameter function (16) has a UMVUE if and only if

$$c_j = c_i^*, \quad j \in I_i, \quad i = 1, \dots, m \tag{17}$$

where the sets I_1, \dots, I_m are the index sets of V_1, \dots, V_m from Lemma 2.

A statistic $\gamma(X)$ is a UMVUE if and only if

$$\gamma(x_j) = \gamma_i, \quad x_j \in I_i, \quad i = 1, \dots, m. \tag{18}$$

Proof. Let $\chi(X)$ be an unbiased estimator of zero,

$$E_{\theta} [\chi(X)] = \sum_1^N \chi(x_i) p_i(\theta) = 0, \theta \in \Theta. \quad (19)$$

On looking at $p_i(\theta)$ as a vector v_i , the unbiased estimator of zero from (19) becomes a contrast $\sum_1^N c_i v_i$ with $c_i = \chi(x_i)$.

By Lemma 1, a statistic $\gamma(X)$ is a UMVUE if for any unbiased estimator of zero $\chi(X)$,

$$E_{\theta} [\gamma(X)\chi(X)] = \sum_1^N \gamma(x_i)\chi(x_i)p_i(\theta) = 0, \theta \in \Theta. \quad (20)$$

The relation (20) means that $\gamma = (\gamma(x_1), \dots, \gamma(x_N))^T \in \mathbb{R}^N$ is such a vector that for any contrast $\sum_1^N c_j v_j$, the sum $\sum_1^N h_i v_i$ with $h_i = \gamma(x_i)c_i$, $i = 1, \dots, N$ is also a contrast.

- By Lemma 4, such a γ must be constant on each I_i thus proving (18). Conversely, suppose that (18) holds. By virtue of Corollary 1, $\sum_{j \in I_i} c_j v_j$ is a contrast for $i = 1, \dots, m$, i. e.,

$$\sum_{j \in I_i} \chi(x_j) p_j(\theta) = 0, \theta \in \Theta, i = 1, \dots, m$$

and thus,

$$E_{\theta} [\gamma(X) \chi(X)] = \sum_1^N \gamma(x_j) \chi(x_j) p_j(\theta) = \sum_{i=1}^m \gamma(x_i) \left[\sum_{j \in I_i} \chi(x_j) p_j(\theta) \right] = 0, \theta \in \Theta.$$

This proves the second claim of Theorem 1.

For $\gamma(X)$ satisfying (18),

$$E_{\theta} [\gamma(X)] = g(\theta) = \sum_{i=1}^m \gamma_i \left[\sum_{j \in I_i} p_j(\theta) \right]$$

proving (17). \square

Example

Return to the first example from Section 1 with $X \in \{x_1, x_2, x_3, x_4\}$ and

$$p_1(\theta) = \theta/2, p_2(\theta) = \theta/3, p_3(\theta) = \theta^2, p_4(\theta) = 1 - 5\theta/6 - \theta^2, \theta < 2/3.$$

The minimal sufficient partition is

$$\mathcal{X}_1 = \{x_1, x_2\}, \mathcal{X}_2 = \{x_3\}, \mathcal{X}_4 = \{x_4\},$$

the same as the partition (5) from Lemma 2. Any quadratic polynomial in θ admits a UMVUE.

Another example

In the second example $X \in \{x_1, x_2, x_3, x_4, x_5\}$ with

$$p_1(\theta) = \theta/2, p_2(\theta) = \theta/6, p_3(\theta) = \theta/6 + \theta^2/2, p_4(\theta) = \theta^2/2, p_5(\theta) = 1 - 5\theta/6 - \theta^2/2, \theta < 2/3$$

the minimal sufficient partition is

$$\mathcal{X}_1 = \{x_1, x_2\}, \mathcal{X}_2 = \{x_3\}, \mathcal{X}_3 = \{x_4\}, \mathcal{X}_4 = \{x_5\}.$$

The vectors $v_1 = \theta/2$, $v_2 = \theta/6$, $v_3 = \theta/6 + \theta^2/2$, $v_4 = \theta^2/2$ are linearly dependent and independent of $v_5 = 1 - 5\theta/6 - \theta^2/2$ so that the partition from Lemma 2 is

$$V_1 = \{v_1, v_2, v_3, v_4\}, V_2 = \{v_5\}.$$

Any quadratic polynomial in θ is estimable but only polynomials of the form $a + b(5\theta/6 + \theta^2)$ admit UMVUEs.