

BAYESIAN INFERENCE IN LINEAR RELATIONS

BY C. VILLEGAS¹

Simon Fraser University

1. Introduction. The statistical analysis of a linear relation among several unobserved variables, when the observations are all subjected to error, has a long history, going back to a paper by R. J. Adcock published in 1877. The early writers on this subject, notably R. J. Adcock (1878), C. H. Kummel (1879) Karl Pearson (1901) and M. J. van Uven (1930) were mainly concerned with the derivation of least squares estimators. Modern statistical methods were used for the first time in the analysis of linear relations by A. Wald in paper published in 1940. As was pointed out by J. Neyman in 1937, if no replications are available and the errors and the unobserved variables are independent and have Gaussian distributions, then the linear relation may be unidentifiable in the sense that even a complete knowledge of the sampling distributions of the observed random variables is not sufficient to determine the linear relation, because the number of parameters in the linear relation model is, in such cases, greater than the number of parameters which determine the sampling distributions. If we try to solve this problem by Bayesian methods, we cannot expect that the posterior distribution will be consistent in the sense of converging to the true values of the parameters when the sample size increases indefinitely. However, the assumption that the unobserved variables are independent is unrealistic in many cases, particularly in time series analysis, and better alternative models are needed. Of course, under certain additional conditions there are no problems of identifiability. Thus, J. Kiefer and J. Wolfowitz (1956) showed that, under certain conditions of identifiability, when no replications are available and the unobserved variables have probability distributions, the method of maximum likelihood, if properly applied, will yield not only consistent estimators of the linear relation, but, what is even more remarkable, it will yield also consistent estimators of the probability distribution of the unobserved random variables. However, Kiefer and Wolfowitz do not give explicit expressions for the maximum likelihood estimators.

In experimental work it is usually possible to replicate the observations. Replicated experiments can be analyzed without great difficulties, because we can easily obtain from them estimators of the experimental errors which have known distributions. The case in which replicated observations are available was considered systematically for the first time by G. W. Housner and J. F. Brennan (1948), but the first important work on this case was done by John W.

Received February 11, 1970; revised April 10, 1972.

¹ This paper was supported in part by the Systems Analysis Program, The University of Rochester, under U.S. Bureau of Naval Personnel Contract Number N 00022-69-C-0085, while the author was visiting at The University of Rochester.

Tukey (1951), who showed how estimators of the linear relation may be easily derived from a variance component analysis. In a series of recent papers Villegas proved that an estimator previously proposed by Acton (1959) was the maximum likelihood estimator (1961); obtained invariant least squares estimators (1963) and confidence regions (1964); proved that the least squares estimators are asymptotically efficient (1966) and considered the case of non-linear relations (1969). The problems of testing the linearity of a relation and testing for simultaneous linear relations have been considered, respectively, by C. R. Rao (1965) and A. P. Basu (1969). A more complex model with strongly correlated observations was considered by Sprent (1966). The case when no replications are available has been considered recently by Lindley and El-Sayyad (1968) using Bayesian methods and by Kalbfleisch and Sprott (1970), using likelihood methods. For a more detailed survey of the extensive literature available on the analysis of linear relations, the interested reader is referred to a paper by Madansky (1959).

In this paper a Bayesian statistical theory of linear relations will be developed for the case in which replicated observations are available, but assuming no additional prior knowledge about the linear relation. The prior to be used must be therefore a prior representing ignorance. In a previous paper [29] it was argued that, if the parameter is an element of a group of transformations of the sample space, and if this group has an invariant Haar measure, then this Haar measure is the prior which represents ignorance. However, the prior used in this paper is not a single Haar measure but rather a product of Haar measures of several groups of transformations of the sample space. The results obtained in this paper remain valid when no replications are available, provided that there is some additional prior knowledge compensating for the lack of replication.

Some results of a purely mathematical nature which are used in this paper have been collected in an Appendix. The first Section of this Appendix is devoted to the theory of invariant measures in homogeneous spaces. In this section we review the more relevant results of this theory, giving a new derivation of the invariant measure for hyperplanes. For a more detailed account of this important and interesting theory, the reader is referred to Nachbin [16].

2. Notation and model. An affine linear relation

$$(2.1) \quad \beta_1 x^{(1)} + \dots + \beta_p x^{(p)} + \delta = 0$$

among p real variables $x^{(1)}, \dots, x^{(p)}$ may be represented geometrically, in a p -dimensional vector space \mathcal{R} , by a hyperplane, or, in other words, by a $(p - 1)$ -dimensional affine subspace. To simplify our analysis, it will be convenient to assume that the vector space \mathcal{R} has been equipped with an inner product and with an orthonormal basis o_1, \dots, o_p arbitrarily chosen. If the coefficients of (2.1) are normalized by $\sum \beta_j^2 = 1$, then δ is the (signed) distance from the origin to the hyperplane.

In general, an m -dimensional affine subspace, or, briefly, an m -dimensional

flat is the image of an m -dimensional subspace H under a translation $x \rightarrow x + a$, and may be denoted by $H + a$. Thus

$$(2.2) \quad H + a = \{x + a : x \in H\}.$$

Clearly, if $x \in H + a$, then $H + x$ denotes the same flat. Therefore the vector a in the notation $H + a$ is not uniquely determined. However, there is one and only one vector lying on the flat and orthogonal to it, and, if a is this vector, then the flat will be denoted by $H \oplus a$, where the special symbol \oplus is used to emphasize the fact that a is orthogonal to H and is therefore uniquely determined. We can think of \oplus as a relation from the space \mathcal{H} of all m -dimensional subspaces to the vector space \mathcal{R} . The graph of this relation, that is, the set of all points (H, a) of the Cartesian product $\mathcal{H} \times \mathcal{R}$ such that a is orthogonal to H , will be denoted by $\mathcal{H} \oplus \mathcal{R}$. There is a natural bijective (i.e., one-to-one) correspondence between $\mathcal{H} \oplus \mathcal{R}$ and the space of all m -dimensional flats, and we shall use the same notation $H \oplus a$ to denote a hyperplane and the corresponding point of $\mathcal{H} \oplus \mathcal{R}$. The space of all m -dimensional subspaces of \mathcal{R} , equipped with its natural topology, is called a *Grassmann manifold*. In the particular case $m = 1$, we have the space of all lines going through the origin, which is called the $(p - 1)$ -dimensional projective space. Arbitrary m -dimensional flats will be considered only in the Appendix; in all the remaining sections of this paper we shall consider only the case $m = p - 1$ in which the flats are called hyperplanes. By duality, the space \mathcal{H} of all hyperplanes H going through the origin is also a projective space.

Suppose that, in order to obtain information about a hyperplane $H \oplus a$, a replicated experiment has been performed, and that, from a preliminary statistical analysis, we have derived estimates y_1, \dots, y_n of n points x_1, \dots, x_n lying on $H \oplus a$. Then, more precisely, our model will be

$$(2.3) \quad \Gamma(y_i - x_i) = v_i \quad (i = 1, \dots, n),$$

where: (i) the observed random vectors y_i have unknown mean values x_i which lie on a hyperplane $H \oplus a$ but do not lie on any other flat of smaller dimension; (ii) the v_i are independently and identically distributed random vectors which have the standard p -dimensional Gaussian distribution; and (iii) the unknown parameter Γ is, with respect to the given basis, a positive upper triangular transformation, i.e., it has, with respect to the given basis, a matrix which is upper triangular and has positive diagonal elements. The parameter Γ will be called, simply, the *linear parameter*. Note that, with probability 1, the observed vectors y_i do not lie on any proper flat; and, in what follows, we assume that this is true for the realized vectors.

We assume, in addition, that from the same preliminary statistical analysis we have derived also a positive definite transformation W which is independent from the y_i and has the usual Wishart distribution with $q \geq p$ degrees of freedom and parameter $\Sigma = (\Gamma'\Gamma)^{-1}$, where Γ' denotes the adjoint of Γ . This transformation W may be called a *Wishart transformation*. When no replications are

available, the transformation W does not exist. However, the results obtained in this paper will also be valid in this case, provided that the lack of replications is compensated by an additional prior knowledge equivalent to the observation of a Wishart transformation W .

In some cases it is known that the affine linear relation is homogeneous, i.e., that $a = 0$. This happens for instance when the experiment has a block design and the y_i are the estimates of the intrablock effects (see [24] page 1056). Although the estimates of the intrablock effects are not independent, this case can be brought back to the case of independent y_i considered in this paper by a linear transformation.

