J. Appl. Prob. **15**, 514–522 (1978)
Printed in Israel

GAUSSIAN FAMILIES AND A THEOREM ON PATTERNED MATRICES

WAYNE W. BARRETT*, *University of Wisconsin-Madison*
PHILIP J. FEINSILVER**, *University of Utah*

Abstract

In this paper we use the properties of the covariance matrix of a Gaussian Markovian family to give a probabilistic proof of a theorem about inverses of tridiagonal matrices.

COVARIANCE MATRIX; GAUSSIAN FAMILY; MARKOV PROPERTY

1. Introduction

In the study of families of Gaussian random variables, matrices, determinants, inverses and quadratic forms appear in the density function of the family and thus play a fundamental role. As such, one can expect interesting facts about linear algebra to arise, and indeed formulas for determinants and inverses of symmetric positive definite matrices in terms of Gaussian integrals are well known. By introducing the additional concepts of conditional expectation and the Markov property, which are elementary for Gaussian families, we give a probabilistic proof of a theorem about inverses of tridiagonal (Jacobi) matrices, which is the 'natural' proof for a result that is purely algebraic and interesting as such. Our result appears new, although very similar theorems appear in Gantmakher and Krein (1950) and Graybill (1969). Our approach leads to a complete and simple characterization of such matrices, while illustrating and utilizing their probabilistic significance.

2. Gaussian families

A Gaussian random variable with mean zero has a density function

$$\frac{1}{\sqrt{2\pi t}}e^{-x^2/2t} \quad \text{with } t>0.$$

A Gaussian family $(X_1,\cdots,X_N)$ where each X_k has mean zero has a density function

Received 12 August 1977; revision received 2 November 1977.
* Now at Texas A&M University.
**Now at Southern Illinois University.

$$p(x) = \frac{1}{\sqrt{(2\pi)^N \det R}} e^{-\frac{1}{2}A(x)}$$

where $A = R^{-1}$, $A(x) = (Ax, x) = xAx^T$, $x = (x_1, \cdots, x_N)$, R is positive definite and symmetric and thus A is also. R is called the covariance matrix and $R_{ij} = \langle X_i X_j \rangle$, where the angle brackets denote expected value.

We have from Feller (1971) the following result.

Basic Theorem. R is a positive definite symmetric matrix if and only if it is the covariance matrix of a Gaussian family; i.e., there are linearly independent Gaussian random variables $X_1, \cdots, X_N$ such that

$$R_{ij} = \langle X_i X_j \rangle.$$

The fact that $\int_{R^N} p(x)dx = 1$ gives

$$\frac{1}{\sqrt{\det A}} = \frac{1}{\sqrt{(2\pi)^N}} \int_{R^N} e^{-\frac{1}{2}A(x)}\, dx \tag{1}$$

or

$$\det A = \frac{(2\pi)^N}{\left(\int_{R^N} e^{-\frac{1}{2}A(x)}\, dx\right)^2}, \tag{2}$$

a formula for the determinant of A in terms of a Gaussian integral.

On the other hand

$$R_{ij} = \langle X_i X_j \rangle = \int_{R^N} x_i x_j p(x) dx$$

gives

$$(A^{-1})_{ij} = \frac{\sqrt{\det A}}{(2\pi)^N} \int_{R^N} x_i x_j e^{-\frac{1}{2}A(x)}\, dx \tag{3}$$

a formula for the inverse of A in terms of a Gaussian integral.

Applying $\partial/\partial A_{ij}$ to both sides of (1)

$$-\tfrac{1}{2}(\det A)^{-\frac{3}{2}} \operatorname{cof} A_{ij} = \frac{1}{\sqrt{(2\pi)^N}} \int_{R^N} (-\tfrac{1}{2}x_i x_j) e^{-\frac{1}{2}A(x)}\, dx$$

since $\operatorname{cof} A_{ij}$ (the cofactor of A_{ij}) equals $(\partial/\partial A_{ij}) \det A$ from the expansion of $\det A$ in minors. Thus

$$\frac{\operatorname{cof} A_{ij}}{\det A} = \frac{\sqrt{\det A}}{\sqrt{(2\pi)^N}} \int_{R^N} x_i x_j e^{-\frac{1}{2}A(x)}\, dx = (A^{-1})_{ij}.$$

Noting $\operatorname{cof} A_{ij} = \operatorname{cof} A_{ji}$ since A is symmetric, we have recovered the classical inverse formula, albeit for a restricted class of matrices.

