New Results on the Spectral Problem for Sample Auto-Covariance Matrices

Reimer Kühn

Disordered Systems Group
Department of Mathematics, King's College London

joint work with Peter Sollich

IEEE CAMAD 2015, Guildford, Sept 7 2015

Details in Europhys. Lett. 99 20008 (2012), available from http://www.mth.kcl.ac.uk/∼kuehn

Outline

Outline

Sample Auto-Covariance Matrices

- Consider a (second order stationary) stochastic process

$$X = (X_t)_{t\in\mathbb{Z}} = \ldots, X_{-2}, X_{-1}, X_0, X_1, X_2, \ldots$$

- Sample auto-covariance matrix of a realization:

$$C_{ij} = \frac{1}{M}\sum_{t=1}^{M} x_{i+t}x_{j+t}\ , \quad 1 \le i,j \le N\ .$$

- Can write this in terms of a matrix X with entries $X_{it} = x_{i+t}$ as

$$C_{ij} = \frac{1}{M}(XX^T)_{ij}\ .$$

- Expect finite sample fluctuation around mean

$$C_{ij} = \langle x_i x_j \rangle \pm \mathcal{O}(1/\sqrt{M}) = \bar{C}(|i-j|) \pm \mathcal{O}(1/\sqrt{M})$$

- $\Rightarrow$ C is a symmetric randomly perturbed Toeplitz matrix.

Sample Auto-Covariance Matrices – Spectrum

- We are interested in the spectrum of C as $N \to \infty$ & $M \to \infty$.
- Expect that the spectrum depends on the aspect ratio $\alpha = N/M$.
- Take limit $N \to \infty$ & $M \to \infty$ @ fixed α.
- Known results
  - Existence of limiting spectral density for auto-covariance matrices of moving average processes with i.i.d. driving @ $\alpha = 1$ (Basak, Bose, Sen 2011).
  - Universality of results @ $\alpha = 1$: independence of statistics of i.i.d. driving (numerical, Sen 2010)
  - Existence of limiting spectral density for random Toeplitz, Hankel and Markov matrices with i.i.d. entries (Bryc, Dembo Jiang, 2007)
  - In the $\alpha \to 0$-limit $\Leftrightarrow$ no sampling noise: Szegö's Theorem

$$\rho_0(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi}\, \delta(\lambda - \hat{C}(q))\ , \qquad \hat{C}(q) = \sum_{n\in\mathbb{Z}} \bar{C}(n)\mathrm{e}^{\mathrm{i}qn}$$

Outline

Comparison with Wishart-Laguerre Ensemble

- Sample covariances for N-dimensional data — sample-size M,

$$C_{ij} = \frac{1}{M} \sum_{t=1}^{M} x_{it} x_{jt}$$

- Express in terms of $N \times M$ matrices $X = (x_{it})$ as

$$C_{ij} = \frac{1}{M} (XX^T)_{ij} \; .$$

- Expect finite sample fluctuation around mean. For i.i.d. entries x_{it}

$$C_{ij} = \langle x_i x_j \rangle \pm \mathcal{O}(1/\sqrt{M}) = \delta_{ij} \pm \mathcal{O}(1/\sqrt{M})$$

- Spectrum of C as $N \to \infty$, $M \to \infty$ @ fixed $\alpha = N/M$
$\Rightarrow$ Marčenko Pastur-Law

$$\rho_\alpha(\lambda) = \frac{\sqrt{4\alpha - (\lambda - (1+\alpha))^2}}{2\pi\alpha\lambda}$$

Principal Differences

- Columns of X for the covariance problem are i.i.d. random vectors in $\mathbb{R}^N$

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1M} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2M} \\ \vdots & & & \ddots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{NM} \end{pmatrix}$$

Principal Differences

- Columns of X for the auto-covariance problem are sections of a single time series $(x_t)_{t\in\mathbb{Z}}$

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_M \\ x_2 & x_3 & x_4 & \dots & x_{1+M} \\ \vdots & & & \ddots & \vdots \\ x_N & x_{N+1} & x_{N+2} & \dots & x_{N-1+M} \end{pmatrix}$$

Principal Differences

- Columns of X for the auto-covariance problem are sections of a single time series $(x_t)_{t\in\mathbb{Z}}$

$$X = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_M \\ x_2 & x_3 & x_4 & \dots & x_{1+M} \\ \vdots & & & \ddots & \vdots \\ x_N & x_{N+1} & x_{N+2} & \dots & x_{N-1+M} \end{pmatrix}$$

- In the auto-covariance problem, X is a rectangular Hankel matrix.
- Number of random variables in the problem is $\mathcal{O}(N)$, rather than $\mathcal{O}(N^2)$ as in the Wishart Laguerre ensemble.
- Extensive body of knowledge about the Wishart-Laguerre ensemble and its variants (applications in multivariate statistics, signal-processing, finance, ...)
- Comparatively little is known about the auto-covariance problem (although many applications in time-series analysis, signal processing, information theory).

