The Structure of a Linear Model: Sufficiency, Ancillarity, Invariance, Equivariance, and the Normal Distribution

Wolfgang Bischoff

University of Karlsruhe, D-76128 Karlsruhe, Germany
E-mail: wolfgang.bischoff@math.uni-karlsruhe.de

Received January 2, 1998

Consider a general linear model \( Y = X\beta + Z \) where \( \text{Cov} Z \) may be known only partially. We investigate carefully the notions of sufficiency, ancillarity, invariance, and equivariance and related notions for projectors in a general linear model. In this way we can prove a Basu type theorem. This result can be used to give the relation between the sufficiency of the generalized least-squares estimator and the assumption that \( Z \) is normally distributed. So we can generalize the well-known result that the generalized least-squares estimator is sufficient for \( \beta \) if \( Z \) is normally distributed. Further we can solve the converse problem as well.

Key words and phrases: linear model, sufficiency, specific sufficiency, ancillarity, invariance, equivariance, normal distribution, partially known covariance matrices.

1. INTRODUCTION

Let us consider a linear model

\[
Y = (Y_1, \ldots, Y_n)^\top = X\beta + Z,
\]

where

- \( X \) is a known real \( n \times m \) matrix with rank(\(X\)) = \( m \),
- \( \beta \in \mathbb{R}^m \) is the unknown parameter vector,
- \( Z = (Z_1, \ldots, Z_n)^\top \) is an \( \mathbb{R}^n \)-random vector with \( \text{EZ} = 0, \text{Cov} Z = \sigma^2 \Sigma \) positive definite.

For this model we investigate the structure of invariance and equivariance carefully. In this way we get equivalent statements that a projector is ancillary and sufficient, respectively, and a Basu type theorem can be proved. Especially, these results imply the following characterization of the multivariate normal distribution by sufficiency of the generalized least-squares estimator.
Let \( \sigma^2 \Sigma \) be known. Then it is well known that the generalized least-squares estimation for \( X\beta \) and for \( \beta \) is sufficient for \( X\beta \) and for \( \beta \), respectively, if \( Z \) is normally distributed; see for example Arnold [1].

Conversely, let us assume additionally that

\[
Z_1, ..., Z_n \text{ are stochastically independent.} \tag{1.2}
\]

Then for the special case \( X = I_n := (1, ..., 1) \trans \in \mathbb{R}^n \) and \( \Sigma = I_n \) (\( I_n \) the unity matrix of \( \mathbb{R}^n \)) it is known that if the least-squares estimation for \( \beta \) is sufficient for \( \beta \), that is, the sample mean \( \frac{1}{n} \sum_{i=1}^{n} y_i \) is sufficient for \( \beta \), then \( Z \) is normally distributed. For that result and some generalizations of this one-dimensional case (\( m = 1 \)) see Kelker and Matthes [10], Kagan et al. [9], Bartfai [3], and Eberl [7, 8].

Among other things Bischoff et al. [5] showed for the linear model (1.1) given above: If, in addition, condition (1.2) is fulfilled and \( \text{Cov} Z = \text{diag}(\sigma_1^2, ..., \sigma_n^2), \sigma_j^2 > 0 \) for \( j = 1, ..., n \) known, then the following result holds true:

The generalized least-squares estimation for \( X\beta \) is sufficient for \( X\beta \) if and only if

\[
\forall i \in \{1, ..., n\} : [e_i \notin V \cup V^\perp = Z_i \text{ is normally distributed}],
\]

where

\[
e_1 = (1, 0, ..., 0) \trans, e_2 = (0, 1, 0, ..., 0) \trans, ..., e_n = (0, ..., 0, 1) \trans \in \mathbb{R}^n,
\]

\[
V = \text{range}(X), V^\perp = \{ y \in \mathbb{R}^n | \forall v \in V : v \trans y = 0 \}.
\]

The results of Bischoff et al. [5] generalize the one-dimensional results stated in the papers mentioned above. In the present paper we generalize the above characterization of the multivariate normal distribution by allowing

\[
\text{Cov} Z = \sigma^2 \Sigma \text{ is any positive definite } n \times n \text{ matrix where}
\]

\[
\sigma^2 > 0 \text{ is unknown and } \Sigma \text{ may be partially unknown.}
\]

Note that the last condition can imply the following problem. If \( \text{Cov} Z \) is not a diagonal matrix, then the assumption (1.2) is violated. Condition (1.2) was assumed in all papers mentioned above. We show that only a relaxed corresponding condition is necessary. This condition is stated in Section 2.3. It depends on the model (design) matrix and on the covariance structure. For that it is necessary to develop the structure of covariance matrices. This is considered for some models in Section 2.1, where we discuss stochastic models for covariance matrices. Moreover, we are interested in \( X\beta \) or \( \beta \) but not in \( \sigma^2 \Sigma \). That means that \( \sigma^2 \Sigma \) is a nuisance parameter if it is not known.
Therefore we are looking for specifically sufficient statistics for $X\beta$ or $\beta$. In Section 2.2 the structure of the linear model is investigated by treating (specific) sufficiency, ancillarity, invariance, and equivariance, and we state a Basu type theorem there. Results of Sections 2.1 and 2.2 are used in Section 2.3 to show a generalization of the characterization of the multivariate normal distribution mentioned above. Further we apply our results to some examples there. In Section 3.1 we prove a Basu type theorem for a general group-structured model. Thus we can derive the results of Section 2.2. In Section 3.2 the results of Section 2.3 are proved more generally for coordinate-free linear models in a Hilbert space as described in Eaton’s book (1983). It is most convenient to develop the results for such models in a Hilbert space because then the inner product and the bases can be chosen suitably.

