

INMA(1)型随机矩阵的极限谱分布

李雨静, 李晨龙*

太原理工大学数学学院, 山西 太原

收稿日期: 2023年9月25日; 录用日期: 2023年10月19日; 发布日期: 2023年10月26日

摘要

整值时间序列数据在多个领域中广泛存在, 如金融学、无线通讯网络、犯罪学等。对整值时间序列模型的研究在理论和应用上都有着重要价值。然而, 面对当前海量的高维数据, 经典的极限理论不再适用, 亟需发展分析大维整值时间序列模型极限性质的理论与方法。从文章结构和研究意义整体上来看, 本文研究了一类具有整值时间序列结构的大维随机矩阵的极限谱分布, 即INMA(1)型随机矩阵的极限谱分布。先介绍了INMA(1)型随机矩阵的定义, 并证明INMA(1)型随机矩阵样本协方差矩阵的极限谱分布的存在性; 其次, 针对文章中的内容和结果, 利用Stieltjes变换给出了该模型的样本协方差矩阵的极限谱密度; 最后通过数值模拟, 验证了本文方法的有效性。

关键词

极限谱分布, INMA(1)过程, 随机矩阵, Stieltjes变换

Limiting Spectral Distribution of INMA(1) Random Matrix

Yujing Li, Chenlong Li*

School of Mathematics, Taiyuan University of Technology, Taiyuan Shanxi

Received: Sep. 25th, 2023; accepted: Oct. 19th, 2023; published: Oct. 26th, 2023

Abstract

Integer time series data are widely used in many fields; such as finance, wireless communication networks, criminology, etc. The study of integer-valued time series models has important value in both theory and application. However, in the face of the current massive high-dimensional data, classical bounds theory is no longer applicable, and there is an urgent need to develop theories and methods for analysing the bounds properties of high-dimensional integer time series models.

*通讯作者。

From the perspective of the overall structure and research significance of the article, this paper investigates the limit spectral distribution of a class of high-dimensional random matrices with integer-valued time-series structure, namely the limit spectral distribution of INMA(1) type random matrices. First, the definition of INMA(1) type random matrix is introduced, and the existence of the limit spectral distribution of the covariance matrix of the INMA(1) type random matrix sample is proved; Second, based on the contents and results of the article, the Stieltjes transform was used to obtain the maximum spectral density of the sample covariance matrix of the model; Finally, the effectiveness of the proposed method was verified by numerical simulation.

Keywords

Limiting Spectral Distribution, INMA(1) Process, Random Matrix, Stieltjes Transform

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

整值时间序列是对接时间顺序发生的事件的计数, 出现在许多实际场景中。例如, 每年飓风发生的数量、物种每年的繁殖数量、保险公司每月的理赔数量、股票的交易数量等。Mckenzie [1]提出了基于二项稀疏算子的建模思路, 用一个细化运算符来代替经典时间序列模型中的标量乘法, 从而实现对整值数据建模与分析的目的。20 世纪 80 年代末, Al-Osh 等[2]基于二项稀疏算子建立了一阶整数值自回归 (INAR(1))过程及一阶整数值移动平均(INMA(1))过程, 开启了整数值时间序列研究的先河。随后发展出了各类整值时间序列模型[3] [4]。

随着现代科学技术的飞速发展, 收集和储存的数据规模越来越庞大。在处理海量数据时, 构造的数据矩阵的维数也越来越大, 从而极大地促进了大维数据矩阵的研究。经典的中心极限定理不再适用于大维随机矩阵, 而降维又会导致数据信息的丢失。经过众多学者的努力, 发展出了能有效处理大维数据矩阵的随机矩阵理论。近年来, 随机矩阵理论在核物理、流体力学、图像处理等领域得到了广泛应用[5]。

在处理大维海量数据时, 一个重要的研究方向是大维随机矩阵的谱分析。上世纪 50 年代, Wigner [6] [7]将随机矩阵与量子物理结合起来研究核物理, 发现了随机 Wigner 矩阵, 并证明了 Wigner 矩阵标准化后的经验谱分布函数的期望收敛到半圆律。随后 Marcenko 和 Pastur [8]在随机矩阵所有元素互相独立的条件下推导出了大维样本协方差矩阵的极限谱分布。为弱化矩阵中元素之间的独立性假设, Silverstein [9]考虑了 $X = T^{1/2}Y$ 的情况, 其中 T 表示 Hermitian 矩阵, Y 表示元素相互独立的随机向量。Bai 和 Zhou [10]利用 Stieltjes 变换研究了只有列向量相互独立的大维样本协方差矩阵的极限谱分布。胡江[11]等人将此方法应用于推导具有 AR(1)过程依赖结构的随机矩阵的极限谱分布。Pfaffel 和 Schlemm [12]将此方法进一步扩展到具有线性元素的随机矩阵的极限谱分布的研究中, 推导出了样本协方差极限谱分布的 Stieltjes 变换形式。文献[13]利用 Pfaffel 和 Schlemm 发展出的理论, 给出了具有 ARMA(p, q)过程依赖结构的随机矩阵的极限谱密度。然而, 整数值时间序列模型都是非线性的, 无法直接利用文献[13]的理论给出 INMA(1)型随机矩阵的极限谱分布。