Outline

Main Result

- For stationary processes with true auto-covariance $\bar{C}(n) \in \ell_1(\mathbb{Z})$

$$\rho_\alpha(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi} \frac{1}{\hat{C}(q)} \rho_\alpha^{(0)}\left(\frac{\lambda}{\hat{C}(q)}\right) \qquad (1)$$

- **Note 1**: As $\hat{C}(q) \equiv 1$ for uncorrelated data the scaling function $\rho_\alpha^{(0)}(x)$ must be associated with the spectral density for auto-covariance matrices of i.i.d. data.
- **Note 2**: We have a good analytic approximation for $\rho_\alpha^{(0)}(x)$.
- **Note 3**: Our tests suggest that the result Eq. (1) is exact, but so far no complete proof.

Main Result

- **Note 4**: Our scaling result is essentially a generalization of Szegö's result about spectra of Toeplitz matrices to the case including sampling noise.
  - Recall (Szegö)

$$\rho_0(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi}\, \delta(\lambda - \hat{C}(q))$$

  - Can write this as

$$\rho_0(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi}\, \frac{1}{\hat{C}(q)}\, \rho_0^{(0)}\left(\frac{\lambda}{\hat{C}(q)}\right) ,$$

    in which

$$\rho_0^{(0)}(\lambda) = \delta(\lambda - 1)$$

    is indeed the spectral density of the auto-covariance matrix for i.i.d sequences in the $\alpha \to 0$-limit of vanishing sampling noise.

Outline

Spectral Density and Resolvent

- Spectral density of sample covariance matrix from resolvent

$$\rho(\lambda) = \lim_{N\to\infty} \frac{1}{\pi N} \mathrm{Im}\,\mathrm{Tr}\left\langle \left[\lambda_\varepsilon \mathbb{1} - C\right]^{-1} \right\rangle, \qquad \lambda_\varepsilon = \lambda - \mathrm{i}\varepsilon$$

- Express as (S F Edwards & R C Jones, JPA, 1976)

$$\begin{aligned} \rho_\alpha(\lambda) &= \lim_{N\to\infty} \frac{1}{\pi N}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\,\mathrm{Tr}\left\langle \ln\left[\lambda_\varepsilon \mathbb{1} - C\right]\right\rangle \\ &= \lim_{N\to\infty} -\frac{2}{\pi N}\,\mathrm{Im}\,\frac{\partial}{\partial\lambda}\left\langle \ln Z_N \right\rangle, \end{aligned}$$

where Z_N is a Gaussian integral:

$$Z_N = \int \prod_{k=1}^{N} \frac{\mathrm{d}u_k}{\sqrt{2\pi/\mathrm{i}}}\ \exp\left\{-\frac{\mathrm{i}}{2}\sum_{k,\ell} u_k(\lambda_\varepsilon \delta_{k\ell} - C_{k\ell})u_\ell\right\}$$

Closed Form Approximation & Scaling

- ...several approximations needed (not all controlled (!)) to manage the calculation.
- $\Rightarrow$ allow closed form expression of $\langle Z_N \rangle$, and hence $\rho_\alpha(\lambda)$

$$\langle Z_N \rangle = \prod_{\nu=0}^{(N-1)/2} \left\{ \frac{2\,\mathrm{i}}{\alpha \hat{C}(p_\nu)} \int_0^\infty \mathrm{d}y\, \frac{\mathrm{e}^{-\mathrm{i}y\lambda_\varepsilon 2/(\alpha \hat{C}(p_\nu))}}{\left(1-\mathrm{i}y\right)^{2/\alpha}} \right\}$$

- Gives

$$\rho_\alpha(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi}\, \frac{1}{\hat{C}(q)}\, \rho_\alpha^{(0)}\left(\frac{\lambda}{\hat{C}(q)}\right)$$

- Our approximations give

$$\rho_\alpha^{(0)}(\lambda) = -\lim_{\varepsilon\to 0} \frac{1}{\pi} \mathrm{Im} \frac{\partial}{\partial \lambda} \ln I_\alpha\left(\frac{2}{\alpha}\lambda_\varepsilon\right)$$