2. THE STRUCTURE OF A LINEAR MODEL

2.1. Stochastic Models for Covariance Matrices

In this section we show firstly how certain covariance matrices may occur in stochastic models. In Section 2.3 we explain our results for special linear models with error vector $Z$ coming from such stochastic models. For all models we assume that

$$\eta_1, \ldots, \eta_n$$

are real random variables with

$$E(\eta_j) = 0, \quad j = 1, \ldots, n,$$

$$\text{Cov}(\eta_i, \eta_j) = 0, \quad 0 \leq i < j \leq n,$$

and

$$\text{Var}(\eta_j) = \sigma^2 > 0, \quad j = 1, \ldots, n.$$  

Further we use the notation

$$\eta = (\eta_1, \ldots, \eta_n)^\top.$$

Model 1. For $j = 1, \ldots, n$ let

$$Z_1 := \eta_1, Z_2 := \alpha Z_1 + \eta_2, \ldots, Z_n := \alpha Z_{n-1} + \eta_n.$$
where \( \alpha \in \mathbb{R} \) is arbitrary. Then we have \( Z = B \eta \), where

\[
B = \begin{pmatrix}
1 & 0 & \cdots & 0 \\
\alpha & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
\alpha^{n-1} & \alpha^{n-2} & \cdots & 1
\end{pmatrix},
\]

implying that

\[
\text{Cov} Z = \sigma^2 BB^T =: \sigma^2 C_1(\alpha).
\]

**Model 2.** For \( j = 1, \ldots, n \) and \( \alpha \in \mathbb{R} \) let

\[
Z_1 := \alpha \eta_n + \eta_1, \quad Z_2 := \alpha \eta_1 + \eta_2, \quad Z_3 := \alpha \eta_2 + \eta_3, \ldots, \quad Z_n := \alpha \eta_{n-1} + \eta_n;
\]

that is, we have \( Z = B \eta \), where

\[
B = \begin{pmatrix}
1 & 0 & \cdots & 0 & \alpha \\
\alpha & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & 1 & \alpha \\
0 & \cdots & 0 & \alpha & 1
\end{pmatrix}
\]

Let \( C = (1 + \alpha^2) I_n + \alpha \sum_{i=1}^{n-1} (e_i e_i^T + \alpha e_i e_i^T) \); then Cov Z is given by

\[
\text{Cov} Z = \sigma^2 BB^T = \sigma^2 (C + \alpha (e_1 e_1^T + c \alpha e_1 e_1^T)) =: \sigma^2 C_2(\alpha).
\]

**Model 3.** For \( j = 1, \ldots, n \) and \( \alpha \in \mathbb{R} \setminus \{1\} \) let

\[
Z_j := \eta_j + \alpha \sum_{i=1, i \neq j}^{n} \eta_i, \quad j = 1, \ldots, n.
\]

Then we have \( Z = B \eta \), where

\[
B = \begin{pmatrix}
1 & \alpha & \cdots & \alpha \\
\alpha & 1 & \cdots & \alpha \\
\vdots & \vdots & \ddots & \vdots \\
\alpha & \alpha & \cdots & 1
\end{pmatrix}
\]

Note that the covariance of \( Z \) is given by

\[
\text{Cov} Z = \sigma^2 BB^T = \sigma^2 (1 + (n-1) \alpha^2)(I_n + \gamma 1_n 1_n^T),
\]
where

\[ \gamma = \frac{2a + (n - 2) \sigma^2}{1 + (n - 1) \sigma^2}. \]

2.2. The Group-Structure of a Linear Model

We consider the linear model (1.1)

\[ Y = X\beta + Z, \quad \beta \in \mathbb{R}^n, \quad EZ = 0, \quad \text{Cov} Z = \sigma^2\Sigma \text{ positive definite, } \quad \sigma^2 > 0, \]

where \( X \) is known. We are interested in \( \beta \) and not in \( \sigma^2 \Sigma \). Thus \( \sigma^2 \Sigma \) is a nuisance parameter. But at the first glance let \( \sigma^2 \Sigma \) be known. The above model is given in coordinatized form. However, it is easier to understand, state, and prove the results and it is statistically more natural to consider the corresponding coordinate-free model,

\[ Y = \theta + Z, \quad \theta \in V, \quad (2.1) \]

where \( V = \text{range}(X) \) with \( \dim(V) = m \in \{1, \ldots, n\} \).

Next let us define some general notions. Let \( \mathcal{H} \) be a separable metric space equipped with its Borel-\( \sigma \)-algebra \( \mathcal{B} = \mathcal{B}(\mathcal{H}) \). We consider a group \( G \) acting from the left on \( \mathcal{H} \) by

\[ G \times \mathcal{H} \to \mathcal{H}, \quad (g, x) \mapsto gx. \]

The group acts measurably and continuously if the mapping

\[ \mathcal{H} \to \mathcal{H}, \quad x \mapsto gx \]

is measurable and continuous, respectively. If a group acts continuously, then it acts measurably.

If the group acts measurably we can define the class of probability measures induced by \( G \) and by a given probability measure \( P_0 \) on \( \mathcal{B} \),

\[ \mathcal{P} = \{ gP_0 \mid g \in G \}, \]

where \( gP_0 \) is the image measure of \( P_0 \) with respect to \( g \). Thus we have for \( g \in G \)

\[ \forall B \in \mathcal{B} : gP_0(B) = P_0(g^{-1}(B)). \]

We define

\[ \forall g \in G : P_g := gP_0 \]
and
\[ \forall g, h \in G : hP_0 = h(gP_0) = (h \cdot g) P_0 = P_{h \cdot g}. \]

Note that if \( e \) is the neutral element in \( G \) then \( P_0 = P_e \). We can also say that \( G \) acts transitively on \( \mathcal{P} \). As an example we consider the linear model (2.1).

