Eigenvectors and Eigenvalues of Stationary Processes

Overview

1. Toeplitz matrices
2. Szegö's theorem
3. Circulant matrices
4. Matrix Norms
5. Approximation via circulants

Toeplitz and circulant matrices

Toeplitz matrix A banded, square matrix Γ_n (subscript n for the $n \times n$ matrix) with elements $[\Gamma_n]_{jk} = \gamma_{j-k}$,

$$\Gamma_n = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots & \gamma_{1-n} \\ \gamma_1 & \gamma_0 & \gamma_{-1} & \cdots & \gamma_{2-n} \\ \gamma_2 & \gamma_1 & \gamma_0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \gamma_{-1} \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \tag{1}$$

Symmetry Toeplitz matrices don't have to be symmetric or real-valued, but ours will be since we'll set $\gamma_{-h} = \gamma_h = \text{Cov}(X_{t+h}, X_t)$ for some stationary process X_t. From now on, Γ_n is the covariance matrix of a stationary process.

Assumptions We will assume that the covariances are absolutely summable,

$$\sum_h |\gamma_h| < \infty$$

Breadth The results shown here are in the form that's most accessible, without searching for too much generality. All of these extend more generally to other types of sequences, such as those that are square summable, and other matrices that need not be symmetric.

Szegö's Theorem

Fourier transform Since the covariances are absolutely summable, it is (relatively) easy to show that we can define a continuous function from the γ_h,

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma_h e^{i2\pi\omega h}, \quad -\tfrac{1}{2} < \omega \leq \tfrac{1}{2}. \tag{2}$$

If γ_h are the covariances of a stationary process, then $f(\omega)$ is known as the *spectral density function* or spectrum of the process. (See Section 4.1-4.3 in SS.)

Inverse transform The Fourier transform is invertible in the sense that we can recover the sequence of covariances from the spectral density function $f(\omega)$ by integration

$$\gamma_h = \int_{-1/2}^{1/2} f(\omega) e^{-i2\pi h\omega} d\omega \ . \tag{3}$$

Heuristically, the expression for the variance $\gamma_0 = \int_{-1/2}^{1/2} f(\omega) d\omega$ suggests that the spectral density decomposes the variance of the process into a continuous mix of frequencies.

Szegö's theorem Define the eigenvalues of Γ_n as

$$\tau_{n,0}, \tau_{n,1}, \ldots, \tau_{n,n-1} \ .$$

These are all positive if Γ_n is positive definite (as we often require or assume). Szegö's theorem shows that we can use the spectum to approximate various various functions of the eigenvalues. (An interesting question in the analysis of Toeplitz matrices in general is what happens when Γ_n is not full rank.)

Let G denote an arbitrary continuous function. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{h=0}^{n-1} G(\tau_{n,h}) = \int_{-1/2}^{1/2} G(\, f(\omega)\,) d\omega \tag{4}$$

That is, we get to replace a sum by an integral. The sum resembles a Riemann approximation to an integral, as if $\tau_{n,h} = f(2\pi h/n)$.

Basic examples

Trace The trace of a square matrix is the sum of the diagonal values of the matrix, which equals the sum of the eigenvalues of the matrix.

$$\frac{1}{n}\text{trace}(\Gamma_n) = \frac{1}{n}\sum_h \tau_{n,h} \approx \int_{-1/2}^{1/2} f(\omega)d\omega$$

Though it follows directly from (3) that $\gamma_0 = \int_{-1/2}^{1/2} f(\omega)d\omega$, it is also a consequence of (4) as well.

Determinant The product of the eigenvalues of a matrix.

$$\log|\Gamma_n|^{1/n} = \frac{1}{n}\sum \log\tau_{n,h} \approx \int_{-1/2}^{1/2} \log f(\omega)d\omega\ .$$

Prediction application Here's a surprising application of Szegö's theorem to prove a theorem of Kolmogorov. If X_t is a staionary process, how well is it possible to predict X_{n+1} linearly from its n predecessors? In particular, what is

$$V_n = \min_a \mathbb{E}\left(X_{n+1} - a_0X_n - a_2X_{n-1} - \cdots - a_{n-1}X_1\right)^2$$

In the Gaussian case, the problem is equalvalent to finding the variance of the conditional expectation of X_{n+1} given $X_{1:n}$. Szegö's theorem provides an elegant answer:

$$\lim_n V_n = \exp\left(\int_{-1/2}^{1/2} \log f(\omega)d\omega\right)\ . \tag{5}$$