3. Inverses of tridiagonal matrices

We now turn to our main objective, to give a characterization of the family of matrices R such that $A = R^{-1}$ is a (positive definite, symmetric) tridiagonal matrix, i.e., $A_{ij} = 0$ for $|i-j|>1$,

Definition. We say that a positive definite matrix R has the triangle property if

$$R_{ij} = \frac{R_{ik}R_{kj}}{R_{kk}} \quad \text{for all } i \leqq k \leqq j;$$

or equivalently if

$$R_{ij} = \frac{R_{i\,i+1}R_{i+1\,i+2}\cdots R_{j-1\,j}}{R_{i+1\,i+1}\cdots R_{j-1\,j-1}} \quad \text{for all } i<j.$$

According to the second formula all elements R_{ij} are determined by the elements on the main diagonal and superdiagonal. This formula can be remembered by the diagram in Figure 1.

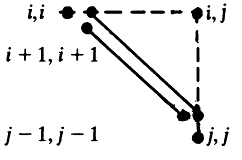

Figure 1

To find the element R_{ij} in the matrix R draw a horizontal and vertical line back to the main diagonal and then multiply along the superdiagonal to find the numerator and along the main diagonal to find the denominator as indicated by the figure. Hence the name triangle property.

A sufficient condition for R to have the triangle property is that R 'factor'. We say that an $N \times N$ symmetric matrix R 'factors' when there exist numbers $a_1, \cdots, a_N$, $b_1, \cdots, b_N$ such that $R_{ij} = a_i b_j$ for $i \leqq j$. This is easily seen to be sufficient since

$$\frac{R_{ik}R_{kj}}{R_{kk}} = \frac{a_i b_k a_k b_j}{a_k b_k} = R_{ij}.$$

It is not necessary since a diagonal matrix has the triangle property, but does not 'factor.'

Theorem. A positive definite symmetric matrix R has the triangle property if and only if its inverse A is tridiagonal.

Before giving the proof, we recall a few facts about Gaussian families and conditional expectations.

For a Gaussian family, $(X_1, \cdots, X_N)$, let $\mathscr{F}_n$ be the σ-field generated by $(X_1, \cdots, X_n)$ where $n \leqq N$. Let Y be $\mathscr{F}_N$-measurable.

The conditional expectation $E(Y \mid \mathscr{F}_n)$ is that function $\phi(X_1, \cdots, X_n)$ such that

$$\langle (Y - \phi)\Psi \rangle = 0 \quad \text{for } \Psi \text{ any function of } X_1, \cdots, X_n.$$

The basic properties of the conditional expectation are:

(i) $\langle E(Y \mid \mathscr{F}_n) \rangle = \langle Y \rangle$

(ii) $E(E(Y \mid \mathscr{F}_n) \mid \mathscr{F}_m) = E(E(Y \mid \mathscr{F}_m) \mid \mathscr{F}_n) = E(Y \mid \mathscr{F}_{m \wedge n})$, $m \wedge n = \min(m, n)$.

(iii) For $\alpha(X_1, \cdots, X_n)$, $E(\alpha Y \mid \mathscr{F}_n) = \alpha E(Y \mid \mathscr{F}_n)$.

For Gaussian families, conditional expectations have a particularly simple structure.

Proposition 1. If $1 \leqq k < j \leqq N$, $E(X_j \mid \mathscr{F}_k)$ is the linear projection of X_j on to the linear subspace generated by $X_1, \cdots, X_k$.

Since $E(X_j \mid X_k)$ is the projection of X_j on X_k,

$$E(X_j \mid X_k) = \frac{\langle X_j X_k \rangle}{\langle X_k^2 \rangle} X_k = \frac{R_{kj}}{R_{kk}} X_k. \tag{4}$$

We need one last fact, an explicit formula for $E(X_N \mid \mathscr{F}_{N-1})$, Feller (1971).

Proposition 2.

$$E(X_N \mid \mathscr{F}_{N-1}) = -\frac{A_{1N}}{A_{NN}} X_1 - \frac{A_{2N}}{A_{NN}} X_2 - \cdots - \frac{A_{N-1N}}{A_{NN}} X_{N-1}.$$

We give a proof here since this proposition is basic to the theorem and Feller's argument depends on a different approach to conditional expectations.