**Example.** Let \( P_0 \) be the distribution of \( Z \) defined on \((\mathcal{H}, \mathcal{B}(\mathcal{H})) = (\mathbb{R}^n, \mathcal{B}^n)\), and let \( G \) be the additive group \((V, +)\) which is a subgroup of \((\mathbb{R}^n, +)\). Then \( V \) is acting from the left on \( \mathbb{R}^n \) continuously by
\[ (v, x) \mapsto v + x. \]

If \( P_0 \) is the distribution of \( Z \), then the class of possible distributions of \( Y = \theta + Z \) is given by
\[ \mathcal{P} = \{ P_\theta | \theta \in \mathcal{V} \} \]
and we have
\[ \forall \theta \in \mathcal{V} \quad \forall B \in \mathcal{B} : P_\theta(B) = P_0(B - \theta). \]

\( \mathcal{P} \) is called a \( \mathcal{V} \)-location family.

Before the example is continued let us consider some further general definitions and remarks. Let \((\mathcal{Y}, \mathcal{D})\) be a second measure space. Then a statistic \( S : \mathcal{H} \to \mathcal{Y} \) is called \( G \)-invariant if
\[ \forall g \in G \quad \forall x \in \mathcal{H} : S(gx) = S(x). \]

The set
\[ \mathcal{B}_G := \{ B \in \mathcal{B} | 1_B \text{ is } G\text{-invariant} \} \]
is called the \( \sigma \)-algebra of \( G \)-invariant sets. Note that a \( G \)-invariant statistic is \( \mathcal{B}_G \)-measurable. A statistic \( S \) is called ancillary for \( \mathcal{P} \) if the distribution of \( S \) is the same for each \( P \in \mathcal{P} \). Obviously, a \( G \)-invariant statistic is ancillary for \( \mathcal{P} = \{ P_g | g \in G \} \).

Let
\[ \mathcal{H}/G = \{ Gx | x \in \mathcal{H} \}, \quad \text{where} \quad Gx = \{ gx | g \in G \} \]
denote the sets of orbits. A statistic \( S : \mathcal{H} \to \mathcal{Y} \) is called maximal \( G \)-invariant if
\[ \forall Gx, Gy \in \mathcal{H}/G \quad \text{with} \quad Gx \neq Gy : S(x) \neq S(y). \]
We call a set $C_0 \in \mathcal{B}(\mathcal{H})$ with
\[ \forall x \in \mathcal{H} : C_0 \cap Gx \neq \emptyset, \]
\[ \forall a, b \in C_0 \quad \text{with} \quad a \neq b : Ga \neq Gb \]
a cut. $C_0$ is called a nice cut if
\[ \forall x, y \in C_0 : Gx = Gy =: G_0, \] (2.2)
where $G_x = \{ g \in G | gx = x \}$. In the sequel we assume that $G_0 = \{ e \}$ without loss of generality. Note that if $C_0$ is a nice cut, then $gC_0$ is a nice cut as well. A statistic $S$ is called nice $G$-equivariant with respect to $C$, if $C$ is a nice cut and
\[ \forall g, h \in G : S(gx) = S(gy), \]
A nice $G$-equivariant statistic $S$ is called maximal if
\[ \forall g, h \in G : S(gx) = S(hx), \]
Obviously, we have $\mathcal{H} = \bigcup_{x \in G} gC$. Let us define
\[ \mathcal{H}/C := \{ gC | g \in G \} \quad \text{where} \quad gC = \{ gc | c \in C \}.
\]
$\mathcal{H}/C$ is the set of nice cuts induced by $G$ and $C$. By these considerations it is obvious that a (maximal) nice $G$-equivariant and a (maximal) $G$-invariant statistic are defined in a dual way.

**Example (Continued).** Let $W$ be a subspace of $\mathbb{R}^n$ with $V \oplus W = \mathbb{R}^n$, and let $\text{pr}_{W \mid V}$ be the projection onto $W$ along $V$. The class of $V$-invariant Borel sets is given by
\[ \mathcal{B}_{V} := \{ A \in \mathcal{B} | \exists B \in \mathcal{B} : A = \text{pr}_{W \mid V}(B) \} = \{ A \in \mathcal{B} | \forall \theta \in V : A - \theta = A \}.
\]
Thus each $\text{pr}_{W \mid V}$ is a maximal invariant statistic and so an ancillary statistic. For all $x \in \mathbb{R}^n$ we have $G_x = \{ 0 \}$; therefore each $W$ with $V \oplus W = \mathbb{R}^n$ is a nice cut. Thus each $\text{pr}_{W \mid V}$ is a maximal nice $V$-equivariant statistic.

Next let us consider the case where $P_0$ is not fully known. This means that
\[ P_0 = P_{0, \rho} \in \{ P_{0, \rho} | \rho \in A \}, \]
where $A$ is a parameter space. Then we consider the following class of probability measures on $\mathcal{B}$:
\[ \mathcal{P}_{G, A} = \{ gP_{0, \rho} | g \in G, \rho \in A \}. \]
We define \( gP_{\alpha,\rho} := P_{\alpha,\rho} \) and assume that \( \rho \) is a nuisance parameter. In the sequel we write \( \mathcal{G}_0 \) for \( \{ P_{g} \mid g \in G \} \).

**Example (Continued).** Let us consider model 1 given in Section 2.1. Let the distribution of \((\eta_1, \ldots, \eta_n)^\top\) be known up to variance \( \sigma^2 \), \( j = 1, \ldots, n \). Then the distribution of \( Z \) is parameterized by \( \rho = (\sigma^2, \alpha) \in (0, \infty) \times \mathbb{R} = A \) if \( \alpha \) is unknown. Therefore the distribution of the linear model \( Y = \theta + Z \), \( \theta \in V \), is parameterized by \( \mathcal{P}_{V, A} = \{ P_{\theta, \rho} \mid \theta \in V, \rho \in A \} \) for the class of distributions of \( Y \). For the other models in Section 2.1 the class \( \mathcal{P}_{V, A} \) is defined accordingly.