3. The likelihood function. As was said before, it will be convenient to assume that the vector space \mathcal{R} has been equipped with an arbitrarily chosen inner product, becoming in this way a Euclidean space. As usual, the inner product of two vectors $x, y \in \mathcal{R}$ will be denoted by (x, y) and the norm of a vector x will be denoted by $\|x\|$. If $A: \mathcal{R} \rightarrow \mathcal{R}$ is a bijective (i.e., a one-to-one) linear transformation, then a new inner product $(x, y)_A$ and a new norm $\|x\|_A$ called the A -inner product and the A -norm, may be defined by

$$(3.1) \quad (x, y)_A = (Ax, Ay), \quad \|x\|_A = \|Ax\|.$$

It can be shown that any inner product can be defined in this way, using a suitably chosen A . The vector space \mathcal{R} , equipped with the new A -inner product, becomes a new Euclidean space \mathcal{R}_A and (3.1) shows that the transformation $A: \mathcal{R}_A \rightarrow \mathcal{R}$ is an isometry. The usual Euclidean measures in \mathcal{R} and in \mathcal{R}_A of a measurable set E will be denoted by $|E|$ and by $|E|_A$ and will be called, respectively, the measure and the A -measure of E . Clearly

$$(3.2) \quad |E|_A = |AE| = |A| |E|,$$

where $|A|$ denotes the absolute value of the determinant of A . We shall say that a random vector y has the A -standard Gaussian distribution in \mathcal{R} if the random vector $v = Ay$ has the standard Gaussian distribution in \mathcal{R} . The differential form of the A -standard Gaussian distribution is proportional to

$$\exp[-\frac{1}{2}\|y\|_A^2] dy|_A$$

where $dy|_A$ denotes the differential form of the A -measure in \mathcal{R} . By (3.2) we have

$$dy|_A = |A| dy$$

and therefore the density of the A -standard Gaussian distribution is proportional to

$$|A| \exp -\frac{1}{2}\|y\|_A^2.$$

This distribution may be called also the Gaussian distribution with zero mean and linear parameter A . It is clear now that the joint density of the random variables y_i in the linear relation model is proportional to

$$(3.3) \quad |\Gamma|^n \exp -\frac{1}{2}Q_\Gamma,$$

where

$$(3.4) \quad Q_{\Gamma} = \sum_{i=1}^n \|y_i - x_i\|_{\Gamma}^2.$$

By assumption, the transformation W has a Wishart distribution with parameter $\Sigma = (\Gamma'\Gamma)^{-1}$ and $q \geq p$ degrees of freedom. Let us denote by $\text{tr } A$ the trace of a linear transformation $A: \mathcal{R} \rightarrow \mathcal{R}$. Then the sampling density of W is proportional to

$$(3.5) \quad |W|^{(q-p-1)/2} |\Gamma|^q \exp -\frac{1}{2} \text{tr } \Gamma' \Gamma W$$

for W positive definite and is equal to zero otherwise. The likelihood is proportional to the product of (3.3) and (3.5) or, equivalently, it is proportional to

$$(3.6) \quad |\Gamma|^N \exp -\frac{1}{2}(Q_{\Gamma} + \text{tr } W \Gamma' \Gamma),$$

where $N = q + n$. It is well known that the likelihood is also the posterior density, with respect to any given prior measure, of the corresponding posterior distribution of the parameters $\Gamma, H \oplus a, x_1, \dots, x_n$.

4. Prior measures. It will be convenient to substitute new parameter vectors \tilde{x}_i defined by

$$(4.1) \quad \tilde{x}_i = \Gamma x_i \quad (i = 1, \dots, n).$$

Then the model equation (2.3) becomes

$$(4.2) \quad \Gamma y_i - \tilde{x}_i = v_i \quad (i = 1, \dots, n).$$

The new parameters \tilde{x}_i lie on a hyperplane, which we shall denote by $\tilde{H} \oplus \tilde{a}$. Let \mathcal{M} be a $(p-1)$ -dimensional subspace which has been arbitrarily chosen, and let O be an orthogonal transformation such that

$$(4.3) \quad O\mathcal{M} = \tilde{H}.$$

Then the vectors $O'\tilde{x}_i$ lie on the hyperplane $\mathcal{M} \oplus c$, where

$$(4.4) \quad c = O'\tilde{a}$$

belongs to the orthogonal complement \mathcal{M}^{\perp} of \mathcal{M} , and we can write

$$(4.5) \quad O'\tilde{x}_i = u_i + c$$

where u_i is the projection of $O'\tilde{x}_i$ on \mathcal{M} . Since the p -dimensional standard Gaussian distribution is invariant under orthogonal transformations, we can write our model equation as

$$(4.6) \quad Ay_i - u_i - c = v_i$$

where

$$(4.7) \quad A = O'\Gamma$$

is a bijective linear transformation. Let \mathcal{N} be an auxiliary n -dimensional

Euclidean space, with an orthonormal basis e_1, \dots, e_n and define the linear transformations Y, V from \mathcal{N} to \mathcal{R} and $U: \mathcal{N} \rightarrow \mathcal{M}$ by

$$(4.8) \quad Ye_i = y_i, \quad Ve_i = v_i, \quad Ue_i = u_i.$$

Then our model equation becomes

$$(4.9) \quad AY - U - C = V$$

where $C: \mathcal{N} \rightarrow \mathcal{M}^\perp$ is the constant transformation whose value is the point c of the orthogonal complement \mathcal{M}^\perp of \mathcal{M} , which, of course, is a line going through the origin and orthogonal to \mathcal{M} . Let \mathcal{V} be the space of all linear transformations $Y: \mathcal{N} \rightarrow \mathcal{R}$, let \mathcal{A} be the (multiplicative) group of all bijective linear transformations $A: \mathcal{R} \rightarrow \mathcal{R}$, let \mathcal{U} be the additive group of all linear transformations $U: \mathcal{N} \rightarrow \mathcal{M}$, and let \mathcal{C} be the additive group of all constant transformations $C: \mathcal{N} \rightarrow \mathcal{M}^\perp$. The three groups of parameters \mathcal{A} , \mathcal{U} and \mathcal{C} can be considered as groups of operators acting on the sample space \mathcal{V} according to the obvious composition laws $(A, Y) \rightarrow AY$, $(U, Y) \rightarrow Y - U$ and $(C, Y) \rightarrow Y - C$. The groups \mathcal{U} and \mathcal{C} are commutative and therefore unimodular, i.e., each of them has, up to an arbitrary scale factor, only one Haar measure [16]. It is well known that, although the group \mathcal{A} is not commutative, it is also unimodular. The Haar measures on the three groups of parameters are then the unique measures which are naturally associated with these groups. Therefore, for a person who does not have any additional knowledge concerning the values of the parameters, it would be natural to choose as a joint prior measure for the parameters A, U and C , the product of the Haar measures on the three groups. Clearly the Haar measure on the group \mathcal{U} has differential form $du_1 \dots du_n$, where du_i is the differential form of the usual measure on \mathcal{M} . By Proposition 2.1 of [29], the Haar measure for $A = \Gamma'O$ is the product of the invariant measure φ for O and a right invariant Haar measure τ^+ for Γ . Thus, the differential form of the prior measure for all parameters $\Gamma, O, c, u_1, \dots, u_n$ is

$$(4.10) \quad \tau^+(d\Gamma)\varphi(dO) dc du_1 \dots du_n.$$

We have to derive from (4.10) the corresponding prior measure for the former parameters $\Gamma, H \oplus a, x_1, \dots, x_n$. If $h = H^\perp$ is the line that goes through the origin and is orthogonal to a hyperplane $H \in \mathcal{H}$, and E is any measurable set in \mathcal{R} , we shall define

$$(4.11) \quad \nu(H, E) = |h \cap E|^{(1)}$$

where $|h \cap E|^{(1)}$ denotes the usual one-dimensional measure of $h \cap E$ considered as a subset of the line h . For a fixed E , $\nu(\cdot, E)$ will denote the function defined on \mathcal{H} whose value at $H \in \mathcal{H}$ is $\nu(H, E)$. Similarly, for a fixed $H \in \mathcal{H}$, $\nu(H, \cdot)$ denotes the function defined on the σ -algebra of measurable sets of \mathcal{R} whose value at the measurable set E is $\nu(H, E)$. Clearly, for any measurable set E , $\nu(\cdot, E)$ is a measurable function in \mathcal{H} , and, for any $H \in \mathcal{H}$, $\nu(H, \cdot)$ is a measure defined on the σ -algebra of measurable sets of \mathcal{R} , which is concentrated on the line

$h = H^\perp$. A function ν of two arguments, which is a measurable function when it is considered only as a function of one of them (for a fixed value of the other) and is a measure when it is considered as a function of the second argument (for a fixed value of the first), will be called, in this paper, a *partial measure*.

Consider the substitutions $c \rightarrow \bar{a} = Oc$, $u_i \rightarrow \tilde{x}_i = O(u_i + c)$. Then, clearly, the prior measure for the parameters Γ , O , \bar{a} , $\tilde{x}_1, \dots, \tilde{x}_n$ will be

$$(4.12) \quad \tau^+(d\Gamma)\varphi(dO)\nu(\tilde{H}, d\bar{a})|d\tilde{x}_1|^{\tilde{H} \oplus \bar{a}} \dots |d\tilde{x}_n|^{\tilde{H} \oplus \bar{a}},$$

where \tilde{H} is defined by (4.3), and, for a given $\tilde{H} \oplus \bar{a}$, $|d\tilde{x}_i|^{\tilde{H} \oplus \bar{a}}$ denotes the differential form of the usual measure on the hyperplane $\tilde{H} \oplus \bar{a}$. Let $\tilde{X}: \mathcal{N} \rightarrow \mathcal{R}$ be the linear transformation defined by $\tilde{X}e_i = \tilde{x}_i$. Then the prior measure (4.12) can be written in a more compact notation as

$$(4.13) \quad \tau^+(d\Gamma)\varphi(dO)\nu(\tilde{H}, d\bar{a})|d\tilde{X}|^{\tilde{H} \oplus \bar{a}}$$

where $|d\tilde{X}|^{\tilde{H} \oplus \bar{a}} = |d\tilde{x}_1|^{\tilde{H} \oplus \bar{a}} \dots |d\tilde{x}_n|^{\tilde{H} \oplus \bar{a}}$.