Proof. Let $\phi(X_1, \cdots, X_{N-1}) = E(X_N \mid \mathscr{F}_{N-1})$. Since $\phi \in \operatorname{Span}(X_1, \cdots, X_{N-1})$ by Proposition 1

$$\phi = \sum_{k=1}^{N-1} c_k X_k. \tag{5}$$

By the definition of conditional expectation,

$$\langle \psi(\phi - X_N) \rangle = 0 \quad \text{for any } \psi(X_1, \cdots, X_{N-1}).$$

Taking ψ successively equal to $X_1, \cdots, X_{N-1}$ gives

$$\sum_{k=1}^{N-1} R_{jk} c_k - R_{jN} = 0, \quad j = 1, \cdots, N-1.$$

Multiplying by $-A_{NN}$,

$$\sum_{k=1}^{N-1} R_{jk}(-A_{NN}c_k)+R_{jN}A_{NN}=0, \quad j=1,\cdots,N-1. \tag{6}$$

Since $RA=I$,

$$\sum_{k=1}^{N} R_{jk}A_{kN}=0 \quad \text{for } j=1,\cdots,N-1. \tag{7}$$

Subtracting (6) from (7) we get

$$\sum_{k=1}^{N-1} R_{jk}(A_{kN}+A_{NN}c_k)=0 \quad j=1,\cdots,N-1.$$

This can be written $R'\alpha=0$ where α is the $N-1$ vector with components $\alpha_k=A_{kN}+A_{NN}c_k$ and R' is the covariance matrix of $X_1,\cdots,X_{N-1}$. R' is non-singular by the basic theorem so $\alpha\equiv 0$; i.e.,

$$c_k=-\frac{A_{kN}}{A_{NN}}.$$

Substituting these values of c_k in (5) gives the result.

The main feature of the proof of the theorem is the Markov property of the associated Gaussian family.

Definition. We say that a Gaussian family $X_1,\cdots,X_N$ is Markov if

$$E(X_j\mid\mathcal{F}_k)=E(X_j\mid X_k) \quad \text{for all } 1\leqq k\leqq j\leqq N.$$

Remark. In general the Markov property means the stronger condition

$$E(e^{izX_j}\mid\mathcal{F}_k)=E(e^{izX_j}\mid X_k), \quad z\in R$$

or equivalently

$$E(\phi(X_j)\mid\mathcal{F}_k)=E(\phi(X_j)\mid X_k)$$

for bounded continuous functions ϕ, for $1\leqq k\leqq j\leqq N$.

However, this is not really stronger in the Gaussian case.

The proof of the theorem is in two lemmas.

Lemma 1. The family $X_1,\cdots,X_N$ is Markov if and only if the covariance matrix has the triangle property.

Proof. (a) Assume $X_1,\cdots,X_N$ is Markov. Let $i\leqq k\leqq j$. Then

$$R_{ij}=\langle X_iX_j\rangle=\langle E(X_iX_j\mid\mathcal{F}_k)\rangle=\langle X_iE(X_j\mid\mathcal{F}_k)\rangle$$

$$=\langle X_iE(X_j\mid X_k)\rangle=\langle X_i\frac{R_{kj}}{R_{kk}}X_k\rangle=\frac{R_{ik}R_{kj}}{R_{kk}}$$

by the properties of conditional expectation, the Markov property and the remark following Proposition 1.

(b) Suppose $R_{ij} = R_{ik}R_{kj}/R_{kk}$ for all $i \leqq k \leqq j$. We wish to show that

$$E(X_j \mid \mathcal{F}_k) = E(X_j \mid X_k) = \frac{R_{kj}}{R_{kk}} X_k.$$

If $1 \leqq i \leqq k$,

$$\left\langle \left(E(X_j \mid \mathcal{F}_k) - \frac{R_{kj}}{R_{kk}} X_k \right) X_i \right\rangle = \langle E(X_i X_j \mid \mathcal{F}_k) \rangle - \frac{R_{ik}R_{kj}}{R_{kk}}$$

$$= \langle X_i X_j \rangle - R_{ij} = 0.$$

Since $E(X_j \mid \mathcal{F}_k) - (R_{kj}/R_{kk})X_k$ is a linear combination of $X_1, \cdots, X_k$ (by Proposition 1) and orthogonal to each of them, it equals 0. Therefore, $E(X_j \mid \mathcal{F}_k) = (R_{kj}/R_{kk})X_k = E(X_j \mid X_k)$ and the family is Markov.