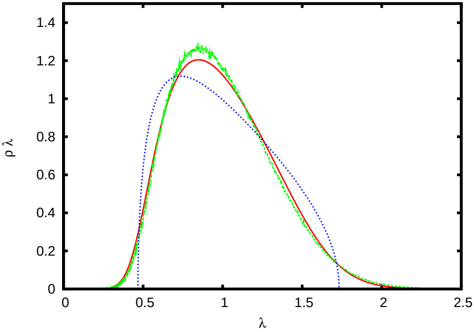
with

$$I_\alpha(x) = \mathrm{i}\,(-x)^{-1+2/\alpha}\, \mathrm{e}^{-x}\, \Gamma(1-2/\alpha, -x)\,, \quad \mathrm{Im}\, x < 0\,.$$

Outline

The Scaling Function — Spectrum for i.i.d. Data

- Spectral density for $x_n \sim \mathcal{N}(0,1)$ i.i.d. @ $\alpha = 0.1$

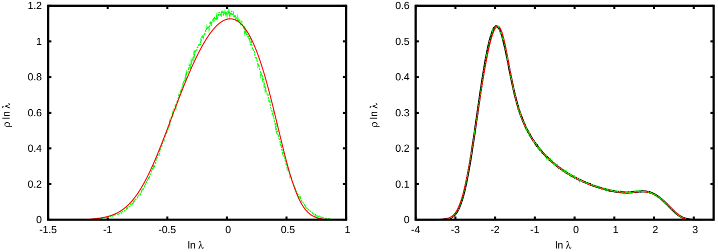


Simulation results (green); analytic approximation for $\rho_\alpha^{(0)}(\lambda)$ (red), Marčenko-Pastur law (blue-dashed).

AR-1 Process @ $\alpha = 0.1$

- (Logarithmic) Spectral density for AR-1 process @ $\alpha = 0.1$

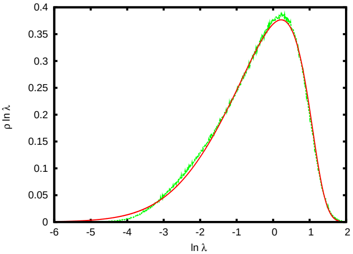
$$x_n = a_1 \, x_{n-1} + \sqrt{1 - a_1^2} \, \xi_n$$

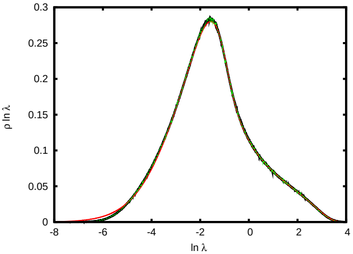


Left: i.i.d. data, simulation (green) and analytic result (red).
Right: $a_1 = 0.8$. Comparing scaling based on the empirical scaling function (black) with that based on the analytic result (red) and simulations (green).

AR-1 Process @ $\alpha = 0.8$

- (Logarithmic) Spectral density for AR-1 process @ $\alpha = 0.8$





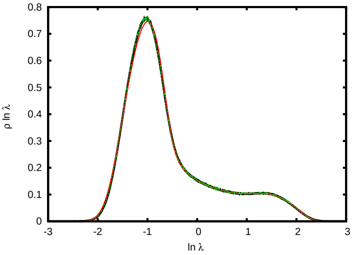
Left: i.i.d. data, simulation (green) and analytic result (red). **Right** $a_1 = 0.8$. Comparing scaling based on the empirical scaling function (black) with that based on the analytic result (red) and simulations (green).

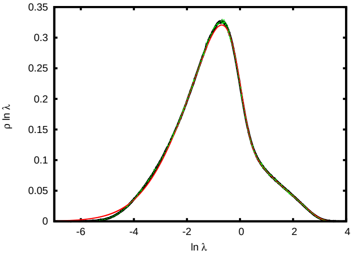
AR-2 Process (Two Real Eigenvalues)

- (Logarithmic) Spectral density for AR-2 process

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} = \sigma\xi_n$$

$$a_1 = 0.5,\ a_2 = -3/16\ , \qquad \sigma \text{ such that } \bar{C}(0) = 1\ .$$





Comparing scaling based on the empirical scaling function (black) with that based on the analytic result (red) and simulations (green).