Let \mathcal{S} be the unit sphere in \mathcal{R} , that is, the set of all vectors b such that $\|b\| = 1$. For any measurable subset E of \mathcal{S} , let $|E|^{(p-1)}$ denote the usual $(p-1)$ -dimensional measure of E . Let \mathcal{L} be the space of all lines l going through the origin. Any set E in \mathcal{L} can be considered also in a natural way as a set in \mathcal{R} (i.e., as the set of all points of all lines of E), and therefore it has an intersection $E \cap \mathcal{S}$ with \mathcal{S} . If this intersection is a measurable set, then we shall say that E is measurable. We shall define a measure λ in the projective space \mathcal{L} by

$$(4.14) \quad \lambda(E) = \frac{|E \cap \mathcal{S}|^{(p-1)}}{|\mathcal{S}|^{(p-1)}}.$$

This measure is clearly normalized by $\lambda(\mathcal{L}) = 1$, and, since \mathcal{L} is a $(p-1)$ -dimensional projective space, λ may be called the *normalized projective measure* in \mathcal{L} . Given a set E in \mathcal{H} , consider the set E^\perp in \mathcal{L} defined by

$$E^\perp = \{l: l = H^\perp, H \in E\}.$$

We shall say that the set $E \subset \mathcal{H}$ is measurable if E^\perp is a measurable set, and we shall define a measure μ on \mathcal{H} by

$$(4.15) \quad \mu(E) = \lambda(E^\perp).$$

This measure is also normalized by $\mu(\mathcal{H}) = 1$, and may be called the *normalized projective measure* in \mathcal{H} . By Proposition A1.3 of the Appendix, this measure is the only normalized measure in \mathcal{H} which is invariant under orthogonal transformations. By Proposition A1.6 of the Appendix, there is a partial measure $\chi(\tilde{H}, dO)$ such that

$$(4.16) \quad \varphi(dO) = \mu(d\tilde{H})\chi(\tilde{H}, dO).$$

Substituting in (4.13), and introducing \tilde{H} as an additional parameter, it follows that the prior measure for all the parameters Γ , \tilde{H} , \bar{a} , \tilde{X} , O is

$$(4.17) \quad \tau^+(d\Gamma)\mu(d\tilde{H})\nu(\tilde{H}, d\bar{a})|d\tilde{X}|^{\tilde{H} \oplus \bar{a}}\chi(\tilde{H}, dO).$$

Let $\mu \times \nu$ be the measure in the product space $\mathcal{H} \times \mathcal{R}$ whose differential form is

$$(4.18) \quad \mu \times \nu(d(H, a)) = \mu(dH)\nu(H, da) .$$

Given a measure in the product space $\mathcal{H} \times \mathcal{R}$ a factorization like (4.18) is not uniquely determined. In effect, if we do not require μ to be normalized, then we can multiply $\mu(dH)$ by any factor $f(H)$, where f is an arbitrary positive continuous function, provided that we divide $\nu(H, da)$ by the same quantity. If we require that μ be normalized, then the function f must satisfy the condition

$$\int f(H)\mu(dH) = 1 .$$

In the particular case considered, from the definition (4.11) it follows that, for any $H \in \mathcal{H}$, $\nu(H, \mathcal{R})$ is infinite, and therefore there is no marginal measure on \mathcal{H} and we may refer simply to the measure μ as the prior measure in \mathcal{H} .

Clearly the measure $\mu \times \nu$ defined on the Cartesian product $\mathcal{H} \times \mathcal{R}$ vanishes on the complement of $\mathcal{H} \oplus \mathcal{R}$, and therefore it can be considered also as a measure in $\mathcal{H} \oplus \mathcal{R}$, or, equivalently, as a measure on the space of all hyperplanes. As will be shown in the Appendix, this measure is, up to an arbitrary scale factor, the only measure invariant under Euclidean displacements, and was found by Blaschke [5] using different methods. When we consider $\mu \times \nu$ as a measure on $\mathcal{H} \oplus \mathcal{R}$ we shall write its differential form as $\mu \times \nu(d(H \oplus a))$. Thus, the prior measure for the parameters $\Gamma, \tilde{H} \oplus \tilde{a}, \tilde{X}, O$ becomes

$$(4.19) \quad \tau^+(d\Gamma)\mu \times \nu(d(\tilde{H} \oplus \tilde{a}))|d\tilde{X}|^{\tilde{n} \oplus \tilde{a}}\chi(\tilde{H}, dO) .$$

Clearly, if it is known beforehand that the relation is homogeneous, then $\tilde{a} = 0$ and the prior measure will be simply

$$(4.20) \quad \tau^+(d\Gamma)\mu(d\tilde{H})|d\tilde{X}|^{\tilde{n}}\chi(\tilde{H}, dO) .$$

If the linear parameter Γ is known, then the prior measure for the affine case will be

$$(4.21) \quad \mu \times \nu(d(\tilde{H} \oplus \tilde{a}))|d\tilde{X}|^{\tilde{n} \oplus \tilde{a}}\chi(\tilde{H}, dO)$$

and for the homogeneous case it will be, simply,

$$(4.22) \quad \mu(d\tilde{H})|d\tilde{X}|^{\tilde{n}}\chi(\tilde{H}, dO) .$$

Let us define, for any measurable set $E \subset \mathcal{H}$, the measure μ_Γ by

$$(4.23) \quad \mu_\Gamma(E) = \mu(\Gamma E) .$$

Similarly, let $|dx|_\Gamma^{H \oplus a}$ be the differential form of the measure defined, for any measurable set $E \subset \mathcal{R}$, by

$$|E|_\Gamma^{H \oplus a} = |\Gamma(E \cap (H \oplus a))|^{(p-1)} ,$$

where the second member denotes the usual $(p-1)$ -dimensional measure of the set $\Gamma(E \cap (H \oplus a))$. Let $X: \mathcal{N} \rightarrow \mathcal{R}$ be the linear transformation defined by

$Xe_i = x_i$ and let $|dX|_{\Gamma}^{H \oplus a} = |dx_1|_{\Gamma}^{H \oplus a} \cdots |dx_n|_{\Gamma}^{H \oplus a}$. Then the prior measure for the old parameters Γ , $H \oplus a$, X , O will be, in the homogeneous case,

$$(4.24) \quad \tau^+(d\Gamma)\mu_{\Gamma}(dH)|dX|_{\Gamma}^H\chi(\Gamma H, dO).$$

Let us define also, for any measurable set $E \subset \mathcal{H} \oplus \mathcal{R}$, the measure $(\mu \times \nu)_{\Gamma}$ by

$$(4.25) \quad (\mu \times \nu)_{\Gamma}(E) = \mu \times \nu(\Gamma E).$$

Then the prior measure for the old parameters Γ , $H \oplus a$, X , O in the affine case will be

$$(4.26) \quad \tau^+(d\Gamma)(\mu \times \nu)_{\Gamma}(d(H \oplus a))|dX|_{\Gamma}^{H \oplus a}\chi(\Gamma H, dO).$$

5. Linear parameter known. In this Section we assume that the linear parameter Γ is known. The results obtained in this Section are valid also when no replications are available. They will also be valid for the case in which Γ is unknown, provided that they are interpreted only as properties of the conditional distribution of the remaining parameters given the value of Γ . We shall consider first the case of an affine relation. Since Γ is known, the likelihood is clearly given by (3.3) and (3.4). As is well known, the likelihood is proportional to the posterior density of the parameters $H \oplus a$, x_1, \dots, x_n , O with respect to the prior measure (4.21). Since the likelihood does not depend on O , it follows that the parameter O , given the values of the remaining parameters, is unidentified. We can integrate out O immediately, and it follows that (3.3) is also the posterior density of the remaining parameters, $H \oplus a$, x_1, \dots, x_n with respect to the measure

$$(5.1) \quad (\mu \times \nu)_{\Gamma}(d(H \oplus a))|dx_1|_{\Gamma}^{H \oplus a} \cdots |dx_n|_{\Gamma}^{H \oplus a}.$$

In this Section it will be convenient to use the Γ -inner product of the space \mathcal{R}_{Γ} . The orthogonal projections will be different in the spaces \mathcal{R} and \mathcal{R}_{Γ} , and they will be called, respectively, the projections and the Γ -projections. The projection and the Γ -projection of a point x on an affine subspace M will be denoted by x^M and x_{Γ}^M . Let $y_{i\Gamma}^{H \oplus a}$ be the Γ -projection of the point y_i on the hyperplane $H \oplus a$. Clearly

$$(5.2) \quad \|x_i - y_i\|_{\Gamma}^2 = \|x_i - y_{i\Gamma}^{H \oplus a}\|_{\Gamma}^2 + \|y_i, H \oplus a\|_{\Gamma}^2,$$

where $\|y_i, H \oplus a\|_{\Gamma}$ is the distance from the point y_i to the hyperplane $H \oplus a$ in the Euclidean space \mathcal{R}_{Γ} , and may be called briefly the Γ -distance from y_i to $H \oplus a$.

We shall say that a random variable x has the $(p - 1)$ -dimensional Γ -standard Gaussian distribution on a hyperplane $H \in \mathcal{H}$ if it has the $(p - 1)$ -dimensional standard Gaussian distribution on H relative to the space \mathcal{R}_{Γ} . We shall say that a random variable x has a $(p - 1)$ -dimensional Gaussian distribution on a hyperplane $H \oplus a$, with mean value $y \in H \oplus a$ and linear parameter Γ , if $x - y$ has the $(p - 1)$ -dimensional Γ -standard Gaussian distribution on H .

