Metrika, Volume 35, 1988, page 49–52

A Minimal Characterization of the Covariance Matrix

By R. Grübel[1]

Summary: Let X be a k-dimensional random vector with mean vector μ and non-singular covariance matrix Σ. We show that among all pairs (a, Δ), $a \in \mathbb{R}^k$, $\Delta \in \mathbb{R}^{k \times k}$ positive definite and symmetric and $E(X-a)'\Delta^{-1}(X-a) = k$, (μ, Σ) is the unique pair which minimizes $\det \Delta$. This motivates certain robust estimators of location and scale.

Let X be a k-dimensional random vector with $E\|X\|^2 < \infty$, let $\mu = EX$ denote its mean vector and $\Sigma = E(X-\mu)(X-\mu)'$ its covariance matrix (we regard vectors $a \in \mathbb{R}^k$ as columns and write a' for the transpose of a). It is well known that μ is the best L^2-approximand for X in the sense that $a \to E\|X-a\|^2$ has a unique (global) minimum in $a = \mu$, a simple argument being

$$E\|X-a\|^2 = E(X-\mu+\mu-a)'(X-\mu+\mu-a) = E(X-\mu)'(X-\mu) + \|\mu-a\|^2.$$

Since $E(X-\mu)'(X-\mu) = \operatorname{tr}(E(X-\mu)(X-\mu)')$ the minimum is the sum of the diagonal elements of Σ, so there is no interpretation of the full covariance matrix in this approach.

Σ is positive semi-definite and symmetric. Assume for the rest of the paper that X is genuinely k-dimensional, i.e. not concentrated on a hyperplane. Then Σ is positive definite so that the ellipsoid $E(\mu, \Sigma)$ is non-degenerate where for all $a \in \mathbb{R}^k$ and all positive definite symmetric $k \times k$-matrices Δ

$$E(a, \Delta) = \{x \in \mathbb{R}^k : (x-a)'\Delta^{-1}(x-a) \leqslant 1\}.$$

---

[1] R. Grübel, Department of Mathematics, Imperial College, 180 Queen's Gate, London SW7 2BZ, UK.

Research supported by the Nuffield Foundation.

0026–1335/88/1/49–52 $2.50

Note that the volume of $E(a, \Delta)$ is a strictly increasing function of det Δ. Within a family $\{E(a, \alpha\Delta)\colon \alpha > 0\}$ the unique ellipsoid which has X on its boundary is obtained for $\alpha = (X-a)'\Delta^{-1}(X-a)$. As the following theorem shows $E(\mu, \Sigma)$ is the unique minimum volume ellipsoid which assigns mean k to this factor. It thereby provides us with a simultaneous characterization of mean and covariance.

Theorem: Let X be a k-dimensional random vector with $E\|X\|^2 < \infty$, mean μ and non-degenerate covariance matrix Σ. Then among all pairs (a, Δ), where $a \in \mathbb{R}^k$, Δ positive definite and symmetric and

$$E(X-a)'\Delta^{-1}(X-a) = k \tag{1}$$

(μ, Σ) is the unique pair which minimizes det Δ.

Proof: Let $\Sigma^{-1/2}$ be a symmetric matrix with $\Sigma^{-1/2}\Sigma^{-1/2}\Sigma = I$ where I denotes the $k \times k$ identity matrix ($\Sigma^{-1/2}$ exists since Σ is positive definite and symmetric). Put $Y = \Sigma^{-1/2}(X-\mu)$. Then

$$\begin{aligned} E(X-\mu)'\Sigma^{-1}(X-\mu) &= EY'Y = \operatorname{tr}(EYY') \\ &= \operatorname{tr}(\Sigma^{-1/2}E[(X-\mu)(X-\mu)']\Sigma^{-1/2}) \\ &= \operatorname{tr}(I) = k, \end{aligned}$$

so (μ, Σ) satisfies (1).

This transformation also shows that we now may assume $\mu = 0$, $\Sigma = I$.

