Bayes Estimators of Covariance Parameters and the Influence of Designs

Younis Fathy and Christine Mueller

Department of Mathematics and Computer Science, University of Kassel, D-34109 Kassel, cmueller@mathematik.uni-kassel.de

Summary. It is assumed that the covariance matrix of N observations has the form $C_\theta = \sum_{r=1}^{R} \theta_r U_r$ where $U_1, \ldots, U_R$ are known covariance matrices and $\theta_1, \ldots, \theta_R$ are unknown parameters. Estimators for $\sum_{r=1}^{R} \theta_r b_r$ with known $b_1, \ldots, b_R$ are characterized which minimize the Bayes risk within all invariant quadratic unbiased estimators. In this characterization, the matrix A, which determines the quadratic form of the estimator, is given by a linear equation system which is not of full rank. It is shown that some solutions of the equation system provides asymmetric matrices A. Therefore, sufficient conditions are presented which ensures the symmetry of the matrix A. Having this result, the influence of designs on the Bayes risk is studied.

Key words: Bayes invariant quadratic unbiased estimator, quadratic form, time dependence, spatial covariance, one and two dimensional designs

1 Introduction

We assume that an observation $z(x) \in \Re$ at the experimental condition $x \in \Re^q$ is given by

$$z(x) = f(x)^\top \beta + e(x),$$

where $f : \Re^k \longrightarrow \Re^p$ is a known regression function, $\beta \in \Re^p$ an unknown parameter vector and $e(x)$ the measurement error with expectation equal to 0. Several observations $z_1(x_1), \ldots, z_N(x_N)$ at $x_1, \ldots, x_N$ are given by the vector $Z = (z_1(x_1), \ldots, z_N(x_N))^\top$ and the design matrix $F = (f(x_1), \ldots, f(x_N))^\top$ so that we have the linear model $Z = F\beta + E$.

In many linear models it is assumed that the covariance matrix of the error vector $E = (e(x_1), \ldots, e(x_N))^\top$ is the identity matrix times an unknown variance parameter. This means in particular that normal distributed observations $z_1(x_1), \ldots, z_N(x_N)$ are stochastically independent. But if the experimental conditions are time points or points in the space it is more likely that points which are close together provide similar observations than points with large distance. Then the errors $e(x_1), \dots, e(x_N)$ and the observations $z_1(x_1), \dots, z_N(x_N)$ are not anymore uncorrelated. Then in general an arbitrary covariance matrix of E is possible.

Since a general covariance matrix of E has more parameters than observations are available, some assumptions are made for the covariance. Here we will consider the case that the covariance matrix is given by a linear combination of known covariance matrices $U_1, \dots, U_R$ so that

$$C_\theta = \sum_{r=1}^{R} \theta_r \, U_r,$$

where only $\theta = (\theta_1, \dots, \theta_R)^\top$ is unknown. Hence we consider a mixed linear model with variance components $\theta_1, \dots, \theta_R$. In many cases one matrix U_r will be the identity matrix and another matrix will have components $U_r(x_i, x_j)$ which are a decreasing function of the distance between x_i and x_j. Then the identity matrix represents the effect which is known as nugget effect in spatial statistics (see e.g. Cressie (1993)).

There are several possibilities to estimate the variance components $\theta_1, \dots,$ θ_R (see e.g. Rao and Kleffe (1988) or Koch (1999)). If a linear combination $\alpha = \sum_{r=1}^{R} b_r \theta_r$ with given $b = (b_1, \dots, b_R)^\top$ shall be estimated, then this can be done by an estimator given by the quadratic from $\hat{\alpha}_0(Z) = Z^\top Q Z$. Such estimators should be invariant with respect to linear transformations of the linear model, i.e. they should satisfy $\hat{\alpha}_0(Z + F\beta) = \hat{\alpha}_0(Z)$ for any $\beta \in \Re^p$. This invariance property is in particular satisfied if the estimator has the form $\hat{\alpha}_0(Z) = Z^\top M \, A \, M Z$ where $M = I - F(F^\top F)^- F^\top$ is the projection matrix. Then one can work also with observation vectors $Y = M \, Z$ with expectation 0 and covariance matrix

$$\tilde{C}_\theta = M \, C_\theta \, M^\top = \sum_{r=1}^{R} \theta_r \, M \, U_r \, M^\top = \sum_{r=1}^{R} \theta_r \, V_r.$$