本文研究了具有 INMA(1)结构的大维随机矩阵的极限谱分布。主要研究思路如下: 首先通过对 INMA(1)过程的研究得, INMA(1)过程表示为一类特殊的线性过程; 然后基于线性表示此证明具有 INMA(1)结构的大维随机矩阵样本协方差矩阵的极限谱分布存在; 其次利用 Stieltjes 变换给出了具有

INMA(1)结构的大维随机矩阵样本协方差矩阵的极限谱密度函数; 最后通过数值模拟, 绘制直方图并与理论结果对比, 验证本文方法的有效性。

2. 基本概念及性质

2.1. INMA(1)过程及其样本协方差矩阵

20 世纪 80 年代末, Al-Osh 等[2]基于二项稀疏算子构建了 INMA(1)过程。具体而言, 称满足如下形式的随机变量序列 $\{Y_t\}_{t \geq 1}$ 是 INMA(1)过程,

$$\begin{aligned} Y_t &= \varepsilon_t + \alpha \circ \varepsilon_{t-1} \\ &:= \varepsilon_t + \sum_{i=1}^{\varepsilon_{t-1}} B_i, \end{aligned} \quad (1)$$

其中, $\{B_i\}_{i \geq 1}$ 是相互独立的伯努利随机变量序列, $P(B_i=1)=1-P(B_i=0)=\alpha$, $\alpha \in (0,1)$, $\{\varepsilon_t\}_{t \geq 1}$ 是独立同分布的泊松随机变量序列, 满足 $\{\varepsilon_t\}_{t \geq 1} \sim \text{Poisson}(\lambda)$, 且 $\{\varepsilon_t\}_{t \geq 1}$ 与 $\{B_i\}_{i \geq 1}$ 独立, 记号 “ \circ ” 表示二项稀疏算子[14]。

性质 1.1. [15]满足上述定义的随机变量序列 $\{Y_t\}_{t \geq 1}$ 是平稳的, 其均值、方差、自协方差函数分别如下所示:

$$\begin{aligned} \mu &:= E(Y_t) = (1+\alpha)\lambda, \\ \sigma^2 &:= \text{Var}(Y_t) = (1+\alpha)\lambda, \\ \gamma(k) &:= \text{Cov}(Y_t, Y_{t-k}) = \begin{cases} \alpha\lambda, & k=1 \\ 0, & k \geq 2 \end{cases} \end{aligned}$$

性质 1.2. 令 $f(w)$ 表示 $\{Y_t\}_{t \geq 1}$ 的功率谱密度函数, 则

$$f(w) = 2\alpha\lambda \cos w + (\alpha+1)\lambda, \quad w \in [0, 2\pi].$$

证明 由性质 1.1 可得

$$f(w) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i w k} = \alpha\lambda (e^{-i w} + e^{i w}) + \gamma(0) = 2\alpha\lambda \cos w + (\alpha+1)\lambda.$$

证毕。□

称 $\mathbf{X} = (X_{i,j})_{\substack{1 \leq i \leq p, 1 \leq j \leq n}}$ 为具有 INMA(1)结构的随机矩阵, 若 \mathbf{X} 的每一行 $X_i = (X_{i,1}, \dots, X_{i,n})'$, $(i=1, \dots, p)$ 是由 INMA(1)过程生成的

$$X_{i,j} = \varepsilon_{i,j} + \alpha \circ \varepsilon_{i,j-1}$$

其中, 各行之间元素是相互独立的。 \mathbf{X} 的样本协方差矩阵为:

$$S_n = \frac{1}{p-1} \sum_{i=1}^p (X_i - \bar{X})(X_i - \bar{X})',$$

其中, $\bar{X} = p^{-1} \sum_{i=1}^p X_i$ 。

2.2. 极限谱密度及 Stieltjes 变换

记 $\lambda_1, \dots, \lambda_p$ 为 p 阶样本协方差阵 S_n 的 p 个特征根, 称

$$F^{S_n}(x) = p^{-1} \sum_{i=1}^p I(\lambda_i \leq x)$$

是矩阵 S_n 的经验谱分布函数。当 $n \rightarrow \infty, p \rightarrow \infty$ 时, 若 $F^{S_n}(x)$ 存在且是非随机的, 则称其为 S_n 的极限谱分布。 $F^{S_n}(x)$ 的 Stieltjes 变换为

$$m_{\hat{F}}(z) := \int \frac{1}{x-z} dF^{S_n}(x) = p^{-1} \text{tr} \left\{ (S_n - zI)^{-1} \right\},$$