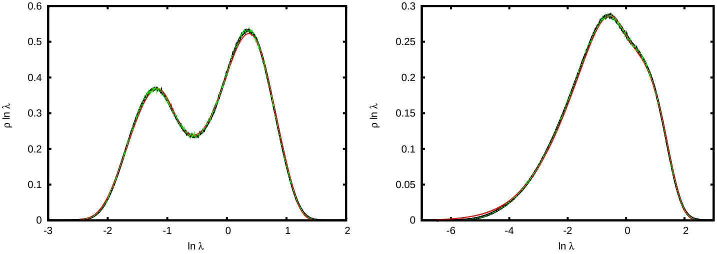
Left: $\alpha = 0.1$, **Right**: $\alpha = 0.8$.

AR-2 Process (Complex Conjugate Eigenvalues)

- (Logarithmic) Spectral density for AR-2 process

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} = \sigma\xi_n$$

$$a_1 = 0.5,\ a_2 = 5/16\ , \qquad \sigma \text{ such that } \bar{C}(0) = 1\ .$$



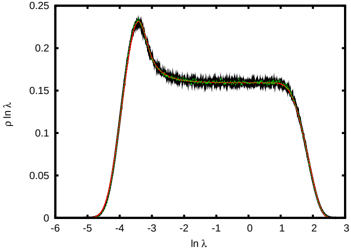
Comparing scaling based on the empirical scaling function (black) with that based on the analytic result (red) and simulations (green).

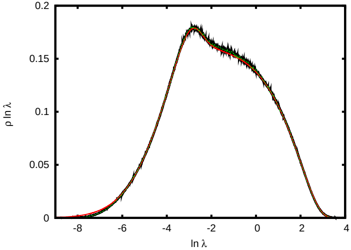
Left: $\alpha = 0.1$, **Right**: $\alpha = 0.8$.

A Process with Long Range Auto-Correlation

- A process with power-law decay of auto-correlation

$$\bar{C}(n) = \frac{1}{1 + (n/2)^2} \,,$$





Comparing scaling based on the empirical scaling function (black) with that based on the analytic result (red) and simulations (green).

Left: $\alpha = 0.1$, **Right**: $\alpha = 0.8$.

Outline

Applications

- Spectral estimation & time series analysis: Use

$$\rho_\alpha(\lambda) = \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi} \, \frac{1}{\hat{C}(q)} \, \rho_\alpha^{(0)}\left(\frac{\lambda}{\hat{C}(q)} \right)$$

  to fit a functional form of $\hat{C}(q)$ to sample data $\rho_\alpha(\lambda)$.
- For finitely parameterized $\hat{C}(q)$ (e.g. ARMA(p,q) models), can use standard fitting routines.
- Differential entropy rate of a Gaussian process $X = (X_t)_{t \in \mathbb{Z}}$:

$$h(X) = \frac{1}{2} \log(2\pi \mathrm{e}) + \int_0^{2\pi} \frac{\mathrm{d}q}{2\pi} \log \hat{C}(q)$$

  Estimate using sample auto-covariances of aspect ratio α:

$$\Rightarrow \quad h_\alpha(X) = h(X) + \langle \log \lambda \rangle_\alpha^{(0)}$$

Applications – contd.

- Lossy compression & Shannon rate distortion theory: Parametric form of rate distortion function

$$D_\theta = \int_0^{2\pi} \mathrm{d}q \min\left(\theta, \hat{C}(q)\right)$$

$$R_\theta = \int_0^{2\pi} \mathrm{d}q \max\left(0, \frac{1}{2}\log\frac{\hat{C}(q)}{\theta}\right)$$

  Influence of sampling noise when estimated using auto-covariance matrices of aspect ratio α:

$$D_\theta(\alpha) = \left\langle D_{\theta/\lambda}\right\rangle_\alpha^{(0)}$$

$$R_\theta(\alpha) = \left\langle R_{\theta/\lambda}\right\rangle_\alpha^{(0)}$$

- Would need inversion formula to to use this as an estimator ???
- One-step prediction error δ_1 of best linear predictor for weakly stationary stationary processes

$$\delta_1(\alpha) = \delta_1 \times \exp\left(\langle \log\lambda\rangle_\alpha^{(0)}\right)$$

Outline

Summary

- Computed DOS of sample auto-covariance matrices.
- Equivalent to finding distribution of singular values of random rectangular Hankel matries.
- [Key ingredient: Szegö's theorem for Toeplitz matrices]
- Obtain a scaling form for DOS in terms of DOS for i.i.i data.
    - amounts to generalization to Szegö's result to situations with sampling noise.
    - results suggest that scaling is exact
    - ideas for an independent proof
- Applications: time-series analysis, signal processing, information theory, finance . . .
- **Thanks!** K. Anand, L. Dall'Asta, P. Vivo