The aim is then to determine the matrix A in the quadratic form $\hat{\alpha}(Y) = Y^\top A \, Y$. Assuming an a-priori distribution for the unknown parameters $\theta_1, \dots,$ θ_R, in Section 2 we present a representation of the matrix A of the estimator which minimizes the Bayes risk within all invariant quadratic unbiased estimators. It turns out that the matrix A is given by an equation $G \, O = W$ where the vector O contains the matrix A in vectorized form and the matrix G and the vector W depend on $V_1, \dots, V_R$ and b, respectively. Thereby the form of A depends very much which g-inverse of G is used to solve the equation. It is shown that sometimes a solution is obtained where the matrix A is not a symmetric matrix. Since quadratic forms basing on a symmetric matrix have good inference properties as usually their distribution is given by χ^2 distributions, a symmetric matrix can be obtained from an asymmetric A by using $(A + A^\top)/2$. Another possibility is to find symmetric solutions A satisfying the equation $GO = W$. Section 3 provides a condition which ensures this. Having a way to calculate a reasonable estimate $\hat{\alpha}(Y) = Y^\top AY$, we study in Section 4 the influence of the design on the Bayes risk of this estimator by examples.

2 Bayes invariant quadratic unbiased estimators (BAIQUE)

Unbiased quadratic estimators $\hat{\alpha}(Y) = Y^\top AY$ for $\alpha = \sum_{r=1}^{R} b_r\theta_r = b^\top\theta$ are characterized by Rao (1972) as follows.

Lemma 1. *The quadratic estimator $Y^\top AY$ is unbiased estimator for the linear function $\alpha = b^\top\theta$ if and only if*

$$trAV_r = b_r \quad for \quad r = 1, \dots, R. \tag{1}$$

Let p_Θ be a prior distribution for the vector of parameters $(\theta_1, \theta_2, \dots, \theta_R)$ which has second order moments of the form

$$E(\Theta_i\Theta_j) = \int \theta_i\theta_j \; p_\Theta(\theta)d\theta = C_{ij} \geq 0; \qquad i, j = 1, \dots, R.$$

We assume here that the matrix of these second order moments is positive semidefinite so that we can write

$$\mathcal{C} = (C_{ij})_{i,j=1,\dots,R} = S\, S^\top = (\sum_{r=1}^{R} s_{ir}s_{jr})_{i,j=1,\dots,R}. \tag{2}$$

The Bayes risk of the estimator $\hat{\alpha}(Y) = Y^\top AY$ is given by

$$r(\hat{\alpha}) = E(E_\Theta(\hat{\alpha}(Y) - \alpha)^2) = \int E_\theta(\hat{\alpha}(Y) - \alpha)^2 p_\Theta(\theta)d\theta.$$

Definition 1. *(Gnot and Kleffe (1983)) A quadratic form $\hat{\alpha}(Y) = Y^\top AY$ is called a Bayes invariant quadratic unbiased estimate (BAIQUE) if it minimizes $E(E_\Theta(\hat{\alpha}(Y) - \alpha)^2)$ subject to all invariant unbiased quadratic estimators.*

Theorem 1. *(Compare Fathy and Qassim (2002))*
The quadratic form $Y^\top AY$ is BAIQUE if and only if the matrix A satisfies

$$\begin{pmatrix} vecV_1 & vecV_2 & \dots & vecV_R & \vdots & T \\ \dots & \dots & \dots & \dots & \vdots & \dots \\ 0 & 0 & \dots & 0 & \vdots & (vecV_1)^\top \\ 0 & 0 & \dots & 0 & \vdots & (vecV_2)^\top \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \vdots & (vecV_R)^\top \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_R \\ \dots \\ vecA \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \dots \\ b_1 \\ \vdots \\ b_R \end{pmatrix} \tag{3}$$

where $\lambda_1, \ldots, \lambda_R$ are Lagrange multipliers corresponding to the constraint (1), vec stands for the vec operator, $T = \sum_{r=1}^{R} T_r \bigotimes T_r$, $T_r = \sum_{i=1}^{R} s_{ir} V_i$ with s_{ir} given by (2).