其中, I 为 p 阶单位阵。 $\text{tr}(\cdot)$ 表示矩阵的迹。

3. 主要结果及证明

3.1. 具有 INMA(1)结构的随机矩阵的极限谱密度

由于均值不影响 S_n 的极限谱分布[16], 所以对 S_n 极限谱分布的求解可以简化为对 $p^{-1} \sum_{i=1}^p X_i X_i'$ 极限谱分布的求解。在下文中, 我们仍令 $S_n = p^{-1} \sum_{i=1}^p X_i X_i' = p^{-1} X X'$ 。记 $\underline{S}_n = p^{-1} X X'$, 其对应极限谱分布的 Stieltjes 变换形式为 $\underline{m}_{\hat{F}}(z)$ 。下文简记 $m = m_{\hat{F}}(z)$, $\underline{m} = \underline{m}_{\hat{F}}(z)$ 。下面给出本文主要结论。

定理 2.1. 设 S_n 是 INMA(1)型随机矩阵的样本协方差矩阵, 若 $\{\varepsilon_t\}_{t \geq 1}$ 是相互独立的, 其均值为 0, 方差为 σ^2 , 当 $n \rightarrow \infty$ 时, $p/n = c$, F^{S_n} 以概率 1 收敛到一个非随机概率分布 F , 该分布 Stieltjes 变换满足

$$z = -\frac{1}{\underline{m}} + \frac{c}{2\pi} \int_0^{2\pi} \frac{2\alpha\lambda \cos w + (\alpha+1)\lambda}{2\alpha\lambda \underline{m} \cos w + (\alpha+1)\lambda \underline{m} + 1} dw, \quad (2)$$

其中, $\underline{m} = -(1-c)z^{-1} + cm$ 。

定理 2.2. INMA(1)型随机矩阵的样本协方差矩阵 S_n 的极限谱分布 F 的 Stieltjes 变换是关于 z 的函数。记

$$f_i(z) = \frac{1}{\pi} \lim_{z_2 \rightarrow 0} \text{Im} [m_i(z)], \quad i = 1, \dots, 4,$$

其中 $z = z_1 + iz_2 \in C^+$, $c = \lim_{n \rightarrow \infty} p/n \in (0, \infty)$, $m_i = c^{-1} \underline{m}_i + (1-c)/(yz)$, $\underline{m}_i, (i=1, \dots, 4)$ 为定理 2.1 中式(2)的四个解, 则 INMA(1)型随机矩阵的样本协方差矩阵 S_n 的极限谱密度为

$$f(x) = \sum_{i=1}^4 f_i(z) \mathbf{1}\{f_i(z) > 0\}.$$

3.2. 定理 2.1 的证明

为了证明定理 2.1, 我们需要用到如下引理。

引理 2.1. [8] 设 $X = (X_{i,j})_{1 \leq i \leq p, 1 \leq j \leq n}$ 是 $p \times n$ 的随机矩阵, 且 X 的第 i 行满足

$$X_{i,j} = \sum_{k=0}^{\infty} c_k Z_{i,j-k}, \quad i = 1, \dots, p, \quad j = 1, \dots, n, \quad c_k \in R,$$

其中 $(Z_{i,j})_{1 \leq i \leq p, 1 \leq j \leq n}$ 是独立随机变量数组且满足:

$$EZ_{i,j} = 0, \quad EZ_{i,j}^2 = 1, \quad \sigma_4 = \sup_{i,j} EZ_{i,j}^4 < \infty$$

以及对所有的 $\varepsilon > 0$, 有当 $n \rightarrow \infty$ 时,

$$\frac{1}{np} \sum_{i=1}^p \sum_{j=1}^n E \left(Z_{i,j-k}^2 \mathbf{1}\{Z_{i,j-k}^2 \geq \varepsilon n\} \right) \rightarrow 0 \quad (\text{Lindeberg 型条件}).$$

则 $\{X_i\}_{i=1, \dots, p}$ 可以表示为线性过程, 并且具有连续可微的谱密度 f 。若进一步满足下述条件:

存在正数 C 和 δ , 对所有 $k \geq 0$ 有 $|c_k| \leq C(k+1)^{-1-\delta}$;

对所有 $\gamma \in R$, 只有有限个 $w \in [0, 2\pi]$ 是 $f(w) = \gamma$ 的根;

对几乎所有的 $w \in [0, 2\pi]$, $f(w) \neq 0$ 。

则当 $n \rightarrow \infty$, $p/n \rightarrow c \in (0, \infty)$ 时, S_n 的经验谱分布几乎必然收敛到一个具有有界支撑的非随机概率分布 F , 且存在正数 λ_- 和 λ_+ , 使得 F 的 Stieltjes 变换 $z \mapsto m_{\hat{F}}(z)$ 是满足如下方程的唯一 $C^+ \rightarrow C^+$ 的映射

$$\frac{1}{m} = -z + \frac{c}{2\pi} \int_{\lambda_-}^{\lambda_+} \frac{\lambda}{1 + \lambda m} \sum_{w \in [0, 2\pi]} \frac{1}{f'(w)} d\lambda。$$

