Quadratically Asymmetric Distributions and Their Application to Chromosome Classification

Gunter Ritter

Fakultät für Mathematik und Informatik, Universität Passau, 94030 Passau, Germany

1 Quadratically Asymmetric Distributions

A distribution on Euclidean space $\mathbf{R}^d$ is called *spherically symmetric* if it is invariant with respect to the orthogonal group $SO(d)$ and an $\mathbf{R}^d$-valued random vector is *elliptically symmetric* if it is the affine image of a spherically-symmetric random variable S (cf., e.g., Fang et al., 1990). The successes of elliptical symmetry in the context of high-dimensional classification problems suggest to consider also more complex functions of spherically-symmetric random variables S, e.g., quadratic ones. This leads to several types of stochastic models which we call *quadratically asymmetric* (cf. Ritter). One of them is of the form

(1) $X = g(S) = b + AS + QS^2,$

where $S^2 = (S_1^2, ..., S_d^2)$, b is an element of $\mathbf{R}^d$, and $A > 0$ and Q are $d \times d$ - matrices. We call the matrix Q the *quadratic asymmetry* of the model X.

It can be proved that g is one-to-one on the open centered ball of radius $(2\|A^{-1}Q\|_{1,2})^{-1}$, where $\|\cdot\|_{1,2}$ denotes the norm of a matrix as a linear operator $(\mathbf{R}^d, \|\cdot\|_1) \to (\mathbf{R}^d, \|\cdot\|_2)$. Therefore, the quadratically asymmetric model can be reasonably handled if one assumes $r := ess\,sup\|S\| < (2\|A^{-1}Q\|_{1,2})^{-1}$. (In practical applications, observations of X lying outside the g-image of the centered ball of radius r must be considered as outliers with respect to the quadratically asymmetric model.)

2 Estimation of the Parameters b, A, and Q

As in the normal model, the overall number of real parameters to be estimated is again *quadratic* in the dimension d. For estimating the vector b and the matrix A, the following relations, which are easily proved, can be used

(2) $EX = b + Q\mathbf{1}$

(3) $Var\,X = A^2 + Q(Var\,S^2)Q^T;$

here, **1** is the constant vector one. Assuming that the radial function φ of S is chosen in advance, for estimating the quadratic asymmetry several algorithms are available. We sketch here an algorithm based on the third central moment $E\hat{X}(\hat{X}^2)^T$. Writing $R \in \mathbf{R}^{d^2 \times d^2}$ for the matrix defined by

$$R_{(ij)(kl)} = 2\delta_{jk} ES_i S_j (S_l^2 - 1) + \delta_{ik} E(S_j^2 - 1)(S_l^2 - 1)$$

and abbreviating $B = A^{-1}Q$, we have

(4) $$E[A^{-1}\hat{X}][(A^{-1}\hat{X})^2]^T = RB + E[B(S^2 - \mathbf{1})][(B(S^2 - \mathbf{1}))^2]^T$$

if $ES^6 < \infty$. Moreover, from (**??**) we infer

(5) $$A^{-1}(Var\, X)A^{-1} = I + B(Var\, S^2)B^T.$$

We have the following statements:

(a) If S is not zero then R is invertible and, under mild conditions on S, A, and B, the parameters A and Q in (**??**) are essentially the only solutions of the nonlinear equations (**??**), (**??**).

(b) If the central moments of X appearing in (**??**) and (**??**) are replaced by estimates based on an iid-sequence (X_k) of observations of (**??**) then the resulting sequence of estimators for A und B (and hence A and Q) is *strongly consistent.*

If B is small the second and third powers of B appearing in (**??**) and (**??**) can be neglected. Then we have

(6) $$A \approx \sqrt{Var\, X}, B \approx R^{-1}E[A^{-1}\hat{X}][(A^{-1}\hat{X})^2]^T,$$

i.e., estimation of A und Q is reduced to operations of linear algebra.

Applied to the large Copenhagen data set Cpr, the quadratically asymmetric model reduces the error rate of automatic chromosome classification by 20 % relative to the best-fitting elliptically-symmetric models (cf. Ritter and Gaggermeier, Ritter et al., 1995).

References

Fang, K.-T., S. Kotz and K.-W. Ng (1990). *Symmetric Multivariate and Related Distributions.* Chapman and Hall, London, New York

Ritter, G. "Quadratic asymmetry - distributions beyond elliptical symmetry", in preparation

Ritter, G., K. Gaggermeier. "Automatic classification of chromosomes by means of quadratically asymmetric statistical distributions", in preparation

Ritter, G., M.T. Gallegos, and K. Gaggermeier (1995). "Automatic context-sensitive karyotyping of human chromosomes based on elliptically symmetric statistical distributions". Pattern Recognition 28, 823-831