MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

THE CONDITIONAL DISTRIBUTION OF RANDOM ERRORS FOR GIVEN OBSERVED VALUES WITH APPLICATION TO FACTOR ANALYSIS

EJNAR LYTTKENS
University of Uppsala, Uppsala, Sweden

1. Scalar variables

The simplest assumption adhered to in the theory of errors is that the random error is distributed independently of the true value. It is supposed that the systematic errors are eliminated. For the random error a normal distribution with zero mean and the variance σ^2 is assumed.

The observable variable will be denoted by X, the true value by Y and the random error by Z, so that $X = Y+Z$. The conditional mean of the random error for a given observed value is given by Eddington and published by Dyson [2] and in a later paper (Eddington [3]) also the conditional variance is given:

$$E(Z|X = x) = -\sigma^2 \frac{\log f(x)}{dx}, \quad \operatorname{var}(Z|X = x) = \sigma^2+\sigma^4 \frac{d^2 \log f(x)}{dx^2},$$

where $f(x)$ is the frequency function of X. Earlier the same formulae were given by Malmqist ([8], [9]) in connection with the conditional distribution of the absolute magnitudes of stars for a given apparent magnitude. In the fifties psychometric application of these formulae were announed by McHuge ([11]) and Lord ([4]). The present author (Lyttkens [5]) has shown that the conditional characteristic function of the random error for a given observed value is $e^{-(1/2)\sigma^2t^2}f(x-i\sigma^2t)$ and from the expansion of the logarithm of this function the conditional semi-invariants are obtained. The conditional characteristic function and the conditional semi-invariants of Y for $X = x$ are then directly obtained because of the relation $Y = X-Z$. The conditional characteristic function of Z for $X \leqslant x$ is $e^{-(1/2)\sigma^2t^2}F(x-i\sigma^2t)$, where $F(x)$ is the distribution function of X, but the corresponding function pertaining to Y is more complicated (Lyttkens [5]). A simple application is the case where the non-normed distribution function of X has the form $Ke^{\beta x}$. Reference is made to Malmquist ([10]) as regards the distribution of apparent magnitudes of stars, and to Neuman and Scott ([12]) as regards field galaxies.

2. Variables in p dimensions

Let X be the observable vector, Y the true vector and Z the residual vector. The residual is assumed to be distributed independently of the true vector and the residual vector is supposed to have a multivariate normal distribution with zero mean and covariance matrix Σ. Let $f(x)$ be the frequency function of the observable vector. The conditional characteristic function of Z for $X = x$ is $e^{-(1/2)t'\Sigma t}f(x-i\Sigma t)$ and from the relation $Y = X-Z$ it follows that the characteristic function of Y for $X = x$ is $e^{it'x-(1/2)t'\Sigma t}f(x+i\Sigma t)$ (Lyttkens [6]). In view of the application aimed at in this paper, we write down the conditional mean vector and covariance matrix of Y for $X = x$:

$$E(Y|X = x) = x+\Sigma\frac{\partial \log f(x)}{\partial x}, \qquad \mathrm{Cov}(Y|X = x) = \Sigma+\Sigma\frac{\partial^2 \log f(x)}{\partial x\,\partial x'}\Sigma,$$

where $\mathrm{Cov}(Y|X = x)$ stands for the conditional covariance matrix of Y for $X = x$.

3. The factor analysis model with equal variances

Henceforth it will be assumed that all variables have zero mean. In factor analysis it is assumed that Σ is diagonal and that $Y = \Gamma V$, where the vector V has $k < p$ elements, referred to as (common) factors. At first the case $\Sigma = \sigma^2 I$ will be dealt with. Let Ω be the covariance matrix of X; then the covariance matrix of Y is $\Omega-\sigma^2 I$ and this matrix is singular with the rank k.

4. Principal components and factors

Let us replace X, Y and Z by the orthogonal transforms Ξ, H, and $\mathscr{Z}$ so that

$$\Xi = B'X, \qquad H = B'Y, \qquad \mathscr{Z} = B'Z,$$

where B is an orthogonal matrix. For any orthogonal matrix B the transformed residual vector retains the covariance matrix $\sigma^2 I$. Here B will be chosen so that Ξ is the vector of principal components of X (cf. Anderson [1]). Let Λ be the diagonal matrix of the eigenvalues of Ω. The diagonal matrix of the eigenvalues of $\Omega-\sigma^2 I$ is then $\Lambda-\sigma^2 I$ and this matrix has only k non-vanishing elements, which implies that only k elements of H have non-vanishing variances. Let H_1 be the vector of these k elements of H, while the vectors of corresponding elements of Ξ and $\mathscr{Z}$ are denoted by Ξ_1 and $\mathscr{Z}_1$.