由引理 2.1 可知, 当随机矩阵内部满足线性结构时, 其生成样本协方差阵的极限谱分布存在并可以用 Stieltjes 变换表示。对于具有 INMA(1)结构的随机矩阵来说, 由于 INMA(1)中的二项稀疏算子是非线性的, 我们不可以直接使用引理 2.1。下面首先证明 INMA(1)过程可以进行线性表示。

证明 对任意 $j \geq k$, $j \in N$, 令

$$Z_{i,j-k} = \begin{cases} c_k^{-1}(\alpha_k \circ \varepsilon_{j-k} - \alpha_k \lambda), & k=0,1 \\ \xi_{j-k}, & k \geq 2 \end{cases},$$

其中 $\alpha_0=1$, $\alpha_1=\alpha$, $c_k = \sqrt{\alpha_k^2 \lambda^2 + \alpha_k \lambda}$, $\xi_1 \sim N(0,1)$, $\xi_j, j \geq 1$ 是独立同分布的。

则

$$X_{i,j} = \begin{cases} c_k Z_{i,j-k}, & k=0,1 \\ 0, & k \geq 2 \end{cases}。$$

由文献[12]可知, 如果潜在的线性过程具有非零平均值, 即 $X_{i,j} = \mu + \sum_{k=0}^{\infty} c_k Z_{i,j-k}$, $\mu \neq 0$, 引理 2.1 仍然有效。所以 $X_{i,j}$ 可以表示成线性过程。下面证明 $X_{i,j}$ 满足引理 2.1 的条件。

当 $k=0,1$ 时, 则有

$$\begin{aligned} EZ_{j-k} &= E\left[c_k^{-1}(\alpha_k \circ \varepsilon_{j-k} - \alpha_k \lambda)\right] = c_k^{-1}(\alpha_k \lambda - \alpha_k \lambda) = 0, \\ EZ_{j-k}^2 &= E\left[c_k^{-1}(\alpha_k \circ \varepsilon_{j-k} - \alpha_k \lambda)\right]^2 = c_k^{-2}(\alpha_k^2 \lambda^2 + \alpha_k \lambda) = 1, \\ EZ_{j-k}^4 &= E\left[c_k^{-1}(\alpha_k \circ \varepsilon_{j-k} - \alpha_k \lambda)\right]^4 \leq 8c_k^{-4}E\left[(\alpha_k \circ \varepsilon_{j-k})^4 + (\alpha_k \lambda)^4\right] \\ &= 8c_k^{-4}E\left[E\left(\sum_{i=1}^{\varepsilon_{j-k}} B_i\right)^4 \middle| \varepsilon_{j-k}\right] + 8c_k^{-4}E(\alpha_k \lambda)^4 \\ &\leq 8c_k^{-4}E\left[\varepsilon_{j-k} \times E\left(\sum_{i=1}^{\varepsilon_{j-k}} B_i^4 \middle| \varepsilon_{j-k}\right)\right] + 8c_k^{-4}E(\alpha_k \lambda)^4 \\ &= \frac{8(\alpha_k \lambda + \alpha_k^4 \lambda^4)}{\alpha_k^4 \lambda^4 + 2\alpha_k^3 \lambda^3 + \alpha_k^2 \lambda^2} < \infty. \end{aligned}$$

当 $k \geq 2$ 时, 则 $EZ_{j-k} = 0$, $EZ_{j-k}^2 = 1$, $EZ_{j-k}^4 = 3 < \infty$ 。

对所有的 $\varepsilon > 0$, 当 $n \rightarrow \infty$ 时, 由 Holder 不等式和切比雪夫不等式得

$$(np)^{-1} \sum_{i=1}^p \sum_{j=1}^n E\left(Z_{i,j-k}^2 1_{\{Z_{i,j-k} \geq \sqrt{\varepsilon n}\}}\right) \leq (np)^{-1} \sum_{i=1}^p \sum_{j=1}^n \sqrt{EZ_{i,j-k}^4} \times \sqrt{1/\varepsilon n} \leq nC(np)^{-1} \sum_{i=1}^p \sqrt{1/\varepsilon n} = C\sqrt{1/\varepsilon n} \rightarrow 0,$$

其中 $C^2 = \max\left\{3, \frac{8(\alpha\lambda + \alpha^4\lambda^4)}{\alpha^4\lambda^4 + 2\alpha^3\lambda^3 + \alpha^2\lambda^2}, \frac{8(\lambda + \lambda^4)}{\lambda^4 + 2\lambda^3 + \lambda^2}\right\}$ 。

由上述证明知 \mathbf{X} 的每一行元素可表示为随机线性序列。注意到

$$|c_k| = \begin{cases} (\lambda^2 + \lambda)^{1/2}, & k=0 \\ (\alpha^2 \lambda^2 + \alpha \lambda)^{1/2}, & k=1 \\ 0, & k \geq 2 \end{cases}.$$