PROPOSITION 5.1. *A posteriori, given $H \oplus a$, the parameters x_i are independent random variables. The posterior distribution of x_i , given $H \oplus a$, is a $(p - 1)$ -dimensional Gaussian distribution on the hyperplane $H \oplus a$, which has linear parameter Γ and is centered at the Γ -projection of the point y_i .*

PROOF. By substitution of (5.2) in the posterior density (3.3) it follows that a posteriori, given $H \oplus a$, the x_i are independent, and the conditional density of x_i with respect to the $(p - 1)$ -dimensional Γ -measure on $H \oplus a$ is proportional to

$$(5.3) \quad \exp -\frac{1}{2} \|x_i - y_{i\Gamma^{H \oplus a}}\|_{\Gamma}^2,$$

and the conclusion follows immediately.

Let

$$(5.4) \quad y. = \frac{1}{n} \sum y_i$$

be the baricenter of the points y_i , and, if $A: \mathcal{R} \rightarrow \mathcal{R}$ is a linear transformation, let $\|A\|_h$ be the norm of the restriction of A to the line $h = H^\perp$.

PROPOSITION 5.2. *The conditional posterior distribution of the parameter a , given H , is a Gaussian distribution on the line $h = H^\perp$ which goes through the origin and is orthogonal to H . The mean of this distribution is the projection of the baricenter $y.$ on the line h , and its variance is $n^{-1} \|\Gamma'^{-1}\|_h^2$.*

PROOF. Integrating out the parameters x_1, \dots, x_n it follows that the posterior density of $H \oplus a$ with respect to the measure $(\mu \times \nu)_\Gamma$ is proportional to

$$(5.5) \quad \exp -\frac{1}{2} Q_\Gamma(H \oplus a)$$

where

$$(5.6) \quad Q_\Gamma(H \oplus a) = \sum_{i=1}^n \|y_i, H \oplus a\|_{\Gamma}^2$$

is the moment of inertia of the system of points y_i with respect to the hyperplane $H \oplus a$ in the metric of \mathcal{R}_Γ , or, more briefly, is the Γ -moment of inertia of the points y_i with respect to $H \oplus a$. As is well known, this is equal to the sum of the Γ -moment of inertia, with respect to $H \oplus a$, of a mass n placed at the baricenter $y.$ plus the Γ -moment of inertia of the same system of points with respect to the hyperplane $H + y.$. Thus,

$$(5.7) \quad Q_\Gamma(H \oplus a) = n \|y., H \oplus a\|_{\Gamma}^2 + Q_\Gamma(H + y.),$$

where

$$(5.8) \quad Q_\Gamma(H + y.) = \sum_{i=1}^n \|y_i, H + y.\|_{\Gamma}^2.$$

On the other hand, by Proposition A3.3 of the Appendix, the measure $(\mu \times \nu)_\Gamma$ is equal to $\mu_\Gamma \times \nu_\Gamma$, and by Proposition A3.4 of the Appendix, ν_Γ is proportional to ν . Therefore, the conditional posterior distribution of a , given H , is concentrated on the line $h = H^\perp$ and has a density proportional to

$$\exp -\frac{n}{2} \|y., H \oplus a\|_{\Gamma}^2.$$

By Proposition A2.2 of the Appendix it follows that this density is proportional to

$$(5.9) \quad \exp -\frac{n}{2} \frac{\|a - y_{\cdot}^h\|^2}{\|\Gamma'^{-1}\|_h^2}.$$

PROPOSITION 5.3 *The marginal posterior density of H , with respect to the measure μ_{Γ} is proportional to*

$$(5.10) \quad \exp -\frac{1}{2} Q_{\Gamma}(H + y_{\cdot}),$$

where $Q_{\Gamma}(H + y_{\cdot})$ is given by (5.8).

PROOF. We have to integrate out a from the joint density (5.5). By Proposition A3.4 of the Appendix, the integral of the conditional density (5.9) with respect to the measure $\nu_{\Gamma}(H, \cdot)$ does not depend on H and the conclusion follows immediately.

REMARK. Since the space \mathcal{H} of all hyperplanes going through the origin is a projective space, the distribution (5.10) may be called a $(p-1)$ -dimensional projective Gaussian distribution. Let us associate with each $H \in \mathcal{H}$ a Γ -unit vector b Γ -orthogonal to H (that is, a vector b , which, relatively to \mathcal{R}_{Γ} is a unit vector orthogonal to H). Then, by (5.8),

$$Q_{\Gamma}(H + y_{\cdot}) = \sum_{i=1}^n (y_i - y_{\cdot}, b)_{\Gamma}^2.$$

Define a transformation $\Delta Y: \mathcal{N} \rightarrow \mathcal{R}_{\Gamma}$ by

$$(5.11) \quad \Delta Y e_i = y_i - y_{\cdot}.$$

Let $\Delta Y^{\Gamma}: \mathcal{R}_{\Gamma} \rightarrow \mathcal{N}$ be the adjoint of $\Delta Y: \mathcal{N} \rightarrow \mathcal{R}_{\Gamma}$. The transformation ΔY^{Γ} will be called the Γ -adjoint of ΔY to emphasize its dependence on the Γ -inner product. Clearly

$$\begin{aligned} Q_{\Gamma}(H + y_{\cdot}) &= \sum_{i=1}^n (e_i, \Delta Y^{\Gamma} b)_{\Gamma}^2 \\ &= \|\Delta Y^{\Gamma} b\|_{\Gamma}^2 \\ &= (b, \Delta Y \Delta Y^{\Gamma} b)_{\Gamma}. \end{aligned}$$

Let $\delta_1, \dots, \delta_p$ be a Γ -orthonormal basis whose elements are eigenvectors of the Γ -selfadjoint transformation $\Delta Y \Delta Y^{\Gamma}$, and let

$$b = \sum_{j=1}^p \beta_j \delta_j.$$

Then

$$(5.12) \quad Q_{\Gamma}(H + y_{\cdot}) = \sum_{j=1}^p \eta_j \beta_j^2,$$

where the η_j are the eigenvalues corresponding to the eigenvectors δ_j . For the particular case $p = 2$, if we write $\beta_1 = \sin(\theta/2)$, then (5.10) is proportional to

$$(5.13) \quad \exp K \cos \theta \quad -\pi < \theta \leq \pi$$

where $K = (\beta_1 - \beta_2)/4$. This is the well-known *circular normal distribution*, which was introduced by von Mises (1918), tabulated by Gumbel, Greenwood and Durand (1953) and generalized to higher dimensions [6]. Returning now to the

general case, the vectors $\Delta y_i = y_i - y$ do not lie on any proper subspace because, by assumption, the vectors y_i do not lie on any single hyperplane. Therefore, by Theorem 3.5 of [27], the transformation $\Delta Y \Delta Y^\Gamma$ is Γ -positive definite and its eigenvalues are all positive. We shall assume that the basis $\delta_1, \dots, \delta_p$ has been ordered in such a way that $\eta_1 \geq \dots \geq \eta_p$ and we shall consider only the case, which happens with probability 1, in which $\eta_{p-1} > \eta_p$.

PROPOSITION 5.4. *If $\eta_1 \geq \dots \geq \eta_{p-1} > \eta_p$, then there is one and only one hyperplane \hat{H} which maximizes the posterior density (5.10). This hyperplane \hat{H} , which will be called the most likely hyperplane in \mathcal{H} , is the hyperplane that goes through the origin and is Γ -orthogonal to the eigenvector δ_p corresponding to the smallest eigenvalue η_p of the transformation $\Delta Y \Delta Y^\Gamma$.*

PROOF. Maximizing the posterior density (5.10) is equivalent to minimizing (5.12). Since $\sum \beta_j^2 = 1$, it follows that $Q_\Gamma(H + y)$ is minimized when $\beta_p = 1$ and all the other β_j are equal to zero, i.e., when $b = \delta_p$.

REMARK. Note that \hat{H} , the most likely hyperplane in \mathcal{H} , is also the maximum likelihood estimate [24].

PROPOSITION 5.5. *If it is known that the relation is homogeneous, i.e., that $a = 0$, then:*

(i) *a posteriori, given H , the parameters x_i are independent random variables and have $(p - 1)$ -dimensional Gaussian distributions on H , with linear parameter Γ and mean values equal to the Γ -projections of the points y_i ;*

(ii) *the posterior density of H with respect to the measure μ is proportional to*

$$(5.14) \quad \exp -\frac{1}{2} Q_\Gamma(H),$$

where

$$(5.15) \quad Q_\Gamma(H) = \sum_{i=1}^n \|y_i, H\|_\Gamma^2.$$

PROOF. The proof is entirely similar to the proof for the affine case and is omitted.

REMARK. Let $Y^\Gamma: \mathcal{R}_\Gamma \rightarrow \mathcal{N}$ be the Γ -adjoint of Y . Since the vectors y_i do not lie on any proper subspace, it follows that the transformation $Y Y^\Gamma$ is Γ -positive definite. Let $\delta_1, \dots, \delta_p$ be a Γ -orthonormal basis whose elements are eigenvectors of $Y Y^\Gamma$. Let us associate with each hyperplane $H \in \mathcal{H}$ a Γ -unit vector

$$b = \sum_{j=1}^p \beta_j \delta_j$$

which is Γ -orthogonal to H . Then

$$(5.16) \quad Q_\Gamma(H) = \sum_{j=1}^p \eta_j \beta_j^2,$$

where η_1, \dots, η_p are the eigenvalues of $Y Y^\Gamma$.

6. Linear parameter unknown. In this section we shall consider the case in

which the linear parameter is unknown. As before, we consider first an affine linear relation $H \oplus a$. The results obtained in the previous section are still valid in the case in which the linear parameter is unknown, provided that they are interpreted only as properties of the conditional distribution of the remaining parameters given Γ . We summarize these results in the following Proposition 6.1.