The answer follows from applying (4) to the ratio of determinants, once you note that

$$V_n = \frac{|\Gamma_{n+1}|}{|\Gamma_n|}\ . \tag{6}$$

(Long aside: To see that (6) holds, covariance manipulations of the multivariate normal distribution show you that the variance of the scalar r.v. Y given the vector r.v. X is $\text{Var}(Y|X) = \sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}$, where the variance matrix is partitioned as

$$\Sigma = \text{Var}\begin{pmatrix} Y \\ X \end{pmatrix} = \begin{pmatrix} \sigma_{yy} & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_{xx} \end{pmatrix}\ .$$

To find the determinant $|\Sigma|$, postmultiply Σ by a matrix with 1s on the diagonal, obtaining

$$|\Sigma| = \left| \Sigma \begin{pmatrix} 1 & 0 \\ \Sigma_{xx}^{-1}\Sigma_{xy} & I \end{pmatrix} \right| = (\sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy})|\Sigma_{xx}|$$

The expression (6) follows. Note that the regression vector $\beta = \Sigma_{xx}^{-1}\Sigma_{xy})$.

Now back to the prediction problem. From Szegö's theorem, $\frac{1}{n+1}\sum \log \tau_{n+1,j} \approx \frac{1}{n}\sum \log \tau_{n,j}$ since both approximate the same integral. Now plug in $|\Gamma_{n+1}|^{1/(n+1)} \approx |\Gamma_n|^{1/n}$ and (5) follows.

Circulant matrices

Circulant matrix is a special type of Toeplitz matrix constructed by rotating a vector $c_0, c_1, \ldots, c_{n-1}$, say, cyclically by one position to fill successive rows of a matrix,

$$C_n = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & c_1 & \\ \vdots & & \ddots & \ddots & c_1 \\ c_1 & c_2 & \cdots & c_{n-1} & c_0 \end{bmatrix} \qquad (7)$$

Circulant matrices are an important type of Toeplitz matrix for our purpose (they have others!) because we can easily find their eigenvectors and eigenvalues.

Eigenvectors These are relatively easy to find from the basic definition of an eigenvector and a great guess for the answer. An eigenvector u of C_n satisfies $C_n\, u = \tau\, u$, which gives a system of n equations:

$$\begin{aligned} c_0u_0 + c_1u_1 + \cdots + c_{n-1}u_{n-1} &= \tau u_0 \\ c_{n-1}u_0 + c_0u_1 + \cdots + c_{n-2}u_{n-1} &= \tau u_1 \\ \vdots \quad &\quad \vdots \\ c_1u_0 + c_2u_1 + \cdots + \ c_0\ \ u_{n-1} &= \tau u_{n-1} \end{aligned}$$

The equation in the rth row is, in modular form,

$$\text{row } r: \quad \sum_{j=0}^{n-r-1} c_j u_{j+r} + \sum_{j=n-r}^{n-1} c_j u_{(j+r)|n} = \tau u_r$$

or in more conventional form as

$$\text{row } r: \quad \sum_{j=0}^{n-r-1} c_j u_{j+r} + \sum_{j=n-r}^{n-1} c_j u_{j+r-n} = \tau u_r \ .$$

Guess $u_j = \rho^j$ (maybe reasonable to guess this if you have been studying differential equations.). Then

$$\sum_{j=0}^{n-r-1} c_j \rho^{j+r} + \sum_{j=n-r}^{n-1} c_j \rho^{j+r-n} = \tau \, \rho^r \ .$$

If $\rho^n = 1$, it works! These are the n roots of unity; any $\rho = \exp(i2\pi j/n)$ works for $j = 0, 1, \ldots, n-1$. The eigenvectors have the form $u' = (\rho^0, \rho^1, \rho^2, \ldots, \rho^{n-1})$:

$$u_r' = (1, e^{i2\pi r/n}, e^{i2\pi 2r/n}, \ldots, e^{i2\pi(n-1)r/n})$$

as seen in the discrete Fourier transform.

Eigenvalues These are the discrete Fourier transforms of the sequence that defines the circulant,

$$\tau_{n,k} = \sum_{j=0}^{n-1} c_j \, e^{i2\pi jk/n}$$

If the $c_j = \gamma_j$, then we've got the first n terms of the sum that defines the spectral density (2).

Implications for covariance matrices We can now anticipate the results. Asymptotically in n, the vectors that define the discrete Fourier transform are eigenvectors of *every* covariance matrix. The eigenvalues are then the transform coefficients of the covariances, namely values of the spectral density function. If we define an orthogonal matrix $U = (u_0, u_1, \ldots, u_{n-1})$ from the eigenvectors u_n, then we obtain

$$U' \, \Gamma_n \, U \approx \text{diag}(f(\omega_j))$$

To confirm these guesses, we need to show that a circulant matrix provides a "good" approximation to the covariance, good enough so that we can use the results for circulants when describing covariance matrices. To describe what we mean by a good approximation, we need a way to measure the distance between matrices, a norm.