$\forall \delta > 0$ 和 $\forall C \geq 2^{1+\delta} \sqrt{\lambda^2 + \lambda}$ 都有 $|c_k| \leq C(k+1)^{-1-\delta}, \forall k \geq 0$ 成立。因此引理 2.1 中的条件 1 满足。由性质 1.2 可知, $f(w)$ 是有理函数, 对于 $\gamma \in \mathbb{R}$ 满足 $f(w) = \gamma$ 的根只有有限多个, 引理 2.1 中的条件 2、3 满足。

由引理 2.1 可知, 当 $n \rightarrow \infty, p/n \rightarrow c \in (0, \infty)$ 时, 样本协方差矩阵 S_n 的极限谱分布 F 存在, 且存在正数 $a < b$ 使得 F 的 Stieltjes 变换是满足如下方程的唯一映射

$$\frac{1}{m} = -z + \frac{c}{2\pi} \int_a^b \frac{\lambda}{1 + \lambda m} \sum_{w \in [0, 2\pi]} \frac{1}{f'(w)} d\lambda.$$

由文献[17]可知, 上式等价于

$$z = -\frac{1}{m} + \frac{c}{2\pi} \int_0^{2\pi} \frac{f(t)}{1 + mf(t)} dt,$$

将谱密度 $f(w)$ 代入, 得

$$z = -\frac{1}{m} + \frac{c}{2\pi} \int_0^{2\pi} \frac{2\alpha\lambda \cos w + (\alpha+1)\lambda}{2\alpha\lambda m \cos w + (\alpha+1)\lambda m + 1} dw,$$

其中, m 是 S_n 对应极限谱分布的 Stieltjes 变换, \underline{m} 是 \underline{S}_n 对应极限谱分布的 Stieltjes 变换, 满足

$$\begin{aligned} cm - \frac{1-c}{z} &= \frac{p}{n} \cdot \frac{1}{p} \sum_{i=1}^p \frac{1}{\lambda_i - z} - \left(1 - \frac{p}{n}\right) \frac{1}{z} \\ &= \frac{1}{n} \sum_{i=1}^p \frac{1}{\lambda_i - z} + \frac{1}{n} (n-p) \frac{1}{0-z} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{\lambda_i - z} \\ &= \underline{m}. \end{aligned}$$

证毕。□

3.3. 定理 2.2 的证明

根据留数定理, (2)式中积分部分可表示为:

$$\frac{c}{2\pi} \int_0^{2\pi} \frac{2\alpha\lambda \cos w + (\alpha+1)\lambda}{2\alpha\lambda m \cos w + (\alpha+1)\lambda m + 1} dw = \frac{c}{m} - \frac{c}{\underline{m} \sqrt{[\underline{m}(\alpha+1)\lambda + 1]^2 - 4(\underline{m}\alpha\lambda)^2}}.$$

则 $\underline{m} \in \mathbb{C}^+$ 是如下方程的唯一的解

$$[c - (z\underline{m} + 1)] \sqrt{[\underline{m}(\alpha+1)\lambda + 1]^2 - 4(\underline{m}\alpha\lambda)^2} = c.$$

上式可转化为如下的一元四次方程:

$$\begin{aligned} \underline{m}^4 + \left[\frac{2(\alpha+1)}{\lambda(1+2\alpha-3\alpha^2)} - \frac{2(y-1)}{z} \right] \underline{m}^3 + \left[\frac{1}{\lambda^2(1+2\alpha-3\alpha^2)} - \frac{4(\alpha+1)(y-1)}{\lambda z(1+2\alpha-3\alpha^2)} + \frac{(y-1)^2}{z^2} \right] \underline{m}^2 \\ + \frac{2(\alpha+1)(y-1)^2 \lambda - 2z(y-1)}{\lambda^2 z^2 (1+2\alpha-3\alpha^2)} \underline{m} + \frac{1-2y}{\lambda^2 z^2 (1+2\alpha-3\alpha^2)} = 0. \end{aligned}$$

由文献[11]可知方程的4个根为

$$\begin{aligned} \underline{m}_1 &= \frac{1}{2} \left(\sqrt{2t-u} + \sqrt{-2t-u - \frac{2v}{\sqrt{2t-u}}} \right) - \frac{a}{4}, \\ \underline{m}_2 &= \frac{1}{2} \left(\sqrt{2t-u} - \sqrt{-2t-u - \frac{2v}{\sqrt{2t-u}}} \right) - \frac{a}{4}, \\ \underline{m}_3 &= \frac{1}{2} \left(-\sqrt{2t-u} + \sqrt{-2t-u + \frac{2v}{\sqrt{2t-u}}} \right) - \frac{a}{4}, \\ \underline{m}_4 &= \frac{1}{2} \left(-\sqrt{2t-u} - \sqrt{-2t-u + \frac{2v}{\sqrt{2t-u}}} \right) - \frac{a}{4}. \end{aligned}$$