Let $\varphi(\xi_1)$ be the frequency function of Ξ_1. The condition $X = x$ is equivalent to the condition $\Xi = \xi$, where $\xi = B'x$. However, only the components of Ξ_1 effects the conditional distribution. The conditional characteristic function of H_1 for $\Xi_1 = \xi_1$ is

$$e^{-it_1'\xi_1-(1/2)\sigma^2 t_1' t_1}\varphi(\xi_1+i\sigma^2 t).$$

Thus,

$$E(H_1|X = x) = \xi_1+\sigma^2\frac{\partial \log\varphi(\xi_1)}{\partial \xi_1}, \qquad \mathrm{Cov}(H_1|X = x) = \sigma^2 I+\sigma^4\frac{\partial^2 \log\varphi(\xi_1)}{\partial\xi_1\,\partial\xi_1'}.$$

5. Normalization of factors

Let N be the covariance matrix of H_1. This covariance matrix is diagonal and its jth diagonal element is $\lambda_j-\sigma^2$, $j = 1, \ldots, k$. Let V be the factor vector, normalized in such a way that all its components have unit variance. With $Y = \Gamma V$, we write

$$U = N^{-1}\Gamma' X, \qquad V = N^{-1}\Gamma' Y, \qquad W = N^{-1}\Gamma' Z.$$

The covariance matrix of W is $\Sigma^* = \sigma^2 N^{-1}$. Let $\psi(u)$ be the frequency function of U. With $u = N^{-1}\Gamma' x$ we have

$$E(V|X = x) = u+\Sigma^*\frac{\partial \log\psi(u)}{\partial u}, \qquad \mathrm{Cov}(V|X = x) = \Sigma^*+\Sigma^*\frac{\partial^2 \log\psi(u)}{\partial u\,\partial u'}\Sigma^*$$

or in terms of the partial derivatives of $f(x)$

$$E(V|X = x) = u+\Sigma^*\Gamma'\frac{\partial \log f(x)}{\partial x}, \qquad \mathrm{Cov}(V|X = x) = \Sigma^*+\Sigma^*\Gamma'\frac{\partial^2 \log f(x)}{\partial x\,\partial x'}\Gamma\Sigma^*.$$

Returning to the vector Y we get

$$E(Y|X = x) = \Gamma E(V|X = x), \qquad \mathrm{Cov}(Y|X = x) = \Gamma\,\mathrm{Cov}(V|X = x)\Gamma'.$$

6. Normalized factors when the residual variances are unequal

Remembering that Σ is diagonal, we introduce the new vectors

$$X^* = \Sigma^{-1/2}X, \qquad Y^* = \Sigma^{-1/2}Y, \qquad Z^* = \Sigma^{-1/2}Z.$$

The covariance matrix of X^* is $\Omega^* = \Sigma^{-1/2}\Omega\Sigma^{-1/2}$. The principal components analysis is performed with respect to X^*. Let N be the diagonal matrix of the non-vanishing eigenvalues of Ω^*-I. With $Y = \Gamma V$, where V is the normalized factor, Γ we write

$$U = N^{-1}\Gamma'\Sigma^{-1}X, \qquad V = N^{-1}\Gamma'\Sigma^{-1}Y, \qquad W = N^{-1}\Gamma'\Sigma^{-1}Z.$$

Let the frequency function of U be denoted by $\psi(u)$. With $u = N^{-1}\Gamma'\Sigma^{-1}x$ and $\Sigma^* = N^{-1}$ the conditional mean and the conditional covariance matrix of V for $X = x$ are obtained with the aid of the formulae which are given in the previous section.