Proof. The Bayes risk for unbiased $\hat{\alpha}$ satisfies

$$r(\hat{\alpha}) = E(Var_{\Theta}(Y^{\top}AY)) = E(2trA\tilde{C}_{\Theta}A\tilde{C}_{\Theta}) = 2(\sum_{i}^{R}\sum_{j}^{R} E(\Theta_i\Theta_j)trAV_iAV_j)$$

$$= 2(\sum_{i}^{R}\sum_{j}^{R} C_{ij}trAV_iAV_j) = 2\sum_{i=1}^{R}\sum_{j=1}^{R}\sum_{r=1}^{R} s_{ir}s_{jr}trAV_iAV_j = 2\sum_{r=1}^{R} trAT_rAT_r.$$

Since

$$\frac{\partial \sum_{r=1}^{R} trAT_rAT_r}{\partial A} = 2\sum_{r=1}^{R} T_rAT_r, \qquad \frac{\partial (trAV_r - b_r)}{\partial A} = V_r,$$

there exists Lagrange multipliers $\lambda_1, \ldots, \lambda_R$ such that

$$(\sum_{r=1}^{R} T_r \bigotimes T_r)vecA + \sum_{r=1}^{R} \lambda_r vecV_r = 0.$$

This implies the assertion together with the constraint (1).

Example (Asymmetric A)
Let consider the one dimensional design of observations at $1, 3, 5$ with the regression function $f(x) = 1$. Assume that the covariance model is

$$Var(Z) = \theta_1 \exp(-D) + \theta_2 I_{3\times 3}$$

where matrix $D = (h_{ij})_{i,j=1,2,3}$ represents the matrix of Euclidean distances and $exp(-D)$ stands for the matrix with components $exp(-h_{ij})$, so one can find

$$D = (h_{ij})_{i,j=1,2,3} = \begin{pmatrix} 0 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 0 \end{pmatrix}.$$

Suppose the uniform prior information on the parameters is given by

$$p_1(\theta_1) = 0.25;\ \ 1 \leq \theta_1 \leq 5, \qquad p_2(\theta_2) = 0.33;\ \ 0 \leq \theta_2 \leq 3,$$

and Θ_1 and Θ_2 are independent. Then equation (3) can be written as $G\,O = W$ with $G \in \Re^{11\times 11}$, $O \in \Re^{11}$, $W \in \Re^{11}$. A solution O of it can be obtained by $O = G^{-}W$ where G^{-} stands for a generalized inverse of matrix G. For the case $b = (1,1)^{\top}$, one obtains from O the Lagrange multipliers $\lambda_1 = -15.3284$, $\lambda_2 = 12.8534$ and the asymmetric matrix A given by

$$A = \begin{pmatrix} 0.6217 & 0.2478 & -0.8696 \\ 0.1217 & 0.1304 & -0.2522 \\ -0.7435 & 0 & 0 \end{pmatrix}.$$

3 Sufficient conditions for a symmetric matrix A

In this section we assume $R = 2$, $U_1 = J$ and $U_2 = I$, where J is an arbitrary covariance matrix and $I = I_{N\times N}$ denotes the identity matrix. Then we denote by $U = MJM$ and $V = MIM = M$ the given matrices in the linear representation of the covariance matrix of Y. Let be $u = vecU$ and $v = vecV$ and T be defined as in Theorem 1. Moreover, we assume that the prior satisfies $E(\Theta\Theta^\top) = I_{2\times 2}$ so that $T = U \bigotimes U + V \bigotimes V$. Then equation (3) can be written as $GO = W$ where

$$G = \begin{pmatrix} u & v & \vdots & T \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & u^\top \\ 0 & 0 & \vdots & v^\top \end{pmatrix}, \quad W = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \cdots \\ b_1 \\ b_2 \end{pmatrix}, \quad O = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \cdots \\ vecA \end{pmatrix}.$$

Theorem 2. *Let $O = (G^\top G)^- G^\top W$ satisfy $GO = W$, where $(G^\top G)^-$ denotes the Moore-Penrose inverse of $G^\top G$, and let be $M = KK^\top$ with $K^\top K = I_{l\times l}$, where l is the rank of M. If the maximum eigenvalue of $K^\top JK$ is less than 1 and $vecA$ is given by O, then A is a symmetric matrix.*

For the proof of Theorem 2, we need the following lemmas.

