NORMAL SAMPLES WITH LINEAR CONSTRAINTS AND GIVEN VARIANCES

D. A. S. FRASER

1. Summary. In Biometrika (1948) a paper [1] by H. L. Seal contained a theorem applying to "n random variables normally distributed about zero mean with unit variance, these variables being connected by means of k linear relations." Arising from this is the question of how to obtain a set of normal variates connected by k linear relations and such that each variate has unit variance; or, more generally, connected by k linear relations and such that each variate has a given variance. The procedure for obtaining such a set of variates when existent from a set of independent normal deviates with unit variances is given in §5. In §§2, 3 and 4, we shall consider various conditions necessary for the existence and construction of such a set.

2. Normal distribution in a linear subspace. Let \((x_1, x_2, \ldots, x_n)\) designate a point in an \(n\)-dimensional Euclidean space \(R^n\). A set of variates \(x_1, x_2, \ldots, x_n\) can be considered as a random point in \(R^n\). In the present problem we shall assume the set has a multivariate normal distribution.

Consider \(k\) homogeneous and independent linear relations

\[ \sum_{j=1}^{n} a_{pj}x_j = 0 \quad (p = n - k + 1, \ldots, n). \]

A variate satisfying these relations and these only will belong to an \((n - k)\)-dimensional linear subspace. By taking linear combinations of the above \(k\) relations, an equivalent set of \(k\) relations can be obtained such that they are orthogonal and normalized:

\[ \sum_{j=1}^{n} b_{pj}x_j = 0 \quad (p = n - k + 1, \ldots, n), \]

and

\[ \sum_{j=1}^{n} b_{pj}b_{qj} = \delta_{pq}. \]

By adding \(n - k\) rows, the matrix \(|b_{pj}|\) can be completed to an \(n\) by \(n\) matrix \(|b_{ij}|\) \((i, j = 1, 2, \ldots, n)\) satisfying the orthogonality conditions

\[ \sum_{k=1}^{n} b_{ik}b_{jk} = \delta_{ij}. \]

Received July 8, 1950.

*It is to be noted that the statement of the theorem in [1] is incorrect. The theorem applies to the residuals of \(n\) normal variables after fitting \(k\) linear constraints.*

363
This matrix can now be considered as the matrix of an orthogonal rotation of \( n \)-space. Consider coordinates \( y_1, y_2, \ldots, y_n \) with respect to the new axes: then

\[
y_i = \sum_{j=1}^{n} b_{ij} x_j.
\]

Since normality is invariant under linear transformations, a set of normally distributed \( x \) variates yields a set of normally distributed \( y \) variates, and conversely. A set of variates \((x_1, x_2, \ldots, x_n)\) satisfying \( k \) linear relations

\[
\sum_{j=1}^{n} a_{pj} x_j = 0
\]

is transformed by the above to set a of variates \((y_1, y_2, \ldots, y_{n-k})\) satisfying no linear constraints where \( y_{n-k+1}, \ldots, y_n \) are identically zero.

3. Conditions on the variance. We have solved the problem of linear constraints by working in an \((n - k)\)-dimensional subspace. How do we interpret in this subspace the original variance conditions:

\[
\text{var} \{x_i\} = \sigma_i \quad (i = 1, 2, \ldots, n),
\]

with

\[
y_r = \sum_{j=1}^{n} b_{rj} x_j,
\]

\[
x_i = \sum_{r=1}^{n-k} b_{ri} y_r.
\]

Each \( x_i \) is seen to be a linear combination of the \( n - k \) variates \( y_r \) and consequently the variance of \( x_i \) can be expressed in terms of the elements of the variance covariance matrix of the \( y_r \). Consider now a multivariate normal distribution in the subspace with covariance matrix \( \| \tau_{rs} \| \) with respect to the axes \( y_1, y_2, \ldots, y_{n-k} \).