7. Further transformation of the factors

When the normalized factors are multiplied by a non-singular matrix A, the transformed residuals obtain the covariance matrix $\Sigma^* = AN^{-1}A'$. With this redefinition of the factors we write again $Y = \Gamma V$. In terms of the original variables we have

$$U = \Sigma^*\Gamma'\Sigma^{-1}X, \qquad V = \Sigma^*\Gamma'\Sigma^{-1}Y, \qquad W = \Sigma^*\Gamma'\Sigma^{-1}Z.$$

With $\Sigma^* = AN^{-1}A'$ and $\boldsymbol{u} = \Sigma^*\Gamma'\Sigma^{-1}\boldsymbol{x}$ the conditional mean and conditional covariance matrix of V for $X = \boldsymbol{x}$ can be obtained with aid of the same formulae as before, remembering that Σ^* is in general non-diagonal for the transformation now considered. In the special case, where the matrix A is orthogonal, only a rotation of the factors is performed.

8. The special case, where the observable variable vector X has a multivariate normal distribution

When the vector X has a p-dimensional normal distribution, the vector Y has a singular normal distribution with the rank k. In this case the principal components are independent of each other. When $\Sigma = \sigma^2 I$ the following formulae are obtained:

$$E(H_j|X = \boldsymbol{x}) = \sigma^2(\lambda_j - \sigma^2)\xi_j/\lambda_j, \quad \mathrm{var}(H_j|X = \boldsymbol{x}) = \sigma^2(\lambda_j - \sigma^2)/\lambda_j$$

and for the normalized factors

$$E(V_j|X = \boldsymbol{x}) = \sigma^2(\lambda_j - \sigma^2)u_j/\lambda_j, \quad \mathrm{var}(V_j|X = \boldsymbol{x}) = \sigma^2/\lambda_j.$$

If the residual variances are not equal, the principal components analysis is performed on the variable $X^* = \Sigma^{-1/2}X$ and then the preceding formulae are valid if σ^2 is put equal to one.

When correlated factors are obtained by multiplying the orthogonal formalized factors by a non-singular matrix A, the following formulae are obtained for the correlated factors

$$E(V|X = \boldsymbol{x}) = AA'(AA' + \Sigma^*)^{-1}\boldsymbol{u}, \quad \mathrm{Cov}(V|X = \boldsymbol{x}) = \Sigma^*(AA' + \Sigma^*)^{-1}AA'.$$

It should be recalled that in the case where the matrix A is orthogonal, that is $AA' = I$, only a rotation of the factors is performed.

References

[1] T. W. Anderson, *An introduction to multivariate statistical analysis*, Wiley, New York 1958.

[2] F. Dyson, *A method for correcting series of parallax observations*, Monthly Notices of the Royal Astronomical Society 86 (1926), pp. 686–706.

[3] A. S. Eddington, *The correction of statistics for accidental errors*, ibid. 100 (1940), pp. 354–361.

[4] F. M. Lord, *An approach to mental test theory*, Psychometrica 24 (1959), pp. 283–302.

[5] E. Lyttkens, *On a class of conditional distributions, connected with some astronomical problems*, Ark. Astronom. 1 (1949), pp. 11–25.

[6] — *On a class multidimensional conditional characteristic functions and semi-variants*, ibid. 1 (1949), pp. 27–45.

[7] — *On the fix-point property of Wold's iterative estimation method for principal components*, Paper in P. R. Krishnaiah (Ed.) *Multivariate analysis*, pp. 335–350, Academic Press, New York 1966.

[8] K. G. Malmquist, *A study of the distribution of stars of spectral type A*, Meddelande från Lunds observatorium Ser II N 1920.

[9] — *On some relations in stellar statistics*, Ark. Mat. Astronom. Fys. 16 23 (1921).

[10] — *The elimination of the effects of the accidental errors of measurements in statistical investigations*, ibid. 27 A 24 (1921).

[11] R. B. McHuge, *The interval estimation of a true score*, Psychol. Bull. (1957), pp. 73–74.

[12] J. Neyman and E. Scott, *Field galaxies: luminosity, redshift and abudance of types*, in: *Theory*, Proceedings of the fourth Berkeley Symposium on Mathematical Statistics and Probability, vol. III, pp. 261–276, University of California Press, Berkeley, Cal. 1961.

[13] P. Whittle, *On principal components and least square methods of factor analysis*, Skandinavisk Aktuarietidskrift 35 (1953), pp. 223–239.

[14] H. Wold, *On the estimation of principal components and related models by iterative least squares*, Paper in P. R. Krishnaiah (Ed.) *Multivariate analysis*, pp. 391–420, Academic Press, New York 1966.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976