PROPOSITION 6.1. (i) *A posteriori, given Γ and $H \oplus a$, the x_i are independent random variables, which have $(p - 1)$ -dimensional Gaussian distributions in $H \oplus a$, with linear parameter Γ and mean values equal to the Γ -projections of the points y_i .*

(ii) *The conditional posterior distribution of a , given Γ and H , is a Gaussian distribution on the line $h = H^\perp$, with variance $n^{-1} \|\Gamma'^{-1}\|_h^2$ and with mean equal to the projection of the baricenter y .*

(iii) *The conditional posterior distribution of H given Γ is a projective Gaussian distribution, whose density with respect to μ_Γ , is proportional to (5.10).*

(iv) *The marginal posterior joint density of the parameters Γ , $H \oplus a$ with respect to the measure $\tau^+ \times (\mu \times \nu)_\Gamma$ is proportional to*

$$(6.1) \quad |\Gamma|^N \exp -\frac{1}{2}(Q_\Gamma(H \oplus a) + \text{tr } W\Gamma'\Gamma)$$

where $Q_\Gamma(H \oplus a)$ is given as before by (5.6).

(v) *The marginal posterior joint density of Γ and H with respect to the measure $\tau^+ \times \mu_\Gamma$ is proportional to*

$$(6.2) \quad |\Gamma|^N \exp -\frac{1}{2}(Q_\Gamma(H + y) + \text{tr } W\Gamma'\Gamma) .$$

PROPOSITION 6.2. *Let T be the positive upper triangular transformation defined by*

$$(6.3) \quad T'T = NW^{-1} .$$

Then the posterior density of $H \oplus a$, with respect to the measure $(\mu \times \nu)_T$ is proportional to

$$(6.4) \quad [N + Q_T(H \oplus a)]^{-\frac{1}{2}(N+1)}$$

where $Q_T(H \oplus a)$ is obtained from (5.6) by substitution of T for Γ .

PROOF. Let

$$(6.5) \quad Z = \Gamma T^{-1} .$$

By Proposition A3.5 of the Appendix, applied separately to $(\mu \times \nu)_\Gamma$ and $(\mu \times \nu)_T$, we get

$$(\mu \times \nu)_\Gamma(d(H \oplus a)) = \frac{|Z^{-1}|(\mu \times \nu)_T(d(H \oplus a))}{\|Z'^{-1}\|_{T\bar{h}}^{p+1}}$$

where \bar{h} is the line that goes through the origin and is T -orthogonal to H . The posterior joint density of the parameters Γ and $H \oplus a$, with respect to the measure $\tau^+ \times (\mu \times \nu)_T$ is therefore proportional to

$$(6.6) \quad \frac{|\Gamma|^{N-1}}{\|Z'^{-1}\|_{T\bar{h}}^{p+1}} \exp -\frac{1}{2}(Q_\Gamma(H \oplus a) + \text{tr } W\Gamma'\Gamma) .$$

The posterior density of $H \oplus a$, with respect to the measure $(\mu \times \nu)_T$ is the integral of (6.6) with respect to the measure τ^+ . In order to evaluate this integral, we shall make the substitution (6.5). By Proposition A2.2 of the Appendix applied separately to $Q_r(H \oplus a)$ and $Q_T(H \oplus a)$ we have

$$Q_r(H \oplus a) = \frac{Q_T(H \oplus a)}{\|Z'^{-1}\|_{T\hbar}^2}.$$

Then we have to integrate

$$(6.7) \quad \frac{|Z|^{N-1}}{\|Z'^{-1}\|_{T\hbar}^{p+1}} \exp -\frac{1}{2} \left[\frac{Q_T(H \oplus a)}{\|Z'^{-1}\|_{T\hbar}^2} + N \operatorname{tr} Z'Z \right],$$

considered as a function of Z , with respect to the measure τ^+ . When carrying out this integration, H is considered as a constant. Let $\tilde{o}_1, \dots, \tilde{o}_p$ be a new orthonormal basis in \mathcal{R} such that \tilde{o}_p is parallel to $T\tilde{h}$. It is well known that there is a unique decomposition

$$(6.8) \quad Z = O\tilde{Z}$$

where O is an orthogonal transformation and \tilde{Z} is a positive upper triangular transformation with respect to the new basis $\tilde{o}_1, \dots, \tilde{o}_p$. The transformations O and \tilde{Z} are well-defined functions of Z . The matrices $\{\tilde{z}_{jj'}\}$ and $\{\tilde{z}^{jj'}\}$ of \tilde{Z} and \tilde{Z}^{-1} with respect to the new basis are positive upper triangular and the diagonal elements are related by $\tilde{z}^{jj} = \tilde{z}_{jj}^{-1}$. Clearly

$$(6.9) \quad |Z| = |\tilde{Z}| = \prod_{j=1}^p \tilde{z}_{jj};$$

$$(6.10) \quad \operatorname{tr} Z'Z = \operatorname{tr} \tilde{Z}'\tilde{Z} = \sum_{j' \geq j} \tilde{z}_{jj'}^2.$$

We also have

$$\begin{aligned} Z'^{-1} &= O\tilde{Z}'^{-1} \\ \tilde{Z}'^{-1}\tilde{o}_p &= \tilde{z}^{pp}\tilde{o}_p \end{aligned}$$

and therefore

$$(6.11) \quad \|Z'^{-1}\|_{T\hbar} = \|\tilde{Z}'^{-1}\|_{T\hbar} = \tilde{z}_{pp}^{-1}.$$

Substituting in (6.7) we have

$$(6.12) \quad \tilde{z}_{pp}^{p+1} \prod_{j=1}^p \tilde{z}_{jj}^{N-1} \exp -\frac{1}{2} [Q_T(H \oplus a) \tilde{z}_{pp}^2 + N \sum_{j' \geq j} \tilde{z}_{jj'}^2].$$

By Proposition A3.7 of the Appendix, the corresponding measure for \tilde{Z} is again the right Haar measure τ^+ . Therefore, in order to find the posterior density of $H \oplus a$, we have to integrate (6.12) with respect to the measure τ^+ whose differential form is (see [8] page 209)

$$(6.13) \quad \prod_{j=1}^p \tilde{z}_{jj}^{-j} d\tilde{Z}.$$

The variables $\tilde{z}_{jj'}$, different from \tilde{z}_{pp} can be integrated out immediately, and in order to integrate out \tilde{z}_{pp} we have to integrate

$$\tilde{z}_{pp}^N \exp -\frac{1}{2} [N + Q_T(H \oplus a)] \tilde{z}_{pp}^2$$

as a function of \tilde{z}_{pp} . The conclusion follows immediately by the substitution

$$\tilde{z}_{pp} \rightarrow t = (N + Q_T(H \oplus a))^{\frac{1}{2}} \tilde{z}_{pp}^2.$$

PROPOSITION 6.3. *The posterior density of H with respect to the measure μ_T is proportional to*

$$(6.14) \quad (N + Q_T(H + y.))^{-\frac{1}{2}N}$$

where $Q_T(H + y.)$ is obtained from (5.8) by substitution of T for Γ .

PROOF. We have to integrate out Γ from the joint posterior distribution of Γ and H , given by Proposition 6.1. By Proposition A3.2 of the Appendix, applied, separately to μ_Γ and μ_T ,

$$\mu_\Gamma(dH) = \frac{|Z^{-1}| \mu_T(dH)}{\|Z'^{-1}\|_{T_h}^{\frac{p}{2}}},$$

where, as before, Z is defined by (6.5) and \tilde{h} is the line that goes through the origin and is T -orthogonal to H . It follows that the posterior joint density of the parameters Γ and H , with respect to the measure $\tau^+ \times \mu_T$, is proportional to

$$(6.15) \quad \frac{|\Gamma|^{N-1}}{\|Z'^{-1}\|_{T_h}^{\frac{p}{2}}} \exp -\frac{1}{2}(Q_\Gamma(H + y.) + \text{tr } W\Gamma'\Gamma).$$

The remaining part of the proof is similar to the corresponding part of the proof of Proposition 6.2 and is omitted.

PROPOSITION 6.4. *If it is known that the relation is homogeneous, i.e., that $a = 0$, then*

(i) *a posteriori given Γ and H , the x_i are independent random variables, which have $(p - 1)$ -dimensional Gaussian distribution in H , with linear parameter Γ and mean values equal to the Γ -projections of the points y_i ;*

(ii) *the conditional posterior distribution of H given Γ is a projective Gaussian distribution, whose density, with respect to μ_Γ , is proportional to (5.14);*

(iii) *the posterior density of H with respect to the measure μ_T is proportional to*

$$(6.16) \quad (N + Q_T(H))^{-\frac{1}{2}N}$$

where $Q_T(H)$ is obtained from (5.15) by substituting T for Γ .

PROOF. It is omitted.

7. No replications available. When no replications are available and we use the same prior representing ignorance, we have, instead of (6.12),

$$(7.1) \quad \tilde{z}_{pp}^{p+1} \prod_{j=1}^p \tilde{z}_{jj}^{n-1} \exp -\frac{1}{2}Q_T(H \oplus a)\tilde{z}_{pp}^2,$$

and therefore we cannot integrate out the variables \tilde{z}_{jj} . However, we can assume that we have an additional prior knowledge, equivalent to the knowledge obtained from the observation of a Wishart transformation W . More precisely, the prior measure in the affine case will be, instead of (4.26), the product of

$$(7.2) \quad |\Gamma|^q \exp -\frac{1}{2} \text{tr } W\Gamma'\Gamma$$

and (4.26), and, in the homogeneous case, it will be, instead of (4.24), the product of (7.2) and (4.24). Clearly, then, the corresponding posterior distributions, when replications are not available, will be exactly the same as before, i.e., they will be the same as if a Wishart distribution W had really been observed and no additional prior knowledge was available.