A statistic \( S : H \to H \) is called sufficient for \( \mathcal{G}_0 \) if and only if
\[
\forall g \in G \quad \forall B \in \mathcal{B} : P_{g}(B \mid S) = P_{0}(B \mid S) =: P(B \mid S),
\]
where \( P_{g}(B \mid S) \in L_1 \) is the conditional probability of \( B \) under \( S \) with respect to \( P_{g} \). Accordingly, the statistic \( S \) is called specifically sufficient for \( P_{G, A} \) if and only if
\[
\forall g \in G \quad \forall \rho \in A \quad \forall B \in \mathcal{B} : P_{g, \rho}(B \mid S) = P_{0, \rho}(B \mid S) =: P_{\rho}(B \mid S).
\]
Note that a specifically sufficient statistic \( S \) must be independent of the nuisance parameter \( \rho \). A statistic \( S : H \to H \) is called \( G \)-invariantly sufficient for \( \mathcal{G}_0 \) if and only if
\[
\forall g \in G \quad \forall \rho \in A \quad \forall B \in \mathcal{B} : P_{g}(B \mid S) = P_{0}(B \mid S) =: P(B \mid S);
\]
see Eberl [8] and Bischoff et al. [5] for special cases of the last definition. Accordingly, a statistic \( S \) is called specifically \( G \)-invariantly sufficient for \( \mathcal{G}_{G, A} \) if and only if
\[
\forall g \in G \quad \forall \rho \in A \quad \forall B \in \mathcal{B} : P_{g, \rho}(B \mid S) = P_{0, \rho}(B \mid S) =: P_{\rho}(B \mid S).
\]
Thus sufficiency implies \( G \)-invariant sufficiency.

In the sequel we use the notation
\[
\mathcal{B}(S) := S^{-1}(\mathcal{D}),
\]
where \( S : (\mathcal{X}, \mathcal{B}) \to (\mathcal{Y}, \mathcal{D}) \) is a statistic. Now we are in a position to state the main results of this section, which are special examples of Theorem 3.1. Theorem 3.1 is a Basu type theorem. A proof of Corollary 2.1 is given after the proof of Theorem 3.1. In the sequel we write \( (V \text{-invariantly}) \) sufficient for \( V \) instead of \( (V \text{-invariantly}) \) sufficient for \( \mathcal{G} \) as well as specifically \( (V \text{-invariantly}) \) sufficient for \( G \mid A \) instead of specifically \( (V \text{-invariantly}) \) sufficient for \( \mathcal{G}_{G, A} \).
Corollary 2.1. Let the linear model (2.1) be given, let the distribution of $Z$ be known, let $V^\perp = \{ w \in \mathbb{R}^n \mid w^T e = 0 \}$ for all $e \in V$. Then the following statements are equivalent:

(a) $\text{pr}_{V|ZV^\perp}$ is sufficient for $V$.
(b) $\text{pr}_{V|ZV^\perp}$ is $V$-invariantly sufficient for $V$.
(c) $\mathcal{B}(\text{pr}_{V|ZV^\perp})$ and $\mathcal{B}_V$ are independent with respect to $P_0$.
(d) $\forall \theta \in V : \mathcal{B}(\text{pr}_{V|ZV^\perp})$ and $\mathcal{B}_V$ are independent with respect to $P_{\theta}$.

Moreover, $\text{pr}_{V|ZV^\perp}$ is the only projector onto $V$ which can be sufficient for $V$.

Next we state an analogous result for specific sufficient. Note that if for each fixed nuisance parameter the same statistic $S$ is sufficient, then $S$ is specifically sufficient. Therefore Corollary 2.1 implies Corollary 2.2.

Corollary 2.2. Let the linear model (2.1) be given, let the distribution of $Z$ be known up to $\text{Cov}(Z)$ which is parametrized by $\gamma$, and let $\text{pr}_{V|ZV^\perp}$ be independent of $\gamma$. Then the following statements are equivalent:

(a) $\text{pr}_{V|ZV^\perp}$ is specifically sufficient for $V|A$.
(b) $\text{pr}_{V|ZV^\perp}$ is specifically $V$-invariantly sufficient for $V|A$.
(c) $\mathcal{B}(\text{pr}_{V|ZV^\perp})$ and $\mathcal{B}_V$ are independent with respect to $P_{\theta, \rho}$ for all $\rho \in A$.
(d) $\forall \theta \in V \forall \rho \in A : \mathcal{B}(\text{pr}_{V|ZV^\perp})$ and $\mathcal{B}_V$ are independent with respect to $P_{\theta, \rho}$.

Moreover, $\text{pr}_{V|ZV^\perp}$ is the only projector onto $V$ which can be specifically sufficient for $V|A$.

Example (Continued). Consider the linear model (2.1) with $\mathbf{1}_n \in V$.

and let $\text{Cov}(Z) = \sigma^2(I_n + \gamma \mathbf{1}_n \mathbf{1}_n^T) = \sigma^2 \Sigma$; see Model 3 in Section 2.1. Then $\text{pr}_{V|ZV^\perp} = \text{pr}_{V|V^\perp}$ is independent of $\rho = (\sigma^2, \gamma)$.

Remark 2.1. Let Model (1.1) be given with $\text{rank}(X) = m$, then the class of distributions of $Y$ can also parameterized by $\mathbb{R}^m$ and $\mathbb{R}^m \times A$ instead of $V$ and $V \times A$, respectively. Further it is more natural then to consider the generalized least squares estimator $(X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}$ for $\beta$ instead of the projector $\text{pr}_{V|ZV^\perp}$. Hence we can state a result which is analogous to Corollary 2.1.
Corollary 2.3. Let the linear model (1.1) be given, let the distribution of \( Z \) be known, and let \( \text{Cov}(Z) = \Sigma \). Then the following statements are equivalent;

(a) \((X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}\) is sufficient for \( \mathbb{R}^m \).

