

Modeling covariance matrices in a systematic, data-driven manner is impeded by the positive-definiteness constraint and high-dimensionality; however, similar (albeit simpler) hurdles in modeling the mean vector μ of the distribution of a random vector $Y = (y_1, \dots, y_M)'$ has been successfully handled in the context of regression analysis. The resulting techniques have lead to the framework of generalized linear models (GLM), which enjoys a rich and extensive theoretical foundation. The success of GLMs is in most part due to the use of a link function $g(\cdot)$ and a linear predictor $g(\cdot) = X\beta$, which induces an unconstrained parameterization and reduces the parameter space dimension simultaneously. Since the covariance matrix of a random vector Y , defined by $\Sigma = E(Y - \mu)(Y - \mu)$, is a mean-like parameter, one would like to exploit the idea of GLM along with the experience and progress in fitting the mixed-effects and time series models in developing a systematic, data-based procedure for covariance matrices.

Approaches to modeling covariances with the explicit use covariates has been extensively explored in the time series literature, while the implicit use of covariates for covariance modeling has been the focus of many in the area of variance components; see ? and ?. Time series techniques based on spectral and Cholesky decompositions provide the necessary tools for handling the cumbersome positivedefiniteness constraint on a stationary covariance matrix or covariance function. In the GLM setting, simply applying a link function componentwise to the potentially constrained mean vector μ permits its unconstrained estimation. Unfortunately employing the same precise approach to covariance matrices isn't viable since positive-definiteness is a simultaneous constraint on all entries of a matrix. Successfully modeling a general covariance structure almost necessitates decomposing a covariance matrix into its "variance" and "dependence" components because of its inherent complicated structure. The three major methods for performing such decompositions include the variance-correlation decomposition, the spectral decomposition, and the Cholesky decomposition. Section ?? touched on the attractive properties of the latter that lead to advantages over the other two covariance parameterizations.

0.0.1 Linear models for covariance

Gabriel (1962) was among the first to implicitly parameterize a multivariate normal distribution in terms of entries of the precision matrix Ω^{-1} . Dempster (1972) who recognized the entries of $\Sigma^{-1} = (\sigma^{ij})$ as the canonical parameters of the exponential family of normal distributions with mean zero and unknown covariance matrix Σ :

$$\log f(Y, \Sigma^{-1}) = -\frac{1}{2} \text{tr} \Sigma^{-1} (Y'Y) + \log |\Sigma|^{-1/2} - M \log \sqrt{\pi}$$

Soon thereafter, the simple structures of time series and variance components models motivated ? to define the class of linear covariance models:

$$\Sigma = \sum_{i=1}^q \alpha_i U_i \quad (1)$$

where the U_i s are known symmetric matrices and the α_i s are unknown parameters, restricted to ensure that Σ is positive definite. This class of models is general enough to include all linear mixed

effects models as well as certain time series and graphical models. In, for q large enough, any covariance matrix admits representation of the form ??, since one can decompose every covariance matrix as

$$\Sigma = \sum_{i=1}^M \sum_{j=1}^M \sigma_{ij} U_{ij}, \quad (2)$$

where U_{ij} is an $M \times M$ matrix with a 1 in the (i, j) position, and zeros everywhere else. The linear model (??) can be viewed as modeling the link-transformed covariance $g(\Sigma) = \sum_{i=1}^q \alpha_i U_i$, where $g(\cdot)$ is the identity link. Despite the convenience of parameterization, the positive definite constraint (??) makes estimation an arduous task.

Inducing sparsity by setting certain elements of the covariance matrix or its inverse to zero is a common approach to reducing the dimensionality of a covariance structure. Inspection of model (??) and the covariance parameterization given in (??) makes it easy to see that this can be achieved by eliminating certain U_{ij} from the covariates in the linear covariance model. On the extreme end of the sparsity spectrum is the case of independent observations and Σ is diagonal, eliminating all U_{ij} from the linear model covariates for $i \neq j$. Connection between the linear covariance model and other models for covariance discussed in previous sections can be established if we consider intermediary cases, such as classes of stationary moving average (MA) and autoregressive (AR) models introduced in the early times series literature. The $MA(q)$ model corresponds to a banded covariance matrix, setting

$$\sigma_{ij} = 0 \quad \text{for } |i - j| > q, \quad (3)$$

while the $AR(p)$ model corresponds to a banded inverse:

$$\sigma^{ij} = 0 \quad \text{for } |i - j| > p. \quad (4)$$

Of course, there are the nonstationary analogues to these classes of models, some of which were discussed in Section ??. We will review others which are related to antedependence models and Gaussian graphical models. Random variables y_1, \dots, y_M , which correspond to observation times t_1, \dots, t_M , with multivariate normal joint distribution said to be p^{th} -order antedependent or $AD(p)$ if y_t and y_{t+s+1} are independent given the intervening values y_{t+1}, \dots, y_{t+s} for $t = 1, \dots, p$ and all $s \geq p$. A random vector $Y = (y_1, \dots, y_p)$ is $AD(p)$ if and only if its covariance matrix satisfies (??). Closely connected are the classes of variable order AD models and varying order, varying coefficient autoregressive models in which the coefficients and order of antedependence depend on time.

0.0.2 Log-linear covariance models

The constraint on the α_i s in (??) was eliminated with the introduction of log-linear covariance models (?. ?.) For a general covariance matrix having spectral decomposition

$$\Sigma = P \Lambda P', \quad (5)$$

its matrix logarithm, denoted $\log \Sigma$, and defined by $\log \Sigma = P \log \Lambda P'$ is a symmetric matrix with unconstrained entries taking values on \Re . Application of the log-link function leads to the log-linear model for Σ :

$$g(\Sigma) = \log \Sigma = \sum_{i=1}^q \alpha_i U_i, \quad (6)$$

where the U_i s are as before in ?? and the α_i s are now unconstrained. The α_i s, however, now lack statistical interpretation since $g(A) = \log A$ is a highly nonlinear operation. But for diagonal Σ , $\log \Sigma = \text{diag}(\sigma_{11}, \dots, \sigma_M M)$, and model ?? reduces to modeling of heterogeneous variances, which has been extensively studied. Detailed presentation is given in ?, ?m and in references therein.

? were the first to pursue nonparametric estimation of the spectral decomposition for functional data, which arise from experiments which produce observed responses in the form of curves. See ?, ?. The covariance structure is estimated via functional principal component analysis (fPCA); principal components of functional data are estimated using penalized least squares of the normalized eigenvectors, subject to the orthogonality constraint. Additionally, ? proposed kernel-based PCA, but maintaining orthogonality of the smooth principal components remains a major computational challenge in both approaches.