APPENDIX

A1. Invariant measures in homogeneous spaces. We say that a group \mathcal{G} operates (to the left) on a space \mathcal{H} , or is a *group of (left) operators* of \mathcal{H} , if for any $A \in \mathcal{G}$ and any $x \in \mathcal{H}$ there is an element $Ax \in \mathcal{H}$ such that

- (i) if $A, B \in \mathcal{G}$ then for any $x \in \mathcal{H}$ we have $(AB)x = A(Bx)$;
- (ii) if I is the unit element of \mathcal{G} , then for every $x \in \mathcal{H}$, $Ix = x$.

Then, clearly, any element $A \in \mathcal{G}$ induces a bijective (i.e., one-to-one) transformation of \mathcal{H} . If \mathcal{G} is a group of operators on a space \mathcal{H} , we say that \mathcal{G} *operates transitively* on \mathcal{H} if for any two points $x, y \in \mathcal{H}$ there is at least one $A \in \mathcal{G}$ such that $y = Ax$. If a topological group \mathcal{G} is a group of operators of a topological space \mathcal{H} , then we shall say that \mathcal{G} *operates continuously* on \mathcal{H} , if the mapping $(A, x) \rightarrow Ax$ of $\mathcal{G} \times \mathcal{H}$ into \mathcal{H} is continuous. Finally, we shall say that a topological space \mathcal{H} is an *homogeneous (topological) space* with respect to a topological group of operators \mathcal{G} if \mathcal{G} operates continuously and transitively on \mathcal{H} , and for any $x \in \mathcal{H}$, the transformation $A \rightarrow Ax$ from \mathcal{G} into \mathcal{H} transforms any neighborhood of the unit element in \mathcal{G} into a neighborhood of x . We shall say that a measure μ on a (measurable) space \mathcal{H} is *invariant* under a group of (measurable) operators \mathcal{G} , if for any $A \in \mathcal{G}$ we have

$$(A1.1) \quad A\mu = \mu,$$

where $A\mu$ is the measure defined, for any measurable set $E \subset \mathcal{H}$, by

$$(A1.2) \quad A\mu(E) = \mu(A^{-1}E).$$

PROPOSITION A1.1. *If the locally compact space \mathcal{H} is homogeneous under the locally compact group \mathcal{G} , then, up to an arbitrary scale factor, there is at most one measure μ on \mathcal{H} which is invariant under \mathcal{G} .*

PROOF. This Proposition is an immediate consequence of Theorem 1, page 138 of [16].

PROPOSITION A1.2. *If a locally compact space \mathcal{H} is homogeneous under the compact group \mathcal{G} , then there is one and only one (up to an arbitrary scale factor) measure μ on \mathcal{H} which is invariant under \mathcal{G} .*

PROOF. See the remark on page 140 to [16].

COROLLARY A1.1. *If a compact space \mathcal{H} is homogeneous under a compact group \mathcal{G} , then there is one and only one measure μ on \mathcal{H} which is invariant under \mathcal{G} and is normalized by $\mu(\mathcal{H}) = 1$.*

PROPOSITION A1.3. *There is only one normalized measure μ on the space \mathcal{H} of all m -dimensional subspaces of \mathcal{R} , which is invariant under the group \mathcal{O} of orthogonal transformations.*

PROOF. The Grassmann manifold \mathcal{H} is an homogeneous space under the group of orthogonal transformation \mathcal{O} (see [16] page 131, Example 7 and page 143, Example 5), and since both \mathcal{O} and \mathcal{H} are compact, by Corollary A1.1 the conclusion follows immediately. The definition of a measure $\mu \times \nu$ on the space of all hyperplanes, made in Section 4, can be generalized in a natural way to the space of m -dimensional flats. If $h = H^\perp$ is the $(p - m)$ -dimensional orthogonal complement of an m -dimensional subspace $H \in \mathcal{H}$, and E is any measurable set in \mathcal{R} , then we shall define the partial measure ν by

$$(A1.3) \quad \nu(H, E) = |h \cap E|^{(p-m)},$$

where $|h \cap E|^{(p-m)}$ denotes the $(p - m)$ -dimensional Euclidean measure of $h \cap E$ considered as a subset of h .

The measure $\mu \times \nu$ was obtained, using different methods, by Blaschke [5], who proved that it is, up to an arbitrary scale factor, the only invariant measure with respect to Euclidean displacements (see Proposition A1.5 below). Similar problems have been considered for a long time as problems of Integral Geometry [20].

LEMMA A1.1. *The measure $\mu \times \nu$ on the space of m -dimensional flats is invariant with respect to translations.*

PROOF. Given an m -dimensional subspace $H \in \mathcal{H}$, let h be the orthogonal complement, and let $\Pi_h: \mathcal{R} \rightarrow \mathcal{R}$ be the projection on h . Define a transformation $\Pi: \mathcal{H} \times \mathcal{R} \rightarrow \mathcal{H} \times \mathcal{R}$ by

$$\Pi(h, x) = H \oplus \Pi_h x.$$

If $t \in \mathcal{R}$, then the translation $x \rightarrow x + t$ induces in $\mathcal{H} \times \mathcal{R}$ a transformation T defined by

$$T(H, x) = (H, x + t),$$

and it induces in $\mathcal{H} \oplus \mathcal{R}$ a translation ΠT defined by

$$\Pi T(H \oplus a) = H \oplus \Pi_h(a + t).$$

Let \mathcal{B} be the family of Borel sets in $\mathcal{H} \times \mathcal{R}$. Then the family

$$\Pi\mathcal{B} = \{\Pi^{-1}E: E \in \mathcal{B}\}$$

is a σ -algebra of sets in $\mathcal{H} \times \mathcal{R}$ which is closed under translations. Let $\Pi^{-1}\mu \times \nu$ be the measure on the σ -algebra $\Pi\mathcal{B}$ defined for any $E \in \Pi\mathcal{B}$ by

$$\Pi^{-1}\mu \times \nu(E) = \mu \times \nu(\Pi E).$$

Clearly the measure $\Pi^{-1}\mu \times \nu$ is invariant with respect to translations and the conclusion follows immediately.

PROPOSITION A1.4. *The measure $\mu \times \nu$ on the space of m -dimensional affine subspaces is invariant with respect to the group \mathcal{D} of Euclidean displacements.*

PROOF. The partial measure ν is invariant with respect to the group \mathcal{O} of orthogonal transformations in the sense that

$$\nu(H, E) = \nu(OH, OE)$$

for any $O \in \mathcal{O}$. Since μ is invariant with respect to \mathcal{O} , it follows that $\mu \times \nu$ is invariant with respect to \mathcal{O} , and therefore, by Proposition A1.3, the conclusion follows immediately.

PROPOSITION A1.5. *The measure $\mu \times \nu$ on the space of m -dimensional affine subspaces is, up to an arbitrary scale factor, the only invariant measure with respect to the group \mathcal{D} of Euclidean displacements.*

PROOF. The conclusion follows immediately from Propositions A1.1 and A1.5 because the space of m -dimensional affine subspaces is an homogeneous space under \mathcal{D} .

Let M be an m -dimensional subspace of \mathcal{R} and let \mathcal{G} be the group of orthogonal transformations in \mathcal{R} which leave M invariant. Let H be another m -dimensional subspace of \mathcal{R} and let O_H be an orthogonal transformation such that

$$(A1.4) \quad O_H M = H.$$

The set of all orthogonal transformations which have the same property is the left coset $O\mathcal{G}$. Thus there is a natural correspondence between left cosets of \mathcal{G} , that is, elements of the homogeneous space \mathcal{O}/\mathcal{G} , and m -dimensional subspaces, which is an isomorphism with respect to the group of operators \mathcal{O} . Since the groups \mathcal{O} and \mathcal{G} are compact, they are unimodular, that is, they have, up to an arbitrary scale factor, only one Haar measure. Let \mathcal{H} be the space of all m -dimensional subspaces and let \mathcal{O}_σ be the σ -algebra of Borel sets in \mathcal{O} . Define a partial measure χ in $\mathcal{H} \times \mathcal{O}_\sigma$ by

$$(A1.5) \quad \chi(H, E) = \phi[(O_H^{-1}E) \cap \mathcal{G}]$$

where O_H is any orthogonal transformation satisfying the condition (A1.4), and ϕ is the unique normalized invariant measure on \mathcal{G} . Clearly, then, χ is also normalized in the sense that $\chi(H, \mathcal{O}) = 1$ for every $H \in \mathcal{H}$. Notice that since ϕ is an invariant measure on \mathcal{G} , the definition (A1.5) is independent of the choice of the orthogonal transformation O_H . Clearly the partial measure χ is invariant with respect to \mathcal{O} in the sense that, for any $O \in \mathcal{O}$,

$$\chi(OH, OE) = \chi(H, E).$$

PROPOSITION A1.6. *Let \mathcal{G} be the subgroup of the orthogonal transformations which leave invariant an m -dimensional subspace M , let ϕ be the normalized Haar measure on \mathcal{G} , and let χ be the normalized partial measure defined by (A1.5). If μ*

is the normalized measure in \mathcal{H} which is invariant with respect to the group of orthogonal transformations \mathcal{O} , then the normalized Haar measure φ on \mathcal{O} is given by

$$(A1.6) \quad \varphi(E) = \int \chi(H, E) \mu(dH),$$

and we shall write symbolically the differential form $\varphi(dO)$ as in (4.16).