Matrix norms

Norms A norm on a vector space $\mathcal{V}$ is any function $\|x\|$ for $x \in \mathcal{V}$ for which ($\alpha \in \mathbb{R}$)

1. $\|x\| > 0$, $x \neq 0$.
2. $\|\alpha x\| = |\alpha|\|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)

Operator norm Define the operator norm of a square matrix T as the maximum ratio of quadratic forms,

$$\|T\| = \max_x \frac{x'Tx}{x'x} = \max \tau_j$$

where τ_j are the eigenvalues (singular values) of T.

Hilbert-Schmidt or weak norm. For an $n \times n$ matrix T,

$$\begin{aligned} |T|^2 &= \tfrac{1}{n} \sum_{j,k} |t_{jk}|^2 \\ &= \tfrac{1}{n} \text{trace}(T'T) \\ &= \tfrac{1}{n} \sum \tau_j^2 \ . \end{aligned}$$

Connections It is indeed a weaker norm,

$$|T|^2 \leq \|T\|^2 = \tau_{max}^2$$

The two allow you to handle products,

$$|T\,S| \leq \|T\||S| \ .$$

Approximation via circulants

An approximation Consider the following circulant that approximates Γ_n, obtained by "flipping" the covariances and running them "both ways":

$$G_n = \begin{bmatrix} \gamma_0 & \gamma_1+\gamma_{n-1} & \gamma_2+\gamma_{n-2} & \cdots & \gamma_{n-1}+\gamma_1 \\ \gamma_1+\gamma_{n-1} & \gamma_0 & \gamma_1+\gamma_{n-1} & \cdots & \gamma_{n-2}+\gamma_2 \\ \gamma_2+\gamma_{n-2} & \gamma_1+\gamma_{n-1} & \gamma_0 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \gamma_{-1} \\ \gamma_{n-1}+\gamma_1 & \gamma_{n-2}+\gamma_2 & \cdots & \gamma_1+\gamma_{n-1} & \gamma_0 \end{bmatrix} \tag{8}$$

The norm of the difference $\Gamma_n - G_n$ is a familiar type of sum,

$$\begin{aligned} |\Gamma_n - G_n|^2 &= \tfrac{2}{n}\sum_{h=1}^{n-1}(n-h)\gamma_{n-h}^2 \\ &= \tfrac{2}{n}\sum_{h=1}^{n-1} h\,\gamma_h^2 \qquad \text{reverse the sum} \\ &= 2\sum_{h=1}^{n-1}\tfrac{h}{n}\,\gamma_h^2\ . \end{aligned}$$

Since we assume that $\sum|\gamma_j| < \infty$, the sum converges to 0,

$$\lim_{n\to\infty}|\Gamma_n - G_n| = 0\ .$$

Eigenvalues of close matrices If two matrices are close in the weak norm, then the averages of their eigenvalues are close. In particular, given two matrices A and B with eigenvalues α_j and β_j, then

$$|\tfrac{1}{n}\sum_j \alpha_j - \tfrac{1}{n}\sum_j \beta_j| \le |A-B|$$

Proof: If $D = A - B$, then

$$\sum_j \alpha_j - \sum_j \beta_j = \text{trace}(A) - \text{trace}(B) = \text{trace}(D).$$

Now use Cauchy-Schwarz in the form

$$(\sum_j a_j)^2 = (\sum_j a_j\ 1)^2 \le (\sum_j a_j^2)(n) \tag{9}$$

to show that

$$|\text{trace}(D)|^2 = |\sum_j d_{jj}|^2 \le n \sum_j d_{jj}^2 \le n \sum_{j,k} d_{jk}^2 = n^2 |D|^2 \ .$$

The loose bounds in the proof suggest you can do a lot better. In fact, you can move the absolute value inside the sum (so that the actual eigenvalues are getting close, not just on average).

Powers of eigenvalues A messier argument produces a similar result. Given two matrices A and B with $|A - B| \to 0$ and eigenvalues α_j and β_j, then for any power s,

$$\lim \tfrac{1}{n} \sum_j (\alpha_j^s - \beta_j^s) = 0 \ .$$

Extension Now that we've found that powers of the eigenvalues converge for close matrices in the limit, its not hard to get the result for a continuous function. The argument is a common one: polynomials can approximate any continuous function g. The result implies that

$$\lim \tfrac{1}{n} \sum_j g(\alpha_j) - g(\beta_j) = 0 \ .$$