Lemma 2. *Let $M = KK^\top \in R^{p\times p}$ be a symmetric idempotent matrix of rank r, where $K^\top K = I \in R^{r\times r}$ is the identity matrix, and let $B \in R^{p\times p}$ be a nonsingular symmetric matrix. Then the Moore-Penrose inverse of $(MBM)^2$ is given by $((MBM)^2)^- = K(K^\top BK)^{-2}K^\top = K[(K^\top BK)^{-1}]^2K^\top$.*

Lemma 3. *See Kincaid and Cheney (1991), p.172-173.*
If A is an $n \times n$ matrix such that the maximum eigenvalue of $A^\top A$ is less than 1, then $I - A$ is invertible, and

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

Lemma 4. *See Mirsky (1990), p.337-338. Let $\phi_p, \psi_p, \chi_p : \Re^{p\times p} \to \Re^{p\times p}$ are defined by*

$$\phi_p(z) = \sum_{m=0}^{\infty} a_m z^m, \qquad \psi_p(z) = \sum_{m=0}^{\infty} b_m z^m, \qquad \chi_p(z) = \sum_{m=0}^{\infty} c_m z^m,$$

with $c_m = a_0 b_m + a_1 b_{m-1} + \ldots + a_m b_0$ *for* $m = 0, 1, 2, \ldots$. *Assume that* ϕ_1, ψ_1, χ_1 *are convergent for* $z \in \Re$ *with* $|z| < \rho$ *and suppose that* $\phi_1(z)\psi_1(z) = \chi_1(z)$ *for* $|z| < \rho$. *If all characteristic roots of* $A \in \Re^{p \times p}$ *are less than* ρ, *then*

$$\phi_p(A)\psi_p(A) = \chi_p(A).$$

Proof of Theorem 2. At first note that

$$O = (G^\top G)^- G^\top W = \begin{pmatrix} -(D^- BC^-)L \\ (C^- + C^- B^\top D^- BC^-)L \end{pmatrix},$$

where

$$D = A - BC^- B^\top, \;\; A = \begin{pmatrix} u^\top u & u^\top v \\ v^\top u & v^\top v \end{pmatrix}, \;\; B_{2\times p^2} = \begin{pmatrix} u^\top T \\ v^\top T \end{pmatrix},$$

$C = T^2 + uu^\top + vv^\top$ and $L = b_1 u + b_2 v$. Only $(C^- + C^- B^\top D^- BC^-)L$ is important for determination of A. An extension of the Sherman-Morrison-Woodbury lemma (see e.g. Henderson and Searle (1981)) provides

$$C^- = \frac{cH^- - H^- uu^\top H^-}{c} - \frac{(cH^- - H^- uu^\top H^-)vv^\top(cH^- - H^- uu^\top H^-)}{c^2 d},$$

where $H = T^2$, $c = 1 + u^\top H^- u$ and $d = 1 + v^\top (H + uu^\top)^- v$. Using this, we obtain

$$C^- L = H^-(\eta_1 u + \eta_2 v),$$

and thus after some calculations

$$vecA = (C^- + C^- B^\top D^- BC^-)L = H^-(\zeta_1 u + \zeta_2 v + \xi_1 Tu + \xi_2 Tv)$$

where $\eta_1,\ \eta_2,\ \zeta_1,\ \zeta_2,\ \xi_1,\ \xi_2, \in \Re$. To determine the Moore-Penrose inverse of $H = T^2$, we use the fact that the maximum eigenvalue of $K^\top JK \bigotimes K^\top JK$ is less than 1 if the maximum eigenvalue of $K^\top JK$ is less than 1. Hence, since $T = M \bigotimes M(I \bigotimes I + J \bigotimes J)M \bigotimes M$, Lemma 2 and Lemma 3 provide

$$\begin{aligned}(T^2)^- &= K \bigotimes K \left(\left(I \bigotimes I + K^\top JK \bigotimes K^\top JK \right)^{-1} \right)^2 K^\top \bigotimes K^\top \\ &= K \bigotimes K \left(\sum_{k=0}^{\infty} (-1)^k (K^\top JK)^k \bigotimes (K^\top JK)^k \right) \\ &\quad \left(\sum_{l=0}^{\infty} (-1)^l (K^\top JK)^l \bigotimes (K^\top JK)^l \right) K^\top \bigotimes K^\top.\end{aligned}$$

Since $\rho = 1$ satisfies the condition of Lemma 4 for $\phi_1(z) = \psi_1(z) = \sum_{k=0}^{\infty} (-1)^k z^k$ and $\chi_1(z) = \sum_{k=0}^{\infty} (-1)^k (k+1)\, z^k$, the order of the summation can be exchanged so that we obtain

$$H^- = (T^2)^- = \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}(-1)^{k+l}(K(K^\top JK)^{k+l}K^\top \bigotimes K(K^\top JK)^{k+l}K^\top).$$

The assertion follows now from the fact that for any symmetric matrices Q and P, we have $Q \bigotimes Q\ vecP = vec(QPQ)$.