PROOF. Since the measure defined by the second member of (A1.6) is normalized, we have to prove only that $\varphi(E) = \varphi(OE)$ for any $O \in \mathcal{O}$. Since χ is invariant,

$$\varphi(OE) = \int \chi(O^{-1}H, E) \mu(dH)$$

and by the substitution $H \rightarrow \tilde{H} = O^{-1}H$ the conclusion follows immediately.

A2. Transport of inner products. Let \mathcal{R} be as before a p -dimensional vector space which has been equipped with an arbitrarily chosen inner product π . Let $A: \mathcal{R} \rightarrow \mathcal{R}$ be a bijective (i.e., one-to-one) linear transformation. Then we can define a new inner product $A\pi$ by

$$(A2.1) \quad A\pi(x, y) = \pi(A^{-1}x, A^{-1}y).$$

The same vector space \mathcal{R} equipped with this new inner product is a new Euclidean space $A\mathcal{R}$, and the transformation $A: \mathcal{R} \rightarrow A\mathcal{R}$ is an isometry. We shall say that $A\pi$ is the inner product transported by A from \mathcal{R} into $A\mathcal{R}$. The A -inner product defined by (3.1) is the inner product transported by A^{-1} and the Euclidean space \mathcal{R}_A (equipped with the A -inner product) is equal to $A^{-1}\mathcal{R}$. The orthogonality relations in \mathcal{R}_A are, of course, different, in general, from those in \mathcal{R} , and therefore the orthogonal complements, in \mathcal{R} and in \mathcal{R}_A of the same subspace M are, in general, different subspaces M^\perp , M_A^\perp . We shall say that M_A^\perp is the A -orthogonal complement of M , and, in general, we shall say that two vectors x, y are A -orthogonal if $(x, y)_A = 0$. A subspace M is the set of all points x such that $(x, y)_A = 0$ for all $y \in M_A^\perp$. Equivalently, a subspace M is the set of all points x such that $(x, A'y) = 0$ for all $y \in M_A^\perp$. Then clearly,

$$(A2.2) \quad M^\perp = A'AM_A^\perp.$$

Since $A: \mathcal{R}_A \rightarrow \mathcal{R}$ is an isometry, it follows that AM_A^\perp is the orthogonal complement of AM , or, in symbols,

$$AM_A^\perp = (AM)^\perp$$

and, by (A2.2),

$$(A2.3) \quad (AH)^\perp = A'^{-1}H^\perp.$$

Therefore, if b is a unit vector orthogonal to H , then $A'^{-1}b$ is orthogonal to AH , $\|A'^{-1}\|_h = \|A'^{-1}b\|$ is the norm of the restriction of A to h , and

$$(A2.4) \quad \tilde{b} = \frac{A'^{-1}b}{\|A'^{-1}\|_h}, \quad h = H^\perp$$

is a unit vector orthogonal to AH .

PROPOSITION A2.1. *The image of a hyperplane $H \oplus x$ under a linear transformation $A: \mathcal{R} \rightarrow \mathcal{R}$ is the hyperplane $AH \oplus A_H x$, where $A_H: \mathcal{R} \rightarrow \mathcal{R}$ is the linear transformation defined by*

$$(A2.5) \quad A_H x = \frac{A'^{-1}x}{\|A'^{-1}\|_h^2}, \quad h = H^\perp.$$

PROOF. Let $A(H \oplus x) = \tilde{H} \oplus \tilde{x}$. Clearly $\tilde{H} = AH$, and, by (A2.4), for some scalar λ , $\tilde{x} = \lambda A'^{-1}x$. Since $A'^{-1}\tilde{x} \in H \oplus x$, it follows that $\lambda(A'A)^{-1}x \in H \oplus x$ and therefore

$$\|x\|^2 = (\lambda(A'A)^{-1}x, x) = \lambda\|A'^{-1}x\|^2, \quad \lambda = \|A'^{-1}\|_h^{-2}.$$

REMARK. The transformation $A: \mathcal{H} \oplus \mathcal{R} \rightarrow \mathcal{H} \oplus \mathcal{R}$ induced by $A: \mathcal{R} \rightarrow \mathcal{R}$ may be extended to a transformation $A: \mathcal{H} \times \mathcal{R} \rightarrow \mathcal{H} \times \mathcal{R}$ defined by $A(H, x) = (AH, A_H x)$ which is clearly bijective.

PROPOSITION A2.2. *The distance from a point x to a hyperplane $H \oplus y$ in the metric of \mathcal{R}_A , also called the A -distance from x to $H \oplus y$, is*

$$\|x, H \oplus y\|_A = \frac{\|x, H \oplus y\|}{\|A'^{-1}\|_h}, \quad h = H^\perp.$$

PROOF. Clearly

$$\begin{aligned} \|x, H \oplus y\|_A &= \|x - y, H\|_A = \|A(x - y), AH\| \\ &= (A(x - y), \tilde{b}) \end{aligned}$$

where \tilde{b} is a unit vector orthogonal to AH . By (A2.4)

$$\|x, H \oplus y\|_A = \frac{(A(x - y), A'^{-1}\tilde{b})}{\|A'^{-1}\|_h}$$

and the conclusion follows immediately

A3. Transport of measures. Given a measure μ defined on a measurable space \mathcal{H} , and a measurable transformation T from \mathcal{H} into another measurable space \mathcal{K} , we can define a measure $T\mu$ on \mathcal{K} by

$$(A3.1) \quad T\mu(E) = \mu(T^{-1}E)$$

where E is any measurable set in \mathcal{K} . The measure $T\mu$ is called the measure transported to \mathcal{K} by T . The measure μ_r defined by (4.23) is the measure transported to \mathcal{R}_r by Γ^{-1} .

LEMMA A3.1. *Let λ be the measure on the $(p - 1)$ -dimensional projective space \mathcal{L} defined by (4.14). If $A: \mathcal{R} \rightarrow \mathcal{R}$ is a bijective linear transformation then, for any measurable $E \subset \mathcal{R}$*

$$(A3.2) \quad |AE| = \int \lambda_A(dl) \int_E \|Ax\|^{p-1} \rho_A(l, dx)$$

where the first integral is over \mathcal{L} , λ_A is the measure transported by A^{-1} , and ρ_A is the partial measure defined by

$$(A3.3) \quad \rho_A(l, E) = |A(l \cap E)|^{(1)}.$$

PROOF. As in a substitution of polar coordinates we have

$$|AE| = \int \lambda(d\bar{l}) \int_{AE} \|\bar{x}\|^{p-1} \rho(\bar{l}, d\bar{x})$$

where ρ is the partial measure obtained from (A3.3) by substituting the identity transformation I for A . The conclusion follows immediately by the substitution $x \rightarrow \bar{x} = Ax$, $l \rightarrow \bar{l} = Al$.

PROPOSITION A3.1. *The density of the transported measure λ_A with respect to the measure λ is*

$$(A3.4) \quad \frac{d\lambda_A}{d\lambda} = \frac{|A|}{\|A\|_l^p}.$$

PROOF. Let B be a measurable set in \mathcal{L} , let $C = \{x: 1 \leq \|x\| \leq 1 + \Delta r\}$, where Δr is an arbitrary positive number, and let $E = B \cap C$. Then, by Lemma A3.1,

$$|AE| = \int_B \lambda_A(dl) \int_C \|Ax\|^{p-1} \rho_A(l, dx).$$

From the definition of ρ_A it follows that the density of ρ_A with respect to ρ is $\|A\|_l$. Therefore

$$|AE| = \int_B \|A\|_l \lambda_A(dl) \int_C \|Ax\|^{p-1} \rho(l, dx)$$

and

$$(A3.5) \quad \lim_{\Delta r \rightarrow 0} \frac{|AE|}{\Delta r} = \int_B \|A\|_l^p \lambda_A(dl).$$

Substituting the identity I for A ,

$$\lim_{\Delta r \rightarrow 0} \frac{|E|}{\Delta r} = \lambda(B).$$

Since $|AE| = |A| |E|$, we have

$$(A3.6) \quad \lim_{\Delta r \rightarrow 0} \frac{|AE|}{\Delta r} = |A| \lambda(B).$$

The conclusion follows immediately by a comparison of (A3.5) and (A3.6).