其中,

$$\begin{aligned} a &= \left[\frac{2(\alpha+1)}{\lambda(1+2\alpha-3\alpha^2)} - \frac{2(y-1)}{z} \right], \quad b = \frac{1}{\lambda^2(1+2\alpha-3\alpha^2)} - \frac{4(\alpha+1)(y-1)}{\lambda z(1+2\alpha-3\alpha^2)} + \frac{(y-1)^2}{z^2}, \\ c &= \frac{2(\alpha+1)(y-1)^2 \lambda - 2z(y-1)}{\lambda^2 z^2 (1+2\alpha-3\alpha^2)}, \quad d = \frac{1-2y}{\lambda^2 z^2 (1+2\alpha-3\alpha^2)}, \end{aligned}$$

$$u = b - \frac{3a^2}{8}, \quad v = c - \frac{ab}{2} + \frac{a^3}{8}, \quad r = \frac{a^2 b}{16} - \frac{ac}{4} + d - \frac{3a^4}{256}, \quad a_1 = \frac{-u}{2}, \quad b_1 = -r, \quad c_1 = \frac{4ur - v^2}{8}, \quad m_1 = b_1 - \frac{a_1^2}{3},$$

$$n_1 = c_1 - \frac{a_1 b_1}{3} + \frac{2a_1^3}{27}, \quad t = \frac{1}{6} \left(-108n_1 + 12\sqrt{12m_1^3 + 81n_1^2} \right)^{1/3} - 2m_1 / \left(-108n_1 + 12\sqrt{12m_1^3 + 81n_1^2} \right)^{1/3} - \frac{a_1}{3}.$$

为了证明定理 2.2, 我们需要用到如下引理。

引理 2.2. [17] 对于每一个 $z_1 > 0$, 都存在

$$\lim_{z_2 \rightarrow 0^+} m_F(z_1 + iz_2) = m_0(z_1)$$

其中, 函数 m_0 在 R^+ 上是连续的, 并且 F 在 R^+ 上有连续导函数 f , $f(z_1) = 1/\pi \operatorname{Im}[m_0(z_1)]$ 。密度函数 f 对于每一个 $z_1 > 0$, $f(z_1) > 0$ 都是解析的。并且其中的 z_1 , $\pi f(z_1)$ 是满足等式(2)的唯一 $m \in C^+$ 的虚部。

根据上述留数定理的推导结果, 可以得到关于 \underline{m} 的四个解。通过公式 $\underline{m} = -(1-c)z^{-1} + cm$, 容易得到 m_i , ($i=1, \dots, 4$), 将 m_i , ($i=1, \dots, 4$) 代入引理 2.2 的公式 $f(z_1) = 1/\pi \operatorname{Im}[m_0(z_1)]$ 中, 得到密度函数的四个解 f_i , ($i=1, \dots, 4$), 选取 f_i , ($i=1, \dots, 4$) 中函数值大于 0 的密度函数组成新密度函数 $f(x)$ 得到 INMA(1) 型随机矩阵的样本协方差矩阵 S_n 的极限谱密度。即 $f(x) = \sum_{i=1}^4 f_i(z) 1\{f_i(z) > 0\}$ 。

证毕。□

4. 数值模拟

本节通过数值模拟对本文结论进行检验。基于式(1), 模拟生成 5000 个服从 INMA(1)过程的随机数

序列。\$\{B_i\}_{i \geq 1}\$ 是相互独立的伯努利随机变量序列, \$\{\varepsilon_i\}_{i \geq 1} \sim \text{Poisson}(\lambda)\$, 且 \$\{\varepsilon_i\}_{i \geq 1}\$ 与 \$\{B_i\}_{i \geq 1}\$ 独立。INMA(1)型随机矩阵的样本协方差矩阵 \$S_n\$ 的极限谱分布的 Stieltjes 变换是关于参数 \$\alpha\$ 和 \$c\$ 的函数。首先假设 \$\lambda=1\$, 绘制频率直方图和极限谱密度见图 1。图 1 左侧是 \$\alpha=0.1, c=0.3\$ 的图像, 右侧是 \$\alpha=0.7, c=0.9\$ 的图像。阴影部分是基于特征根绘制的频率直方图, 红色曲线是由定理 2.2 得到的 INMA(1)型随机矩阵的极限谱密度函数。从图 1 可以看出, 理论结果与数值模拟结果相一致。进一步地, 我们分别绘制了在 \$\lambda=2\$, \$\alpha=0.3, 0.5, 0.7\$ 和 \$c=0.1, 0.7\$ 时, INMA(1)型随机矩阵的样本协方差矩阵 \$S_n\$ 的极限谱分布的密度函数, 其结果见图 2。从图 2 可知, 随着 \$\alpha\$ 的增大, 图像的峰值会随之上升。