(b) \((X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}\) is \( V \)-invariantly sufficient for \( \mathbb{R}^m \).

(c) \( B((X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}) \) and \( B \) are independent with respect to \( P_0 \).

(d) \( \forall \beta \in \mathbb{R}^m : \mathbb{B}_0((X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}) \) and \( B \) are independent with respect to \( P_\beta \).

Moreover, \((X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1}\) is the only statistic in the class \((X^T C^{-1} X)^{-1} X^T C^{-1} | C \) a positive definite \( n \times n \) matrix\) which can be sufficient for \( V \).

Accordingly, Corollary 2.2 can be stated for Model (1.1).

2.3. Sufficiency of the Generalized Least Squares Estimator

The main result of this section is a consequence of Theorem 3.2. For brevity we state the following results for Model (1.1) only.

Corollary 2.4. Let \( \eta = (\eta_1, \ldots, \eta_n)^T \) be a random vector whose distribution is known with \( E\eta = 0 \), \( \text{Cov}(\eta) = \sigma^2 I_n \); let \( B = (b_1, \ldots, b_n) \) be a real \( n \times n \) matrix with \( \text{rank}(B) = n \) which is known, let \( Z = B\eta \), and let Model (1.1) be given:

\[
Y = X\beta + Z, \quad \beta \in \mathbb{R}^m.
\]

With \( V = \{X\beta \mid \beta \in \mathbb{R}^m\} \), let \( b_1, \ldots, b_r \in V \), \( b_{r+1}, \ldots, b_s \in BB^T(V^\perp) \) and \( b_{s+1}, \ldots, b_n \not\in V \cup BB^T(V^\perp) \). We assume that

\[
\{\eta_1, \ldots, \eta_r\}, \{\eta_{r+1}, \ldots, \eta_s\}, \{\eta_{s+1}\}, \ldots, \{\eta_n\} \text{ are stochastically independent.}
\]

Then the generalized least-squares estimate for \( \beta \) is sufficient for \( \beta \), if and only if

\[
\forall i \in \{s + 1, \ldots, n\} : \eta_i \text{ is normally distributed.}
\]

A proof is given after Theorem 3.2. Note that the generalized least-squares estimator for \( \beta \) is given by

\[
(X^T (BB^T)^{-1} X)^{-1} X^T (BB^T)^{-1} \eta.
\]

Corollary 2.5. Let \( \eta = (\eta_1, \ldots, \eta_n)^T \) be a random vector with \( E\eta = 0 \), \( \text{Cov}(\eta) = \sigma^2 I_n \). Let the distribution of \( \eta \) be known up to \( \sigma^2 \). Let \( B_0 = \)
If \((b_1, \cdots, b_n)\) be a regular, real \(n \times n\) matrix depending on a parameter \(\alpha \in \mathbb{A}\), let \(Z = B_\alpha \eta\), and let Model (1.1) be given:

\[
Y = X\beta + Z, \quad \beta \in \mathbb{R}^m.
\]

With \(V = \{X\beta : \beta \in \mathbb{R}^m\}\) let

\[
\forall i \in \{1, \ldots, r\} \forall \alpha \in A : b_i \in V,
\]

\[
\forall i \in \{r+1, \ldots, s\} \forall \alpha \in A : b_i \in B_\alpha B_\alpha^T(V^\perp),
\]

\[
\forall i \in \{s+1, \ldots, n\} \exists x_1, x_2 \in A : b_{n+1} \notin V \land b_{n+2} \notin B_\alpha B_\alpha^T(V^\perp).
\]

We assume that

\[
\{\eta_1, \ldots, \eta_r\}, \{\eta_{r+1}, \ldots, \eta_s\}, \{\eta_{s+1}\}, \ldots, \{\eta_n\} \text{ are stochastically independent.}
\]

If the generalized least-squares estimate for \(\beta\) does not depend on \(\alpha\), then it is specifically sufficient for \(\beta\) if and only if

\[
\forall i \in \{s+1, \ldots, n\} : \eta_i \text{ is normally distributed.}
\]

**Example 2.1.** Let us consider model 1 of Section 2.1 and the model (design) matrix

\[
X = \begin{pmatrix}
0 & 0 \\
1 & 0 \\
\vdots & \vdots \\
1 & 0 \\
1 & 1
\end{pmatrix} \in \mathbb{R}^{2 \times n}.
\]

If \(\alpha \in \mathbb{R}\) is unknown, then the corresponding generalized least-squares estimator depends on \(\alpha\), whence it cannot be specifically sufficient for \(\beta\).

Let \(\alpha \in \mathbb{R}\) be known, and let \(b_i\) the \(i\)th column of \(B\).

Note that

\[
B^{-1} = \begin{pmatrix}
1 & 0 & \cdots & \cdots & 0 \\
-\alpha & 1 & \cdots & \cdots & \vdots \\
0 & \cdots & \cdots & \cdots & \vdots \\
\vdots & \cdots & \cdots & 1 & 0 \\
0 & \cdots & 0 & -\alpha & 1
\end{pmatrix}.
\]

Then, if \(\alpha \neq 1\),

\[
b_n \in V; \quad b_1 \in BB^T(V^\perp); \quad b_2, b_3, \ldots, b_{n-1} \notin V \cup BB^T(V^\perp).
\]
Let $\sigma^2$ be known, and let $\eta_1, ..., \eta_n$ be stochastically independent. Then the generalized least-squares estimator for $\beta$ is sufficient for $\beta$ if and only if $\eta_2, ..., \eta_{n-1}$ are normally distributed.