Let Δ be an arbitrary symmetric and positive definite $k \times k$-matrix. Then there exist an orthogonal matrix U and $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k > 0$ such that

$$\Delta = U \operatorname{diag}(\lambda_1, \ldots, \lambda_k) U'$$

and det $\Delta = \lambda_1\lambda_2 \ldots \lambda_k$, $\Delta^{-1} = U \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_k^{-1})U'$. Now suppose (a, Δ) is such that (1) holds. We put $Z = U'(X-a)$. Then

$$\begin{aligned} E(X-a)'\Delta^{-1}(X-a) &= EZ' \operatorname{diag}(\lambda_1^{-1}, \ldots, \lambda_k^{-1})Z \\ &= \lambda_1^{-1}EZ_1^2 + \ldots + \lambda_k^{-1}EZ_k^2 \\ &= \sum_{i=1}^{k} \lambda_i^{-1} \operatorname{var} Z_i + \sum_{i=1}^{k} \lambda_i^{-1}(EZ_i)^2. \end{aligned}$$

Since $EZ = -U'a$ and $\operatorname{cov}(Z) = I$ (1) implies $\lambda_1^{-1} + \ldots + \lambda_k^{-1} < k$ if $a \neq 0$ and then $\det \Delta > 1$ by the familiar inequality relating geometric and arithmetic mean. So any solution of (1) with $a \neq 0$ does not minimize $\det \Delta$ since by the first part of the proof $(0, I)$ is a solution. If $a = 0$ then the sum of the λ_i^{-1}'s gives k which by the condition on equality of geometric and arithmetic mean leaves open two possibilities: either they all equal 1 or their product is strictly less than 1. Again, the second case would not minimize $\det \Delta$ and we arrive at $\Delta = UIU' = I$, i.e. $(0, I)$ is the unique solution. □

We may regard the solution of the problem

"find (α, Δ) with $\det \Delta = \min!$ under the constraint $E(X-a)'\Delta^{-1}(X-a) = k$"

as a function T of the underlying distribution function F of the random vector X. The theorem then says that T is well-defined on $M = \{F : \int \|x\|^2 F(dx) < \infty\}$ and that $T(F) = (\mu, \Sigma)$ on M. Given a sequence $X_1, X_2, \ldots$ of independent and identically distributed random variables with distribution function F a sequence of estimators for $T(F)$ can be obtained by putting $T_n = T(F_n)$ where F_n denotes the empirical distribution function corresponding to $X_1, \ldots, X_n$. Applying the theorem to the empirical distribution we see that this leads to the familiar estimates of mean and covariance, namely the sample mean and the sample covariance. These are known to be highly non-robust, e.g. a single wrong observation can move the estimate over an arbitrarily long distance. We may explain this by the form of the above constraint which leads us to try

$$Ef((X-a)'\Delta^{-1}(X-a)) = c$$

instead where $f : \mathbb{R}_+ \to \mathbb{R}_+$ is some function growing less rapidly than the identity and the constant c is chosen appropriately. Obviously, any such procedure – if it makes sense, we may e.g. have to replace "=" by "$\leqslant$" – will result in an affine equivariant estimator, i.e. an estimator which behaves like sample mean and covariance under affine transformation of the data.

Indeed, a radical choice of f such as

$$f(x) = \begin{cases} 0, & x \leqslant 1, \\ 1, & x > 1, \end{cases}$$

leads to the minimum volume ellipsoid estimators introduced by Rousseeuw (see also Hampel/Ronchetti/Rousseeuw/Stahel and Davies), given data $x_1, \ldots x_n$ we estimate location and scale by $\hat{\mu}_n$ and $\hat{\Sigma}_n$ where $E(\hat{\mu}_n, \hat{\Sigma}_n)$ is the ellipsoid of minimum volume containing at least a fraction $1 - c$ of these points.

Acknowledgement: I should like to thank Professor Laurie Davies for many stimulating discussions.

References

Davies PL (1986) Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. Preprint, Essen

Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA (1986) Robust statistics: the approach based on influence functions. Wiley, New York

Rousseeuw PJ (1985) Multivariate estimation with high breakdown point. In: Grossmann W, Pflug G, Vincze I, Wertz W (eds) Mathematical statistics and applications. Reidel Publishing Company, Dordrecht, pp 283–297

Received October 10, 1986
(Revised version February 27, 1987)