4 The influence of the design on the Bayes risk

Let be $b = (1,1)$ and let the covariance matrix J be given by $J = \exp(-D)$ where D is the matrix of distances between the design points, i.e.

$$D = (\|x_i - x_j\|)_{i,j=1,\dots,N}.$$

For all considered designs the maximum eigenvalue of $K^\top J K$ is less than 1, so that the matrix A in the BAIQUE $\hat{\alpha}$ is symmetric if A is determined via $O = (G^\top G)^- G^\top W$. The Bayes risk $r(\hat{\alpha}) = 2\,tr(A\,M\,J\,M\,A\,M\,J\,M) + 2\,tr(A\,M\,A\,M)$ of the BAIQUE is given for the following one and two dimensional simple designs. More examples can be found in Fathy (2006).

4.1 One dimensional designs

Table 1 provides the Bayes risk for different four points designs on $[0,1]$ for the stationary model with $f(x) = 1$ and for the linear regression model with $f(x) = (1,x)^\top$. It can be seen that the influence of the design on the Bayes risk

Table 1. Bayes risks for one dimensional designs

Design points	Bayes risk for $f(x) = 1$	Bayes risk for $f(x) = (1,x)^\top$
0, 0.3, 0.7, 1	2.3145	386.93
0.1, 0.25, 0.85, 1	0.8856	455.77
0.1, 0.4, 0.7, 1	3.9366	213.51
0.1, 0.49, 0.61, 1	4.6978	29.48
0.4, 0.5, 0.6, 1	2.9779	279.05
0.7, 0.8, 0.9, 1	34.0702	219.41

depends very much on the model: the best design for the stationary model is the worst for the model with linear trend and, if the last design is not considered, vice versa.

4.2 Two dimensional designs

Table 2 provides the Bayes risk for different four points designs on $[0,15] \times [0,10]$ for the model with linear trend so that $f((x(1),x(2)) = (1,x(1),x(2))^\top$.

Thereby different forms of designs are compared which satisfy $x_1(1)+x_1(2)+x_2(1)+x_2(2)+x_3(1)+x_3(2)+x_4(1)+x_4(2) = 40$ for the design points $x_i = (x_i(1), x_i(2))$. The distance between design points is measured by L_1, L_2 and L_∞ norm. It turns out that the different distance measure provide

Table 2. Bayes risks for two dimensional designs

		Bayes risk		
Design form	Design points	L_1	L_2	L_∞
square	(1,1), (1,9), (9,1), (9,9)	3.2325	3.8362	3.3503
rectangular	(1,1), (1,6), (12,1), (12,6)	3.3801	3.3805	3.3812
triangular	(1,6), (6,1), (6,6), (8,6)	0.5364	0.5341	0.5352
kite	(2,2.5), (6,1), (6,10), (10,2.5)	3.4884	2.6369	2.3821
trapezoid	(1,5), (3,3), (9,3), (11,5)	2.7287	1.4195	0.5505

similar Bayes risks and that always the triangular design has the smallest risk followed by the trapezoid design. This may caused by the fact that in both designs two design points are rather close which is as in the one dimensional case a favorite property for estimating $\theta_1 + \theta_2$.

References

N.A.C. Cressie. *Statistics for Spatial Data.* Wiley, New York, 1993.

Y.M. Fathy. *Bayes quadratic unbiased estimator of spatial covariance parameters.* PhD thesis, University of Kassel, Germany, 2006.

Y.M. Fathy and M.N.I. Qassim. Bayesian estimation of spatial covariance function with two and three parameters. *Al-Rafedain Journal of Science*, 13:43–50, 2002.

S. Gnot and J. Kleffe. Quadratic estimation in mixed linear models with two variance components. *Journal of Statistical Planning and Inference*, 8: 267–279, 1983.

H.V. Henderson and S.R. Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23:53–60, 1981.

D. Kincaid and W. Cheney. *Numerical Analysis.* Wadsworth Inc., California, 1991.

K. Koch. *Parameter Estimation and Hypothesis Testing in Linear Models.* Springer, Berlin, 1999.

L. Mirsky. *An Introduction to Linear Algebra.* Clarendon Press, Oxford, 1990.

C.R. Rao. Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, 67:112–115, 1972.

C.R. Rao and J. Kleffe. *Estimation of Variance Components and Application.* Elsevier, Amsterdam, 1988.

Index