Let $x = 1$. Then

$$b_2, b_n \in V; \quad b_1, b_3, ..., b_{n-1} \in BB^T(V^\perp).$$

Thus we have for $\sigma^2$ known and for $\{\eta_2, \eta_n\}, \{\eta_1, \eta_3, ..., \eta_{n-1}\}$ stochastically independent:

The generalized least squares estimator is sufficient for $\beta$. If $\sigma^2$ is unknown then the above results hold true provided “sufficient” is replaced by “specifically sufficient.”

**Example 2.2.** Let us consider Model 2 of Section 2.1 and a trigonometric regression model of order $k (2k + 1 \leq n)$ with the design points $t_j = (2\pi/n)j$, $j = 0, ..., n-1$. Then the generalized least-squares estimator is independent of $\alpha$ and $\sigma^2$; see Bischoff [4, p. 37]. Further we have

$$BB^T(V^\perp) = V^\perp.$$ 

By these facts it is easy to see that

$$b_{1x}, ..., b_{nx} \notin V \cup V^\perp \quad \text{for all } x \in \mathbb{R}.$$ 

Provided that $\eta_1, ..., \eta_n$ are stochastically independent we obtain then that the generalized least-squares estimator is specifically sufficient for $\beta$ if and only if $\eta_1, ..., \eta_n$ are normally distributed.

**Example 2.3.** Let us consider Model 3 of Section 2.1 and a linear model with $(1, ..., 1)^T \in V$. Then the generalized least-squares estimator and the ordinary least-squares estimator coincide. Further we have

$$BB^T(V^\perp) = V^\perp.$$ 

So it is easy to see that the same result is true as described in Example 2.2.

3. GENERAL RESULTS

3.1. Invariance, Equivariance, Sufficiency

We continue the general case considered in Section 2.2. There we assumed that

$\mathcal{X}$ is a separable metric space, \hspace{1cm} (3.1)

$G$ is an abelian group, \hspace{1cm} (3.2)

$G$ acts measurably on $\mathcal{H}$ by $G \times \mathcal{H} \to \mathcal{H}$, $(g, x) \mapsto gx$, \hfill (3.3)

$P_0$ is a probability measure on $\mathcal{B}(\mathcal{H})$, \hfill (3.4)

$\mathcal{P} = \{ gP_0 \mid g \in G \} = \{ P_g \mid g \in G \}$. \hfill (3.5)

Let us further assume that

$C$ is a nice cut, and $x_0 \in C$ is arbitrarily fixed. \hfill (3.6)

Without loss of generality we assume that $G_0 = \{ e \}$ (see (2.2)); otherwise we have to consider $G/G_0$ instead of $G$ in the rest of the paper. Then we have

$\mathcal{H} = Gx_0 \times C$.

We assume that

$Gx_0$ is equipped with the relative topology of $\mathcal{H}$, \hfill (3.7)

$C$ is equipped with the relative topology of $\mathcal{H}$, \hfill (3.8)

$\text{pr}_G : \mathcal{H} = Gx_0 \times C \to Gx_0$, $x = (gx_0, c) \mapsto gx_0$ is continuous, \hfill (3.9)

$\text{pr}_C : \mathcal{H} = Gx_0 \times C \to C$, $x = (gx_0, c) \mapsto c$ is continuous, \hfill (3.10)

$Gx_0 \times C \to \mathcal{H}$, $(gx_0, c) \mapsto gc$ is continuous, \hfill (3.11)

$\mathcal{B}(\mathcal{H}) = \mathcal{B}(Gx_0) \otimes \mathcal{B}(C)$. \hfill (3.12)

Note that $\text{pr}_G$ is a maximal nice $G$-equivariant statistic and $\text{pr}_C$ is a maximal $G$-invariant statistic. By the topological assumptions we have a homeomorphism

$\varphi : Gx_0 \times C \to \mathcal{H}$, $(gx_0, c) \mapsto gc$; $\varphi^{-1} = (\text{pr}_G \circ \text{pr}_C)$.

Thus $\varphi^{-1}P_0$ is a probability measure on $(Gx_0 \times C, \mathcal{B}(Gx_0) \otimes \mathcal{B}(C))$. We denote the marginal probability measure of $\varphi^{-1}P_0$ onto $Gx_0$ by $\pi_G$. We write $\pi_G(dg)$ instead of $\pi_G(dg, gx_0)$.

**Theorem 3.1.** Let the assumptions (3.1)–(3.12) be fulfilled. Further let

$S : (\mathcal{H}, \mathcal{B}(\mathcal{H})) \to (\mathcal{Y}, \mathcal{D})$ be a nice $G$-equivariant statistic with respect to $C$,

$G$ act on $\mathcal{B}(S)$ by $G \times \mathcal{B}(S) \to \mathcal{B}(S)$, $(g, B) \mapsto gB$,

$\varphi^{-1}(P_0) = \pi_G \otimes \pi_{C|G}$, where $\pi_{C|G}$ is a transition probability.
Then the following statements are equivalent:

(a) $S$ is $G$-invariantly sufficient for $\mathcal{P}$.

(b) $\mathcal{B}(S)$ and $\mathcal{B}_G$ are independent with respect to $P_0$.

(c) $\forall g \in G: \mathcal{B}(S)$ and $\mathcal{B}_G$ are independent with respect to $P_g$.

If, additionally, $S$ is a maximal nice $G$-equivariant statistic with $\mathcal{B}(S) = \mathcal{B}(\text{pr}_G)$, then each of the three statements given above is equivalent to

(d) $S$ is sufficient for $\mathcal{P}$.