Lemma 2. The family $X_1, \cdots, X_N$ is Markov if and only if the inverse covariance matrix is tridiagonal.

What this lemma is saying is that the inverse covariance matrix is tridiagonal if and only if the density function $p(x) = p(x_1, \cdots, x_N)$ factors into transition densities

$$p(x) = p_1(x_1)p_2(x_1, x_2)p_3(x_2, x_3) \cdots p_N(x_{N-1}, x_N).$$

This is a familiar property of Markovian densities and can be taken here as the heuristic basis for the lemma.

Proof. (a) Assume A is tridiagonal. Then

$$A_{1N} = A_{2N} = \cdots = A_{N-2\,N} = 0$$

and

$$E(X_N \mid \mathcal{F}_{N-1}) = -\left(\frac{A_{N-1N}}{A_{NN}}\right) X_{N-1}$$

from Proposition 2. By the remark following Proposition 1,

$$E(X_N \mid X_{N-1}) = \left(\frac{R_{N-1\,N}}{R_{N-1\,N-1}}\right) X_{N-1}.$$

But for a tridiagonal matrix, setting $j = N-1$ in (7) gives:

$$R_{N-1\,N-1}A_{N-1\,N} + R_{N-1\,N}A_{NN} = 0$$

or

$$\frac{R_{N-1\,N}}{R_{N-1\,N-1}} = -\frac{A_{N-1\,N}}{A_{NN}}.$$

Therefore $E(X_N \mid \mathscr{F}_{N-1}) = E(X_N \mid X_{N-1})$.

To compute $E(X_j \mid X_k)$, $1 \leqq k \leqq j \leqq N$, the variables $x_{k+1}, \cdots, x_{j-1}, x_{j+1}, \cdots, x_N$ can be integrated out of the density $\gamma e^{-\frac{1}{2}A(x)}$ leaving a Gaussian density $\gamma' e^{-\frac{1}{2}A'(x)}$ in the variables $x_1, \cdots, x_k, x_j$. The new matrix A' is also tridiagonal (see computation below). Then the same argument as above gives

$$E(X_j \mid \mathscr{F}_k) = E(X_j \mid X_k)$$

which is the Markov property.

(b) Assume $X_1, \cdots, X_N$ is Markov. From Proposition 2

$$E(X_N \mid \mathscr{F}_{N-1}) = -\frac{A_{N-1\,N}}{A_{NN}} X_{N-1} - \sum_{0<j<N-1} \frac{A_{jN}}{A_{NN}} X_j.$$

Since

$$E(X_N \mid \mathscr{F}_{N-1}) = E(X_N \mid X_{N-1}) = \frac{R_{N-1\,N}}{R_{NN}} X_{N-1}$$

linear independence implies that $A_{jN} = 0$ for $0 < j < N-1$.

Let $a_{N-1} = A$ with Nth row and column replaced by zeroes, and let $A^{(N-1)} =$ inverse covariance matrix of the family $(X_1, \cdots, X_{N-1})$. Then

$$\begin{aligned} &C_{N-1} \exp\{-\tfrac{1}{2} A^{(N-1)}(x_1, \cdots, x_{N-1})\} \\ &= C_N \int \exp\{-\tfrac{1}{2} A(x)\} dx_N \\ &= C_N \int \exp\{-\tfrac{1}{2} a_{N-1}(x) - A_{N-1N} x_{N-1} x_N - \tfrac{1}{2} A_{NN} x_N^2\} dx_N \\ &= \sqrt{2\pi / A_{NN}}\, C_N \exp\{-\tfrac{1}{2} a_{N-1}(x)\} \exp\{\tfrac{1}{2}(A^2_{N-1\,N}/A_{NN}) x^2_{N-1}\} \end{aligned}$$

where the C's are normalizing constants. Notice that the only difference between $A^{(N-1)}$ and the $N-1 \times N-1$ submatrix of A is in the $(N-1, N-1)$ position. Thus, repeating the above Markov argument implies that

$$A_{j\,N-1} = 0 \quad \text{for } 0 < j < N-2.$$

And so induction (and the symmetry of A) implies that A is tridiagonal.