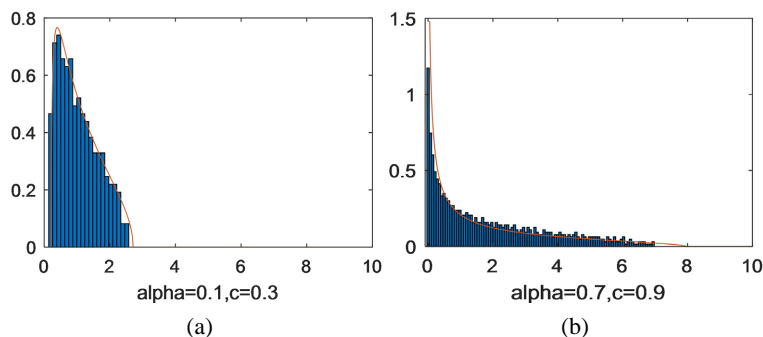


Figure 1. Histograms of eigenvalues and limiting spectral density

图 1. 特征值和极限谱密度的直方图

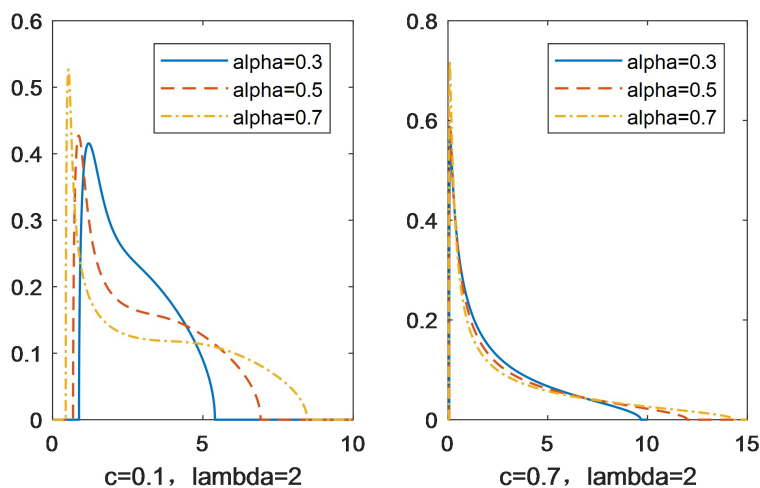


Figure 2. Limiting spectral density of INMA(1) type random matrix under different parameters

图 2. INMA(1)型随机矩阵不同参数下的极限谱密度

本文主要探讨了两个问题, 第一个是 INMA(1)过程的线性表示, 第二个是基于线性表示推导了具有 INMA(1)结构随机矩阵的极限谱密度, 并通过数值模拟验证理论结果。通过对 INMA(1)过程的研究发现, INMA(1)的相关结构与连续情形下的 MA(1)过程类似。特别地, 当 INMA(1)过程 \$\alpha=1, \lambda=1\$ 时, 就变为了 MA(1)过程, 即 \$Y_t = \varepsilon_t + \varepsilon_{t-1}\$。分别绘制了 INMA(1)过程和 MA(1)过程的极限谱密度函数图像见图 3。其中, 左侧设置 INMA(1)参数为 \$c=0.1, \lambda=1, \alpha=0.3, 0.5, 0.7\$, 右侧 MA(1)过程参数为 \$c=0.1, \alpha=0.1, 0.5, 0.7\$。通过比较发现, MA(1)过程相比 INMA(1)过程更靠近原点, INMA(1)过程的峰值比 MA(1)过程低。这说明 INMA(1)过程极限谱分布的尾概率大于 MA(1)过程极限谱分布的尾概率。