Proof. First, let us note that for a $P_0$-integrable function $f$ holds true:

$$\forall g \in G: \int f(x) \, P_0(dx) = \int f(x) \cdot (g^{-1} \cdot g) \, P_0(dx) = \int f(g^{-1}(x)) \, P_g(dx).$$

(a) $\Rightarrow$ (b): Let $A \in \mathcal{B}_G$, $D \in \mathcal{D}$, and $g \in G$ be arbitrarily fixed. Then we have

$$P_0(A \cap S^{-1}(D)) = \int 1_A(x) \cdot 1_{S^{-1}(D)}(x) \, P_0(dx)$$

$$= \int 1_A(g^{-1}x) \cdot 1_{S^{-1}(D)}(g^{-1}x) \, P_0(dx)$$

$$= \int 1_A(x) \cdot 1_{gS^{-1}(D)}(x) \, P_0(dx)$$

$$= \int E(1_A \mid S)(x) \cdot 1_{gS^{-1}(D)}(x) \, P_0(dx)$$

$$= \int h \cdot S(x) \cdot 1_{gS^{-1}(D)}(x) \, P_0(dx)$$

$$= \int h \cdot S(x) \cdot 1_{S^{-1}(D)}(g^{-1}x) \, P_0(dx)$$

$$= \int h \cdot S(gx) \cdot 1_{S^{-1}(D)}(x) \, P_0(dx),$$

where $h \cdot S(x) = E(1_A \mid S)(x)$.
Let \( x \in \mathcal{X} \) be arbitrarily fixed with \( \text{pr}_G x = \ell x_0 \) (\( \ell \in G \)), hence \( \text{pr}_C x = \ell^{-1} x \). For \( y \in \mathcal{X} \) let \( \text{pr}_G y = gx_0 \), \( c = \text{pr}_C y \). By the above observation we get for a \( G \)-equivariant statistic \( S \):

\[
P_0(A) = \int_{\mathcal{X}} h \cdot S(\ell y) \, P_0(dy)
\]

\[
= \int_{\mathcal{G}^0 \times C} h \cdot S(\phi(g x_0, c)) \varphi^{-1}(P_0)(d(g, c))
\]

\[
= \int_{\mathcal{G}^0} h \cdot S(\phi(g x_0, \ell^{-1} x)) \varphi^{-1}(P_0)(d(g, c))
\]

\[
= \int_{\mathcal{G}^0} h \cdot S(\phi(g x_0, \ell^{-1} x)) \pi(c)(dg)
\]

Thus we obtain

\[
P_0(A \cap S^{-1}(D)) = \int_{\mathcal{G}^0} P_0(A \cap S^{-1}(D)) \pi(c)(dg)
\]

\[
= \int_{\mathcal{G}^0} \int_{\mathcal{X}} h \cdot S(g x) \cdot 1_{S^{-1}(D)}(x) \, P_0(dx) \pi(c)(dg)
\]

\[
= \int_{\mathcal{X}} \left( \int_{\mathcal{G}^0} h \cdot S(g x) \pi(c)(dg) \right) 1_{S^{-1}(D)}(x) \, P_0(dx)
\]

\[
= P_0(A) \cdot P_0(S^{-1}(D)).
\]

(b) \(\implies\) (c): For \( B \in \mathcal{G} \), \( A \in \mathcal{B}(S) \), and \( g \in G \) we have

\[
P_g(B \cap A) = P_0(g^{-1}(B \cap A)) = P_0(B \cap g^{-1}A)
\]

\[
= P_0(B) \cdot P(g^{-1}A) = P_0(B) \cdot P_0(A).
\]

(c) \(\implies\) (a): For \( B \in \mathcal{B}_G \) and \( g \in G \) we have

\[
P_g(B \mid \mathcal{B}(S)) = P_0(B) = P_0(B),
\]

implying that \( S \) is \( G \)-invariantly sufficient.

Finally it is sufficient to show (c) \(\implies\) (d): Because \( \phi \) is a homeomorphism we have

\[
\mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathcal{B}(\text{pr}_C) \cup \mathcal{B}(\text{pr}_G)) = \mathcal{B}(\mathcal{B}_G \cup \mathcal{B}(S)),
\]
where $\mathcal{B}(\mathcal{C})$ is the $\sigma$-algebra induced by $\mathcal{C}$. Therefore assumption (c) implies that $S$ is sufficient for $\mathcal{P}$ by Basu’s Theorem; see Barra [2, p. 26, Theorem 3].

Proof of Corollary 2.1. All assumptions of Theorem 3.1 are fulfilled if $G = (V, +)$, and $C = W$ with $V \oplus W = \mathbb{R}^n$; see the example considered in Section 2.2. Further $pr_{V|W}$ is a maximal nice $V$-equivariant statistic. Thus (a)–(d) of Corollary 2.1 are equivalent for $pr_{V|W}$ instead of $pr_{V|X|V^+}$.

For showing the last statement note that for $W$ with $V \oplus W = \mathbb{R}^n$,

$$\mathcal{B}(pr_{V|W}) = \mathcal{B}_V.$$

Next we show that statement (d) with $pr_{V|W}$ instead of $pr_{V|X|V^+}$ cannot be true if $W \neq \Sigma(V^+)$. Note that statement (d) implies then that $pr_{V|W} Y$ and $pr_{W|V} Y$ are uncorrelated; but

$$\text{Cov}(pr_{V|W} Y, pr_{W|V} Y) = pr_{V|W} \Sigma pr_{W|V}^T = 0$$

if and only if $W = \Sigma(V^+)$.  

3.2. Normal Distribution and Sufficiency

In this section we consider an abstract linear model; for more information on such models see Eaton [6]. Let $(H, (\cdot, \cdot))$ be a Hilbert-space of dimension $n$, let $Z$ be a random variable with values in $(H, \mathcal{B})$, where $\mathcal{B} = \mathcal{B}(H)$ is the Borel-$\sigma$-algebra of $H$, and let $V$ be a subspace of $H$ with $0 < \dim(V) \leq n$. Further we assume

$$EZ = 0; \quad \text{that is, } \forall a \in H : E(a, Z) = 0.$$

The expectation of $Z$ is independent of the inner product. Next, we assume that $\text{Cov} Z$ exists. The covariance is the (unique with respect to $(\cdot, \cdot)$) self-adjoint linear mapping $C : H \rightarrow H$ fulfilling $\forall a, b \in H : \text{Cov}((a, Z), (b, Z)) = (a, Cb)$.