Using the basic theorem to produce the family $X_1, \cdots, X_N$ from the covariance matrix R and combining the two lemmas completes the proof of the theorem.

If R is a covariance matrix with the triangle property, the tridiagonal matrix $A = R^{-1}$ can be written down explicitly.

We set

$$d_{ij} = R_{ii} R_{jj} - R_{ij}^2 = \det(\text{covariance matrix of } (X_i, X_j)).$$

Then we have

$$
\begin{aligned}
A_{ij} &= \frac{-R_{ij}}{d_{ij}} \quad \text{for } |i-j| = 1 \\
& R_{ii}\frac{d_{i-1\,i+1}}{d_{i-1\,i}d_{i\,i+1}} \quad \text{for } i = j \neq 1, N \\
& \frac{R_{22}}{d_{12}}, \quad i = j = 1 \\
& \frac{R_{N-1N-1}}{d_{N-1N}}, \quad i = j = N.
\end{aligned}
$$

Corollary. If a covariance matrix R 'factors,' its inverse is tridiagonal.

Proof. This follows immediately from the theorem since R then has the triangle property. In fact, it is easy to show from the theorem that R 'factors' if and only if it has a tridiagonal inverse all of whose diagonal, superdiagonal, and subdiagonal elements are non-zero. This is the theorem in Gantmakher and Krein (1950).

4. Examples

1. If

$$
R = \begin{bmatrix}
1 & 1 & 1 & \cdots & 1 \\
1 & 2 & 2 & \cdots & 2 \\
1 & 2 & 3 & \cdots & 3 \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
\cdot & \cdot & \cdot & & \cdot \\
1 & 2 & 3 & \cdots & N
\end{bmatrix},
$$

$R_{ij} = a_i b_j$ for $i \leqq j$ where $a_i = i$ and $b_j = 1$ for all j. Since R factors, its inverse A must be tridiagonal by the corollary. In fact,

$$
A = \begin{bmatrix}
2 & -1 & & & 0 \\
-1 & 2 & \ddots & & \\
& \ddots & \ddots & \ddots & \\
& & \ddots & 2 & -1 \\
0 & & & -1 & 1
\end{bmatrix}.
$$

This is a special case of the general case $R_{ij} = \min(\tau_i, \tau_j)$ where $0 < \tau_1 < \tau_2 < \cdots < \tau_N$, the covariance matrix for Brownian motion. Again R_{ij} factors with $a_i = \tau_i$ and $b_i = 1$, but the inverse is more complicated.

2. If

$$R = \begin{bmatrix} 1 & r & r^2 & \cdots & r^{N-1} \\ r & 1 & r & \cdots & r^{N-2} \\ r^2 & r & 1 & \cdots & r^{N-3} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ r^{N-1} & r^{N-2} & r^{N-3} & \cdots & 1 \end{bmatrix},$$

then R_{ij} factors into $R_{ij} = r^{-i}r^j$ for $i \leqq j$. Thus its inverse is tridiagonal. This is a special case of the Ornstein–Uhlenbeck process where $R_{ij} = e^{-\sigma|\tau_i - \tau_j|}$ with $0 < \tau_1 < \tau_2 < \cdots < \tau_N$. In this case,

$$A = (1-r^2)^{-1} \begin{bmatrix} 1 & -r & 0 & \cdots & 0 & 0 \\ -r & 1+r^2 & -r & \cdots & 0 & 0 \\ 0 & -r & 1+r^2 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1+r^2 & -r \\ 0 & 0 & 0 & \cdots & -r & 1 \end{bmatrix}.$$

References

FELLER, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. 2, 2nd edn, Wiley, New York, 85–86.

GANTMAKHER, F. R. AND KREIN, M. G. (1950) *Oscillation Matrices and Kernels and Small Vibrations of Mechanical Systems*, 2nd edn, 118. Translated from a publication of the State Publishing House for Technical-Theoretical Literature, Moscow-Leningrad, 1950. United States Atomic Energy Commission, Office of Technical Information.

GRAYBILL, F. A. (1969) *Introduction to Matrices with Applications in Statistics*, p. 179. Wadsworth Publishing Company, Belmont, California.