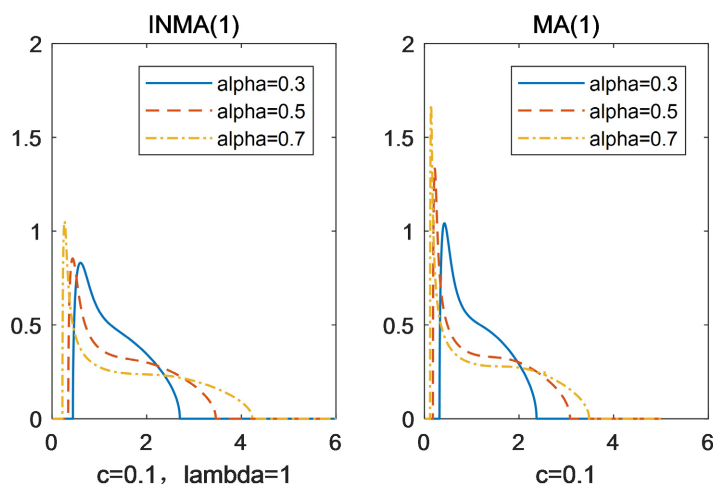


Figure 3. Limiting spectral density of INMA(1) process and MA(1) process

图 3. INMA(1)过程和 MA(1)过程的极限谱密度

结合大维随机矩阵的样本协方差矩阵的极限谱分布函数和经验谱分布函数, 可以构造出相关统计量用于假设检验。例如: 原假设为 $\alpha = 0$ 时, 原观测序列的极限谱密度是 MP 律; 原假设为 $\alpha = 1$ 时, 观测序列的差序列 $\{Y_t - Y_{t-1}\}_{t \geq 1}$ 的极限谱密度是 MP 律。若经验谱密度和 MP 律的差异越小, 则接受原假设的可能性越大。特别地, 以经验谱密度和 MP 律的差的绝对值构造一个积分检验统计量, 结合自助法可以得到该积分检验统计量的经验分布和 P 值, 从而实现假设检验的目的。另一方面, 对于观测序列是否服从 INMA(1) 这一假设检验问题, 可以先利用序列估计 INMA(1) 过程, 结合定理 2.2 可得到该估计 INMA(1) 过程的极限谱密度, 同样的, 利用经验谱密度和极限谱密度的差的绝对值构造一个积分检验统计量, 结合自助法得到该积分检验统计量的经验分布和 P 值, 从而对上述问题进行检验。

基金项目

基于相依整值随机场和深度学习的人类突发性行为建模和预测(RD2300002084)。

参考文献

- [1] McKenzie, E. (1985) Some Simple Models for Discrete Variate Time Series. *JAWRA Journal of the American Water Resources Association*, **21**, 645-650. <https://doi.org/10.1111/j.1752-1688.1985.tb05379.x>
- [2] Al-Osh, M.A. and Alzaid, A.A. (1988) Integer-Valued Moving Average (INMA) Process. *Statistical Papers*, **29**, 281-300. <https://doi.org/10.1007/BF02924535>
- [3] Davis, R.A., Fokianos, K., Holan, S.H., et al. (2021) Count Time Series: A Methodological Review. *Journal of the American Statistical Association*, **116**, 1533-1547. <https://doi.org/10.1080/01621459.2021.1904957>
- [4] Karlis, D. and Khan, N.M. (2023) Models for Integer Data. *Annals Review of Statistics and Its Application*, **10**, 297-323. <https://doi.org/10.1146/annurev-statistics-032921-022516>
- [5] 王成. 大维数据的总体协方差矩阵研究[D]: [博士学位论文]. 合肥: 中国科学技术大学, 2013.
- [6] Wigner, E.P. (1995) Characteristic Vectors Bordered Matrices with Infinite Dimensions. *Annals of Mathematics*, **62**, 548-564. <https://doi.org/10.2307/1970079>
- [7] Wigner, E.P. (1958) On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics*, **67**, 325-327. <https://doi.org/10.2307/1970008>
- [8] Marchenkou, V.A. and Pastur, L.A. (1967) Distribution of Eigenvalues for Some Sets of Random Matrices. *Mathematics of the USSR-Sbornik*, **1**, 457-483. <https://doi.org/10.1070/SM1967v001n04ABEH001994>
- [9] Silverstein, J.W. and Bai, Z.D. (1995) On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices. *Journal of Multivariate Analysis*, **54**, 175-192. <https://doi.org/10.1006/jmva.1995.1051>

-
- [10] Bai, Z.D. and Zhou, W. (2008) Large Sample Covariance Matrices without Independence Structures in Columns. *Statistica Sinica*, **18**, 425-442.
- [11] 胡江, 胡果荣, 史晓磊. 大维 AR(1)模型样本协方差阵的极限谱密度[J]. 东北师大学报(自然科学版), 2010, 42(2): 1-6.
- [12] Pfaffel, O. and Schlemm, E. (2011) Eigenvalue Distribution of Large Sample Covariance Matrices of Linear Process. *Probability and Mathematical Statistics*, **31**, 313-329.
- [13] 王才士, 刘琰. ARMA(p,q)过程生成的大维样本协方差矩阵的极限谱分布[J]. 兰州理工大学学报, 2014, 40(2): 150-153.
- [14] Steutel, F.W. and van Harn, K. (1979) Discrete Analogues of Self-Decomposability and Stability. *Annals of Probability*, **7**, 893-899. <https://doi.org/10.1214/aop/1176994950>
- [15] 杨凯, 张庆庆, 刁亚静, 等. 基于 INMA(q)模型的高频股票交易数据实证分析[J]. 长春工业大学学报, 2020, 41(2): 112-117.
- [16] 赵晓芳. 大维随机矩阵极限谱分布理论研究及其在 AR(1)模型中的应用[D]: [硕士学位论文]. 合肥: 中国科学技术大学, 2011.
- [17] Wang, C., Jin, B. and Miao, B. (2011) On limiting Spectral Distribution of Large Sample Covariance Matrices by VARMA(p,q). *Journal of Time Series Analysis*, **32**, 539-546. <https://doi.org/10.1111/j.1467-9892.2010.00712.x>