We assume that $C$ is positive definite.

Note that the covariance depends on the inner product. Moreover, we can choose the inner product in such a way that $\text{Cov} Z = \text{id}_H$ (identity mapping of $H$).

Therefore we can assume without loss of generality that

$$Z$$

is a random variable with values in $(H, (\cdot, \cdot))$ and with

$$EZ = 0, \quad \text{Cov} Z = \text{id}_H.$$
Let $U$ be a subspace of $H$. In the sequel we use the notation

$$U^\perp = \{ a \in H | (a, u) = 0 \text{ for all } u \in U \}$$

and the orthogonal projector onto $U$ with respect to $(\cdot, \cdot)$.

**Example.** Let $\mathbb{R}^n$ be equipped with the usual inner product $[\cdot, \cdot]$. Given a random vector $Z$ with $\text{Cov} \ Z = C$ (positive definite) then

$$\text{Cov} \ Z = I_n \quad \text{with respect to the inner product} \quad (\cdot, \cdot) = [\cdot, C^{-1} \cdot].$$

Further we have

$$U^\perp = CW \quad \text{where} \quad W = \{ a \in H | \forall u \in U : [a, u] = 0 \}$$

and

$$\text{pr}_U = X(X^\top C^{-1}X)^{-1} X^\top C^{-1} \quad \text{where the columns of} \ X \ \text{are a basis of} \ U.$$

Next we consider the linear model

$$Y := \theta + Z, \quad \theta \in V, \quad \text{with} \ \text{E} Z = 0, \ \text{Cov} \ Z = \text{id}_H,$$

where $\theta$ is the unknown parameter vector. As in Section 2.2 we consider the $V$-location family

$$\mathcal{P} := \{ P_\theta | \theta \in V \}.$$

Now we are in the position to state our main result.

**Theorem 3.2.** Given the linear model in $(H, (\cdot, \cdot))$ described above, let $b_1, \ldots, b_n$ be an orthonormal basis, let

$$b_1, \ldots, b_r \in V,$$

$$b_{r+1}, \ldots, b_s \in V^\perp,$$

$$b_{s+1}, \ldots, b_n \notin V \cup V^\perp,$$

and let $\{(b_1, Z), \ldots, (b_r, Z)\}, \{(b_{r+1}, Z), \ldots, (b_s, Z)\}, \{(b_{s+1}, Z)\}, \ldots, \{(b_n, Z)\}$ be stochastically independent. Then the following statements are equivalent:

(a) $\text{pr}_V$ is sufficient for $\mathcal{P}$,

(b) $\forall i \in \{s+1, \ldots, n\} : (b_i, Z)$ is normally distributed.

**Proof.** First we define

$$H_1 := \text{span} \{b_1, \ldots, b_r, b_{r+1}, \ldots, b_s\},$$

$$H_2 := \text{span} \{b_{s+1}, \ldots, b_n\},$$
and we get the direct sum

\[ H = H_1 \oplus H_2 \quad \text{with} \quad H_1 \perp H_2. \]

Further let

\[ b_1, \ldots, b_r, u_{r+1}, \ldots, u_m \] be an orthonormal basis of \( V \),

\[ b_{r+1}, \ldots, b_s, w_{s+1}, \ldots, w_J \] be an orthonormal basis of \( V^\perp \)

then we have

\[ u_{r+1}, \ldots, u_m, w_{s+1}, \ldots, w_J \] is an orthonormal basis of \( H_2 \).

\[(a) \Rightarrow (b): \text{Let } b_j \notin V \cup V^\perp. \text{ Then it follows that}\]

\[ \exists v \in V \cap H_1^\perp : (b_j, v) \neq 0, \]

\[ \exists w \in V^\perp \cap H_1^\perp : (b_j, w) \neq 0. \]

By Theorem 3.1 we have

\[ L_1 := (v, Z) \quad \text{and} \quad L_2 := (w, Z) \text{ are stochastically independent.} \]

Using that \( b_1, \ldots, b_n \) is an orthonormal basis of \((H, (\cdot, \cdot))\) we have the following description of \( L_1 \) and \( L_2 \):

\[ L_1 = \sum_{k=r+1}^{n} (b_k, v)(b_k, Z), \quad L_2 = \sum_{k=s+1}^{n} (b_k, w)(b_k, Z). \]

Thus by Darmois and Skitovich's theorem it follows that \((b_i, Z)\) is normally distributed.

\[(b) \Rightarrow (a): \text{By the assumptions we have}\]

\[ \Pr_{H_2} Z \text{ is normally distributed with } \text{id}_{H_2} \text{ as covariance.} \]

Therefore \((u_{r+1}, Z)u_{r+1}, \ldots, (u_m, Z)u_m, (w_{s+1}, Z)w_{s+1}, \ldots, (w_J, Z)w_J\) are independent and so

\[ \Pr_{V} Z, \Pr_{V^\perp} Z \text{ are independent.} \]

Thus \((a)\) follows by Theorem 3.1. \qed

**Proof of Corollary 2.4.** We show that the assumptions of Theorem 3.2 are fulfilled. We choose

\[ (H, (\cdot, \cdot)) = (\mathbb{R}^n, [\cdot, (BB^T)^{-1} \cdot]), \]
where \([x, y] = x^T y\), \(x, y \in \mathbb{R}^n\). Then \(\{b_1, ..., b_n\}\) is an orthonormal basis.

Because of

\[
\forall i \in \{1, ..., n\} : b_i^T (BB^T)^{-1} Z = b_i^T B^{-1} = e_i^T \eta = \eta_i
\]

all assumptions of Theorem 2.3 are fulfilled.

REFERENCES