PROPOSITION A3.2. *The density of the transported measure μ_A with respect to the projective measure μ on the space \mathcal{H} is*

$$(A3.7) \quad \frac{d\mu_A}{d\mu} = |A|^{-1} \|A'^{-1}\|_l^{-p}, \quad l = H^\perp.$$

PROOF. Since

$$\mu_A(dH) = \lambda(A'^{-1}dl)$$

by Proposition A3.1 we have

$$\mu_A(dH) = \frac{|A'^{-1}|}{\|A'^{-1}\|_l^p} \lambda(dl) = \frac{|A'^{-1}|}{\|A'^{-1}\|_l^p} \mu(dH).$$

PROPOSITION A3.3. *Let ν_A be the partial measure defined by*

$$(A3.8) \quad \nu_A(H, E) = \nu(AH, A_H E)$$

where A_H is the linear transformation (A2.5). If $\mu \times \nu$ is the measure on $\mathcal{H} \times \mathcal{R}$ defined by (4.18), then the transported measure $(\mu \times \nu)_A$ has differential form

$$(A3.9) \quad (\mu \times \nu)_A(d(H, x)) = \mu_A(dH)\nu_A(H, dx).$$

PROOF. We have to prove that, for any measurable set $E \subset \mathcal{H} \times \mathcal{R}$,

$$(\mu \times \nu)_A(E) = \int_E \mu_A(dH)\nu_A(H, dx).$$

It will be sufficient to prove this for sets $E = A^{-1}(B \times C)$ where $B \subset \mathcal{H}$ and $C \subset \mathcal{R}$ are arbitrary measurable sets. Then

$$(\mu \times \nu)_A(E) = \mu \times \nu(B \times C) = \int_B \mu(d\tilde{H})\nu(\tilde{H}, C).$$

Making the substitution $H \rightarrow \tilde{H} = AH$,

$$\begin{aligned} (\mu \times \nu)_A(E) &= \int_{A^{-1}B} \mu_A(dH)\nu_A(H, A_H^{-1}C) \\ &= \int_{A^{-1}(B \times C)} \mu_A(dH)\nu_A(H, dx). \end{aligned}$$

PROPOSITION A3.4. *The density with respect to ν of the partial measure ν_A defined by (A3.8), is*

$$(A3.10) \quad \frac{d\nu_A}{d\nu}(H, x) = \frac{1}{\|A'^{-1}\|_h}, \quad h = H^\perp.$$

PROOF. Since $A_H h = (AH)^\perp$, by the definition (A3.8) we have

$$\begin{aligned} \nu_A(H, E) &= |A_H(h \cap E)|^{(1)} \\ &= \|A_H\|_h \nu(H, E). \end{aligned}$$

By the definition (A2.5)

$$\|A_H\|_h = \frac{1}{\|A'^{-1}\|_h}$$

and the conclusion follows immediately

PROPOSITION A3.5. *The density of the measure $(\mu \times \nu)_A$ with respect to the measure $\mu \times \nu$ is*

$$\frac{1}{|A| \|A'^{-1}\|_h^{p+1}}, \quad h = H^\perp.$$

PROOF. It follows immediately from Propositions A3.2, A3.3, and A3.4.

PROPOSITION A3.6. *The differential form of the A -measure on a hyperplane $H \in \mathcal{H}$ is*

$$(A3.11) \quad |dx|_A^H = |A| \|A'^{-1}\|_h |dx|^H, \quad h = H^\perp,$$

where $|dx|^H$ is the differential form of the usual measure on H .

PROOF. Let B be an arbitrary measurable set in H , let b be a unit vector orthogonal to H , and let E be the set

$$E = \{x + \lambda b : x \in B, 0 \leq \lambda \leq 1\}.$$

Clearly

$$\begin{aligned}|E| &= |B|^H \\ |E|_A &= |B|_A^H(Ab, \bar{b})\end{aligned}$$

where \bar{b} is a unit vector orthogonal to AH . By (3.2)

$$|B|_A^H = \frac{A}{(Ab, \bar{b})} |B|^H$$

and by (A2.4) the conclusion follows immediately.

PROPOSITION A3.7. *Let \mathcal{X} be the group of linear transformations $Z: \mathcal{R} \rightarrow \mathcal{R}$ which are positive upper triangular with respect to a basis o_1, \dots, o_p and let $\tilde{\mathcal{X}}$ be the group of linear transformations $\tilde{Z}: \mathcal{R} \rightarrow \mathcal{R}$ which are positive upper triangular with respect to another basis $\tilde{o}_1, \dots, \tilde{o}_p$. Let $\phi: \mathcal{X} \rightarrow \tilde{\mathcal{X}}$ be the measurable (but not necessarily linear) transformation defined by*

$$(A3.12) \quad Z = O\phi Z$$

where O is an orthogonal transformation which depends on Z . Let τ be a right invariant Haar measure on \mathcal{X} . Then the transported measure

$$(A3.13) \quad \tilde{\tau} = \phi\tau$$

is a right invariant Haar measure on $\tilde{\mathcal{X}}$.

PROOF. Let \mathcal{A} be the group of all bijective linear transformations $A: \mathcal{R} \rightarrow \mathcal{R}$. Let $\phi: \mathcal{A} \rightarrow \mathcal{X}$ and $\tilde{\phi}: \mathcal{A} \rightarrow \tilde{\mathcal{X}}$ be the measurable (but not necessarily linear) transformations defined by

$$(A3.14) \quad A = O\phi A, \quad A = \tilde{O}\tilde{\phi}A$$

where O and \tilde{O} are orthogonal transformations which depend on A . According to (A3.12) we have

$$(A3.15) \quad \phi A = O^*\phi\phi A$$

where O^* is an orthogonal transformation depending on ϕA , and by the first of the equations (A3.14),

$$(A3.16) \quad A = OO^*\phi\phi A.$$

Since OO^* is an orthogonal transformation a comparison of (A3.16) and the second of the equations (A3.14) shows that

$$(A3.17) \quad \tilde{\phi} = \phi\phi.$$

Let λ be an invariant Haar measure in \mathcal{A} . Then, by Proposition 2.1 of [29] the transported measures $\phi\lambda$ and $\tilde{\phi}\lambda$ are right Haar measures and therefore

$$k\tau = \phi\lambda \quad \tilde{k}\tilde{\tau} = \tilde{\phi}\lambda$$

where k and \bar{k} are positive scalars. Then by (A3.17)

$$\bar{k}\bar{\tau} = \phi\phi\lambda = k\phi\tau$$

and the conclusion follows immediately.

REFERENCES

- [1] ACTON, FORMAN S. (1959). *Analysis of Straight-Line Data*. Wiley, New York.
- [2] ADCOCK, R. J. (1877). Note on the method of least squares. *Analyst* **4** 183–184.
- [3] ADCOCK, R. J. (1878). A problem in least squares. *Analyst* **5** 53–54.
- [4] BASU, A. P. (1969). On some tests for several linear relations. *J. Roy. Statist. Soc. Ser. B* **65**–71.
- [5] BLASCHKE, W. (1935). *Integralgeometrie* **1**. Herman, Paris.
- [6] DOWNS, T. D. (1966). Some relationships among the von Mises distributions of different dimensions. *Biometrika* **53** 269–272.
- [7] GUMBEL, E. J., GREENWOOD, J. A., and DURAND, D. (1953). The circular normal distribution: theory and tables. *J. Amer. Statist. Assoc.* **48** 131–152.
- [8] HEWITT, EDWIN and ROSS, KENNETH A. (1963). *Abstract Harmonic Analysis*. Springer-Verlag, Berlin.
- [9] HOUSNER, G. W. and BRENNAN, J. F. (1948). The estimation of linear trends. *Ann. Math. Statist.* **19** 380–388.
- [10] KALBFLEISCH, JOHN D. and SPROTT, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters. *J. Roy. Statist. Soc. Ser. B* **32** 175–208.
- [11] KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.
- [12] KUMMEL, CHAS H. (1879). Reduction of observed equations which contain more than one observed quantity. *Analyst* **6** 97–105.
- [13] LINDLEY, D. V. and EL-SAYYAD, G. M. (1968). The Bayesian estimation of a linear functional relationship. *J. Roy. Statist. Soc. Ser. B* **30** 190–202.
- [14] MADANSKY, ALBERT (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Soc.* **54** 173–206.
- [15] VON MISES, R. (1918). Über die Ganzzahligkeit der Atomgewichte und verwandte Fragen. *Phys. Z.* **19** 490–500.
- [16] NACHBIN, LEOPOLDO (1965). *The Haar Integral*. Van Nostrand, Princeton.
- [17] NEYMAN, J. (1937). Remarks on a paper by E. C. Rhodes. *J. Roy. Statist. Soc.* **100** 50–57.
- [18] PEARSON, KARL (1901). On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **2** 559–572.
- [19] RAO, C. RADHAKRISHNA (1965). *Linear Statistical Inference and its Applications*. Wiley, New York.
- [20] SANTALÓ, L. A. (1953). *Introduction to Integral Geometry*. Herman, Paris.
- [21] SPRENT, P. (1966). A generalized least-squares approach to linear functional relationships. *J. Roy. Statist. Soc. Ser. B* **28** 278–297.
- [22] TUKEY, JOHN W. (1951). Components in regression. *Biometrics* **7** 33–70.
- [23] VAN UVEN, M. J. (1930). Adjustment of N points (in n -dimensional space) to the best linear $(n-1)$ -dimensional space. *K. Akad. Wetens. Amsterdam, Proc. Sec. Sciences* **33** 143–157; 307–326.
- [24] VILLEGAS, C. (1961). Maximum likelihood estimation of a linear functional relationship. *Ann. Math. Statist.* **32** 1048–1062.
- [25] VILLEGAS, C. (1963). On the least squares estimation of a linear relation. *Fac. Ingen. Agrimens. Montevideo Publ. Didáct. Inst. Mat. Estadist.* **3** 189–203.
- [26] VILLEGAS, C. (1964). Confidence region for a linear relation. *Ann. Math. Statist.* **35** 780–788.

- [27] VILLEGAS, C. (1966). On the asymptotic efficiency of least squares estimators. *Ann. Math. Statist.* **37** 1676–1683.
- [28] VILLEGAS, C. (1969). On the least squares estimation of non-linear relations. *Ann. Math. Statist.* **40** 462–466.
- [29] VILLEGAS, C. (1971). On Haar priors, in *Foundations of Statistical Inference*, ed. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston of Canada, Toronto.
- [30] WALD, ABRAHAM (1940). The fitting of straight lines if both variables are subject to error. *Ann. Math. Statist.* **11** 284–300.

DEPARTMENT OF MATHEMATICS
SIMON FRASER UNIVERSITY
BURNABY 2, BRITISH COLUMBIA