Thus the variance conditions after rotation into the subspace become

\[
\sum_{r, s=1}^{n-k} b_{ri} \tau_{rs} b_{si} = \sigma_i \quad (i = 1, 2, \ldots, n).
\]

4. Existence. Our problem has now reduced itself to that of finding a multivariate normal distribution in \( n - k \) dimensions with covariance matrix \( \| \tau_{rs} \| \) such that

\[
\sum_{r, s=1}^{n-k} b_{ri} \tau_{rs} b_{si} = \sigma_i \quad (i = 1, 2, \ldots, n),
\]

or

\[
\sum_{r, s=1}^{n-k} c_{ri} \tau_{rs} c_{si} = 1 \quad (i = 1, 2, \ldots, n),
\]

where \( c_{ri} = \sigma_i^{-1} b_{ri} \).
We have \( n \) equations with \( (n-k+1) \) unknowns. If \( (n-k+1) \geq n \), a solution will exist and can best be obtained by solving the equation directly. If \( (n-k+1) < n \), an application of linear regression theory would be indicated.

To find a matrix \( ||r_{rs}|| \), if one exists, is equivalent to finding a generalized ellipsoid

\[
\sum_{r, s=1}^{n-k} z_r r_{rs} z_s = 1
\]

passing through the \( n \) points

\((c_{1i}, \ldots, c_{n-k,i})\)

\((i = 1, 2, \ldots, n)\).

This is accomplished using linear regression theory by fitting to the constant 1 the functions \( z_r z_s \) \((r, s = 1, 2, \ldots, n-k)\) for the \( n \) "sample" values given above of the vector \((z_1, z_2, \ldots, z_{n-k})\). If the sum of squares for residuals is zero then a quadratic surface exists. However, to have a solution to our distribution problem, the matrix of the quadratic form must be positive. If it is not positive definite, then our variance conditions have imposed a further linear constraint on the set of variates.

5. Conclusions. The problem may be stated: to find normal variables \( x_1, x_2, \ldots, x_n \) satisfying \( k \) homogeneous and independent linear relations

\[
\sum_{j=1}^{n} a_{pj} x_j = 0 \quad (p = n-k+1, \ldots, n),
\]

and with

\[
\text{var} \{ x_i \} = v_i \quad (i = 1, 2, \ldots, n).
\]

The solution can be described in five steps.

5.1. Find a matrix \( ||b_{p||} \) with \( p = n-k+1, \ldots, n \) and \( j = 1, 2, \ldots, n \) with orthogonal and normalized rows equivalent to \( ||a_{p||} \) as described in §2.

5.2. Complete \( ||b_{p||} \) to an orthogonal matrix \( ||b_{ij||} \) \((i, j = 1, 2, \ldots, n)\).

5.3. Find a quadratic equation

\[
\sum_{r, s=1}^{n-k} z_r r_{rs} z_s = 1
\]

satisfied by the \( n \) points

\((b_{1i} w_i^{-\frac{1}{2}}, \ldots, b_{n-k,i} w_i^{-\frac{1}{2}})\)

\((i = 1, 2, \ldots, n)\),

if it exists, directly or by regression theory as in §4. If the equation does not exist or if it exists with a non-positive matrix then the problem has no solution.

5.4. If the matrix \( ||r_{rs}|| \) is positive, then find random variates \( y_1, y_2, \ldots, y_{n-k} \) with zero means and \( ||r_{rs}|| \) as covariance matrix. (If \( ||r_{rs}|| \) is positive
definite, take the square root matrix of \( ||r_{rs}|| \) and apply as a linear transformation to \( n - k \) independent normal variates with means 0 and variances 1. If positive but not definite, then the previous method will work in a subspace of \( y_1, y_2, \ldots, y_{n-k} \).

5.5. Obtain the set of \( x \) variates, thus solving the problem, by applying the transformation

\[
x_i = \sum_{r=1}^{n-k} b_{ir} y_r
\]

to the \( y \) variates obtained in 5.4.

References


University of Toronto